

NOTES AND CORRESPONDENCE

Does Increased Predicted Warm-Season Rainfall Indicate Enhanced Likelihood of Rain Occurrence?

WILLIAM A. GALLUS JR.

Department of Geological and Atmospheric Science, Iowa State University, Ames, Iowa

MOTI SEGAL

Department of Agronomy, Iowa State University, Ames, Iowa

5 January 2004 and 9 June 2004

ABSTRACT

The likelihood of simulated rainfall above a specified threshold being observed is evaluated as a function of the amounts predicted by two mesoscale models. Evaluations are performed for 20 warm-season mesoscale convective system events over the upper Midwest of the United States. Simulations were performed using 10-km versions of the National Centers for Environmental Prediction Eta Model and the Weather Research and Forecasting (WRF) model, with two different convective parameterizations tested in both models. It was found that, despite large differences in the biases of these different models and configurations, a robust relationship was present whereby a substantial increase in the likelihood of observed rainfall exceeding a specified threshold occurred in areas where the model runs forecast higher rainfall amounts. Rainfall was found to be less likely to occur in those areas where the models indicated no rainfall than it was elsewhere in the domain; it was more likely to occur in those regions where rainfall was predicted, especially where the predicted rainfall amounts were largest. The probability of rainfall relative-operating-characteristic and relative-operating-level curves showed that probabilistic forecasts determined from quantitative precipitation forecast values had skill comparable to the skill obtained using more traditional methods in which probabilities are based on the fraction of ensemble members experiencing rainfall. When the entire sample of cases was broken into training and test sets, the probability forecasts of the test sets evidenced good reliability. The relationship noted should provide some additional guidelines to operational forecasters. The results imply that the tested models are better able to indicate the regions where atmospheric processes are most favorable for convective rainfall (where the models generate enhanced amounts) than they are able to predict accurately the rainfall amounts that will be observed.

1. Introduction

An intriguing quantitative precipitation forecasting (QPF) question asks, Given predicted precipitation uncertainty, can forecasts of rainfall be improved in selective subdomains of the predicted precipitation domain? It has been shown (e.g., Hamill and Colucci 1997) that for ensemble simulations of precipitation, an increase in forecast probability is generally associated with an increase in the probability of occurrence of precipitation. Conceptually, one may attribute the improvement in rainfall forecast quality at grid points within the subdomain generated by the intersection of ensemble members to the fact that the variability in initialization/

physical formulation among members has not affected the prediction of precipitation occurrence there. In this note, an additional independent situation is evaluated focusing on subdomains with enhanced predicted precipitation as elaborated in the hypothesis below.

We will examine the hypothesis that in subdomains consisting of model grid points at which large amounts of precipitation are predicted, the probability of experiencing a lighter rain amount is higher than that valid for the entire simulation domain. In addition, skillful probabilistic forecasts over the entire domain can be issued based on QPF amount. The hypothesis can be viewed as a specific fundamental issue in QPF that has not been addressed for fine-resolution simulations of warm-season convective rainfall (it was partially examined for winter precipitation in coarser-grid European Centre for Medium-Range Weather Forecasts simulations; Atger 2001). Also, if confirmed positively, it may provide some advantage in the operational forecasting of precipitation.

Corresponding author address: William A. Gallus Jr., Department of Geological and Atmospheric Science, 3025 Agronomy Hall, Iowa State University, Ames, IA 50011.
E-mail: wgallus@iastate.edu

The regional forecast skill of summer convective rainfall is relatively low as compared with that of the larger-scale organized precipitation systems that dominate the cooler seasons of the year (e.g., Schneider et al. 1996). Examining the hypothesis for summer mesoscale convective systems should therefore provide the most rigorous test. Stensrud et al. (2000) suggest that, for mesoscale convective systems, varying model physics is a potentially more powerful method in creating an ensemble than is perturbing initial conditions. Wandishin et al. (2001) also demonstrate the advantages of an ensemble using mixed models or mixed physics. Based on their conclusions, the present study will concentrate on variants produced from two different models, the National Centers for Environmental Prediction (NCEP) Eta Model (Mesinger et al. 1988; Janjić 1994) and the Weather Research and Forecasting (WRF) model (Skamarock et al. 2001), with two different cumulus parameterization (CP) schemes used in each: the Betts–Miller–Janjić (Betts and Miller 1986; Janjić 1994; hereinafter BMJ) and Kain–Fritsch (1993; hereinafter KF). Briefly, the BMJ is an adjustment-type scheme that forces soundings at each point toward a reference profile of temperature and specific humidity. The scheme's structure favors activation in cases with substantial amounts of moisture in low and midlevels and positive convective available potential energy (CAPE). The KF scheme removes CAPE (calculated using the traditional, undiluted parcel-ascent method) through vertical reorganization of mass. The scheme consists of a convective trigger function (based on grid-resolved vertical velocity), a mass flux formulation, and closure assumptions.

