

The Dependence of Ensemble Dispersion on Analysis–Forecast Systems: Implications to Short-Range Ensemble Forecasting of Precipitation

STEVEN L. MULLEN AND JUN DU*

Department of Atmospheric Sciences, The University of Arizona, Tucson, Arizona

FREDERICK SANDERS

Marblehead, Massachusetts

9 February 1998 and 3 July 1998

ABSTRACT

The impact of differences in analysis–forecast systems on dispersion of an ensemble forecast is examined for a case of cyclogenesis. Changes in the dispersion properties between two 25-member ensemble forecasts with different cumulus parameterization schemes and different initial analyses are compared. The statistical significance of the changes is assessed.

Error growth due to initial condition uncertainty depends significantly on the analysis–forecast system. Quantitative precipitation forecasts and probabilistic quantitative precipitation forecasts are extremely sensitive to the specification of physical parameterizations in the model. Regions of large variability tend to coincide with a high likelihood of parameterized convection. Analysis of other model fields suggests that those with relatively large energy in the mesoscale also exhibit highly significant differences in dispersion.

The results presented here provide evidence that the combined effect of uncertainties in model physics and the initial state provides a means to increase the dispersion of ensemble prediction systems, but care must be taken in the construction of mixed ensemble systems to ensure that other properties of the ensemble distribution are not overly degraded.

1. Introduction

An ensemble forecast (EF), as originally proposed by Epstein (1969a) and Leith (1974) for weather prediction, denotes the running of multiple forecasts starting from the same time with different but equally plausible initial states consistent with analysis uncertainty. The definition has expanded today to include ensembles run as a function of either plausible uncertainties in the initial state and/or model formulations (e.g., Brooks et al. 1995). A direct outcome from any EF is probabilistic information on predictand uncertainties obtained from the dispersion of ensemble members (Tracton et al. 1996).

Mesoscale predictability and the merit of short-range (1–2 day) ensemble forecasting, or SREF, especially applied to quantitative precipitation forecasts (QPFs) and

probabilistic QPFs (PQPFs), are areas of high interest (Anthes 1986; Brooks et al. 1995; Fritsch et al. 1998). Extreme sensitivity to initial condition uncertainty (ICU) marks QPF. Du et al. (1997, hereafter DMS) find for a case of explosive cyclogenesis that ICU can lead to variations in 6-h QPFs ranging from zero to 3.5" at individual grid points by +12 h, despite relatively little dispersion of the sea level pressure field as judged from cyclone position and central pressure (Mullen and Du 1994). Results from cloud ensemble models (CEMs) are even more spectacular, with differences in temperature of $\pm 1^\circ\text{C}$ or in specific humidity of $\pm 1 \text{ gm kg}^{-1}$, well within realistic estimates of analysis error (Daley and Mayer 1986), making the difference between no storm and a severe thunderstorm (e.g., Brooks et al. 1993; Crook 1996). These studies imply very short predictability limits for QPF and convectively produced precipitation. In the presence of such sensitivity to ICU, and presumably short deterministic limits, it is not surprising that a PQPF from an SREF would be more reliable than a single forecast at higher resolution. In fact, recent results from pilot studies by DMS and Hamill and Colucci (1998) indicate just that.

While these studies establish the potential of SREF with perturbed initial conditions applied to PQPF, Stens-

* Current affiliation: National Centers for Environmental Prediction, Washington, D.C.

Corresponding author address: Dr. Steven L. Mullen, Department of Atmospheric Sciences, PAS Building #81, The University of Arizona, Tucson, AZ 85721.
E-mail: mullen@atmos.arizona.edu

rud and Fritsch (1994a,b) demonstrate that precipitation simulations in a convective environment are strongly impacted by both ICU and changes in physical parameterizations. They argue that SREF should consider the impacts of both sensitivities. An important question not addressed by prior studies is how ICU sensitivity varies with changes in model physics, or in the case of the operational environment varies with changes in the analysis–forecast system. Consider the practical consequences of the hypothetical situation where one candidate SREF system exhibits a negligible bias but a relatively large dispersion, as judged from verification of rank statistics (e.g., Anderson 1996; Hamill and Colucci 1997), while the second candidate has a smaller dispersion but a larger bias. Because of its larger dispersion, the first system requires larger ensemble sizes to obtain reliable estimates of probability density functions (PDFs) and moments thereof. Yet because of computational constraints and the relative ease of mitigating biases through postprocessing, one might consider implementing the second system.

In this note we examine the dependence of ensemble dispersion on the specification of the analysis–forecast system. A limited-area model, with lateral boundary conditions supplied by forecasts from a global model, is used to produce two ensemble forecasts for a case of cyclogenesis. The storm occurred on 14–15 December 1987 over the midwestern United States (see Schneider 1990; Mass and Schultz 1993; Powers and Reed 1993 and DMS for synoptic overviews), where the dense surface rain gauge network allows for an optimal verification of precipitation. Because of their importance and difficulty, we focus as before (DMS) on evaluating forecasts of precipitation accumulated during 6-h periods. We also briefly examine the behavior for a limited number of other model fields and parameters.

2. Methodology

a. Model descriptions and perturbation design

Two numerical forecast models are employed in this study: the global National Center for Atmospheric Research (NCAR) Community Climate Model version 1 (CCM1) and the limited-area Pennsylvania State University–NCAR Mesoscale Model version 4 (MM4). The following is only a brief description of each model including the options we selected. For general information on the models, readers are referred to Williamson et al. (1987) for the CCM1 and to Anthes et al. (1987) and Zhang et al. (1988) for the MM4.

The version of CCM1 used in this study has 12 vertical layers. The global CCM1 uses the spectral transform method to compute horizontal derivatives and perform linear operations. The spectral resolution is triangular 42 (T42), which is roughly equivalent to a uniform 2.8° latitude–longitude gridpoint resolution. The model includes the following parameterized physical

processes: convection, condensation, shortwave and longwave radiative transfers, surface fluxes of heat, moisture and momentum, and interaction with subgrid-scale motions through diffusion. The model forecasts start at 1200 UTC 14 December 1987 and are run for 36 h. Initial analyses for the CCM1 forecasts are obtained by bilinearly interpolating global analyses on a uniform 2.5° latitude–longitude that are produced by the National Meteorological Center (NMC, now the National Centers for Environmental Prediction or NCEP) to the model grid.

Like the global CCM1, the MM4 is run for 36 h from 1200 UTC 14 December 1987. Fifteen vertical layers are used for the MM4 forecasts. The grid spacing is 80 km with mesh of 126 × 101 points, a horizontal resolution and domain size comparable to the operational Nested Grid Model (e.g., Hoke et al. 1989) and the pilot Eta Model ensembles (e.g., Hamill and Colucci 1998; Stensrud et al. 1999). To simulate the operational environment and allow for unbounded predictability error growth (e.g., Errico and Baumhefner 1987; Warner et al. 1997), the lateral boundary conditions for the MM4 are provided by the CCM1 forecasts and updated every 3 h assuming linear variation between updates. The parameterizations of the surface and planetary boundary layers follow Blackadar (1979) and Zhang and Anthes (1982). One ensemble uses a combination of an explicit moisture scheme (Hsie et al. 1984) and an Arakawa–Schubert (1974) cumulus parameterization, as modified by Grell et al. (1991, hereafter referred to as the GEX ensemble). The other uses just a Kuo (1974) cumulus scheme, as modified by Anthes (1977, hereafter referred to as the KUO ensemble) and no explicit scheme. Initial analyses for the GEX ensemble are obtained by first interpolating from the NMC 2.5° latitude–longitude grids to the MM4 grid, then performing a modified Cressman (1959) objective analysis (see Manning and Haagensen 1992) using the NMC values as a first guess. The Cressman step was not done for the KUO ensemble, a choice that yields somewhat smoother analyses over data rich North America with a poorer fit to the radiosonde observations. The use of slightly different basic states upon which to superimpose initial perturbations is patterned after the operational environment, where different analysis–forecast systems always produce different initial analyses because data assimilation depends upon the first-guess forecast and model configuration.

