Removal of Distortion Error from an Ensemble Forecast

JUN DU* AND STEVEN L. MULLEN

Institute of Atmospheric Physics, The University of Arizona, Tucson, Arizona

FREDERICK SANDERS

Marblehead, Massachusetts

ABSTRACT

Large errors developed by 24 h during a 25-member ensemble forecast of quantity of precipitation. The errors could be attributed to an insufficient northeastward motion of the area of precipitation and excessive amounts. This was determined by partitioning of the root-mean-square error into a *distortion error,* the sum of contributions from incorrect position and magnitude, and a *residual error.* The distortion error accounted for more than half of the total error. The distortion error occurs on the synoptic scale and can likely be somewhat ameliorated by future improvements in analysis–forecast systems. The residual error occurs at smaller, less predictable scales, and prospects for its deterministic improvement are not so sanguine.

1. Introduction

A 36-h ensemble quantitative precipitation forecast (QPF) was discussed in detail by Du et al. (1997). The ensemble-mean QPF field produced by the Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model, Version 4 (MM4), with a modified version (Grell et al. 1991) of the Arakawa and Schubert (1974) cumulus parameterization scheme, developed a very large error during the course of the forecast, as illustrated in Figs. 1 and 2. Even larger errors were found in an ensemble using a version (Anthes 1977) of the Kuo (1974) scheme (Mullen et al. 1999). Note that the centroid of the predicted precipitation is about 400 km south-southwest of the centroid of the analyzed area and that the maximum predicted amount is about twice the amount derived from observations. These discrepancies, seen at 24 h after the initial time, apparently increased with time, although a comparison after 24 h could not be made with confidence, because a substantial part of the area had moved into Canada and offshore, where few observations were available.

The reason for this large difference is not clear, al-

though it is related in part to the forecast position of a surface cyclone being situated about 100–200 km south of the analyzed position, as shown by Mullen and Du (1994, their Fig. 2). Neither this error in the position of the sea level cyclone, nor in the large error in prediction of the precipitation shield relative to the cyclone (or in the magnitude of the forecast precipitation), are believed to be a general characteristic of the MM4 model or its later versions.

Even a visual inspection of this case suggests that errors on a relatively large scale quickly played a major role in degrading overall QPF performance. The purpose of this note is to decompose the QPF error of this ensemble forecast into components due to displacement and amplitude, and then examine performance after the removal of these error components. It will be shown that these error components occurred on the synoptic scale and evolved coherently in time, and that they accounted for most of the total error in the QPF field.

2. Computation of distortion error

Hoffman et al. (1995, hereafter referred to as H95) demonstrate that the forecast error can be divided into three components: displacement, amplitude, and residual errors. The *displacement error* measures the part of total error that can be explained by translating the forecast field. The *amplitude error* gives the part of the total error that can be explained by multiplying the displaced forecast field by a constant coefficient. The sum of the displacement and the amplitude errors is termed the *dis-*

---

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

Corresponding author address: Dr. Steven L. Mullen, Department of Atmospheric Sciences, The University of Arizona, 1118 E. 4th Street, Tucson, AZ 85721.
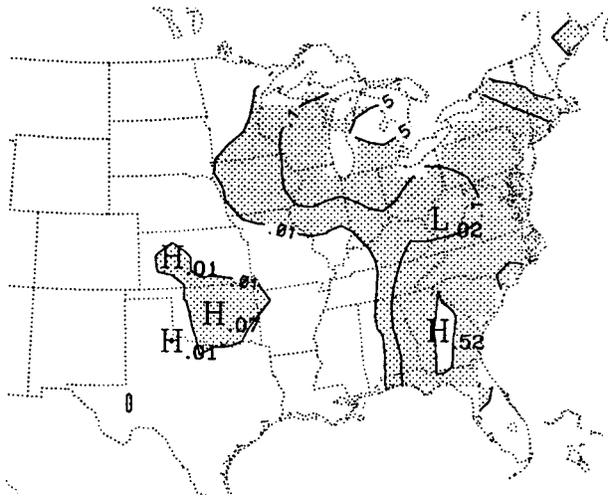E-mail: atmosci@atmo.arizona.edu

FIG. 1. Analysis of precipitation accumulated 18–24 h after the initial time (1200 UTC 14 Dec 1987), based on all rain gauges in the United States bounded by the left border of the figure. This is the verification area for the forecast. Isohyets are for 0.01, 0.1, and 0.5 in.