The BMJ and KF schemes are known to differ in some features of their predicted rainfall, and in the response to atmospheric background conditions. Gallus and Segal (2001, hereinafter GS01), for instance, found large differences in the BMJ and KF bias scores (bias = f/o , where f is the number of points forecast to have rainfall above a specified threshold and o is the number of points observed to exceed that threshold) as a function of rainfall amount. In addition, spread ratios (Wandishin et al. 2001), or equivalently the inverse of correspondence ratios (Stensrud and Wandishin 2000), defined as the union of all grid points at which rainfall in N ensemble members exceeds a specified threshold divided by the intersection of points where the same is true, were found to be large when runs with these two schemes ($N = 2$) were compared. As such, simulations with the BMJ and KF runs are considered to be, to a reasonable extent, independent and thus attractive for a simplified yet comprehensive test of our hypotheses. In addition, the above models and convective schemes have been used widely, thus furthermore providing merit to their adoption in the present study.

2. Data and method

Simulations of 20 warm-season convective events in the upper Midwest have been performed using 10-km

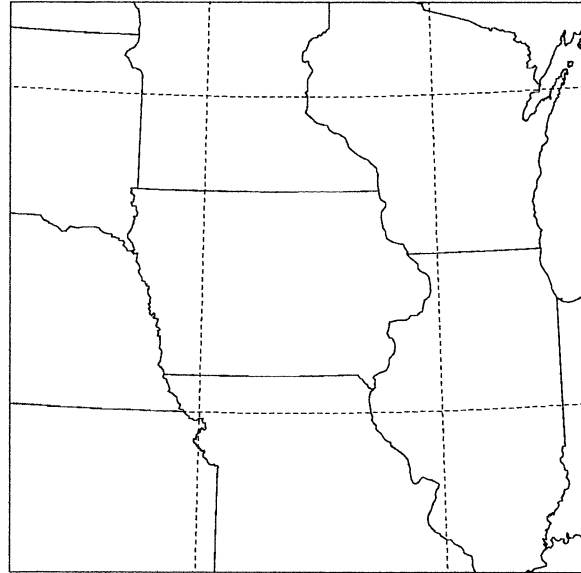


FIG. 1. Model domain used for all simulations.

versions of the Eta and WRF models. Events were chosen if mesoscale boundaries, usually convectively induced, were present at 0000 or 1200 UTC, and significant mesoscale convective system (MCS) rainfall followed within the next 12–18 h over the upper Midwest. Simulations were run for 24 h over a fixed mesoscale domain covering roughly 1000 km \times 1000 km (Fig. 1). It should be noted that Warner et al. (1997) have found that errors from lateral boundary conditions in limited domains can be more serious than initial condition errors. Most of the significant rainfall in our simulated events, however, occurs in the 6–18-h forecast period. The short-range nature of the forecasts in the present study, along with the slower steering flow in the warm season, likely reduces the boundary condition problems somewhat. Note that, during the first 6 h of integration, model spinup of precipitation may also result in less skillful forecasts. GS01 noted that such spinup problems in the cases to be discussed in the present study seemed to affect the Eta-KF runs more than the Eta-BMJ runs, likely a result of NCEP's use of the BMJ scheme during data assimilation for the Eta. Standard initial and boundary condition data were provided by operational NCEP Eta output interpolated to a 40-km grid.

Rainfall forecast skill was evaluated by comparing the model predictions for 6-h periods (within the full 24-h integration period) during which active organized convection was observed within the simulation domain using the 4-km-horizontal-resolution NCEP stage-IV (Baldwin and Mitchell 1997) precipitation observations. The observations were areally averaged onto the Eta Model's 10-km grid (using procedures similar to those used at NCEP; M. Baldwin 2000, personal communication). The WRF model grid boxes closely matched those of the Eta Model, but some bilinear interpolation

TABLE 1. Bias scores for the Eta and WRF models running with both the BMJ and KF convective schemes for the given rainfall thresholds (in.).

Model run	Predicted rainfall threshold (in.)							
	0.01	0.05	0.10	0.15	0.25	0.35	0.50	1.0
Eta-BMJ	1.231	1.404	1.480	1.487	1.445	1.465	1.289	0.558
Eta-KF	0.835	0.890	0.906	0.911	0.925	0.972	1.066	1.521
WRF-BMJ	1.179	1.280	1.318	1.296	1.196	1.112	0.927	0.381
WRF-KF	1.124	1.201	1.278	1.329	1.388	1.480	1.528	1.937

was needed to create an exact match to allow straightforward computation and comparison of objective skill measures. When available, the multisensor (includes both radar and gauge observations) data were used; otherwise, the stage-IV gauge measurements were used. For the cases simulated in the present study, these precipitation data (multisensor and gauge) did not differ substantially. However, Schwartz and Benjamin (2000) have noted that the multisensor data may be wetter for rainfall amounts under 0.50 in. (1.27 cm) in 24 h than the gauge-only data and drier for heavier amounts. In the present study, the stage-IV data were also compared with radar and surface reports to disregard small-scale spurious features that occasionally occurred directly over radar sites (manifested as isolated heavy rainfall amounts that were not part of a larger region of precipitation). A total of 51 cases of accumulated rainfall in 6-h periods were extracted from the 20 simulated events and were analyzed.

The 10-km version of the Eta Model used was similar to that used operationally at NCEP in early 2000 (except with reduced horizontal diffusion to better retain mesoscale features), and included the same physical parameterizations present in the operational model [see Janjić (1994) and Rogers et al. (1998) for more details]. In addition to the BMJ convective scheme used operationally, simulations were repeated using a version of the KF scheme adapted for use in the Eta Model (J. Kain 2000, personal communication). The Eta Model was run with 33 vertical levels. The WRF simulations used version 1.2.1 of the WRF model (details were available at the time of writing online at <http://www.wrf-model.org>), with 35 vertical levels, and both the BMJ and KF convective schemes were again used in the WRF. Note that these two convective schemes are tuned slightly differently in the Eta model than in the WRF model.