An ensemble is obtained for each model configuration by integrating 25 forecasts starting from slightly perturbed initial conditions. An unperturbed forecast and 24 perturbed ones are run for each ensemble. Rather than choosing perturbations that are dynamically conditioned such as singular vectors (e.g., Molteni et al. 1996) or bred modes (Toth and Kalnay 1993), we randomly generate perturbation fields that represent equally probable estimates of truth and are consistent with prior estimates of analysis uncertainty (e.g., Daley and Mayer 1986; Augustine et al. 1991). Anderson (1997) argues

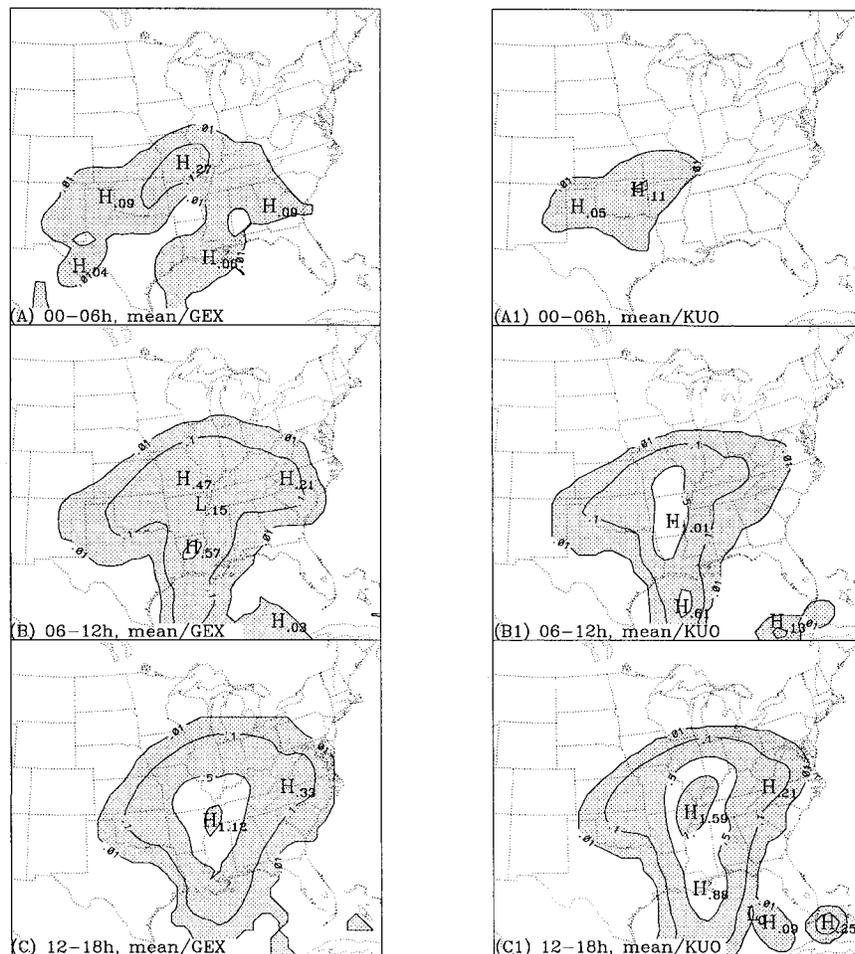


FIG. 1. Spatial distribution of the ensemble mean for 6-h quantitative precipitation forecasts for the GEX ensemble (left panels) and the KUO ensemble (right panels) for each six 6-h period [(a)–(f) in chronological order]. Contour lines are 0.01", 0.1", 0.5", and 1.0", etc. The shading denotes areas of 0.01–0.5", 1.0–2.0", 3.0–4.0", etc.

that unconstrained perturbations can have a number of advantages over dynamically constrained perturbations, such as yielding unbiased estimates of higher moments and PDFs. The method of perturbing initial fields is essentially the same as that used by Errico and Baumhefner (1987), Mullen and Baumhefner (1989, 1994), and DMS. The same perturbation fields are added to the CCM1, GEX, and KUO initial analyses. Because the perturbations are unbalanced, nonlinear normal mode initialization (Errico 1983) is used to remove most of the energy associated with inertial–gravitational modes. Thus, the final, initialized perturbations project strongly on the slow manifold of the models.

The experimental design crudely mimics the type of differences that might arise in the operational environment between two forecast ensembles based on different regional analysis–forecast systems with identical resolutions, time varying lateral boundary conditions, and perturbation strategies. Except for the use of different MM4 cumulus parameterizations and the slightly dif-

ferent basic states upon which the same perturbations (prior to initialization) are superimposed, the two forecast ensembles are identical.

b. Assessment of statistical significance

The statistical problem is to determine whether the use of a different analysis–forecast system yields a significant change in the spatial distributions of the ensemble mean and the standard deviation of forecast fields. The problem can be tested using the Pool-Permutation Procedure (PPP) of Preisendorfer and Barnett (1983) and their SITES and SPREAD measures, metrics that quantify the separation of ensemble means and the ensemble variances, respectively. Their procedure, which is based on Monte Carlo permutation techniques, can be used to estimate the statistical significance of differences in spatial distributions of fields that are intercorrelated, so-called field or global significance. The use of Monte Carlo techniques to estimate field significance

