

Cluster Analysis of Multimodel Ensemble Data over New England

NUSRAT YUSSOUF*

Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma

DAVID J. STENSRUD

NOAA/National Severe Storms Laboratory, Norman, Oklahoma

S. LAKSHMIVARAHAN

School of Computer Science, University of Oklahoma, Norman, Oklahoma

(Manuscript received 10 April 2003, in final form 8 April 2004)

ABSTRACT

An ensemble of 48-h forecasts from 23 cases during the months of July and August 2002, which was created as part of a National Oceanic and Atmospheric Administration pilot program on temperature and air quality forecasting, is evaluated using a clustering method. The ensemble forecasting system consists of 23 total forecasts from four different models: the National Centers for Environmental Prediction (NCEP) Eta Model (ETA), the NCEP Regional Spectral Model (RSM), the Rapid Update Cycle (RUC) model, and the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5). Forecasts of 2-m temperature, 850-hPa *u*-component wind speed, 500-hPa temperature, and 250-hPa *u*-component wind speed are bilinearly interpolated to a common grid, and a cluster analysis is conducted at each of the 17 output times for each of the case days using a hierarchical clustering approach.

Results from the clustering indicate that the forecasts largely cluster by model, with these intramodel clusters occurring quite often near the surface and less often at higher levels in the atmosphere. Results also indicate that model physics diversity plays a relatively larger role than initial condition diversity in producing distinct groupings of the forecasts. If the goal of ensemble forecasting is to have each model forecast represent an equally likely solution, then this goal remains distant as the model forecasts too often cluster based upon the model that produces the forecasts. Ensembles that contain both initial condition and model dynamics and physics uncertainty are recommended.

1. Introduction

The importance of model physics variability in ensemble forecasting is becoming widely accepted for both medium-range (Buizza et al. 1999; Harrison et al. 1999; Evans et al. 2000) and short-range forecasting (Stensrud et al. 2000). However, perhaps even more important and interesting is that model dynamics diversity, not just model physics variability, also is becoming widely accepted as a needed component in ensemble forecasting systems (Atger 1999; Stensrud et al. 2000; Fritsch et al. 2000; Ziehmann 2000; Hou et al. 2001; Wandishin et al. 2001), although questions remain regarding how much benefit is gained through the use

of multiple models versus information from the different operational analyses and model physics diversity for medium-range ensembles (Richardson 2001). In general, for short-range forecasts of sensible weather parameters, these results suggest that it is difficult for a single model, even with strongly perturbed initial conditions, to capture the atmospheric variability. Too often the evolution of the atmosphere lies outside the envelope of the ensemble solutions.

Conceptual models of ensemble forecasting systems suggest that each model within the ensemble has its own envelope of solutions that only partly overlap the solution envelopes from other models in the ensemble, leading to a greater variety of solutions when multimodel ensembles are used. Even when one of the models used as part of an ensemble is clearly inferior, the inclusion of this model yields better overall results (Wandishin et al. 2001). However, the relative importance of model physics variability and model dynamics variability remains unknown, as does the relative importance of model and initial condition uncertainty.

* Additional affiliation: NOAA/National Severe Storms Laboratory, Norman, Oklahoma.

Corresponding author address: Nusrat Yussouf, NOAA/NSSL, 1313 Halley Circle, Norman, OK 73069.
E-mail: Nusrat.Yussouf@noaa.gov

In an attempt to improve our understanding of model diversity in short-range ensemble forecasting, Alhamed et al. (2002) use various cluster analysis techniques to examine data from a 25-member ensemble created as part of the Storm and Mesoscale Ensemble Experiment (SAMEX) during 1998. Clustering groups data based upon their patterns, and as such offers a useful approach to objectively examining ensemble data. Comparisons of the results from the various clustering methods to the results from a subjective clustering indicates that the clustering methods produce results that are largely in agreement with those determined subjectively. This analysis provides confidence that clustering methodologies can be used to examine larger datasets robustly.

Four different models are used to create the ensemble during SAMEX, and the results of the cluster analyses indicate that the forecasts cluster by model (Alhamed et al. 2002). This clustering occurs very quickly, such that within the first few hours of the forecasts these model clusters are seen clearly. This result suggests that model diversity needs to be part of our present ensemble forecasting systems. Unfortunately, Alhamed et al. (2002) only used data from a single case to examine the effects of model diversity on ensembles.

To extend the results of Alhamed et al. (2002), data from another short-range ensemble forecasting experiment are analyzed with the same clustering method. The National Oceanic and Atmospheric Administration (NOAA) began a pilot program on temperature and air quality forecasting over New England during the summer of 2002. As part of this program, a short-range ensemble forecasting system was constructed to evaluate if an ensemble approach can provide improved 2-m temperature and dewpoint temperature forecasts. Data from this pilot program are used in this study.

The ensemble forecasting system is outlined in section 2. Section 3 contains a brief overview of the clustering methods used. Results from the clustering analyses are found in section 4, followed by a final discussion in section 5.

2. Data

The New England Temperature and Air Quality Pilot Project started on 15 July 2002 and ran through 31 August 2002 for a total of 48 days. During this time, four different numerical weather prediction models provided daily forecasts in real time. The models are the National Centers for Environmental Prediction (NCEP) Eta Model (Black 1994), the Regional Spectral Model (RSM; Juang and Kanamitsu 1994), the Rapid Update Cycle model (RUC; Benjamin et al. 1994, 2001), and the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5; Dudhia 1993). The computational domains for each of these models extend well beyond the 48 contiguous states.

Fifteen forecasts are obtained from NCEP. Five of

these are from the 48-km Eta Model (ETA) that uses the breeding of growing modes technique to generate two positive and two negative bred perturbations, in addition to the control analysis, for the initial and boundary conditions (Toth and Kalnay 1993, 1997). The Eta Model uses the Betts–Miller–Janjic convective scheme (Betts and Miller 1993; Janjic 1994), a 1.5-order closure planetary boundary layer scheme (Janjic 1994), and a multilayer soil–vegetation land surface model (Chen et al. 1996). Another five forecasts use a version of the Eta Model (EKF) that incorporates the Kain–Fritsch convective parameterization scheme (EtaKF; Kain et al. 2001) in a separate bred-mode cycling. All other model physical process schemes are the same as in the Eta Model. Note that the control analyses for the ETA and EKF are identical,¹ although the perturbed members are not owing to the separate breeding cycles. The remaining five forecasts are from the 48-km RSM, again using a bred-mode cycling to generate the initial and boundary conditions. The RSM uses a simplified Arakawa–Schubert convective parameterization scheme (Pan and Wu 1995) and the Medium-Range Forecast (MRF) model nonlocal closure planetary boundary layer scheme (Hong and Pan 1996).

