# MCS Rainfall Forecast Accuracy as a Function of Large-Scale Forcing

ISIDORA JANKOV AND WILLIAM A. GALLUS JR.

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

(Manuscript received 12 November 2002, in final form 16 July 2003)

ABSTRACT

The large-scale forcing associated with 20 mesoscale convective system (MCS) events has been evaluated to determine how the magnitude of that forcing influences the rainfall forecasts made with a 10-km grid spacing version of the Eta Model. Different convective parameterizations and initialization modifications were used to simulate these Upper Midwest events. Cases were simulated using both the Betts–Miller–Janjić (BMJ) and the Kain–Fritsch (KF) convective parameterizations, and three different techniques were used to improve the initialization of mesoscale features important to later MCS evolution. These techniques included a cold pool initialization, vertical assimilation of surface mesoscale observations, and an adjustment to initialized relative humidity based on radar echo coverage. As an additional aspect in this work, a morphology analysis of the 20 MCSs was included.

Results suggest that the model using both schemes performs better when net large-scale forcing is strong, which typically is the case when a cold front moves across the domain. When net forcing is weak, which is often the case in midsummer situations north of a warm or stationary front, both versions of the model perform poorly. Runs with the BMJ scheme seem to be more affected by the magnitude of surface frontogenesis than the KF runs. Runs with the KF scheme are more sensitive to the CAPE amount than the BMJ runs. A fairly well-defined split in morphology was observed, with squall lines having trailing stratiform regions likely in scenarios associated with higher equitable threat scores (ETSs) and nonlinear convective clusters strongly dominating the more poorly forecast weakly forced events.

## 1. Introduction

Since rainfall within mesoscale convective systems (MCSs) is a major source of warm season precipitation for the agriculturally important American midwest (Fritsch et al. 1986), and dangerous flash flooding is often associated with these systems (Doswell et al. 1996), the need for accurate precipitation forecasts for these events is clear. Accurate forecasting of MCS rainfall requires good predictions of the occurrence, timing, and location of the systems, and the potentially even greater challenge of forecasting the rainfall amount.

The forecasting of MCS rainfall is complicated by the fact that current grid spacing in operational numerical models is too coarse to explicitly simulate convection, requiring use of a convective parameterization (Stensrud and Fritsch 1994). Convective parameterizations have been found to strongly influence the simulated precipitation patterns (Wang and Seaman 1997) and affect the response of a model to changes in grid spacing (Gallus 1999) or soil moisture (Gallus and Segal 2000). In addition, other mesoscale features possibly acting as the main source for convective forcing (such as outflow boundaries) are poorly resolved in the models (Kain and Fritsch 1993; Stensrud and Fritsch 1994). The dominance of convective system rainfall and the importance of smaller-scale forcing mechanisms in the warm season result in precipitation skill scores remaining much lower in the warm season than in the cold season (e.g., Olson et al. 1995).

The main goal of this paper is to examine the larger-scale forcing associated with MCS events to determine how that forcing may influence the performance of a mesoscale model using different convective parameterizations or initialization modifications. For this purpose high-resolution (10-km grid spacing) Eta simulations of 20 MCSs in the Upper Midwest were examined. Cases were simulated using both the Betts–Miller–Janjić (BMJ) convective parameterization (Betts 1986; Betts and Miller 1986; Janjić 1994) and the Kain–Fritsch (KF) convective parameterization (Kain and Fritsch 1993). In addition, the response of the model to an application of three different techniques to improve the initialization of the mesoscale features important to later MCS evolution (Gallus and Segal 2001) was also examined. These techniques included a cold pool initialization (Stensrud et al. 1999), vertical assimilation of surface mesoscale observations, and an adjustment to initialized relative humidity based on radar echo coverage. As an

*Corresponding author address:* Isidora Jankov, Dept. of Geological and Atmospheric Science, 3010 Agronomy Bldg., Iowa State University, Ames, IA 50011.
E-mail: ijankov@iastate.edu

additional aspect in this work, a morphology analysis of the 20 MCSs was included. Data and methodologies used in the present study are discussed in section 2, results in section 3, with concluding discussion and a summary in the final section.