3. Results

Gallus (1999) and Stensrud et al. (2000), among others, have shown that warm-season rainfall forecasts are strongly sensitive to the convective schemes used. Given this finding, our analysis of observed rainfall likelihood within regions of predicted rainfall uses output from Eta and WRF simulations using both the BMJ and KF schemes. GS201 showed in their study of 14 Eta Model variants that the spread ratio for rainfall when

these two variants (runs using the BMJ and KF schemes) were compared was larger than that associated with the comparison of any two initialization adjustments. Spread ratios between the Eta and WRF runs themselves (not shown) suggest that the impacts from the use of these two different convective schemes are comparable to, if not greater than, the impacts from switching models.

Note also that the bias behavior in both the Eta and WRF runs using the BMJ convective scheme is distinctly different than that in both models using the KF scheme (Table 1). GS01 showed, using the Eta Model, that the tendency for the KF runs to have a low bias at light amounts and a strong tendency toward higher biases at heavier amounts resembles the trends seen when no convective scheme is used in the Eta Model with 10-km grid spacing. On the other hand, that study found that when the BMJ scheme is used, biases are generally higher than 1.0 except for the heaviest amounts, for which the BMJ scheme results in sharply lower biases with large negative bias errors at the heaviest rain amounts. If the hypothesis of greater light-rain frequency being associated with areas of predicted heavier rainfall is true for each of the four model configurations to be evaluated (which have distinctly different bias behavior), the findings are likely to be robust and to apply to most model forecasts of warm-season MCSs, at least in the upper Midwest.

a. Analysis approach

To determine if likelihood of rainfall occurrence varies directly with the amount of rainfall simulated, we compute the hit rate (HR), also called the correct-alarm ratio, for the prediction of rainfall exceeding a specified threshold in subdomains defined by the aggregate of grid points at which rainfall within a specified range of amounts (QPF bin) was simulated. Bins were chosen to match standard QPF thresholds used for operational verification, including <0.01 (no rain), 0.01–0.05, 0.05–0.10, 0.10–0.15, 0.15–0.25, 0.25–0.50, 0.50–1.0, and >1.0 in. (0.01 in. = 0.254 mm). The HR is given by h/f , where f is the total number of grid points in the given subdomain (i.e., the set of grid points at which rainfall is forecast to fall within the range of amounts) and h represents the number of “hits”—those grid points at which the observed rainfall also exceeded specified thresholds. The HRs are used as probability

TABLE 2. Probability (in %) of observed rainfall exceeding thresholds of 0.01, 0.10, and 0.25 in. for simulated rainfall amounts in the specified ranges. The probability of observed rain in the entire domain is shown in parentheses for each threshold. Results are presented for the Eta and WRF models running with the BMJ and KF convective schemes.

Rainfall threshold (in.)	Predicted rainfall amount (in.)							
	<0.01	0.01–0.05	0.05–0.10	0.10–0.15	0.15–0.25	0.25–0.50	0.50–1.0	>1.0
	Eta-BMJ							
0.01 (27.4)	17.4	38.0	43.3	48.6	54.5	58.6	59.1	63.2
0.10 (15.0)	8.2	21.7	25.3	28.2	32.3	35.2	36.8	43.6
0.25 (8.8)	4.5	12.8	14.8	16.3	19.9	22.1	22.7	28.5
	Eta-KF							
0.01 (27.4)	20.8	46.5	52.0	53.1	57.4	61.2	61.9	64.2
0.10 (15.0)	10.6	27.4	31.2	33.3	35.1	38.7	38.8	49.4
0.25 (8.8)	6.1	16.2	18.2	18.9	22.1	23.8	23.9	35.9
	WRF-BMJ							
0.01 (27.4)	17.6	42.6	47.8	51.6	54.8	58.9	64.1	67.0
0.10 (15.0)	8.3	24.2	27.2	30.8	33.7	37.6	42.1	50.4
0.25 (8.8)	4.5	14.6	16.2	18.8	20.1	22.8	27.0	30.0
	WRF-KF							
0.01 (27.4)	18.3	39.8	46.8	50.9	53.5	59.0	60.5	67.0
0.10 (15.0)	8.8	21.9	26.8	30.0	32.8	37.1	39.6	41.9
0.25 (8.8)	4.9	12.2	15.8	17.7	18.8	23.0	24.4	28.9

forecasts for 6-hourly rainfall, and these are determined for observed rainfall exceeding three threshold amounts: 0.01, 0.10, and 0.25 in.