There are six forecasts from National Severe Storms Laboratory (NSSL). Two of these are from the 20-km EtaKF model (NKF) that is started from both the operational Eta Model and the operational “aviation” run of the MRF model initial conditions, but uses a smaller domain than the operational Eta Model (Kain et al. 2001). The remaining four forecasts are from MM5, which uses the Eta Model initial and boundary conditions for the control run and a random coherent structure approach to generate another three initial conditions for the 32-km grid (see Stensrud et al. 2000). These four forecasts also mix the Kain–Fritsch (Kain and Fritsch 1990) and Betts–Miller–Janjic convective schemes, and the MRF and Blackadar (Zhang and Anthes 1982) planetary boundary layer schemes, to provide both initial condition and model physics variability. The control MM5 and one of the NKF runs use the Eta Model control forecast to provide initial and boundary conditions.

Two final forecasts are from Forecast Systems Laboratory (FSL). One forecast is from the 20-km RUC, which uses its own optimal interpolation scheme to produce an initial condition (Benjamin 1989) and the Eta Model forecast for boundary conditions. The RUC uses a Grell (1993) convective scheme, a 1.5-order planetary boundary layer scheme (Burk and Thompson 1989), and a six-layer soil and vegetation model (Smirnova et al. 1997, 2000). The other forecast is an MM5 run (MAQ)²

¹ Although the initial conditions are identical, the postprocessing schemes are not, leading to some slight apparent differences between these control analyses.

² The MAQ forecast does not include the 250-hPa *u*-wind component as an output variable, so there are 22 ensemble members only for this variable.

TABLE 1. List of institutions, models, and their start times for the ensemble members.

Institution	Model	Start time	Members
NCEP	ETA	0900 UTC	ETA1, ETA2, ETA3, ETA4, ETA5
	EtaKF	0900 UTC	EKF1, EKF2, EKF3, EKF4, EKF5
	RSM	0900 UTC	RSM1, RSM2, RSM3, RSM4, RSM5
NSSL	EtaKF	0000 UTC	NKF1, NKF2
	MM5	0000 UTC	MM51, MM52, MM53, MM54
FSL	RUC	1200 UTC	RUC
	MM5 AQ	1200 UTC	MAQ

that uses 27-km grid spacing and is started from the RUC analysis with the Eta Model forecast for boundary conditions. The MAQ uses the Grell convective scheme, the Eta boundary layer scheme, and the Smirnova land surface scheme.

Therefore, the multimodel ensemble evaluated in this study consists of 23 members as listed in Table 1. These 23 members are available once each day depending upon the availability of the various model runs in real time. The unperturbed control forecasts of ETA, EKF, NKF, and MM5 all have initial and boundary conditions based upon the Eta Model analyses and forecasts, although the start times are different (see Table 1). The control forecasts of the RSM and the second NFK similarly have initial and boundary conditions based upon the MRF analyses and forecasts. Because of computer problems, one or more of these members is missing for half of the days. Out of the 48 forecast days, 23 days have complete forecast datasets available, and the cluster analysis that follows is done only on those 23 complete datasets. In addition to storing forecast data for the near-surface fields, the project organizers also stored forecast data for selected fields at several pressure levels from 850 to 250 hPa. These data allow us to evaluate whether or not any of the resulting clusters are changed as a function of atmospheric level.

Results from the 15-member NCEP ensemble indicate

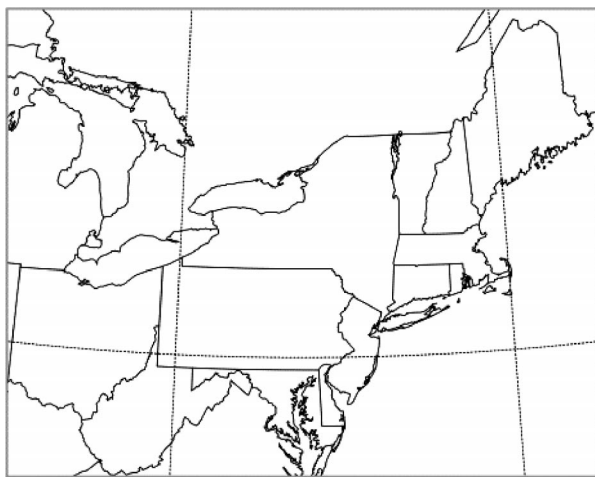


FIG. 1. Common 10-km grid domain used by all the ensemble members. All model forecasts are bilinearly interpolated to this grid prior to input into the clustering algorithm.

that the ensemble median forecasts during September 2002, the month following the conclusion of this NOAA pilot program, are significantly better than the forecasts from the 12-km Eta Model (Du et al. 2003) for six variables investigated. For many of the fields, the root-mean-square errors are 40% less for the ensemble median than for the 12-km Eta forecasts at 48 h. Results further indicate that this ensemble system produces nearly flat rank histograms (Hamill 2001) for precipitation forecasts, suggesting that it has near-perfect spread. Thus, there is reason to believe that the larger 23-member ensemble used in this study provides reasonably accurate and useful forecasts.

The model outputs are bilinearly interpolated to a common 10-km Lambert conformal grid that includes the New England region (Fig. 1) for each of the selected variables and pressure levels. Depending upon local resources, the start times from the model output vary from 0000 to 1200 UTC, but all the model forecasts are available every 3 h from 0 to 48 h beginning at 1200 UTC each day. The dataset for each forecast field studied in this research represents the output from all the model forecasts starting at 1200 UTC each day.

3. Clustering algorithms

Cluster analysis is a multivariate technique that is used to group objects together based on a selected similarity measure. Let $\mathbf{X} = [x_{ij}]$, $1 \leq i \leq m$, $1 \leq j \leq n$ be the data matrix where the n columns represent objects (ensemble members) and m rows represent observations (forecasts). Therefore, X_{ij} represents the i th forecast of the j th ensemble member. Let x_{*j} refer to the j th object, and x_{i*} refer to the i th observation across the n objects. The aim of clustering is to classify this set of n objects according to their similarities as calculated with the chosen measure.