## 2. Data and methodology

Simulations of 20 warm season MCS events over the Upper Midwest were performed using a workstation version of the National Centers for Environmental Prediction (NCEP) Eta Model with 10-km grid spacing and 32 vertical levels. More details about the Eta Model can be found in Mesinger (1998), Janjić (1994), Black (1994), and Rogers et al. (1998). Ten of the 20 cases were initialized at 0000 UTC, and the other 10 at 1200 UTC. Cases were integrated over a 24-h period and over a domain of approximately 1000 km $\times$ 1000 km, centered over Iowa (using a Lambert conformal map projection with a central point latitude and longitude of 42°N and 93°W, respectively). Although Warner et al. (1997) have shown that errors from lateral boundary conditions can cause problems in such small domains, most of the significant precipitation in the 20 cases occurs in the first 6–18 h, so that serious problems from the limited domain should be minimized. For initialization and boundary conditions, output from 40-km NCEP Eta Model gridded binary (GRIB) format files was used. All integrations were performed using both the BMJ and the KF convective parameterizations.

The BMJ is an adjustment-type scheme, which means that it forces the model soundings at each point toward a reference profile of the temperature and specific humidity. The reference temperature profile is mainly defined based on a Betts–Miller (1986) set of observations while the reference-specific humidity profile is calculated based on an established temperature. When the model profiles of temperature and specific humidity are established, the next step compares them with the reference profiles and checks if the conservation of enthalpy ($C_pT + L_vq$, where $C_p$ is the heat capacity of air at constant pressure and $L_v$ is the latent heat of vaporization) is satisfied. Conservation of enthalpy ensures that the latent heat released through convection is proportional to the removal of water vapor. The scheme's structure favors activation in cases with significant amounts of moisture in low and midlevels and positive convective available potential energy (CAPE). Also, it is very important to mention that the activation of this convective scheme at a particular grid point is mainly determined by thermodynamics. Moistening due to vertical motion can favor activation of the scheme, but the scheme does not respond to vertical motion alone. A more thorough overview is provided in Baldwin et al. (2002).

On the other hand, the Kain–Fritsch convective scheme is designed to remove CAPE through vertical reorganization of mass. A CAPE calculation is performed by use of the traditional, undiluted parcel ascent method. This convective scheme consists of three parts: the convective trigger function, the mass flux formulation, and the closure assumptions.

The trigger function is based on checking the parcel for buoyancy at the calculated lifting condensation level (LCL) starting with the lowest 50-mb layer and repeating the same procedure up to 600 mb. The buoyancy criterion is satisfied if $T_{\mathrm{LCL}} + dT > T_{\mathrm{ENV}}$, where $T_{\mathrm{LCL}}$ is the parcel temperature at the LCL, $T_{\mathrm{ENV}}$ is the environmental temperature, and $dT$ is a temperature perturbation added to the parcel based on an assumption that convective development can be supported with a background upward motion. The perturbation formula is $dT = k[w_g - c(z)]^{1/3}$, where $k$ is a unit number with $\mathrm{Ks}^{1/3}$ $\mathrm{cm}^{-1/3}$ dimensions, $w_g$ is the mean grid-resolved vertical velocity at the LCL (cm $\mathrm{s}^{-1}$), and $c(z)$ is a threshold vertical velocity defined as $w_0(Z_{\mathrm{LCL}}/2000)$ for $Z_{\mathrm{LCL}} \leq$ 2000 m ($Z_{\mathrm{LCL}}$ is the height of the LCL above the ground and $w_0 = 2$ cm $\mathrm{s}^{-1}$) and $w_0$ for $Z_{\mathrm{LCL}} > 2000$ m. If a buoyant parcel is found, the scheme estimates the thermodynamic path to the cloud top. This means that surface convergence with the induced vertical motion may have a much bigger impact on the KF convective parameterization than on the BMJ convective parameterization. For example, in a case with a capping inversion suppressing convection and rather low relative humidity above the inversion, there may be an increased chance for the KF scheme to be activated compared to the BMJ. Regarding the mass flux formulation, the convective downdraft is represented with a plume model, which includes detrainment of mass and moisture. Finally, the closure assumption used by the Kain–Fritsch convective parameterization is that the scheme will rearrange mass in the column until 90% of the CAPE is removed. More details and recent updates about the Kain–Fritsch convective scheme can be found in Kain (2003).