In addition to HR, other quantities were computed from the traditional 2×2 contingency table for dichotomous forecasts. Probability of detection (POD) is given by h/o , where o is the total number of grid points in the entire model domain at which observed rainfall exceeded specified thresholds. Probability of false detection (POFD) indicates the subset of all points at which rainfall was not observed to exceed a threshold that were forecast to exceed it, and it can be expressed as $(f - h)/(t - o)$, where t is the total number of grid points in the model domain. Miss ratio (MR) identifies the subset of points not forecast to exceed a threshold that were observed to exceed it, and it can be expressed as $(o - h)/(t - f)$. Using the probability of rainfall values corresponding to the QPF bins, relative operating characteristic (ROC) curves were computed for which POD is plotted as a function of POFD for yes/no forecasts made based on probability thresholds. ROC curves, sometimes referred to as receiver operating characteristic curves, indicate the ability of a forecast to distinguish between events and nonevents, based on various decision thresholds. In addition, based on an assumption of what probability value would cause a forecaster to issue a warning, relative-operating-level (ROL) curves (Mason and Graham 1999) were computed showing HR as a function of MR for rainfall observed to exceed specified thresholds. The skillfulness of each probabilistic forecast was evaluated by computing the areas under the ROC and ROL curves.

It is important to note that this study seeks to explore only one issue: Is rainfall probability of occurrence a function of forecast intensity? A detailed study of pa-

rameters more traditionally used to evaluate a deterministic forecast, such as bias and equitable threat score [$ETS = (h - c)/(f + o - h - c)$, where c is a measure of the number of grid points at which a correct forecast would occur by chance, $c = of/t$], was already performed by GS01 for the Eta runs examined in this note. The examination of likelihood of rainfall occurrence in subdomains defined by the predicted rainfall amount provides insight that supplements that usually supplied by traditional objective skill measures such as threat scores.

b. Analysis results

Wilks (1990) determined that heavier rainfall amounts were more likely to occur when subjective probability-of-precipitation forecasts were high than when the probability forecasts were low. Our analysis is designed to address the reverse question: Is predicted heavy rainfall in a deterministic forecast indicative of greater probability of measurable rainfall being observed than is lighter predicted rainfall?

Hit rates (which can be thought of as probabilities) for observed rainfall amounts exceeding the thresholds 0.01, 0.10, and 0.25 in., evaluated over the subdomains where the Eta and WRF model simulations predicted rainfall amounts within specified bins, are shown in Table 2. In all cases, HRs rise as the subdomain is restricted to those grid points at which increasingly heavy amounts have been forecast. In all four models, the probability of any rainfall being observed when the forecast is for none (less than 0.01 in.) is roughly 20%. The probability for greater than 0.25 in. is around 5%. The probabilities rise rapidly as the forecast amounts increase to between 0.01 and 0.05 in. For heavier forecast amounts, the

changes in probability are more gradual, but for the thresholds being verified in the table, they do continuously increase. For QPF amounts exceeding 1.0 in., the probability of any measurable rainfall ranges from 63% in the Eta-BMJ run to 67% in both WRF runs. The probability of greater than 0.25 in. ranges from 29% in the Eta-BMJ and WRF-KF runs to 36% in the Eta-KF run.

Table 2 also shows (in parentheses) the probability of rainfall above the three specified thresholds for the entire domain, regardless of the forecast. For all model configurations and all thresholds, these values (observed frequency of rainfall) lie between those valid when the QPF amount is none and when the QPF amount is 0.01 inch or more. Thus, rainfall is less likely to occur in those areas where the models indicate no rainfall than it is elsewhere in the domain; it is more likely to occur in those regions where rainfall is predicted, especially where the predicted rainfall amounts are largest.

Although not shown in the table, analysis in this study suggests that the probability values become even larger when the intersection of grid points experiencing simulated rainfall above a specified threshold in two different model versions is used. Questions exist about how to construct a QPF–probability relationship for a combination of models, since the requirement that all models show a rainfall amount within the specified bins results in the neglect of many points for which a range of bins are indicated by different models. Only examining those points for which both Eta configurations show no rain or show rainfall exceeding 1.0 in., the intersection regions have probability values for measurable observed rainfall of 15% and 83%, respectively. For the WRF, the intersection of the BMJ and KF runs yields probability values of 14% and 68% for the same rainfall categories. If the intersections of the Eta and WRF runs are compared, the probabilities of rainfall are 15% and 88%, and 17% and 77%, for the BMJ and KF schemes, respectively. Thus, with the exception of the two WRF runs, there is a much greater increase in probability of rainfall (in comparison with that present in either run by itself) when both runs show greater than 1.0 in. of rainfall at any grid point. These values suggest that more elaborate QPF–probability relationships (e.g., based on ensembles) could be created, yielding a wider range of probabilities that could be forecast.

Probabilistic forecasts generated from ensembles are typically evaluated using ROC diagrams. Figure 2 shows ROC diagrams for verifying thresholds of 0.01 (Fig. 2a) and 0.25 in. (Fig. 2b) for the Eta-BMJ runs using the probability values shown in Table 2 (values for 0.01–0.05, 0.10–0.15, and 0.25–0.50 in. are shown in boldface in Fig. 2). One unusual aspect of the approach is that the probability of rainfall for a forecast of no rain was not zero, but was instead roughly 17%. Thus, no lower probability value could be forecast, and the point in the diagram in Fig. 2a at the upper-right corner would be 17% and not 0. It is possible that, were