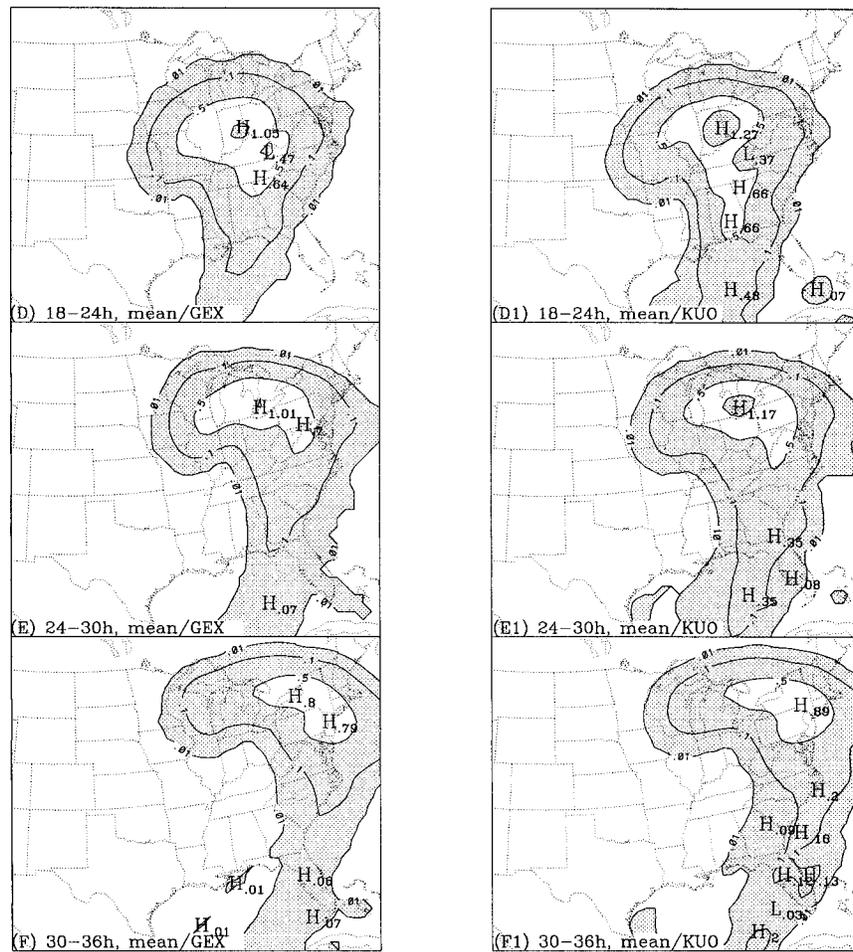


FIG. 1. (Continued)

of geophysical datasets has two primary advantages over parametric tests such as the Student's *t*-test (e.g., Livezey and Chen 1983): it requires neither an a priori assumption of the background distribution of the sample populations (e.g., equal variances), nor an a priori estimate of the number of spatial degrees of freedom.

The PPP technique first pools together all members from both ensembles. Two, 25-member ensembles are then randomly selected from the pool. The SITES (SPREAD) value is next computed based on that sampling from the pool. The randomization procedure is repeated 1000 times, and an empirical distribution of the SITES (SPREAD) values, that is, a background cumulative distribution function (Wilks 1995) or CDF, is formed from the permutations. The SITES (SPREAD) value for the 25-member Grell and Kuo ensembles is finally compared against the background CDF. If the SITES (SPREAD) value for the Grell–Kuo ensemble exceeds the 95% value of the CDF, then the Grell–Kuo means (variances/standard deviations) are deemed different at the 5% significance level, the probability of incorrectly rejecting the null hypothesis. Here, we test

the null hypothesis that two ensemble mean distributions (or standard deviations) are equal. In this study, an a priori significance level of 5% is selected as an acceptable probable error of rejecting the null hypothesis by just a fortuitous random sampling of ensemble states. We apply the testing procedure over the spatial domain shown in Fig. 1.

3. Results

a. Precipitation

Figure 1 shows spatial distributions of the ensemble mean of 6-h accumulated precipitation for the GEX (left panels) and the KUO (right panels) forecast ensembles. While the two ensembles exhibit a comparable total area covered by measurable precipitation, there are obvious differences between them. In particular, the KUO forecasts average more rainfall than the GEX forecasts at all times except the first 6 h. The PPP method applied to the SITES values (results not shown) indicates that differences between the distributions of the ensemble

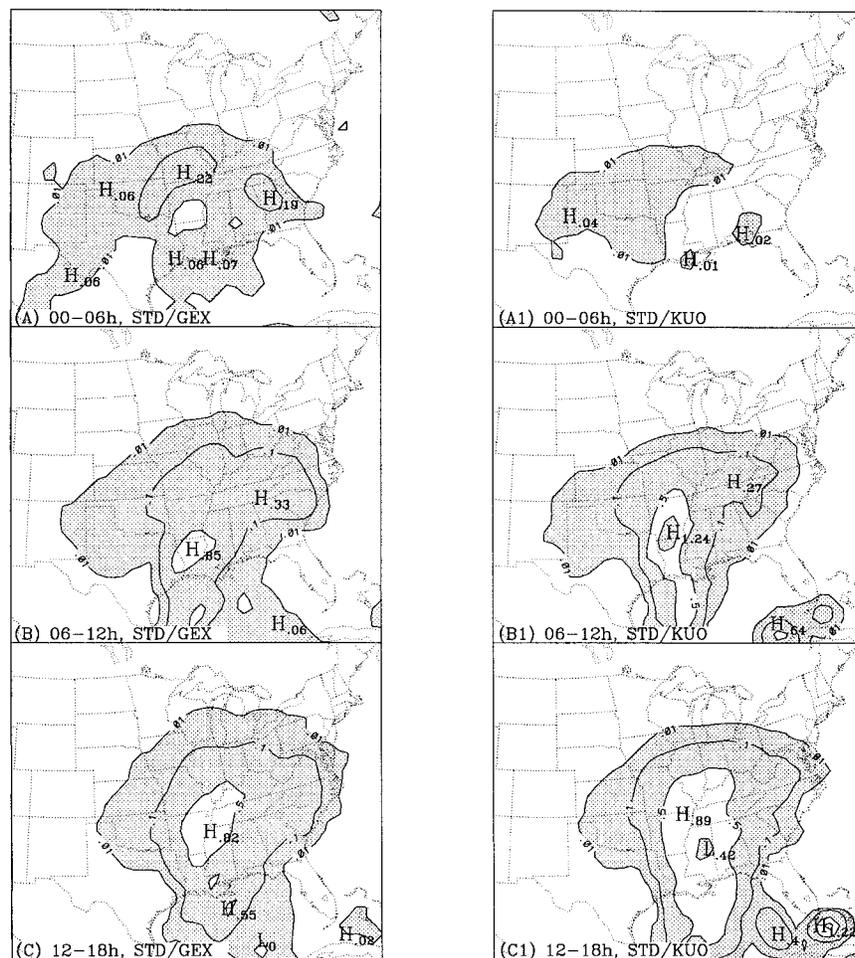


FIG. 2. As in Fig. 1 except for standard deviation.

mean are significant at the 0.1% confidence level at all projections. A high level of significance is not surprising, and is consistent with a plethora of prior studies (e.g., Orlandi and Katzfey 1987; Kuo and Reed 1988; Mullen and Baumhefner 1988; Kuo and Low-Nam 1990, among others) that show model simulations of explosive cyclogenesis can sometimes be extremely sensitive to physics parameterizations in terms of cyclone position and central pressure.

Spatial distributions of the standard deviation of 6-h accumulated precipitation are given in Fig. 2. Both ensembles exhibit a tendency for regions with high variance to coincide with areas with a large mean, a result previously noted in ensemble forecasts of precipitation (Hamill and Colucci 1998). Consistent with that notion, and of particular importance, is the relatively large size of the KUO variance: except for first 6-h period of model "spin up," the KUO values exceed the GEX values in virtually all locations. The SPREAD values (results not shown) indicate that differences between the spatial distributions are also significant at the 0.1% confidence level at all projections. It follows that the forecast dis-

persion and the predictability error growth for the KUO ensemble are significantly faster than for the GEX ensemble, at least for this case of cyclogenesis and choice of model configurations.

These differences in the dispersion characteristics are reflected in extrema for the two ensembles, as shown by volumetric 6-h accumulations (results not shown) and cumulative storm totals (Fig. 3). The data reveal wider spreads for the KUO group after 6 h. The larger spread leads to the desirable outcome of the KUO ensemble encompassing the verification for a longer forecast period than GEX ensemble, but not without a severe penalty: the accuracy of the KUO PPDFs fares worse than the GEX PPDFs at all times. Biases, equitable threat scores (Schaefer 1990), and Ranked Probabilities Scores (Epstein 1969b; Murphy 1971) for five mutually exclusive, collectively exhaustive categories¹ are less

¹ The five categories are no measurable precipitation ($pp < 0.01''$), $0.01'' \leq pp < 0.10''$, $0.10'' \leq pp < 0.50''$, $0.50'' \leq pp < 1.00''$, and $pp \geq 1.00''$.