In addition to varying the convective scheme used, three different types of adjustments to the initial conditions (cold pool initialization, vertical assimilation of mesoscale surface observations, and relative humidity adjustment based on radar echo coverage) were also investigated. The cold pool initialization (Stensrud et al. 1999) adjusts the temperature and moisture in the layer near the surface based on the presence of observed positive mesoscale pressure perturbations. These pressure perturbations are detected by objectively analyzing surface observations using the Barnes scheme (Barnes 1964, 1973) with two different weighting functions. One produces a relatively coarse but smooth pressure field and the other a mesoscale pressure field. The difference of the fields identifies areas of mesoscale pressure perturbation. When positive perturbations are detected in areas of rainfall, the temperature is decreased from the surface upward until the mesoscale surface pressure perturbation hydrostatically produced by the cooling matches the observed pressure perturbation.

A second technique that was used here to improve initial conditions was the vertical assimilation of the mesoscale surface observations. This technique uses surface mesoscale observations to adjust the initialized surface temperature and specific humidity toward the mesoscale observations, and then uses the model's own vertical eddy diffusion over a specified period of time, here chosen to be 3 h, to allow the surface data to influence a deep layer.

The third modification involves adjustment of the relative humidity based on radar echo coverage at initialization. The adjustment eliminates dry layers (relative humidity less than 80%) in the lower and middle troposphere (the layer in which the temperature is higher than $-10°C$) by setting the initial relative humidity to a minimum of 80% at grid points where a radar echo was detected. This technique is conceptually similar to the convective forcing procedure used by Rogers et al. (2000). It generally resulted in rapid activation of the convective scheme in areas where radar echoes existed. More details about all three adjustment techniques are found in Gallus and Segal (2001).

As a measure of forecast accuracy, the equitable threat score (ETS; Schaefer 1990), bias, and Brier score (BS) were used, where

$$\mathrm{ETS} = \frac{(\mathrm{CFA} - \mathrm{CHA})}{(F + O - \mathrm{CFA} - \mathrm{CHA})}, \qquad (1)$$

$$\mathrm{CHA} = O\frac{F}{V}, \qquad (2)$$

$$\mathrm{BIAS} = \frac{F}{O}, \quad \text{and} \qquad (3)$$

$$\mathrm{BS} = \frac{P}{V}. \qquad (4)$$

In the above equations, the right-hand-side variables indicate, respectively, the number of grid points at which (i) rainfall was correctly forecast to exceed the specified threshold (CFA); (ii) rainfall was forecasted to exceed the threshold (F); (iii) rainfall was observed to exceed the threshold (O); (iv) a correct forecast would occur by chance (CHA), where $V$ is the total number of evaluated grid points; and (V) an incorrect forecast was made ($P$). An ETS of 1 would occur with a perfect forecast, with lower values showing a less accurate forecast. BIAS can vary from 0 to $\gg$1. Values of BIAS significantly higher than 1 indicate that the model notably overpredicted areal coverage. Vice versa, BIAS values smaller than 1 mean that the model did not produce enough areas with rainfall exceeding a particular amount. The Brier score typically is used to evaluate probabilistic forecasts (Wilks 1995), and the equation shown is a simplified form that applies in a deterministic forecast. The Brier score has values between 0 and 1. It is important to note that a 0 value indicates a perfect forecast, but unlike ETS, BS is influenced by correct forecasts of ''no rain'' (rainfall less than a threshold), which typically are far more common than correct forecasts of rain.

Because Hamill (1999) has shown that high ETSs are often associated with relatively high BIASs, rigorous hypothesis testing that uses ETSs should include a BIAS-equalization adjustment. The adjustment proposed by Hamill was performed on the output from the 20 cases. ETSs were recalculated after BIASs were equalized. Also following Hamill (1999) a resampling methodology was used to perform the rigorous hypothesis testing. Hamill recommended resampling over the contingency table elements as opposed to resampling over already calculated ETS values. This procedure was followed exactly and repeated 1000 times for both the bias-adjusted and unadjusted contingency table elements.

It should be noted that all objective measures of skill offer only limited insight into the quality of a forecast, and individual measures can be misleading. As already discussed, ETS is often highest in events with a positive BIAS error, because displacement errors are common in rainfall forecasts. The Brier score may be lowest (least error) in easy forecasts where no rain occurs nor is predicted. Kain et al. (2003) have shown an example where ETSs favored one model but a detailed subjective evaluation by expert forecasters regarded another model to be better. An extensive subjective evaluation of the 20 events used in the present study agreed in part with Kain et al. (Jankov and Gallus 2004, hereafter JG). Observed 6-h accumulated precipitation fields for verification were used from the NCEP stage IV multisensor (Baldwin and Mitchell 1997) analysis.