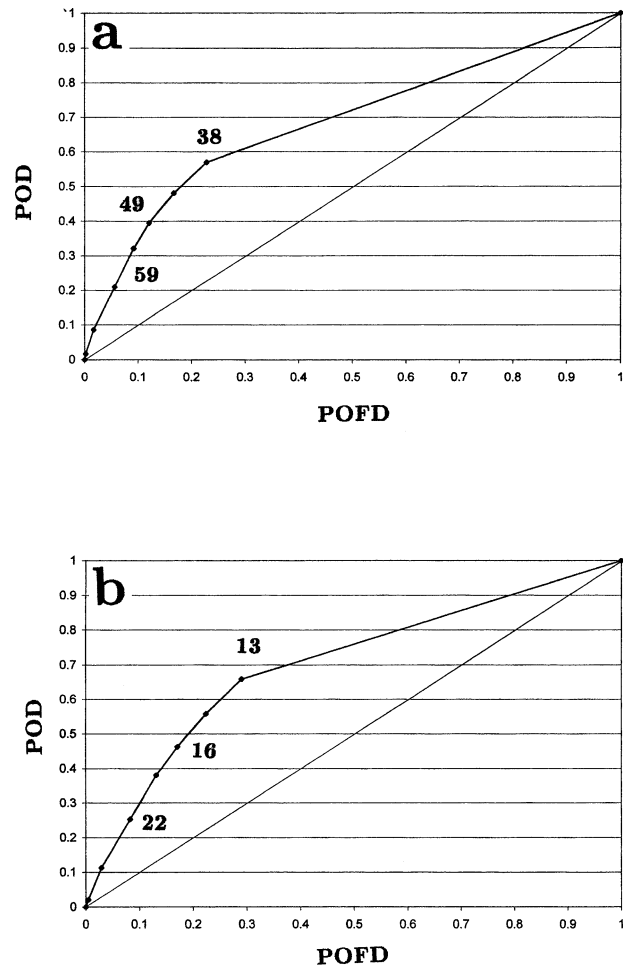


FIG. 2. ROC curve (POD vs POFD) for Eta-BMJ prediction of rainfall exceeding (a) 0.01 and (b) 0.25 in. For area size under the curve, see Table 3. Diagonal line indicates no skill. Probability values corresponding to QPF of 0.01–0.05, 0.10–0.15, and 0.25–0.50 in. are shown in boldface.

more complex criteria (such as determining the HR for those grid points for which no rain was predicted at the grid point and all neighboring points) to be used, lower probability values might be obtained.

It can be seen in the diagram that both curves lie above the diagonal no-skill line (along which false alarms are as likely as hits). Thus, in both cases, the area under the ROC curve, calculated using the trapezoidal method, exceeds 0.5, implying the potential for a skillful forecast. In all four model configurations, the curves look similar (not shown), and the areas under all of the curves are shown in Table 3. In general, the areas are all near 0.7, implying a useful forecast (Buizza et al. 1999) for both rainfall amounts of 0.01 and 0.25 in., with slightly larger values for 0.25 in.

Table 3 also contains information for other probabilistic forecasts made in a more traditional manner from ensembles of various model members. In these ensembles, equal weighting is assigned to each member. The

TABLE 3. Areas under the ROC curves (computed using the trapezoidal method) created by the probability forecasts shown in Table 2 for each model configuration, verified for thresholds of 0.01 and 0.25 in. The asterisk indicates ensemble probability forecasts using common method of equal weighting of ensemble members. Eta/WRF is the four-member ensemble using all four configurations discussed in this paper. Eta-KF/BMJ and WRF-KF/BMJ are two-member ensembles in which the convective scheme has been varied within the Eta and WRF models, respectively. BMJ-Eta/WRF and KF-Eta/WRF are two-member ensembles in which the model has been varied but the convective scheme is the same. GS01-14 member refers to an ensemble using all 14 model variants discussed in GS01.

Model and CP scheme	Rainfall threshold (in.)	
	0.01	0.25
Eta-BMJ	0.6848	0.6997
Eta-KF	0.6387	0.6442
WRF-BMJ	0.6883	0.7004
WRF-KF	0.6759	0.6940
Eta/WRF*	0.7361	0.6565
Eta-KF/BMJ*	0.7029	0.6140
WRF-KF/BMJ*	0.7162	0.6321
BMJ-Eta/WRF*	0.7026	0.6081
KF-Eta/WRF*	0.6889	0.6131
GS01-14 member*	0.7352	0.6670

“Eta/WRF” ensemble uses the four model configurations discussed in this study. That ensemble has a higher area under the curve for the 0.01-in. threshold, but the area is lower for the 0.25-in. threshold. Other ensembles shown include several consisting of two members, representing both Eta configurations, both WRF configurations, both of the BMJ runs, and both KF runs. The areas for these two-member ensembles are not much higher than the QPF–probability technique for the 0.01-in. threshold and are all lower for the 0.25-in. threshold. Even when all 14 model variants discussed in GS01 are used as an ensemble, the area under the ROC curve for the 0.25-in. threshold is less than that of three of the four QPF–probability curves representing model configurations used in the current study. The trend toward increasing skill (areas under the ROC curve) at higher thresholds occurring with the proposed approach is opposite to what universally happens with the ensemble-based probabilistic forecasts shown. The trend toward less skill for heavier thresholds is also present for most ensemble systems evaluated for warm-season precipitation by Mullen and Buizza (2001) and Wandishin et al. (2001). The increase in skill for heavier thresholds with the QPF–probability relationship suggests that the technique may be of most value in assisting forecasters with prediction of heavier rainfall amounts.