It often is useful to transform the data matrix to make each observation contribute equally to the analysis (Romesburg 1984). Since the transformed data matrix \mathbf{Z} is the same size as \mathbf{X} , it can be used in the cluster analysis. Let $\bar{x}_{*j} = 1/m \sum_{i=1}^m x_{ij}$ be the mean of each object. The simplest type of transformation subtracts the object mean from all observations of this object.

$$z_{ij} = x_{ij} - \bar{x}_{*j} \quad \text{for } 1 \leq i \leq m. \quad (3.1)$$

The \mathbf{Z} obtained using this type of transformation is

called the difference field and is used for all the clustering analyses reported in this study.

Since we are interested in the underlying nature of clustering in a multimodel ensemble, and have no desire to predetermine any of the results, a hierarchical method is selected for the cluster analyses [see Alhamed et al. (2002) for a more complete discussion]. This method allows us to derive hypotheses related to the natural clustering of the ensemble members. The first step in a hierarchical cluster analysis is to compute the resemblance coefficient for each pair of variables, and here the Euclidean distance is used as the resemblance coefficient on the difference field. Euclidean distance measures the distance, e_{jk} , between two objects j and k using

$$e_{jk} = \sqrt{\sum_{i=1}^n (z_{ij} - z_{ik})^2} \quad (3.2)$$

where

$$0 \leq e_{jk} \leq \infty,$$

such that e_{jk} is the dissimilarity between objects j and k . Once the matrix is computed, we can use it in conjunction with one of the many hierarchical clustering algorithms. Since several studies indicate that Ward's method is the best among the hierarchical clustering methods (McIntyre and Blashfield 1980; Morey et al. 1983; Jain and Dubes 1988; Breckenridge 1989; Gong and Richman 1995), and Ward's method also produces good results with the SAMEX multimodel ensemble data (Alhamed et al. 2002), it is selected as the clustering method for this study.

For Ward's method, each object is in a separate cluster at the beginning of the clustering procedure. At each iteration, a sum-of-square error S is computed for every possible merger of two clusters. The merger that produces the smallest value of S is taken to be the clustering of this step. This process is repeated until one large cluster, which contains all objects, is obtained. Initially, each cluster contains only one object; hence the value of S at the beginning is zero. Detailed information on clustering methodologies is found in Anderburg (1973), Romesburg (1984), Jain and Dubes (1988), and Alhamed et al. (2002). We now turn to the clustering results from the multimodel ensemble data.

4. Cluster analysis of data

In this section, the results from a cluster analysis of four forecast fields are presented. The fields selected are 2-m temperature, 850-hPa u -wind component, 500-hPa temperature, and 250-hPa u -wind component. The cluster analysis is performed separately for each variable and only for the cases in which we have a complete forecast dataset of all 23 model forecasts. These complete datasets are available for a total of 23 cases. For each forecast field and each case day, 17 different data matrices are constructed, one for each 3-h output time

from 0 to 48 h. The data matrix, \mathbf{X}_{mn} , where $m = 16\,675$ and $n = 23$ represents the values of the model forecast field at each of the 145×115 grid points from the 23 ensemble members. We begin by examining the cluster results for the 2-m temperature field.

a. 2-m temperature

The data matrix for each time interval is first centered using (3.1) and then the Euclidean distance [Eq. (3.2)] matrix $\mathbf{E}_{23 \times 23}$, is computed. To illustrate the hierarchical clustering structures of the 23 ensemble members for 2-m temperature, the case from 9 August 2002 is illustrated. Clustering trees, or dendrograms, are a convenient way to display the results from any cluster analysis. The objects are shown at the bottom and are connected by solid lines. The lower the point (or height) at which two objects connect, the more similar the objects. Thus, the 9-h forecast (Fig. 2a) shows the ETA1 and ETA3 as being the most similar forecasts since they are the first objects to connect. The merged EKF2 and EKF5 are the next most similar forecasts. As you move upward on the diagram, more clusters are formed. Note that as the forecast time increases, the top height for which all the models cluster also increases on average, indicating that the forecasts are becoming less similar.

As in Alhamed et al. (2002), a subjective clustering of several cases also is performed independently by one of the authors (DJS). Results indicate that the clusters derived subjectively and those derived using the clustering algorithm are very similar. The results agreed completely for the 2-m temperature field, and largely agreed for the clusters at the other three vertical levels. At the higher levels in the atmosphere it is at times difficult to determine into which cluster a particular model forecast best fits, and these are the forecast fields for which the subjective clustering and algorithm clustering disagreed. For clusters that are more isolated and distinct, the objective and subjective clustering approaches always agreed. These comparisons provide some confidence that the resulting clusters are realistic and useful for comparing the various ensemble members.

The clustering trees initially are cut to form three clusters (as shown by the straight thick gray line across the dendrogram in Fig. 2a that intersects only three vertical lines on the clustering dendrogram). The first cluster contains ETA and EKF; the second contains RUC, NKF, MM5, and MAQ; and the third cluster contains members from only RSM. From the clustering trees, it is observed that the members from one model build their own cluster first before merging with members from other models. If we choose instead to have four final clusters (thick gray dashed line in Fig. 2a that intersects four vertical lines on the clustering dendrogram), then the first and the third cluster remain the same, but the second cluster breaks into two more clusters with one