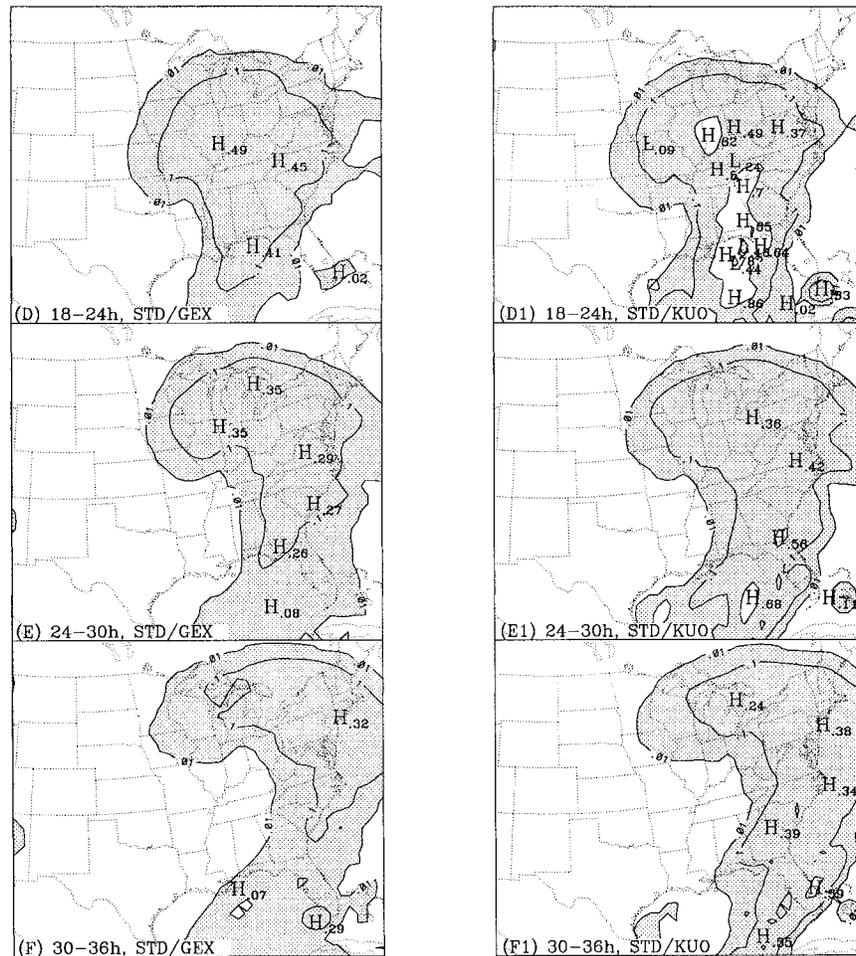


FIG. 2. (Continued)

accurate for the KUO ensemble at all projections (results not shown).

Figure 4 shows the probability p_c that the 6-h precipitation from the cumulus parameterization scheme exceeds 0.01" for the 25-member GEX ensemble. As noted earlier, prior studies with CEMs (Brooks et al. 1993; Crook 1996) show that ensemble forecasts of convection can be very sensitive to ICU. A visual comparison of Figs. 2 and 4 indeed reveals that a high likelihood of convective precipitation tends to coincide with large variance for this case, especially for the 18-h period of 6–24 h when forecast convection is most frequent and 6-h precipitation is largest. To quantify this relationship, we computed a spatial correlation coefficient ρ between (1) the probability of measurable convective rainfall (Fig. 4) and (2) a "chi-square" statistic χ^2 based on probabilities for the same categories used to compute the RPS values

$$\chi^2 = \sum_{i=1}^J p_i(i)^2,$$

where $p_i(i)$ is the probability for rainfall category i and

$J = 5$ is the total number of categories. Note that the limiting values of χ^2 are $1/J$ for a flat distribution (all categories equally populated) and 1.0 for a spike distribution (one category contains all members), so a strong relationship between large spread and high likelihood of convective precipitation would be reflected as large negative correlation. The correlation was computed only over the area where the probability of measurable rainfall exceeded zero (e.g., Fig. 18 of DMS), a more stringent criterion than computing ρ over the entire domain of Fig. 4, which gives much stronger correlations. Values of ρ varied between -0.5 and -0.6 during the 36-h forecast period, with the strongest correlations during the 18-h period of 6–24 h. If the probability fields (Fig. 4) are correlated against the spread (Fig. 2), ρ strengthens to 0.6 to 0.8. A value of $\rho \leq -0.5$ differs significantly from zero at the 5% level based on a one-tailed test of the Student's t-distribution (Spiegel 1975, p. 267) under the conservative assumption of a total of 10 spatial degrees of freedom. This indicates that a significant correlation exists between the dispersion of the precipitation forecasts and the likeli-