Another method of evaluating a probability forecast is to use an ROL curve. A detailed description of the differences between ROC and ROL curves can be found in Mason and Graham (1999). As they discuss, these curves not only show different measures (correct-alarm ratio and miss ratio in an ROL curve, POD and POFD in an ROC curve) but, to allow an evaluation similar to that performed on ROC curves, are constructed in a different way. The definition of a warning is held fixed,

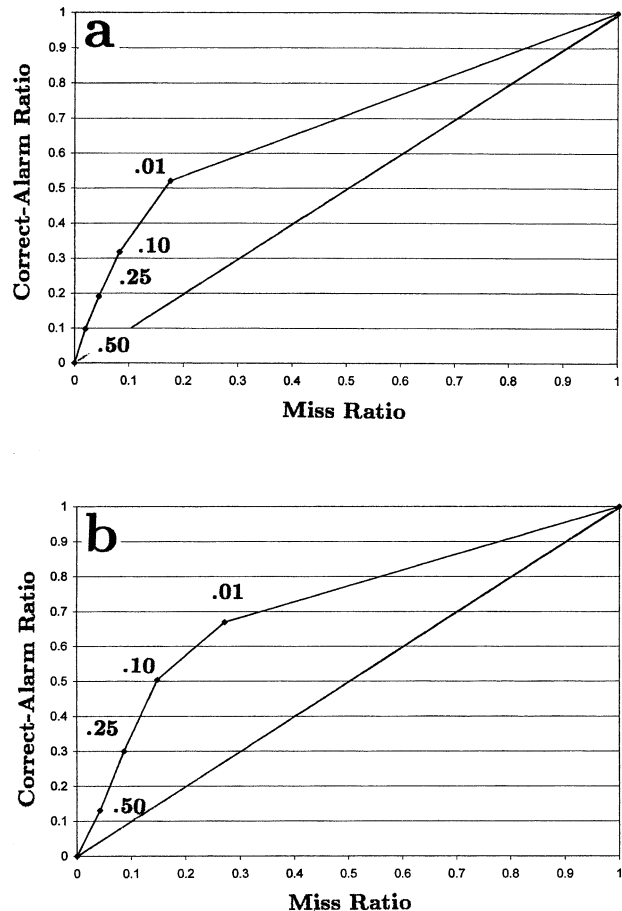


FIG. 3. ROL curve (correct-alarm ratio vs MR) for WRF-BMJ prediction when warnings are issued for rainfall exceeding (a) 0.01 and (b) 1.0 in. For area size under the curve, see Table 4. Diagonal line indicates no skill.

and the curve is thus able to provide an indication of the estimated probabilities of different outcomes given the forecast criteria. In Fig. 3, the ROL curves are shown for the WRF-BMJ run (again all four configurations are similar) assuming a forecaster issues a warning (a forecast of measurable rainfall) based on a rainfall prediction of at least 0.01 in. (the probability forecast for the 0.01–0.05-in. bin was 43%; Fig. 3a) or at least 1.0 in. (probability of 67%; Fig. 3b). The correct-alarm and miss ratios are then plotted for observed rainfall exceeding amounts of 0.01, 0.10, 0.25, and 0.50 in. (the 1.0-in. amount was located very near the lower-left corner). Both curves again lie above the no-skill line. One can see that, when both of these warning criteria are used, the correct-alarm ratios are relatively high for observed rainfall exceeding 0.01 in., with miss ratios that are less than one-half of the correct-alarm ratios. For rainfall exceeding 0.50 in., these warning criteria result in low correct-alarm ratios but even lower miss ratios.

Computed areas under the ROL curves are given for the four model configurations in Table 4. As was implied in Fig. 3, the areas are larger when 1.0 in. is used as

TABLE 4. Areas under the ROL curves (computed using the trapezoidal method) created by plotting correct-alarm ratios and MRs for various QPF amounts. Warning criteria of ≥ 0.01 and ≥ 1.0 in. were used.

Model and CP scheme	Rainfall threshold (in.)	
	0.01	1.0
Eta-BMJ	0.6693	0.6953
Eta-KF	0.6759	0.7120
WRF-BMJ	0.6800	0.7175
WRF-KF	0.6823	0.7110

the threshold for the warning criteria. The areas again are approximately 0.7, similar to the areas under the ROC curves.

As a last evaluation of the performance of the QPF-derived probabilistic forecasts, the reliability of the forecasts was examined. Because the rainfall probabilities assigned to a given QPF bin are based on the observed frequency of rainfall for the cases evaluated, it would be meaningless to examine reliability for the full 51-case sample. Instead, the full sample was divided into 41-case training sets and 10-member test sets, for which the 10 members were chosen randomly and should be generally independent from the training-set cases. Probabilities for rainfall as a function of QPF amount (HRs) were recomputed using the 41 cases, and then the observed frequency from the 10-member test set was compared with the forecast probabilities for observed rainfall amounts exceeding 0.01 in. in the Eta-BMJ (Fig. 4a) and Eta-KF (Fig. 4b) runs. Figure 4 shows the reliability diagram for one of the random tests, with dots showing the scatter present among four other training/test sets also using 41 and 10 cases. For both model configurations, the reliability is good and looks better than that found in some other studies that use ensemble systems to predict warm-season rainfall (e.g., Mullen and Buizza 2001). Although some scatter is present among the five different training and test sets used, the observed frequency is almost always within 10% of the forecast probability.