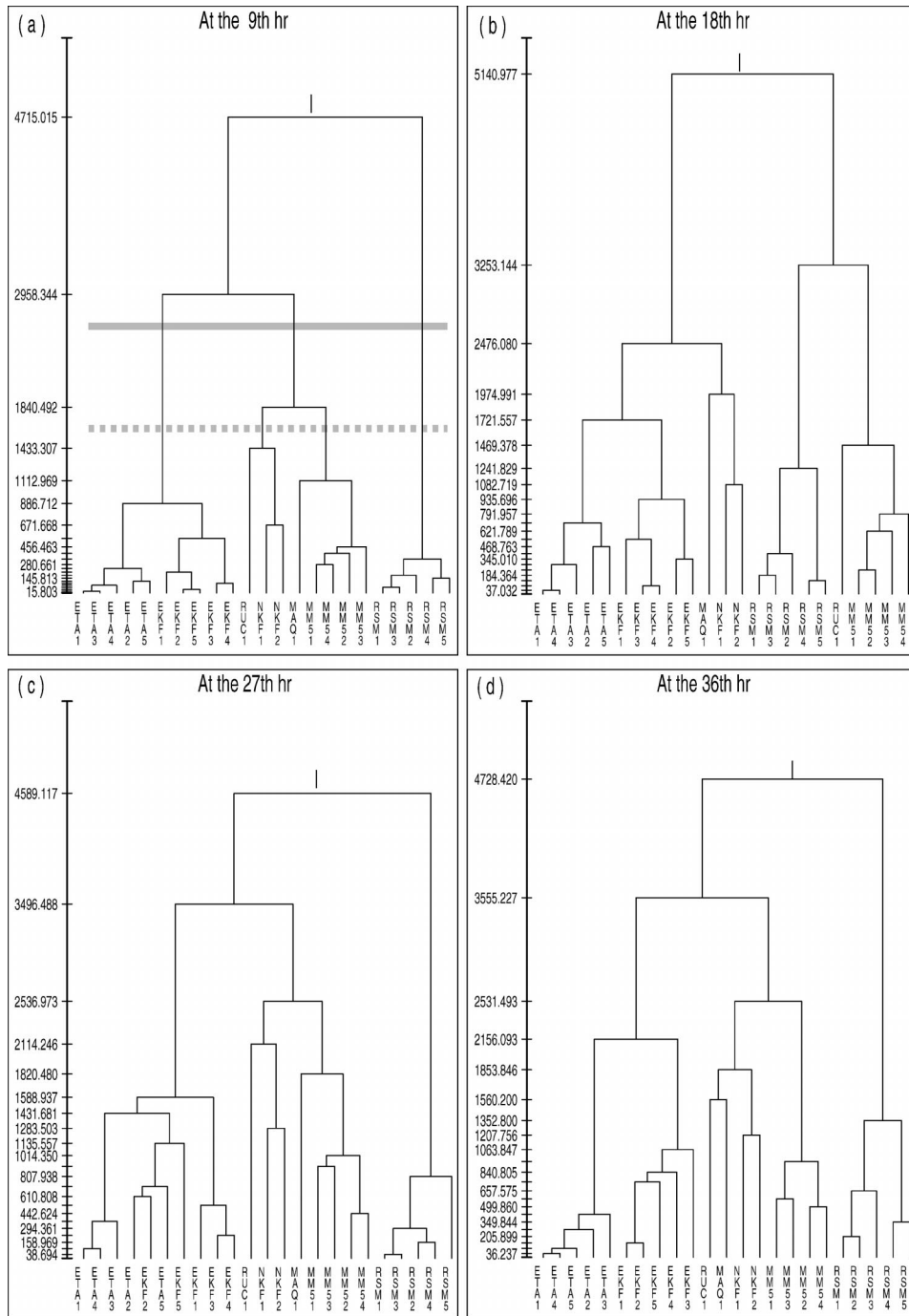


FIG. 2. Ward's dendrogram for 2-m temperature on 9 Aug 2002 based on Euclidean distance for forecast times of (a) 9, (b) 18, (c) 27, and (d) 36 h. Dendrograms, or tree diagrams, show the order in which the model forecasts cluster. The two forecasts that are most alike cluster first and are shown by the lowest connecting line, or branch, on the diagram. The higher the connecting line, or merging point, on the diagram, the lower the level of similarity between the two forecast clusters. Horizontal lines in (a) indicate the groupings for three clusters (thick gray line) or four clusters (dashed gray line).

containing RUC and NKF and the other one containing MAQ and MM5.

To illustrate how the members of the ensemble cluster as they evolve with time, tables are created for each day

with two sets of clustering: one using three clusters, and the other using four clusters. Table 2 summarizes the grouping of the forecast models from the clustering trees on 30 July 2002. By examining the clustering trees from

TABLE 2. Summary of the cluster results for the 2-m temperature field on 30 Jul 2002. Each row represents the clustering of the models at 3-h time intervals, so there are 17 rows for forecast times of 0 through 48 h. The first column displays the hour at which the forecast is valid, the next three columns (C1, C2, and C3) summarize the clustering of the models if the trees are cut to form three clusters, and the remaining four columns (C1, C2, C3, and C4) summarize the four clusters results.

Hour	Three clusters			Four clusters			
	C1	C2	C3	C1	C2	C3	C4
00	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC, MAQ, NKF	MM5	RSM
03	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC	MM5, MAQ, NKF	RSM
06	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC	MM5, MAQ, NKF	RSM
09	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC	MM5, MAQ, NKF	RSM
12	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF, NKF	RUC	MM5, MAQ, NKF	RSM
15	ETA, EKF, NKF	MM5, RUC, MAQ	RSM	ETA, EKF, NKF	NKF	MM5, MAQ, RUC	RSM
18	ETA, EKF, NKF	MM5, RUC, MAQ	RSM	ETA, EKF, NKF	RUC	MM5, MAQ, RUC	RSM
21	ETA, EKF, NKF	MM5, RUC, MAQ	RSM	ETA, EKF, NKF	RUC	MM5, MAQ, RUC	RSM
24	ETA, EKF, NKF, MAQ	MM5, RUC	RSM	ETA	EKF, MAQ, NKF	MM5, MAQ, RUC	RSM
27	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC, NKF	MM5, MAQ, RUC	RSM
30	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC	MM5, MAQ, NKF	RSM
33	ETA, EKF	MM5, RUC, MAQ, NKF	RSM	ETA, EKF	RUC	MM5, MAQ, NKF	RSM
36	ETA, EKF, NKF	MM5, RUC, MAQ	RSM	ETA	EKF, NKF	MM5, MAQ, RUC	RSM
39	ETA, EKF, NKF, MAQ	MM5, RUC	RSM	ETA	EKF, MAQ, NKF	MM5, RUC	RSM
42	ETA, EKF, NKF, MAQ	MM5, RUC	RSM	ETA, EKF	MAQ, NKF	MM5, RUC	RSM
45	ETA, EKF, NKF, MAQ	MM5	RSM, RUC	ETA, EKF	MAQ, NKF	MM5	RSM, RUC
48	ETA, EKF	MM5, MAQ, NKF	RSM, RUC	ETA, EKF	RUC	MM5, MAQ, NKF	RSM

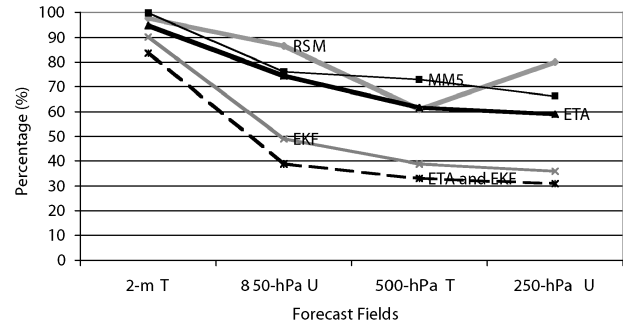


FIG. 3. Percentage of the cases for which the individual model members cluster together when the dendrogram trees are cut to form three clusters for the four different forecast fields and the four main modeling systems (ETA, EKF, MM5, and RSM).

all 17 forecast times from all 23 case days, it is observed that the members from the RSM model have a very strong tendency to build a separate cluster (98% of the cases) regardless of whether or not the trees are cut to form three or four clusters. There are a very few cases in which RSM groups with other models, such as RUC, MM5, or members from NKF (e.g., at 45 and 48 h in Table 2), but these are uncommon. Of all the models, the clustering of the RSM is the most isolated and distinct.