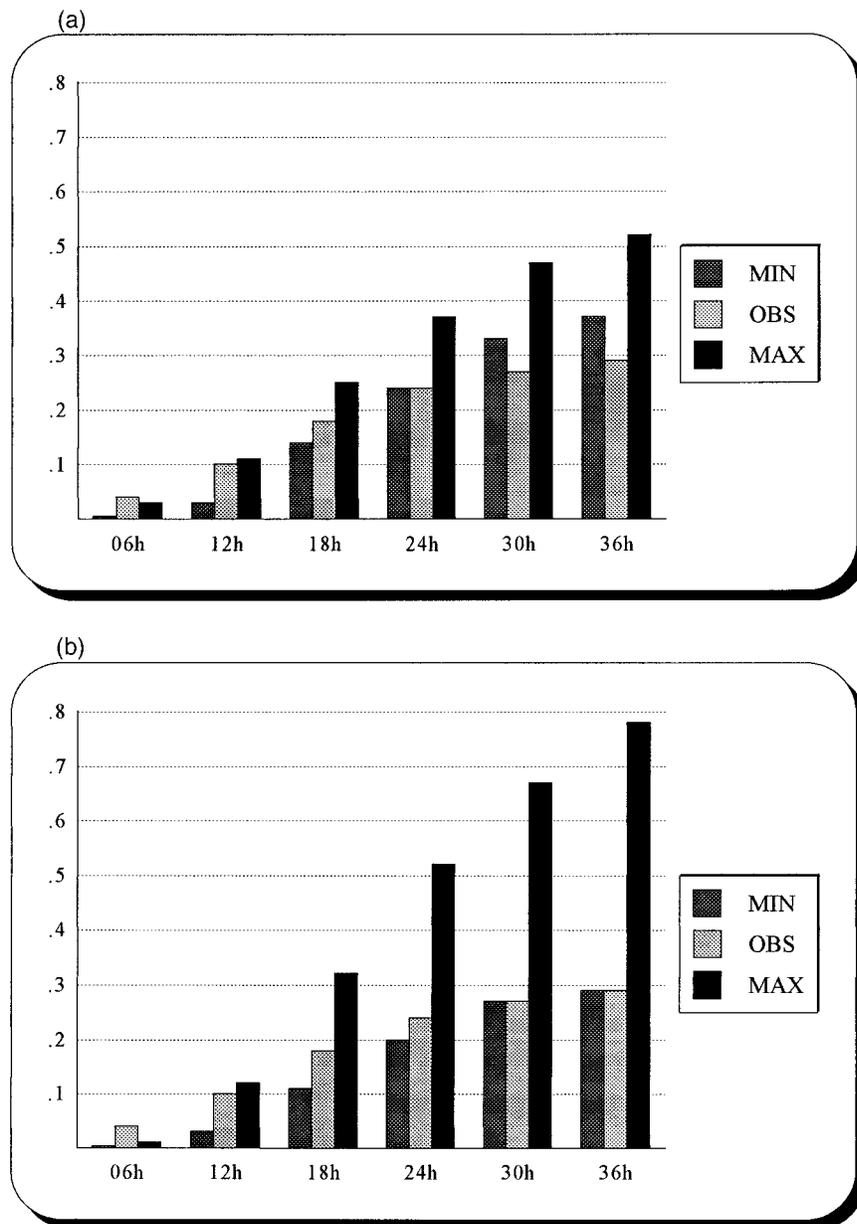


FIG. 3. Spatially averaged, storm total precipitation of forecast extrema for the (a) GEX and (b) KOU ensembles along with the verifying amount.

hood of parameterized convection being activated in the GEX ensemble. It is also consistent with the notion that ensemble forecasts of convectively produced precipitation are extremely sensitive to ICU.

We also examined the distribution of convective precipitation for the KOU scheme and found that no members produced a 6-h accumulation greater than 0.01" at all forecast projections. Wang and Seaman (1997) compared the behavior of six different convective parameterization schemes in a mesoscale model at grid spacings of 36 km and 12 km, much finer than our 80-km

grid. They found that the partitioning of rainfall into subgrid scale and grid resolvable scale was very sensitive to the particular scheme. They noted that the ratio of subgridscale to total rainfall could vary between one and zero for 1-h accumulations, but unlike our results, their implementation of a Kuo–Anthes scheme yielded on average the highest ratios. While differences in the synoptic events and model configurations of Wang and Seaman (1997) prevent meaningful detailed or quantitative comparisons with our results, it does appear that the partitioning of rainfall between grid scale and sub-

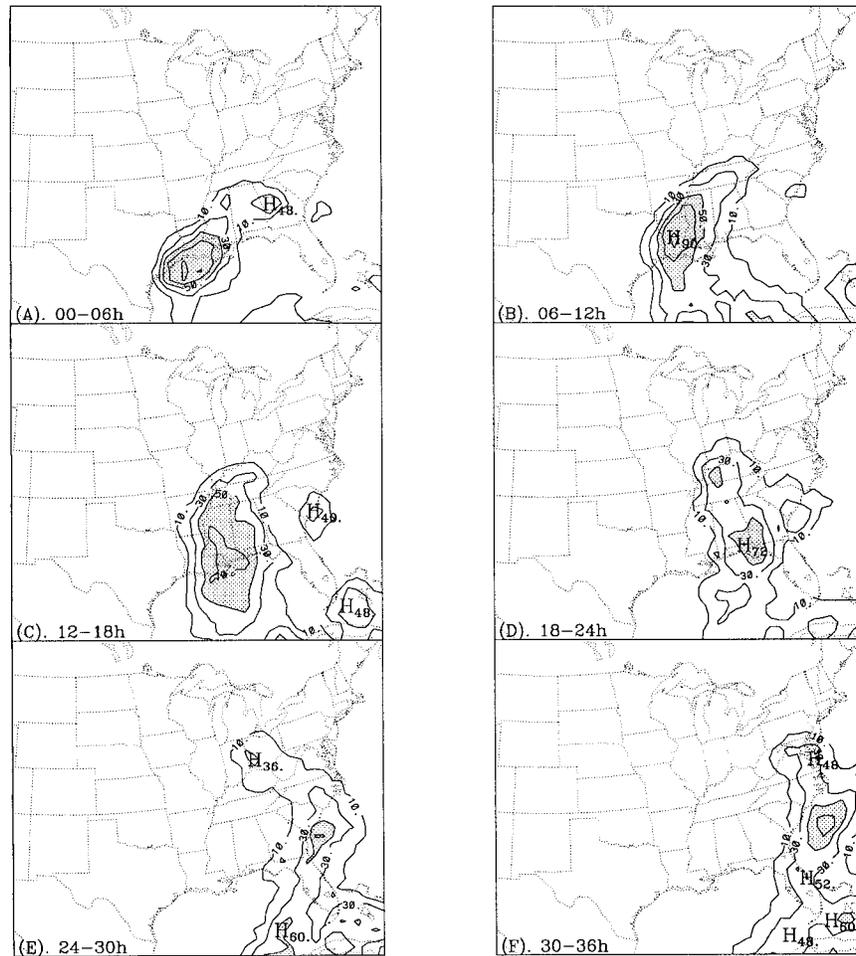


FIG. 4. Probability that the 6-h accumulated precipitation from convective parameterization for the GEX ensemble exceeds 0.01" for each 6-h period [(a)–(f) in chronological order]. Contour lines are 10%, 30%, 50%, 70%, and 90%. Shading denotes regions $\geq 50\%$.

grid scale is an issue that warrants careful attention in the construction of operational ensemble prediction systems (EPSs) that contain cumulus parameterizations.

b. Other fields and parameters

Lorenz (1982) estimates that an upper bound on the predictability of the instantaneous 500-mb geopotential height field, a variable with peak variance in the planetary and large synoptic scales, is about 10 days. He concludes, however, that his predictability estimates for 500-mb geopotential height “may be quite unrealistic for such elements as . . . rainfall.” Our results for the cyclogenesis of 14–15 December 1987 support his conjecture, and the large spreads relative to the ensemble means and flat probability distributions are consistent with much shorter limits for parameterized convective precipitation on an 80-km mesh. Stamus et al. (1992) show that the limits for second-order quantities with relatively greater power at much smaller scales, pre-

sumably those that are most important to forcing convective precipitation (e.g., vertical velocity, stability indices, moisture flux divergence), are also much shorter than 10 days. Although a comprehensive analysis of the predictability characteristics for the two forecast ensembles is beyond the scope of this note, it is also of interest to examine forecast dispersion for secondary variables of operational importance or dynamic/thermodynamic consequence. For this reason, we analyzed the SITES and SPREAD measures for the limited number of fields, pressure levels, and forecast projections shown in Table 1.

The SITES values for the ensemble means (results not shown) indicate significance at the 0.1% level for all fields at all forecast projections. We find this result not surprising. What we find more intriguing is an apparent relationship in the significance of the SPREAD (Table 1) among the various fields. About 85% (18/21) of the parameters of Table 1, which explicitly appear in a water vapor balance and/or the continuity equation for

TABLE 1. Significance levels of difference between the GEX and Kuo ensemble dispersion, as judged from the SPREAD measure and Pool-Permutation Procedure of Preisendorfer and Barnett (1983), for the fields and forecast projections indicated below. The label "NS" indicates not significant at the 5% level, while the label "NA" indicates that the field was not available/not saved and thus its significance level was not computed.