To determine MCS morphology, hourly radar data from the National Climatic Data Center (NCDC) archive were used. Systems were classified based on the predominant radar structure observed over their lifetime. MCS morphology types considered included the following: TS representing trailing stratiform [squall line with stratiform rain following the leading convective line; e.g., Newton (1950), Fujita (1955), Ogura and Liou (1980), and Leary and Rappaport (1987)], LS for leading stratiform [squall line having predominantly nonconvective precipitation in advance of the convective line; e.g., Newton and Fankhauser (1964), Newton (1966), and Parker and Johnson (2000)], SL for a squall line (squall line with no appreciable precipitation behind or ahead of the line), NCs for nonlinear clusters (moderate to intense cells that are embedded within or surrounded by stratiform precipitation), ICs for isolated cells (isolated, individual, moderate to intense single cells), PS for parallel stratiform [a convective line that has stratiform precipitation primarily at one end of the line parallel to it; Parker and Johnson (2000)], and ST for stratiform with embedded cells (generally a nonconvective, large precipitation shield). Although all seven of these types were considered in the analysis, the 20 events in our sample fell into just four of the categories (TS, NC, PS, and SL).

TABLE 1. Observed values of omega, vorticity advection, temperature advection, frontogenesis, and divergence averaged over a 4° × 4° region. Values considered strong (weak) in each category are shown in boldface (italics). Strongly and weakly-forced cases are marked with one and two asterisks, respectively. Here C, c, N, n, R, r, P, and p correspond to cases characterized by high CAPE, low CAPE, low CIN, high CIN, high RH, low RH, high PW, and low PW, respectively. The prevailing morphology of each system is shown in the right column (see text for details).

| Case | 700-mb omega ($\mu$b s$^{-1}$) | 250–850-mb vorticity advection × 10$^{-9}$ (s$^{-2}$) | 850-mb temp advection × 10$^{-5}$ (K s$^{-1}$) | Surface fronto-genesis × 10$^{-1}$ (K m$^{-1}$ s$^{-1}$) | 200-mb divergence × 10$^{-5}$ (s$^{-1}$) | Morphology |
|---|---|---|---|---|---|---|
| 19 May 1998Cnp | −0.04 | **3.20** | *−2.35* | 0.47 | 0.90 | NC |
| 28 Jun 1998*C | **−1.85** | **3.50** | **5.65** | **1.44** | **5.25** | TS |
| 22 Jul 1998cn | −1.51 | 1.00 | 1.64 | **4.49** | *−0.10* | PS |
| 21 Aug 1998Nr | 0.44 | *0.40* | 0.53 | 0.30 | 0.80 | NC |
| 1 Jun 1999**cNR | −0.40 | 2.80 | *−0.13* | 0.71 | *0.60* | NC |
| 4 Jun 1999**r | *0.56* | *0.40* | 1.00 | 0.90 | *−0.50* | TS |
| 5 Jun 1999**Cp | *1.44* | 1.70 | 2.70 | *−0.06* | 1.80 | NC |
| 8 Jun 1999p | −0.25 | 1.20 | *−0.70* | 0.50 | 1.62 | NC |
| 10 Jun 1999*P | **−1.90** | **3.00** | *−0.10* | *0.20* | **2.60** | TS |
| 8 Jul 1999* | **−2.40** | **5.50** | 3.20 | **2.30** | **2.90** | TS |
| 18 Jul 1999P | 0.20 | 1.70 | **6.30** | **3.20** | *0.50* | NC |
| 20 Jul 1999**R | *1.91* | 1.20 | 5.40 | 0.30 | 2.00 | NC |
| 6 Aug 1999cNrp | −0.65 | 1.80 | 2.20 | *−1.66* | 2.50 | NC |
| 11 Aug 1999nR | *0.70* | 2.80 | 3.80 | *−0.14* | **3.20** | NC |
| 12 Aug 1999cnRP | −0.80 | *0.60* | **14.0** | *−0.07* | 0.88 | NC |
| 11 Jun 2000** | *1.60* | **6.10** | *−1.90* | 0.68 | *−0.50* | NC |
| 24 Jun 2000r | 0.00 | *0.30* | 4.56 | 0.60 | 0.80 | NC |
| 26 Jun 2000 | −0.12 | *0.40* | 3.50 | 1.40 | 2.60 | TS |
| 10 Jul 2000*CP | **−1.52** | 1.30 | **6.00** | 1.39 | **3.10** | TS |
| 12 Jul 2000*N | **−3.30** | 2.80 | **6.30** | **4.00** | 0.62 | SL |