The ETA and EKF have a strong tendency to join together to form one cluster (84% of the cases), with the members from the ETA forming their own cluster before merging with members from the EKF (Table 2). However, the ETA and the EKF also are seen to form different clusters (e.g., at 24, 36, and 39 h in Table 2). Here the ETA forms its own cluster, while the EKF merges with some of the other model runs if the trees are cut to form four clusters. There are several days where a few of the members from the ETA and the EKF form one cluster, with the remaining members forming a separate cluster. This happens infrequently, however.

Members from MM5 have the strongest tendency to build their own cluster, occurring in 99.8% of the cases for 2-m temperature, before merging with other models such as MAQ, RUC, and NKF (Table 2). However, unlike the RSM, the MM5 often clusters together with at least one other model forecast member a majority of the time, such that the MM5 cluster is not nearly as isolated as the RSM cluster. These results confirm the conclusions of Alhamed et al. (2002) that the model dynamics is a very important factor in the groupings of the resulting forecasts.

b. Nonsurface fields

The clustering dendrograms resulting from the 850-hPa *u*-component wind speeds, 500-hPa temperatures, and 250-hPa *u*-component wind speeds show the same general tendencies as observed in the previous analysis of the 2-m temperature (Fig. 3). That is, the members of one model tend to build their own cluster before

merging with members from other models. However, intermodel clustering of the model members is more often observed than seen with the 2-m temperature data as indicated by the smaller percentages. Members from the ETA and EKF often remain in a single group, emphasizing the closeness of the members of these two models to each other. However, members from the ETA are more compact, clustering together above 50% of the time for all variables, suggesting that they exhibit less diversity in their solutions than members from the EKF. It is observed that all five members from EKF stay together in only 49% of the cases for 850-hPa u -component wind speeds, decreasing to 37% of the cases for 250-hPa u -component wind speeds. Thus, in many cases one or more members from EKF form clusters with other model members.

The behavior of the RSM is unusual, as the other larger model groups (ETA, EKF, and MM5) show slight declines in the number of cases in which all the model members cluster together as the pressure level of the clustering decreases (Fig. 3). In addition, the initial condition perturbations from the breeding of growing modes technique typically is largest in the mid- to upper troposphere (Toth and Kalnay 1997; Wandishin et al. 2001), suggesting that larger differences between the RSM members should occur at lower tropospheric pressures. This would lead one to expect that the RSM would be less isolated from the other model members and less compact at 500 and 250 hPa than at other levels. That the cluster results do not indicate such a behavior in the RSM suggests the strong role that is played by the model dynamics in determining the cluster results.

In contrast with the RSM, the other models (ETA, EKF, and MM5) show a decline in the frequency at which the individual model members stay together to form clusters as the pressure level is decreased (Fig. 3). This suggests that the initial condition perturbations in the ETA and EKF are playing a role in moving the model solutions apart, and thereby allowing solutions from different models to cluster together. Both initial condition and model physics perturbations are playing a role in the MM5 results. Of the four models evaluated, the EKF appears to be the most dispersive, with all the EKF members clustering together slightly less than half the time by 250 hPa. Yet it is a concern that the models do not develop clusters that are more random selections from the individual models. The lack of these types of clusters again indicates the strong influence that the model dynamics plays in determining the solution for a given initial condition.

If this clustering behavior is examined as a function of forecast time, then the results show that the percentage of cases for which the individual models stay together decreases with forecast time (Fig. 4). This suggests that the solution envelopes from the individual models overlap one another more often at later forecast times. While the clustering for 2-m temperature remains very compact even at 48 h, the clustering for the other

variables indicates that the EKF is most likely to cluster with other models and the RSM is the least likely to cluster with other models. The ETA and MM5 behave in a similar fashion, with one being slightly less compact than the other depending upon the variable examined.

c. Dendrogram analyses

Another approach to analyzing the dendrograms is to compare the height at which each individual model clusters with the height at which all the models cluster. This ratio provides information on the relative amount of variability contained within a single model ensemble compared to the complete multimodel ensemble. The value of this ratio always will be less than or equal to one. The larger the ratio, the less compact (or more variable) is the resulting single model cluster. If a cluster is created to represent a random grouping from all the model forecasts, then this ratio should approach one. Results indicate that the EKF is the most variable of the model clusters at the initial time, with ratios near 0.7 for all nonsurface variables (Fig. 5). As forecast time increases, the EKF cluster remains the most variable, with values approaching 0.9 by 48 h. But all the clusters increase in variability over time with ratio values approaching 0.7 at 48 h for most fields. The 2-m temperature clusters have the least variability at 0 h and remain the least variable, and most compact, out to 48 h. Clearly, increased variability in the near-surface variables is needed within this ensemble.

These results also indicate that the four-member MM5 cluster produces increases in variability over time that are very similar to those seen with the five-member ETA, EKF, and RSM clusters (Fig. 5). Closer inspection of the many dendrograms reveals that the MM5 members often cluster at higher heights on the dendrograms than the members of ETA, EKF, and RSM. Results indicate that this occurs in 59% of the cases for 2-m temperature and 67% of the cases for the 850-hPa u -component wind speed, indicating that the MM5 members often are less similar to each other than are the members of these other models. This suggests that a multiphysics ensemble started using a random coherent structure technique can compete with a bred-mode ensemble using a single model in producing forecast envelopes that overlap those from other models. However, even though the MM5 forecasts use the ETA control analysis as the foundation for all their initial and boundary conditions, it is difficult to evaluate clearly the relative importance of model dynamics versus model initial condition uncertainty in this ensemble using these results alone.