Field	06 h	12 h	18 h	24 h	30 h	36 h
300-mb vorticity	0.1%	NS	NS	NS	5.0%	5.0%
300-mb divergence	0.1%	0.1%	0.1%	0.1%	0.1%	5.0%
500-mb height	NS	1.0%	NS	5.0%	NS	NS
850-mb moisture flux div.	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
850-mb specific humidity	NS	0.5%	0.1%	5.0%	NS	NS
Lowest σ -level vorticity	NA	5.0%	NA	NS	NA	NS
Lowest σ -level divergence	NA	0.1%	NA	0.1%	NA	0.1%
Sea level pressure	NS	NS	0.5%	NS	NS	NS

water vapor (specific humidity, divergence, moisture flux divergence), exhibit significant differences. In contrast, only a third (7/21) of the others (vorticity, sea-level-pressure, 500-mb height) achieve significance at the 5% level. The difference between the proportions is significant at the 1% (5%) level with only eight (three) degrees of freedom (Spiegel 1975, p. 215). Evidently the dispersion of the moisture and divergence fields for these cyclogenesis forecasts is more sensitive to the changes in the analysis–forecast system than the dispersion for the rotational wind and mass fields.

We also examined the variability of the central pressure and position forecasts for the surface cyclone (results not shown). Similar to the QPF results, variability is largest among the Kuo members, especially for cyclone position. Standard deviations run, on average, around 110 km for Kuo, or 70% larger than for GEX. In terms of area confidence bounds for cyclone position (Mullen and Du 1994), this equates to a factor of 3 increase in area for the same level of confidence for the Kuo ensemble.

4. Discussion

Our results can also be interpreted in terms of the reliability test discussed by Tribbia and Baumhefner (1988). They point out that analysis–forecast systems contain two error sources, that due to model deficiencies (the external error source) and that due to the growth of initial data errors (the internal error source). Because the two sources are relatively independent early in the forecast while the error growth is linear or only weakly nonlinear, they propose the use of the internal error growth to judge the significance of any source of external error growth and vice versa. Whenever external error due to a change in the model formulation exceeds the internal error by a certain level (see Fig. 1 of Tribbia and Baumhefner and associated discussion), then the model change can be readily identified against the background "noise" of classic predictability error growth (e.g., Lorenz 1982). Interpreted in terms of their test, the strong significance (0.1% level) of the SITES values by +6 h indicates that the differences due to our changes in cumulus parameterizations (Fig. 1) can also be readily

measured against forecast dispersion owing to initial data error (Fig. 2) at short projections. It further suggests that uncertainties in specification of cumulus parameterizations denote a more important error source for QPF than growth of initial data error, at least for this case and experimental design.

Because the SITES and SPREAD metrics combine the effects of differences in the magnitude and spatial distribution of errors, what cannot be gleaned from them is information on the relative importance of the two effects. To explore this issue, we show CDFs for the root-mean-square (rms) differences for 6-h accumulated precipitation (Fig. 5) and SLP (Fig. 6), computed over the same area as SITES and SPREAD. The Tribbia–Baumhefner reliability test can also be applied to these figures to determine the impact of differences in spatially averaged magnitude only. Comparison of the rainfall CDFs (Fig. 5) indicates that rms differences among individual forecasts ending at 18 h, 24 h, and 30 h do pass the reliability test at the 5% level, but not at the 1% level, whereas the SITES values are significant at the 0.1% confidence level for all times. It follows that differences in the spatial distribution must be playing a crucial role in the high level of significance of SITES. On the other hand, the SLP fields (Fig. 6) are not close to being significant at any projection.

This example shows that assessment of field significance, when applied to the evaluation of ensemble forecasts, augments the Tribbia–Baumhefner reliability test by including the impact of the spatial distribution of the error. We believe that the estimation of field significance can play a useful role in the verification and validation of EPSs, the estimation of predictability limits, and the interpretation of forecast dispersion.

The fact that significant differences occur between the forecast dispersions for the two ensembles has implications on estimates of mesoscale predictability limits. Our results, albeit for only one case, remind us that estimates of predictability error growth depend critically upon the variability characteristics of the model, no matter what the temporal–spatial scale of the phenomenon of interest or forecast projection. These differences also bring to mind the importance of thorough model validation in an era of ensemble forecasting. It is far from

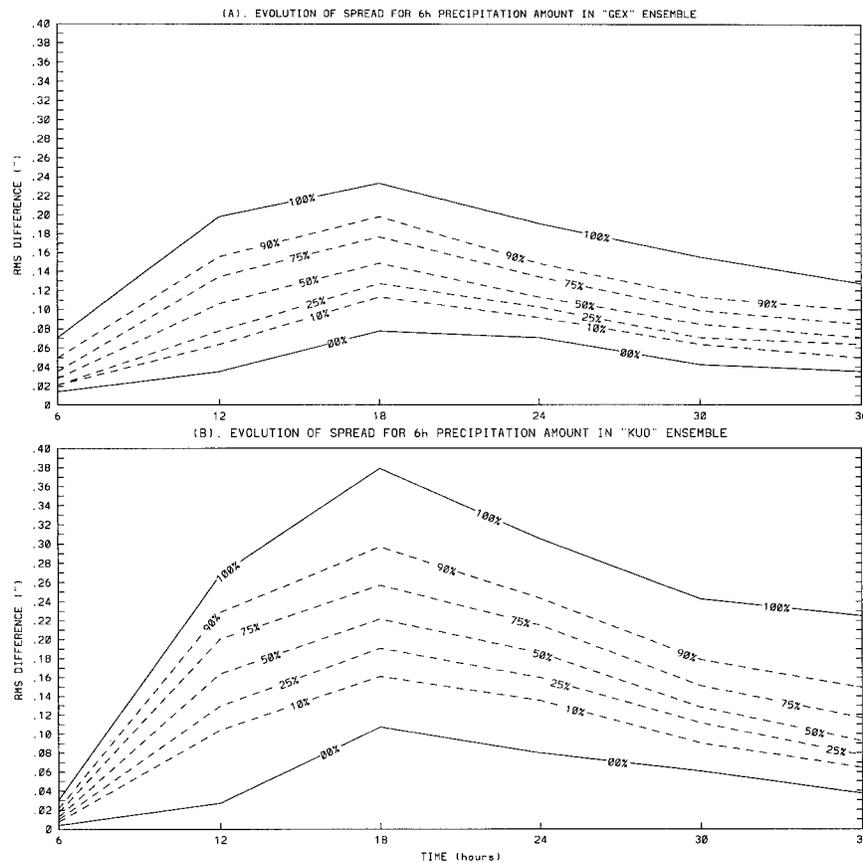


FIG. 5. Cumulative distribution functions for rms differences among individual forecasts of 6-h accumulated precipitation for (a) GEX and (b) KUO ensembles. The 10%, 25%, 50%, 75%, and 90% percentiles are shown as dashed lines, the 0% and 100% as solid lines.

adequate to judge performance of an EPS only in terms of biases or first moments. A viable EPS should produce reliable probability distributions. At a minimum, the EPS model should produce accurate means, variances, and covariances for those scales (both temporal and spatial) and weather phenomena that one wishes to predict, because as Leith (1974) point out, “the mean and variance . . . provides, of course, the lowest two moments which contribute much to the definition of such a probability distribution.” For the same reasons discussed by Palmer (1995) for medium-range and seasonal forecasts, if a candidate mesoscale model for an EPS is to be successful predicting heavy QPF events at projections beyond nowcasting (6–12 h), it should also be able to simulate such events when run as a climate model. Unlike the case of medium-range and seasonal forecasting, however, mesoscale prediction and the operational implementation of EPS faces a far greater challenge in obtaining the observations at the small spatial and short temporal scales needed to verify and improve the models. In the absence of sufficient verifying observations, a distinct possibility for the foreseeable future especially over the data-sparse oceans and for meso- β and meso- γ scales, an operational EPS would have to rely on in-

formation gleaned from much higher resolution models that explicitly resolve processes not resolved by the EPS, such as in the case of using statistics from cloud ensemble model (CEMs) to validate and improve cumulus parameterizations (e.g., Xu 1993).