For heavier verifying thresholds (not shown), the reliability curves are similar, but the range of forecast probabilities is much smaller (see Table 2) and thus the curve itself is smaller. As mentioned earlier, additional calibration might be desired to expand the range of probabilities that could be forecast. Without any calibration, the QPF-derived probabilities for observed rainfall exceeding 0.01 in. are never much less than about 20% or much greater than 70%. Inset into Fig. 4 are histograms showing the frequency with which the forecast probabilities are issued for the test case whose reliability is plotted with the solid line (histograms among all five of the training sets did not differ appreciably). The lowest probability (no model QPF) is overwhelmingly forecast. The frequency for all other bins is much smaller and uniform. This implies that the technique does not demonstrate much sharpness, particularly on the higher-

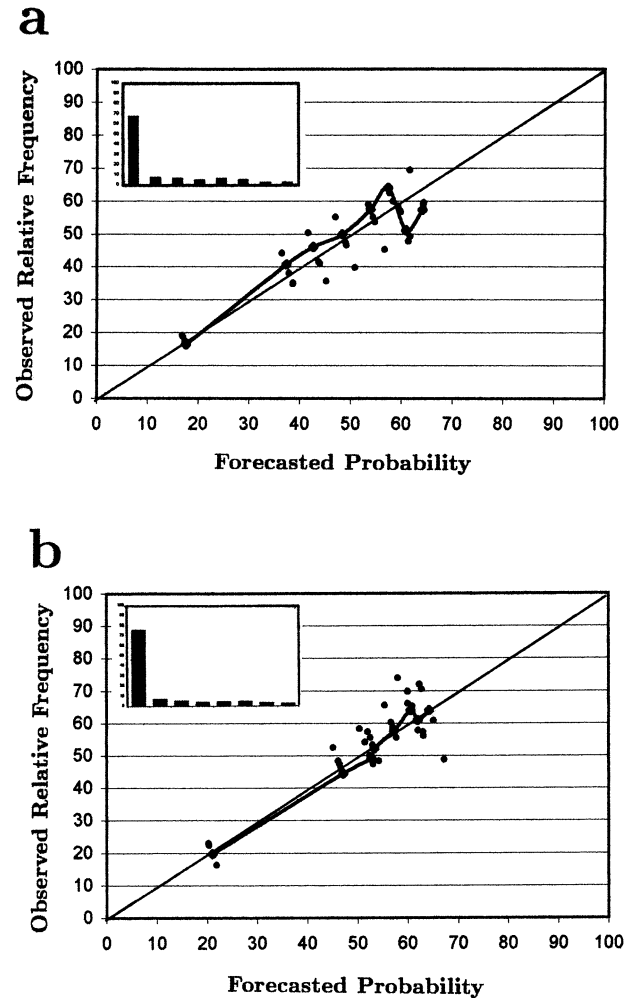


FIG. 4. Reliability diagrams for Eta runs using the (a) BMJ and (b) KF convective schemes for rainfall greater than or exceeding 0.01 in. Forecast probabilities were determined from a 41-event training set. Observed frequency of rainfall is plotted with a solid line for one 10-event test set. Additional points are plotted for four different training and test sets of 41 and 10 events, respectively. Histograms showing relative frequency of forecast probabilities for the eight QPF amount bins shown in Table 2 are inset for the training/test set used to plot the solid line.

probability end. Nonetheless, a wide range of verification parameters suggests that the QPF–probability relationship may be a useful forecast tool, especially if combined with other forms of model guidance.

Finally, the relationship shown in this study could be used by forecasters in their standard issuance of subjectively determined probabilistic precipitation forecasts. It is important to note, however, that nonzero miss ratios indicate a forecaster should not assume that no rain will occur in those areas of the domain in which simulated rainfall does not occur or is light in intensity. ETSS (GS01) for the Eta runs examined in this note were generally low, with no real skill for amounts of 0.50 in. or more and no particular model configuration

consistently earning significantly higher ETSs over a range of thresholds. For the WRF runs, ETSs (not shown) were comparable to those in the Eta runs, with the two convective schemes resulting in simulations with very similar values. Thus, these model forecasts of warm-season precipitation do have substantial errors throughout the domain.

4. Conclusions

It was determined that for mesoscale convective systems in the upper Midwest of the United States the likelihood of rainfall exceeding a specified threshold increases substantially in those subdomains where the Eta and the WRF model predict increasingly heavier rainfall amounts. In other words, HRs increase for increasingly heavy QPF bins. The result appears to be robust, holding true for both the commonly used Eta and WRF models with two different convective schemes, despite the tendency for bias scores to differ substantially among the different model configurations. It therefore appears that forecasters can be more confident of at least light amounts of convective system rainfall occurring if either of these commonly used model runs produces heavy rainfall at a point. The probability of observed rainfall above a specified threshold for the entire domain was found to lie between the values valid when the model QPF amount is none and when the QPF amount is 0.01 in. or more. Thus, rainfall is less likely to occur in those areas where the models indicate no rainfall than it is elsewhere in the domain; it is more likely to occur in those regions where rainfall is predicted, especially where the predicted rainfall amounts are largest.