Thankfully, the control forecasts from the ETA and EKF use identical initial and boundary conditions, and this allows us to evaluate the relative importance of model physics versus model initial condition uncertainty. The ETA and EKF control forecasts cluster at very low heights in the dendrograms at 0 h, with the nonzero

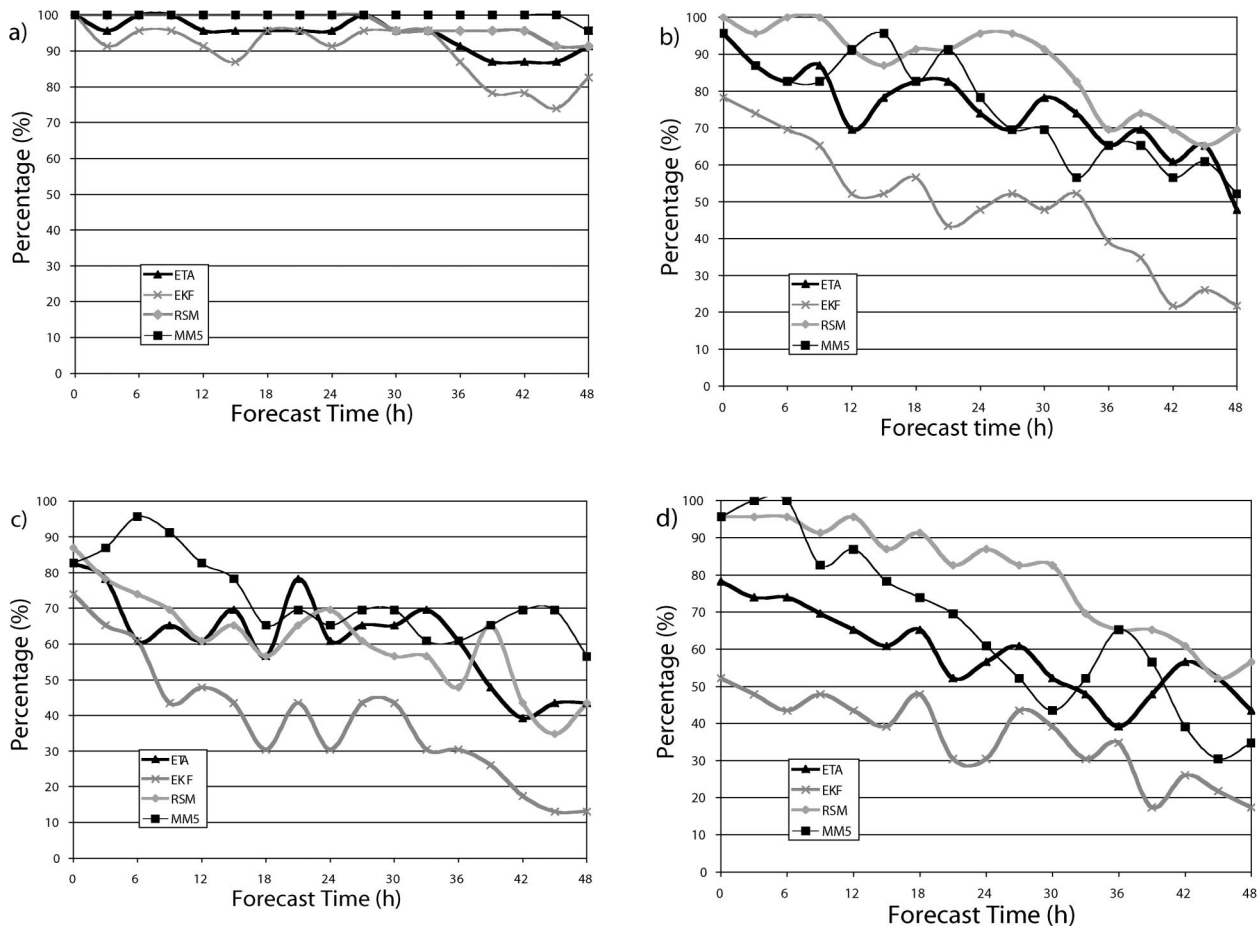


FIG. 4. Percentage of cases for which the individual models cluster together when the dendrogram trees are cut to form three clusters vs forecast time (h) for the four main modeling systems (ETA, EKF, MMS, and RSM). Fields examined are (a) 2-m temperature, (b) 850-hPa *u*-wind component, (c) 500-hPa temperature, and (d) 250-hPa *u*-wind component.

differences attributed to the different postprocessing schemes. If we then compare this height against the height at which all five ETA members cluster for various forecast times, we can examine whether or not the EKF control is contained within the envelope of the ETA members. If the EKF control forecast is contained within this envelope, then model physics diversity is not important to how the forecasts cluster. If it is not contained within this envelope, then model physics diversity is important. Note that the biases from the ETA and EKF control forecasts are very similar for most fields when averaged over many cases and the entire model domain (M. Baldwin 2003, personal communication), suggesting that differences in the biases should not predetermine the outcome of this comparison.

Results indicate that the ratio of the height on the dendrograms at which the ETA–EKF control forecasts cluster to the height at which the five-member ETA forecasts cluster increases from near zero³ at 0 h to over

³ Again, recall that the postprocessing schemes for the ETA and EKF are different, which apparently produce the slightly nonzero ratio values at 0 h.

1.0 by 24 h (Fig. 6). For both 500-hPa temperatures and 850-hPa *u*-component winds, the ratio exceeds 1.0 by 12 h. This indicates that the EKF control forecast quickly moves outside the envelope of the five-member ETA ensemble, created by the breeding of growing modes technique, and strongly suggests that model physics uncertainty plays a relatively larger role in the behavior of the larger multimodel ensemble system than does initial condition uncertainty.