Our results also provide insight into why the use of “mixed” or “grand” ensembles (Richardson et al. 1996), an ensemble constructed by including predictions from all available operational EPSs, improves performance. A ubiquitous feature of current ensemble prediction systems based solely on perturbed initial fields is insufficient dispersion (e.g., Anderson 1996; Buizza 1997; Hamill and Colucci 1997, 1998): verification lies outside the ensemble distribution more frequently than expected by just chance. Comparison of Figs. 1 and 2 reveals that the total dispersion for a mixed ensemble comprised of all 50 members would be greater than the dispersion for just the GEX or KUO ensembles alone. Analysis of variance (Spiegel 1975, p. 307) indicates that the total forecast variance (i.e., the square of the ensemble dispersion about the ensemble mean) equals the sum of the variance within groups and the variance between groups. It follows that the use of different physics parameterizations increases forecast dispersion as

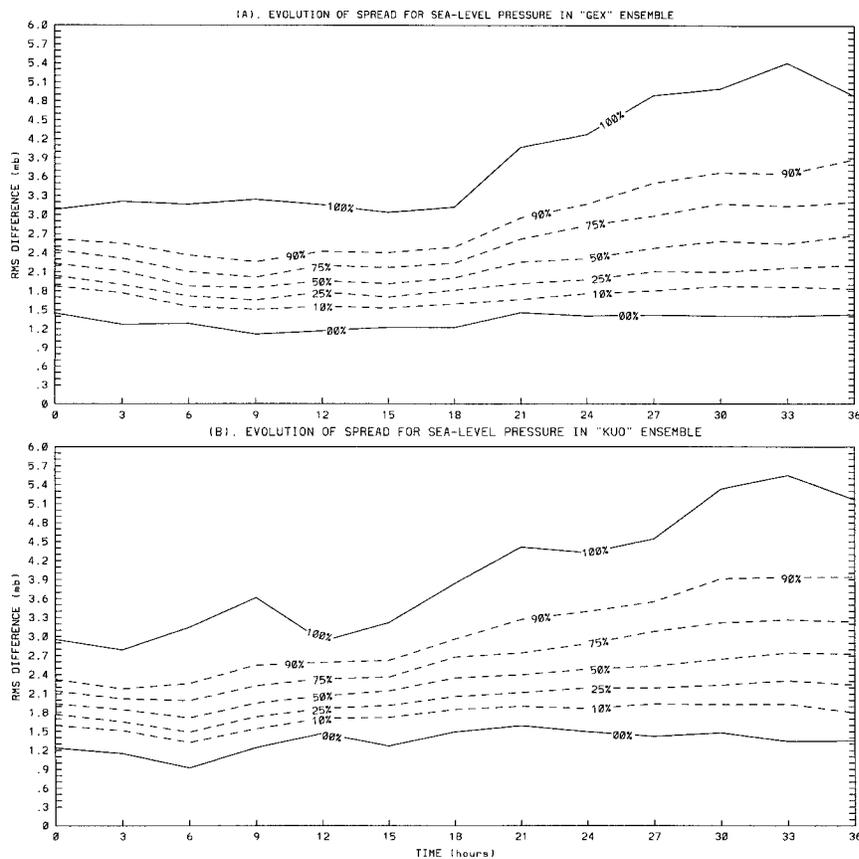


FIG. 6. As in Fig. 5 except for the SLP field.

long as differences among the ensemble means for the individual physics schemes are not zero and the same perturbation strategy is employed for all the schemes. This reasoning can be applied to all models, forecast projections and weather elements, not just the synoptic event and experimental design of this study. It is only for a “perfect ensemble system” (Buizza 1997), one with no model error and initial perturbations that precisely reflect the distribution of analysis error, that one should expect perfect dispersion characteristics.

5. Conclusions

The impact of different analysis–forecast systems on the dispersion of QPFs was examined for a case of cyclogenesis. Significant differences in dispersion were found between two, 25-member ensemble forecasts that used different cumulus parameterization schemes and slightly different basic states upon which the same perturbations were superimposed. QPFs and PQPFs were particularly sensitive to model specification, with regions of large dispersion coinciding with a high likelihood of parameterized convection. Analysis of other model fields suggests that those with relatively large power in the mesoscale also exhibit highly significant differences in dispersion.

Because the results presented here are for only one case, we caution against generalizing our results. We believe that our results, if anything, denote an underestimate of the likely impact of different analysis–forecast systems on dispersion. Recall that we only changed cumulus parameterizations and slightly altered the basic initial state. Inclusion of different model resolutions, boundary layer parameterizations, data assimilation procedures, perturbation strategies, and coupled systems would undoubtedly yield greater differences in dispersion. Clearly more synoptic events need to be examined, and the sensitivity to an array of different analysis–model configurations needs to be considered. Even with these shortcomings in experimental design, our results support the recommendations of prior investigators (Stensrud and Fritsch 1994ab; Brooks et al. 1995; Bresch and Bao 1996; Fritsch et al. 1998) who suggest that the sensitivity to model parameterizations, their tunings or their stochastic representation be considered, especially as a way to increase the dispersion of ensemble prediction systems. However, this study also indicates that care must be taken in the construction of mixed ensemble systems to ensure that other properties of the ensemble distribution are not overly degraded.

Acknowledgments. This project was supported by the

National Science Foundation (NSF) under Grants ATM-9318751 (Sanders), ATM-9419411 (Du, Mullen), and ATM-9612487 (Mullen). The model forecasts were performed at the Scientific Computing Division (SCD) of NCAR. NCAR is supported by NSF.