FIG. 2. Same as Fig. 1 but for ensemble mean predicted precipitation. Isohyet are for 0.01, 0.1, 0.5, and 1 in.

tortion error, while the remainder of the total error is called the *residual error.* The sum of the three components equals the total error.

Application of the technique requires a priori specification of two parameters: 1) the area, or correlation mask as termed by H95, over which the displacement error is computed, and 2) the error metric to be minimized. The technique is qualitatively similar to 2D wavelet analyses (e.g., Briggs and Levine 1997) in the sense that its "basis" function (the mask size and shape vs the mother function for wavelets) is "local" and must be chosen a priori. The decomposition is sensitive to the size of the mask, with increases in its size typically yielding smoother and larger-scale distortion fields and bigger residual errors (H95). The verification area (Fig. 1) serves as our window. Its dimensions are 44 by 38 grid points, which equates to ~3500 km by ~3000 km for the 80-km grid spacing. Regional variations in the distortion error, which are obtained by translating the mask (e.g., Fig. 9 of H95), and changes in area and shape of the mask (e.g., Table 1 of H95), are not examined in this note.

The decomposition is also somewhat sensitive to the choice of error metric (H95). We minimized the root-mean-squared error (rmse) of the ensemble mean fields, but any error measure could be minimized with the H95 technique. Sensitivity of our results to different metrics is not examined in this note.

An underlying premise of the H95 decomposition is that a substantial portion of the total forecast error can be attributed to large-scale, coherent patterns that evolve smoothly in time. For our decomposition of ensemble mean QPF error associated with a strong baroclinic development, the distortion error is predominately large synoptic scale. The remaining residual error is smaller
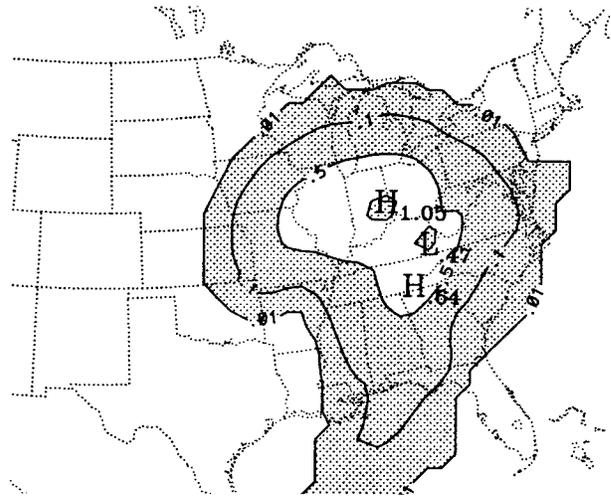
scale, which for an 80-km mesh and a mask size of 3500 km by 3000 km would strongly reflect contributions from meso-α-scale features (wavelengths of 200 to 2000 km).

3. Results

The total rmse, and contributions from the two components and the residual as determined by the H95 method, are shown in Fig. 3 at 6-h intervals for the entire 36-h forecast. These quantities are shown for both the ensemble mean and for the average of the individual ensemble members. The reduction of total rmse (Fig. 3a) after 24 h is due in large part to displacement of much of the analyzed and predicted storm out of the United States, where verification was not possible. Neither component nor the residual can be dismissed as relatively small. The large error growth during the first 24 h is produced by both the displacement and amplitude components, with a noticeably slower growth in the residual. The different growth rates lead to the distortion error exceeding the residual after 12 h. Note that the improvement of the ensemble mean over the average of the members comes primarily from the amplitude error, reflecting the incorrectly predicted magnitude of the precipitation.