5. Discussion

An ensemble of 48-h forecasts from 23 cases during the months of July and August of 2002 over the New England region have been evaluated using a clustering methodology. The ensemble was created as part of a NOAA pilot program on temperature and air quality forecasting. The ensemble forecasting system was constructed using different models: the NCEP Eta Model (five forecasts), the NCEP RSM (five forecasts), the NCEP Eta Model with the Kain–Fritsch convective parameterization (seven forecasts), the RUC (one fore-

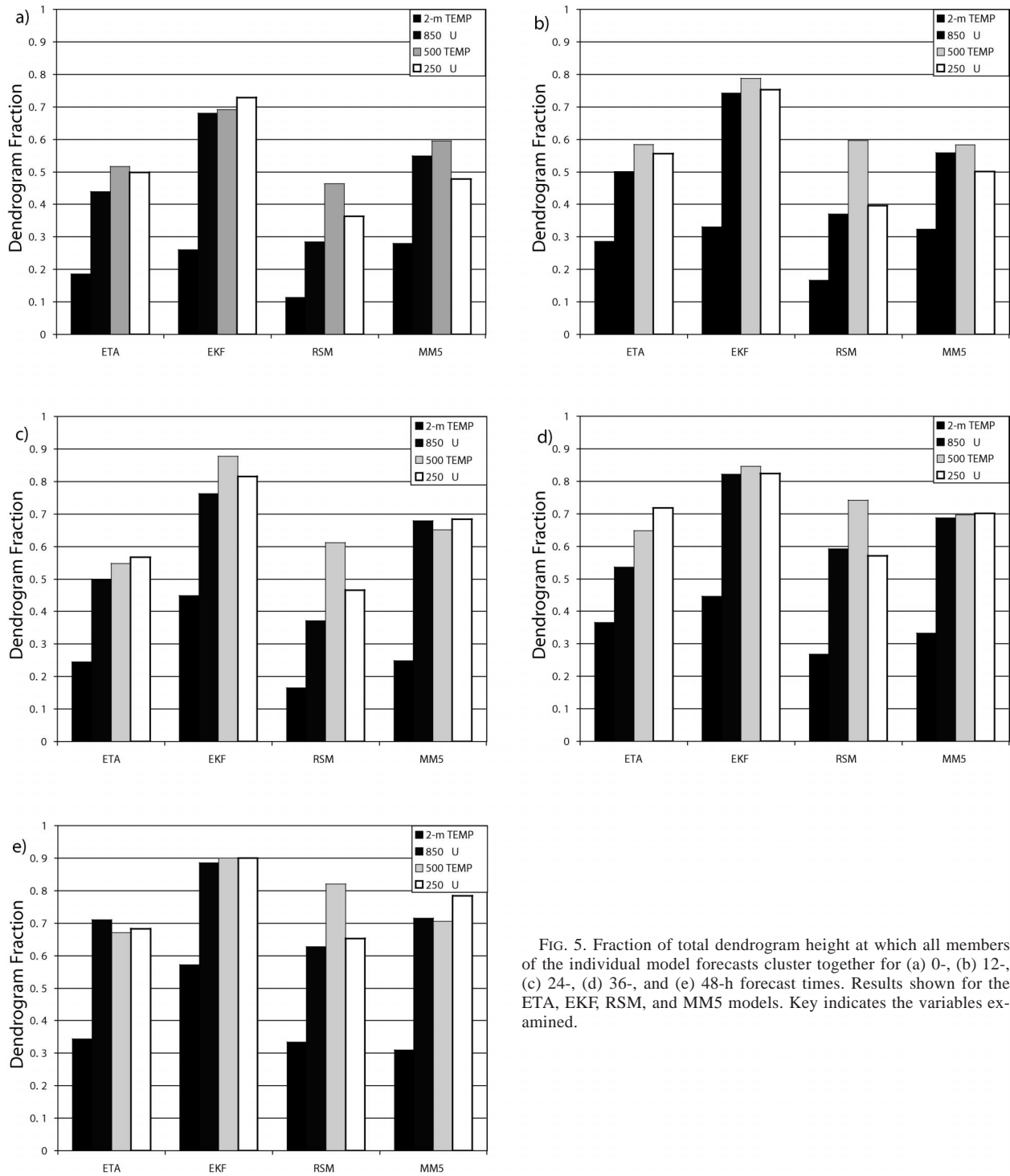


FIG. 5. Fraction of total dendrogram height at which all members of the individual model forecasts cluster together for (a) 0-, (b) 12-, (c) 24-, (d) 36-, and (e) 48-h forecast times. Results shown for the ETA, EKF, RSM, and MM5 models. Key indicates the variables examined.

cast), and the MM5 (five forecasts). Forecasts of 2-m temperature, 850-hPa *u*-component wind speed, 500-hPa temperature, and 250-hPa *u*-component wind speed are bilinearly interpolated to a common grid, and a cluster analysis is conducted at each of the 17 output times for each of the 23 days using the Euclidean distance

dissimilarity measure and Ward's method for hierarchical clustering.

Results from the clustering of all 23 model forecasts indicates that the forecasts largely cluster by model. Of all the models, the RSM members cluster together most frequently and less often form clusters that include

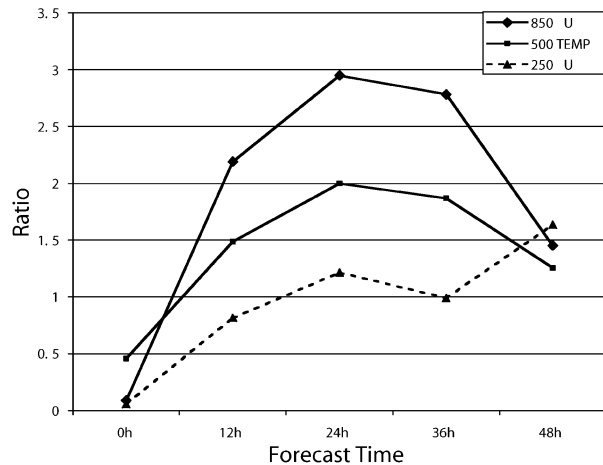


FIG. 6. Ratio of the height of the dendrograms at which the control forecasts from the ETA and EKF cluster to the height at which the five-member ETA forecasts cluster vs forecast time. Value averaged over all 23 cases for the three variables shown. Ratios greater than 1 indicate that the EKF control forecast is outside the envelope of the five-member ETA ensemble started using the bred-mode technique.

members from the other models. Thus, the RSM forecasts tend to be isolated and distinct from the other model forecasts. The ETA and EKF members tend to cluster with their own forecast members first (ETA members cluster together, and EKF members cluster together), and then the ETA and EKF intramodel clusters tend to group together to form one larger intermodel cluster. This joint ETA and EKF clustering occurs over 80% of the time for 2-m temperature, decreasing to slightly over 30% of the time for 250-hPa u -component wind speed. The MM5 forecasts from NSSL also have a strong tendency to cluster together, with nearly 100% of the cases showing the MM5 clustering together for 2-m temperature, and decreasing only to slightly over 65% of the cases for 250-hPa u -component wind speed. However, unlike the RSM, the MM5 members often are in a cluster that contains members from other models, such as the NKF, MAQ, and RUC. Thus, although the MM5 members have a strong tendency to cluster first with themselves, other model forecasts are joined into this cluster.