REFERENCES

- Anderson, J. L., 1996: A method for producing and evaluating probabilistic precipitation forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1529.
- , 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.*, **125**, 2969–2983.
- Anthes, R. A., 1977: A cumulus parameterization scheme utilizing a one-dimensional cloud model. *Mon. Wea. Rev.*, **105**, 963–984.
- , 1986: The general question of predictability. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Amer. Meteor. Soc., 636–656.
- , E.-Y. Hsie, and Y.-H. Kuo, 1987: Description of the Penn State/NCAR Mesoscale Model Version 4 (MM4). NCAR Tech. Note NCAR/TN 282+STR, 66 pp. [Available from National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80303.]
- Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, **31**, 674–701.
- Augustine, S. J., S. L. Mullen, and D. P. Baumhelfner, 1991: Examination of actual analysis differences for use in Monte Carlo forecasting. Preprints, *16th Annual Climate Diagnostics Workshop*, Los Angeles, CA, NOAA/NWS/NMC/CAC, 375–378.
- Blackadar, A. K., 1979: High resolution models of the planetary boundary layer. *Advances in Environmental Science and Engineering*, J. Pfafflin and E. Ziegler, Eds., Vol. 1, Gordon and Beach Science Publications, 50–85.
- Bresch, J. F., and J.-W. Bao, 1996: Generation of mesoscale model ensemble members by varying model physics. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 62–64.
- Brooks, H. E., C. A. Doswell, and L. J. Wicker, 1993: STORMTIPE: A forecasting experiment using a three-dimensional cloud model. *Wea. Forecasting*, **8**, 352–362.
- , M. S. Tracton, D. J. Stensrud, G. J. DiMego, and Z. Toth, 1995: Short-range ensemble forecasting (SREF): Report from a workshop. *Bull. Amer. Meteor. Soc.*, **76**, 1617–1624.
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119.
- Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.
- Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767–1785.
- Daley, R., and T. Mayer, 1986: Estimates of global analysis error from the global weather experiment observational network. *Mon. Wea. Rev.*, **114**, 1642–1653.
- Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459.
- Epstein, E. S., 1969a: Stochastic dynamic prediction. *Tellus*, **21**, 739–759.
- , 1969b: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987.
- Errico, R. M., 1983: A guide to transform software for non-linear normal-mode initialization of the NCAR Community Forecast Model. NCAR/TN-217+IA, 86 pp. [Available from the National Atmospheric Research Center, P.O. Box 3000, Boulder, CO 80303.]
- , and D. P. Baumhelfner, 1987: Predictability experiments using a high-resolution limited-area model. *Mon. Wea. Rev.*, **115**, 488–504.
- Fritsch, J. M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the Eighth Prospectus Development Team, U. S. Weather Research Program. *Bull. Amer. Meteor. Soc.*, **79**, 285–299.
- Grell, G., Y.-H. Kuo, and R. Pasch, 1991: Semiprognostic tests of cumulus parameterization schemes in the middle latitudes. *Mon. Wea. Rev.*, **119**, 5–31.
- Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327.
- , and —, 1998: Evaluation of Eta/RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.*, **126**, 711–724.
- Hoke, J. E., N. A. Phillips, G. J. DiMego, J. J. Tucillo, and J. G. Sela, 1989: The regional analysis and forecast system of the National Meteorological Center. *Wea. Forecasting*, **4**, 323–334.
- Hsie, E.-Y., R. A. Anthes, and D. Keyser, 1984: Numerical simulation of frontogenesis in a moist atmosphere. *J. Atmos. Sci.*, **41**, 2581–2594.
- Kuo, H. L., 1974: Further studies of the parameterization of the effect of cumulus convection on large-scale flow. *J. Atmos. Sci.*, **31**, 1232–1240.
- , and R. J. Reed, 1988: Numerical simulation of an explosively deepening cyclone in the eastern Pacific. *Mon. Wea. Rev.*, **116**, 2081–2105.
- , and S. Low-Nam, 1990: Prediction of nine explosive cyclones over the western Atlantic with a regional model. *Mon. Wea. Rev.*, **118**, 3–25.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Livezey, R. E., and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.
- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513.
- Manning, K. W., and P. L. Haagenson, 1992: Data ingest and objective analysis for the PSU/NCAR modeling system: Programs DATAGRID and RAWINS. NCAR/TN 376+STR+IA, 209 pp. [Available from the National Atmospheric Research Center, P.O. Box 3000, Boulder, CO 80303.]
- Mass, C. F., and D. M. Schultz, 1993: The structure and evolution of a simulated midlatitude cyclone over land. *Mon. Wea. Rev.*, **121**, 889–917.
- Molteni, F., T. N. Palmer, R. Buizza, and T. Petroligias, 1996: The ECMWF ensemble prediction system: Methodology and verification. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–121.
- Mullen, S. L., and D. P. Baumhelfner, 1988: The sensitivity of numerical simulations of explosive oceanic cyclogenesis to changes in physical parameterizations. *Mon. Wea. Rev.*, **116**, 2289–2339.
- , and —, 1989: The impact of initial condition uncertainty on numerical simulations of large scale explosive cyclogenesis. *Mon. Wea. Rev.*, **117**, 2289–2329.
- , and —, 1994: Monte Carlo simulations of explosive cyclogenesis. *Mon. Wea. Rev.*, **122**, 1548–1567.
- , and J. Du, 1994: Monte Carlo forecasts of explosive cyclogenesis with a limited-area, mesoscale model. Preprints, *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 638–640.
- Murphy, A. H., 1971: A note on the Ranked Probability Score. *J. Appl. Meteor.*, **10**, 155–156.
- Orlanski, I., and J. J. Katzfey, 1987: Sensitivity of model simulations for a coastal cyclone. *Mon. Wea. Rev.*, **115**, 2792–2821.
- Palmer, T. N., 1995: The predictability of the atmosphere and oceans: From days to decades. Preprints, *Decadal Climate Variability: Dynamics and Predictability*, Les Houches, France, NATO Advanced Study Institute, 84–181.
- Powers, J. G., and R. J. Reed, 1993: Numerical simulation of the large-amplitude mesoscale gravity-wave event of 15 December 1987 in the central United States. *Mon. Wea. Rev.*, **121**, 2285–2308.

- Preisendorfer, R. W., and T. Barnett, 1983: Numerical model-reality intercomparison tests using small-sample statistics. *J. Atmos. Sci.*, **40**, 1884–1896.
- Richardson, D. S., M. S. J. Harrison, K. B. Robertson, and A. P. Woodcock, 1996: Joint medium-range ensembles using UKMO, ECMWF, and NCEP ensemble systems. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., J26–J28.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.
- Schneider, R. S., 1990: Large-amplitude mesoscale wave disturbances with the intense Midwestern extratropical cyclone of 15 December 1987. *Wea. Forecasting*, **5**, 533–558.
- Spiegel, M. A., 1975: *Probability and Statistics*. McGraw-Hill, 372 pp.
- Stamus, P. A., F. H. Carr, and D. P. Baumhefner, 1992: Application of a scale-separation verification technique to regional forecast models. *Mon. Wea. Rev.*, **120**, 149–163.
- Stensrud, D. J., and J. M. Fritsch, 1994a: Mesoscale convective systems in weakly forced large-scale environment. Part II: Generation of a mesoscale initial condition. *Mon. Wea. Rev.*, **122**, 2068–2083.
- , and —, 1994b: Mesoscale convective systems in weakly forced large-scale environment. Part III: Numerical simulations and implications for operational forecasting. *Mon. Wea. Rev.*, **122**, 2084–2104.
- , H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.*, **127**, 433–446.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- Tracton, M. S., R. Wobus, and Z. Toth, 1996: On the principles and operational utility ensemble prediction. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., J17–J18.
- Tribbia, J. J., and D. P. Baumhefner, 1988: The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. *Mon. Wea. Rev.*, **116**, 2276–2288.
- Wang, W., and N. L. Seaman, 1997: A comparison study of convective parameterization schemes in a mesoscale model. *Mon. Wea. Rev.*, **125**, 252–278.
- Warner, T. T., R. A. Peterson, and R. E. Treadon, 1997: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Williamson, D. L., J. T. Kiehl, V. Ramanathan, R. E. Dickinson, and J. J. Hack, 1987: Description of the NCAR Community Climate Model (CCMM1). NCAR Tech. Note NCAR/TN-285+STR, 112 pp. [Available from National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80303.]
- Xu, K.-M., 1993: Cumulus ensemble simulation. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.*, No. 46, Amer. Meteor. Soc., 221–235.
- 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.
- , E.-Y. Hsie, and M. W. Moncrieff, 1988: A comparison of explicit and implicit prediction of convective and stratiform precipitating weather systems with a meso- β scale numerical model. *Quart. J. Roy. Meteor. Soc.*, **114**, 31–60.