## 3. Results

### a. *Variations in larger-scale forcing among events*

To investigate variations in the accuracy of the precipitation simulations, all of the events in this study were classified into three groups: strongly, moderately, and weakly forced, based on the magnitude of synoptic forcing measures at several levels [700-mb vertical velocity, differential absolute vorticity advection (difference in absolute vorticity advection between 250 and 850 mb), 850-mb temperature advection, surface frontogenesis, and 200-mb divergence]. Although, static stability affects the forcing from the differential absolute vorticity advection, its effects were diagnosed separately in the analysis of CAPE. It should be mentioned that high values of CAPE were almost always accompanied by high values of differential absolute vorticity advection.

The event classification first involved a calculation of a spatially averaged magnitude of the different forcings by using the 0-h operational Eta Model initialization output. The spatial averaging was performed within a 4° × 4° area centered over the centroid of the precipitation system around the time when substantial rainfall was first observed (initialization time in 15 out of 20 events). The magnitudes of all five forcing measures were ranked for all cases such that the five strongest values (25% of events) in each category were given a rank of 3, the weakest five values were given a rank of 1, and the remaining 50% of events given a rank of 2. The ranks for the different measures were summed and the five events with the highest (lowest) sum were considered to be strongly (weakly) forced. In the case of a tie (where several events had the same sum), vertical velocity (which should reflect the net result of the forcing) was used to break the tie. In addition, spatially averaged magnitudes of some thermodynamic characteristics [surface-based CAPE and convective inhibition (CIN), relative humidity (RH) averaged in the 1000–500-mb layer, and precipitable water (PW)] for all events were similarly evaluated. For CAPE and CIN, the deviations from the 1200 or 0000 UTC average were used to classify cases as high or low. Due to the fact that the events were initialized at two different times and that CAPE is usually much larger and CIN smaller at 0000 UTC than at 1200 UTC, the event classification for CAPE and CIN was slightly different. To evenly represent cases with different initialization times, the ''strong'' and ''weak'' categories consisted of the two events from runs initialized at 0000 UTC and two events from 1200 UTC runs having the highest (lowest) CAPE and CIN deviations from the 0000 or 1200 UTC average. In the cases of relative humidity and precipitable water, the highest (lowest) 20% (four cases) were considered to be strong (weak).

The goal of this analysis was to see if certain environments are more predictable than others, and if some favor a specific model configuration. For instance, the KF scheme includes a parameterized downdraft (unlike the BMJ), so that the rainfall patterns produced by the two schemes may differ. The magnitudes of the synoptic forcing mechanisms are presented in Table 1. The deviations from the average for the thermodynamic characteristics of the larger-scale environment are shown for 1200 (Table 2) and 0000 UTC (Table 3). Cases consid-

TABLE 2. Observed deviations from 1200 UTC averages of relative humidity, CAPE, PW, and CIN for the cases initialized at 1200 UTC. Values considered most (least) favorable for convection are shown in boldface (italics). Cases determined in general to be strongly and weakly forced are shown with one and two asterisks, respectively. Average values of each parameter for all events are shown in parentheses with the column headings.

| Case (1200 UTC) | RH (avg 53.7) (%) | CAPE (avg 382) ($m^2\ s^{-2}$) | PW (avg 26.2) (in.) | CIN (avg 92) J $kg^{-1}$ |
|---|---|---|---|---|
| 19 May 1998 | 1.5 | **443** | *−1.1* | *117* |
| 1 Jun 1999** | **21.6** | *−360* | 2.5 | **−86** |
| 4 Jun 1999** | *−9.9* | −232 | −0.5 | −20 |
| 5 Jun 1999** | −0.5 | **372** | *−22.7* | 18 |
| 8 Jun 1999 | −8.8 | 336 | *−2.3* | −3 |
| 8 Jul 1999* | −2.5 | −282 | 0.3 | −45 |
| 18 Jul 1999 | −0.8 | −121 | **16.6** | −62 |
| 20 Jul 1999** | **17.2** | −50 | 11.1 | 78 |
| 6 Aug 1999 | *−24.8* | *−382* | *−9.8* | **−92** |
| 11 Aug 1999 | **7.7** | 281 | 6.1 | *99* |