The vector displacements and amplification factors for the distortion of the ensemble mean forecast are given in Table 1. After correcting each of the 25 ensemble members by removing the distortion errors in Table 1, a new corrected 25-member ensemble is made, the ensemble mean of which is shown in Fig. 4. Comparing Fig. 4, as well as Fig. 2, with Fig. 1, it can be qualitatively seen that the new ensemble mean forecast (Fig. 4) is more accurate overall than the original ensemble mean forecast (Fig. 2), especially in terms of area coverage. While the overall rmse is reduced, some
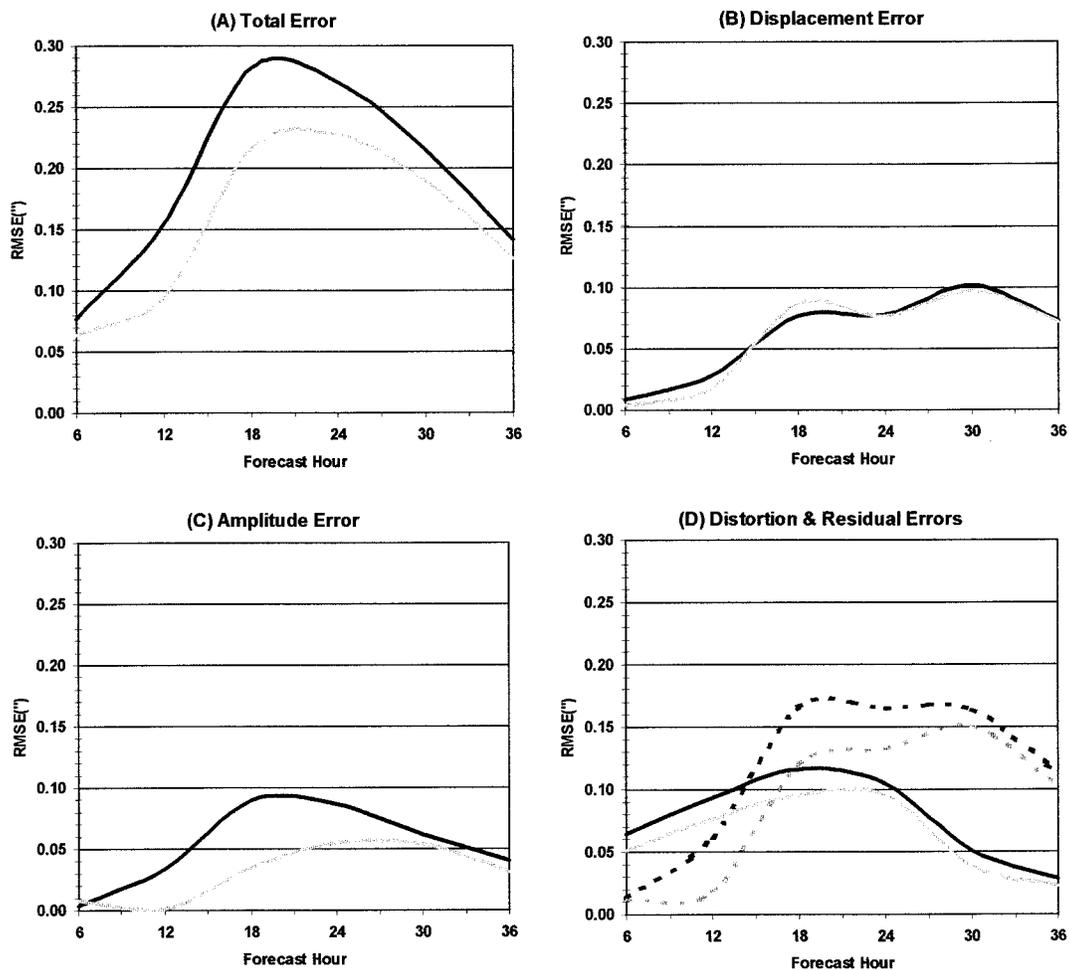
FIG. 3. (a) Total rmse (in.) of 6-h accumulated precipitation at all grid points in the verification area, as a function of time during the forecast period. (b) The same but for the displacement error. (c) The same but for the amplitude error. (d) The same but for the distortion error (dashed) and residual error (solid). The black line shows the average for all individual members of the ensemble, while the gray line shows the ensemble mean.

individual features of the corrected mean show some noteworthy discrepancies with the verifying analysis. For example, Fig. 4 has a 0.4 in. maximum in northern West Virginia that lies within a local region of observed minimal precipitation and is shifted far from the two observed regions (lower Michigan, central Georgia) with greater than 0.5 in. Clearly, the calibrated ensemble forecast is even smoother than the unprocessed ensemble mean.

It is insightful to compare the displacement error vector of Table 1 to average errors in position of the surface cyclone (Mullen and Du 1994, their Fig. 2). As noted in the introduction, both fields tend to be southward of observations, but the precipitation field is erroneously displaced nearly twice as far as the sea level pressure (SLP) field. This difference indicates an improper phasing of the QPF field relative to the predicted SLP field. The difference could be due to a model deficiency, an

TABLE 1. The average vector displacement errors and the average coefficients for the amplitude errors, as defined by Hoffman et al. (1995) for minimization of the rmse, for 6-h total rainfall from the 25-member ensemble forecast. Distance and bearing for the displacement error are to the nearest 10 km and 10°, respectively. Bearing is measured clockwise from north; e.g., a 180° direction denotes a forecast position too far to the south, a 270° direction denotes a forecast position too far to the west, etc. The amplitude coefficient is the factor by which the displaced forecast is multiplied to minimize the rmse. See text.