Closer inspection of the many dendrograms reveals that the MM5 members often cluster at higher heights on the dendrograms than the members of ETA, EKF, and RSM. This result highlights the importance of model physics diversity within a single model, since the MM5 is the only model that has significant model physics diversity contained within the forecasts, and the initial and boundary condition perturbation technique used within the MM5 does not produce modes that grow as quickly as the breeding of growing mode technique (Toth and Kalnay 1993, 1997).

Perhaps the most important comparison is between the control ETA and EKF forecasts that use the identical

initial and boundary conditions. Results indicate that the EKF control forecast quickly moves outside the envelope of forecasts provided by the five-member ETA ensemble that uses the breeding of growing mode technique. This result suggests that model physics uncertainty plays a larger role than initial condition uncertainty in creating a diversity of solutions for this ensemble.

If the goal of ensemble forecasting is to have each model forecast represent an equally likely solution, then these results indicate that we are far from this goal. The model forecasts too often cluster based upon the model that produces the forecasts. Since it is clear that present operational models are far from perfect in reproducing the observed atmospheric structures, the use of ensembles that contain both initial condition and model dynamics and model physics uncertainty are recommended.

Acknowledgments. The authors are thankful to Jun Du, Steve Tracton, Michael Baldwin, Jack Kain, Stan Benjamin, Tracy Lorraine Smith, Steven Peckham, and Georg Grell for providing us with the output from the various forecast models used in this cluster analysis. We further appreciate the local computer support provided by Steven Fletcher, Doug Kennedy, and Brett Morrow. Constructive and helpful reviews from two anonymous reviewers are greatly appreciated. Partial funding for this research was provided under NOAA-OU Cooperative Agreement NA17RJ1227.

REFERENCES

- Alhamed, A., S. Lakshmiarahan, and D. J. Stensrud, 2002: Cluster analysis of multimodel ensemble data from SAMEX. *Mon. Wea. Rev.*, **130**, 226–255.
- Anderberg, M. R., 1973: *Cluster Analysis for Applications*. Academic Press, 359 pp.
- Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.*, **127**, 1941–1953.
- Benjamin, S. G., 1989: An isentropic meso- α scale analysis system and its sensitivity to aircraft and surface observations. *Mon. Wea. Rev.*, **117**, 1586–1605.
- , K. J. Brundage, P. A. Miller, T. L. Smith, G. A. Grell, D. Kim, J. M. Brown, and T. W. Schlatter, 1994: The Rapid Update Cycle at NMC. Preprints, *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 566–568.
- , and Coauthors, 2001: The 20-km version of the RUC. Preprints, *14th Conf. on Numerical Weather Prediction*, Fort Lauderdale, FL, Amer. Meteor. Soc., J75–J79.
- Betts, A. K., and M. J. Miller, 1993: The Betts–Miller scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 46, Amer. Meteor. Soc., 107–121.
- Black, T. L., 1994: The new NMC mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting*, **9**, 265–278.
- Breckenridge, J. N., 1989: Replicating cluster analysis: Method, consistency, and validity. *Multivar. Behav. Res.*, **24**, 147–161.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.
- Burk, S. D., and W. T. Thompson, 1989: A vertically nested regional numerical weather prediction model with second-order closure physics. *Mon. Wea. Rev.*, **117**, 2305–2324.

- Chen, F., and Coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101**, 7251–7268.
- Du, J., G. DiMego, S. Tracton, and B. Zhou, 2003: NCEP short-range ensemble forecasting (SREF) system: Multi-IC, multi-model, and multi-physics approach. Research Activities in Atmospheric and Oceanic Modelling, J. Cote, Ed., CAS/JSC Working Group Numerical Experimentation (WGNE), Rep. 33, WMO Tech. Doc. 1161, 5.09–5.10.
- Dudhia, J., 1993: A nonhydrostatic version of the Penn State–NCAR Mesoscale Model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493–1513.
- Evans, R. E., M. S. J. Harrison, R. J. Graham, and K. R. Mylne, 2000: Joint medium-range ensembles from the Met. Office and ECMWF systems. *Mon. Wea. Rev.*, **128**, 3104–3127.
- Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky, 2000: Model consensus. *Wea. Forecasting*, **15**, 571–582.
- Gong, X., and M. B. Richman, 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *J. Climate*, **8**, 897–931.
- Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 764–787.
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560.
- Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: Two transplant case studies. *Quart. J. Roy. Meteor. Soc.*, **125**, 2487–2516.
- Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX'98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91.
- Jain, A. J., and R. C. Dubes, 1988: *Algorithms For Clustering Data*. Prentice Hall, 320 pp.
- 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.
- Juang, H.-M., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.*, **122**, 3–26.
- Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.
- , M. E. Baldwin, P. Janish, and S. J. Weiss, 2001: Utilizing the Eta Model with two different convective parameterizations to predict convective initiation and evolution at the SPC. Preprints, *Ninth Conf. on Mesoscale Processes*, Fort Lauderdale, FL, Amer. Meteor. Soc., 91–95.
- McIntyre, R. M., and R. K. Blashfield, 1980: A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivar. Behav. Res.*, **2**, 225–238.
- Morey, L. C., R. K. Blashfield, and H. A. Skinner, 1983: A comparison of cluster analysis techniques within a sequential validation framework. *Multivar. Behav. Res.*, **18**, 309–329.
- Pan, H.-L., and W.-S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC Medium-Range Forecast model. NMC Office Note 409, 40 pp. [Available from NCEP, 5200 Auth Rd., Washington, DC 20233.]
- Richardson, D. S., 2001: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.*, **127**, 1847–1864.
- Romesburg, C. H., 1984: *Cluster Analysis For Researchers*. Life Time Learning, 334 pp.
- Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground temperature and surface fluxes. *Mon. Wea. Rev.*, **125**, 1870–1884.
- , —, —, and D. Kim, 2000: Parameterization of cold-season processes in the MAPS land-surface scheme. *J. Geophys. Res.*, **105** (D3), 4077–4086.
- Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 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 multimodel ensemble system. *Mon. Wea. Rev.*, **129**, 729–747.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.
- Zhang, D.-L., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer—Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594–1609.
- Ziehmann, C., 2000: Comparison of a single-model EPS with a multimodel ensemble consisting of a few operational models. *Tellus*, **52A**, 280–299.