The improved HRs, along with some skill shown in ROC and ROL diagrams, imply that the Eta and WRF models are better able to indicate the regions in which atmospheric processes are most favorable for convective rainfall (where the models generate enhanced amounts) than they are able to predict accurately the rainfall amounts that will be observed. The QPF–probability relationship evaluated in this note can be used by forecasters as guidance for issuing probabilistic forecasts from a single deterministic forecast. In addition, forecasters might be able to adjust ensemble-derived probability forecasts using a similar relationship.

The similar trends in HRs and in ROC and ROL curves with very different model configuration bias behaviors implies that the results can likely be generalized for all model forecasts for warm-season convective system rainfall, at least in the upper Midwest. It should be noted, however, that unusually bad bias errors could damage the relationship. GS01 found that in Eta runs not using a convective scheme, biases increased markedly as heavier rainfall thresholds were evaluated, varying from less than 0.5 at 0.01 in., to 1.249 at 0.50 in., and nearly tripling to 3.222 at 1.0 in. Such rapid increases in bias error at heavier thresholds indicate a

serious problem in such a model configuration, evidenced by much worse ETSs than in other configurations. As might be expected with that bias behavior, it was found that improved HRs with increasing QPF amount hold true to a point in an evaluation of those Eta runs, but a slight worsening then occurs when the subdomains are defined using the heaviest QPF bins (not shown). It is unlikely, however, that a model with such large bias errors would be used heavily in an operational setting.

Although the discussion in this note emphasized the QPF–probability relationship within single deterministic runs, additional testing showed that even higher HRs (and thus probability-of-rainfall forecasts) can be obtained if a forecaster examines the subdomain in two or more models where heavy rainfall forecasts are intersecting. The probability of rainfall, for instance, at any point for which both the Eta and WRF models forecast over 1.0 in. reaches 87%. This region, however, generally contains very few grid points, and so the frequency of this forecast would be small.

Future work will include expansion of the study to a much longer time frame that includes both cold-season and warm-season events over the entire United States using the coarser grid resolution present in the operational Eta Model in the late 1990s and early 2000s.

Acknowledgments. We gratefully acknowledge the Ames Laboratory of the U.S. Department of Energy for the use of a PC-cluster to run some of the Eta simulations. We thank Isidora Jankov and Christopher Anderson for their assistance with some of the statistical computations. Software to perform some of the ROC and ROL computations was kindly provided by Matthew Wandishin. The paper was substantially improved by the helpful comments of three anonymous reviewers. This study was supported by NSF Grants ATM-9908932 and ATM-0226059.

REFERENCES

- Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes Geophys.*, **8**, 401–417.
- Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multisensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.
- Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, **112**, 693–709.
- Buizza, R., A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999: Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Wea. Forecasting*, **14**, 168–189.
- Gallus, W. A., Jr., 1999: Eta simulations of three extreme precipitation events: Impact of resolution and choice of convective parameterization. *Wea. Forecasting*, **14**, 405–426.
- , and M. Segal, 2001: Impact of improved initialization of mesoscale features on convective system rainfall in 10-km Eta simulations. *Wea. Forecasting*, **16**, 680–696.

- Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 928–945.
- Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models*, *Meteor. Monogr.*, No. 46, Amer. Meteor. Soc., 165–170.
- Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725.
- Mesinger, F., Z. I. Janjić, S. Nickovic, D. Gavrilov, and D. G. Deaven, 1988: The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev.*, **116**, 1493–1518.
- Mullen, S. L., and R. Buizza, 2001: Quantitative precipitation forecasts over the United States by the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **129**, 638–663.
- Rogers, E., and Coauthors, 1998: Changes to the NCEP operational “early” Eta Analysis/Forecast System. NWS Technical Procedures Bulletin 447, National Oceanic and Atmospheric Administration/National Weather Service, 14 pp. [Available from National Weather Service, Office of Meteorology, 1325 East–West Highway, Silver Spring, MD 20910.]
- Schneider, R. S., N. W. Junker, M. T. Eckert, and T. M. Considine, 1996: The performance of the 29 km Meso Eta Model in support of forecasting at the Hydrometeorological Prediction Center. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J111–J114.
- Schwartz, B. E., and S. G. Benjamin, 2000: Verification of RUC-2 precipitation forecasts using the NCEP multisensor analysis. Preprints, *Fourth Symp. on Integrated Observing Systems*, Long Beach, CA, Amer. Meteor. Soc., 182–185.
- Skamarock, W. C., J. B. Klemp, and J. Dudhia, 2001: Prototypes for the WRF (Weather Research and Forecasting) Model. Preprints, *Ninth Conf. on Mesoscale Processes*, Ft. Lauderdale, FL, Amer. Meteor. Soc., J11–J15.
- Stensrud, D. J., and M. S. Wandishin, 2000: The correspondence ratios in forecast evaluation. *Wea. Forecasting*, **15**, 593–602.
- , 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.
- 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.
- 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 numerical weather prediction. *Bull. Amer. Meteor. Soc.*, **78**, 2599–2617.
- Wilks, D. S., 1990: Probabilistic quantitative precipitation forecasts derived from PoPs and conditional precipitation amount climatologies. *Mon. Wea. Rev.*, **118**, 874–882.