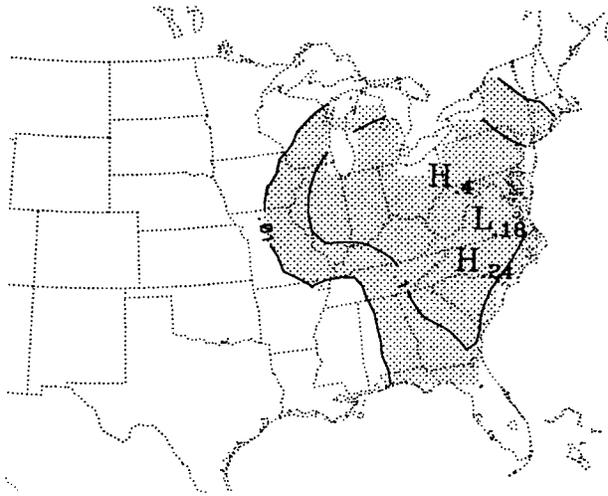|	0–6 h	6–12 h	12–18 h	18–24 h	24–30 h	30–36 h
Displacement	200 km at 240°	260 km at 220°	510 km at 200°	360 km at 240°	520 km at 220°	390 km at 220°
Amplitude coefficient	1.24	0.74	0.46	0.38	0.37	0.55

Fig. 4. Same as for Fig. 1 but for the corrected ensemble mean predicted precipitation. See text.
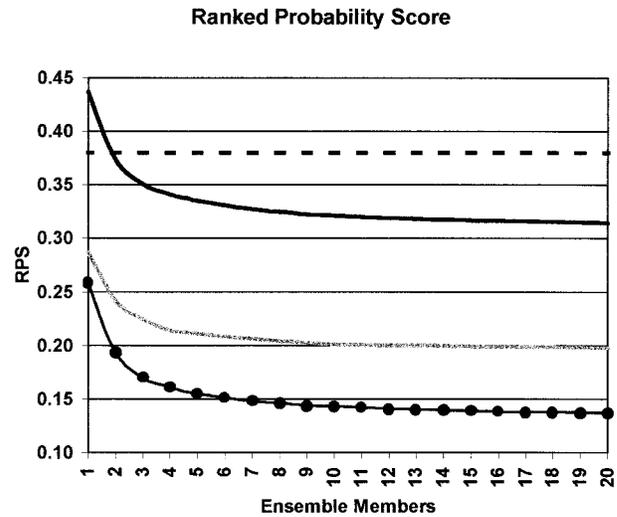
Ranked Probability Score

Fig. 5. Average RPS over the 36-h forecast period, for the entire verification area, as a function of the number of ensemble members. Solid black curve represents the RPS for the raw forecasts. Solid gray curve represents the RPS for the corrected forecasts (see text). The horizontal dashed line represents the RPS for the climatological control forecast. The dotted black line represents the RPS for a "perfect ensemble," where one ensemble member is randomly selected to serve as verification for the remaining 24 members. The results for ensemble sizes between 3 and 20 members are the average of 2000 trials selected from the 25 ensemble members without replacement; results for ensemble sizes of 1 and 2 are the average of all unique combinations.

error in the unperturbed initial analysis upon which perturbations are superimposed, or some combination of the two. For example, mesoscale gravity waves apparently played a crucial role in triggering some of the precipitation (Schneider 1990), and proper simulation of them is utterly beyond the capabilities of an 80-km model (Powers and Reed 1993). Alternatively, errors in the unperturbed moisture field relative to the other fields might also lead to larger displacement errors for precipitation.

A detailed evaluation of probabilistic QPF from the ensemble is made by placing the analyzed precipitation at each grid point in one of five mutually exclusive and exhaustive categories (Du et al. 1997). These categories are no measurable precipitation (less than 0.01 in.) and four categories with lower bounds at 0.01, 0.10, 0.50, and 1.00 in. The forecast amount at each grid point for each ensemble member is then placed in one of these same categories, and a raw frequency is computed. The resulting probability distribution is verified by application of the ranked probability score (RPS; Epstein 1969; Murphy 1971; Wilks 1995, pp. 269–272), while the skill with respect to climatic frequencies (NOAA 1987) is measured by the ranked probability skill score (RPSS).

The area showing positive RPSS for the uncorrected forecast is 76.3% of the total verification area. When the distortion error is removed from the individual ensemble members to generate a calibrated ensemble and forecast probabilities for the five categories are recomputed, the area of positive RPSS increases to 83.8% at 24 h. When verification is restricted to the area of either observed precipitation or nonzero forecast probability of 0.01 in., to simulate a conditional QPF, the area of positive skill drops to 68% for the raw forecast and 70% for the corrected one.

The impact of the distortion error on the RPS, av-

eraged over the entire verification domain and all 6-h forecast periods, is shown in Fig. 5, for ensembles varying in size from 1 member to 20. The RPS of the raw ensemble typically runs ~100% larger than the value for a "perfect ensemble" system, where one member is randomly selected to serve as verification for the remaining members. The perfect ensemble assumption maximizes skill since model error is not considered and a perfect knowledge of analysis error is assumed (Buizza 1997) and, thus, provides an estimate of the upper bound of forecast accuracy. On the other hand, the RPS for the corrected forecasts runs ~40% above the value for the perfect ensemble. Thus, the distortion error for this case accounts for more than half of the RPS error, relative to a perfect ensemble simulation.

4. Discussion

The displacement and amplitude errors, as defined here, are on the synoptic scale. One might hope that future improvements in the analysis–forecast system would reduce these errors. On the contrary, the residual error reflects the smaller scale of the details of the analyzed and predicted patterns and will likely be more difficult to reduce since these scales are less predictable (e.g., Islam et al. 1993). Note that in the early stages of the forecast, the total rmse is dominated by the small-scale residual component and presumably the growth of

the initial conditions errors, which can be ameliorated to some extent by ensemble techniques without improvement in the model.

The distortion errors found in this case are undoubtedly case dependent and cannot be applied mindlessly to another instance in which observations might show a quite different error signature. It is appropriate to refer to them as "model errors for this particular synoptic situation," or "model/situation" errors, where model refers to the total analysis–forecast system and ensemble construction strategy. Our results indicate that these model/situation errors had a deleterious impact on the accuracy and skill of this ensemble forecast for ensemble sizes up to 20, and would likely not be substantially reduced with further increases in ensemble size.

5. Summary

During a 25-member, 36-h, ensemble forecast of quantity of precipitation, large errors developed. These were an insufficient northeastward motion of the area of precipitation and excessive amounts. A five-category verification of the probabilities from the ensemble shows positive skill over a substantial part of the verification area. This skill is increased by correcting the ensemble mean forecast for errors in position and magnitude of the precipitation by application of the method developed by Hoffman et al. (1995). This method allows partitioning of the total root-mean-square error into contributions from incorrect position and magnitude, denoted the *distortion error*. The remaining residual error is denoted *random*. We minimized the rmse for our application of the method, but any error metric can be used.

The results for this case point to the need of pursing a better understanding of the relative roles of model errors and initial data errors on ensemble forecasts of precipitation. The decomposition technique of Hoffman et al. (1995) offers a flexible, synoptically insightful way to decompose coherent, scale-dependent error patterns that may prove useful in that pursuit.

REFERENCES

Anthes, R. A., 1977: A cumulus parameterization scheme utilizing a one-dimensional cloud model. *Mon. Wea. Rev.,* **107,** 963–984.

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.

Briggs, W. M., and R. A. Levine, 1997: Wavelets and field forecast verification. *Mon. Wea. Rev.,* **125,** 1329–1341.

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.

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., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.,* **8,** 985–987.

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.

Hoffman, R. N., Z. Liu, J.-F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.,* **123,** 2758–2770.

Islam, S., R. L. Bras, and K. A. Emanuel, 1993: Predictability of mesoscale rainfall in the Tropics. *J. Appl. Meteor.,* **32,** 297–310.

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.

Mullen, S. L., 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.

——, ——, and F. Sanders, 1999: The dependence of ensemble dispersion on analysis-forecast system: Implications to short-range ensemble forecasting. *Mon. Wea. Rev.,* **127,** 1674–1686.

Murphy, A. H., 1971: A note on the ranked probability score. *J. Appl. Meteor.,* **10,** 155–156.

NOAA, 1987: Monthly relative frequencies of precipitation for the United States for 6-, 12-, and 24-h periods. NOAA Tech. Rep. 39, 262 pp.

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.

Schneider, R. S., 1990: Large-amplitude mesoscale wave disturbances with the intense Midwestern extratropical cyclone of 15 December 1987. *Wea. Forecasting,* **5,** 533–558.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.