TABLE 3. As in Table 2 except for the cases initialized at 0000 UTC.

| Case (0000 UTC) | RH (avg 48.8) (%) | CAPE (avg 1365) ($m^2\ s^{-2}$) | PW (avg 35.2) (in.) | CIN (avg 62) J $kg^{-1}$ |
|---|---|---|---|---|
| 28 Jun 1998* | −0.1 | **810** | −2.5 | −12 |
| 22 Jul 1998 | 5.7 | *−610* | 2.2 | *14* |
| 21 Aug 1998 | *−11.2* | −36 | −0.9 | **−46** |
| 10 Jun 1999 | 8.9 | −180 | **2.9** | 9 |
| 12 Aug 1999 | **12.1** | *−439* | **6.1** | *128* |
| 11 Jun 2000** | −7.7 | −325 | −4.2 | −17 |
| 24 Jun 2000 | *−8.7* | −313 | −3.9 | −9 |
| 26 Jun 2000 | −2.1 | −180 | 0.7 | −16 |
| 10 Jul 2000* | 10.6 | **1538** | **3.9** | −20 |
| 12 Jul 2000* | −4.6 | −268 | −4.7 | **−41** |

ered in further analysis as strongly and weakly forced are marked with one and two asterisks, respectively, in the tables. Also, the last column in Table 1 defines the prevailing morphology type. In addition, the symbols in Table 1—C, c, N, n, R, r, P, and p—correspond to cases characterized by high CAPE, low CAPE, low CIN, high CIN, high RH, low RH, high PW, and low PW, respectively (more detailed information on these parameters is shown in Tables 2 and 3). It can be seen from the top row of Tables 2 and 3 that the average values of CAPE and precipitable water from the 0-h operational Eta Model analysis are higher for the cases initialized at 0000 UTC than for cases initialized at 1200 UTC. Also, as would be expected, the deviation from the average value of CAPE is generally larger for events initialized at 0000 UTC. In all three of these tables, the most (least) favorable 20% or 25% of values are highlighted in boldface (italics) so that the thresholds that apply in our sample can be estimated, allowing possible extension to other datasets. It is beyond the scope of the present study to determine how representative these thresholds are for the warm season over a large area.

### b. *Impact of larger-scale forcing on different model variants*

#### 1) COMPARISON OF CONVECTIVE SCHEMES

To accentuate possible differences in the rainfall forecast accuracy as a function of larger-scale environmental forcing, the analysis focused on the five most strongly and five most weakly forced events. An investigation of the differences in the behavior of runs using the two convective schemes (with standard initialization) under different large-scale forcings concentrated on calculated ETSs, BIASs, and Brier scores for both schemes under both magnitudes of forcing. Table 4 contains both unadjusted ETSs and ETSs adjusted to equalize BIAS for runs using the two convective schemes under both strong and weak forcing. The *p* values, obtained from statistical significance testing of the differences in ETSs between strongly and weakly forced events, are also included in this table. It is important to note that runs using both convective schemes have higher ETSs for strong forcing than for weak forcing, a result consistent with Stensrud et al. (2000) who examined two events. The *p* values for the difference (strong versus weak forcing) in both the adjusted and unadjusted values of ETS for runs with both convective schemes are frequently significant with 95% confidence. The exceptions, in the case of unadjusted values, are the heaviest two thresholds (precipitation greater than 0.25 in.) for the BMJ runs and the heaviest one (precipitation greater than 0.5 in.) for the KF runs. After adjusting to equalize BIAS, the *p* value for the KF runs at the lightest threshold increases to around 0.07.

The average BIAS as a function of the larger-scale forcing and the convective scheme is presented in Table 5. It can be seen that BIASs for runs using both schemes are lower in the strongly forced than in the weakly forced events for all thresholds except the heaviest one (BMJ) or two (KF). The BIAS error (difference from 1.0) is generally better in the KF runs for three out of five strongly forced events and for four out of five weakly forced events.

In addition to the above skill measures, the Brier score as a function of the larger-scale forcing and a different convective parameterization was calculated (Table 6). The BS may behave differently from ETS because it rewards correct forecasts of no rainfall the same as correct yes forecasts. The table shows BS is higher (greater errors) for weakly forced events for both schemes at all thresholds except the heaviest two, a result similar to that obtained using ETS. Again it appears that convective systems occurring with strong synoptic forcing are better simulated than those in weakly forced environments. Note also that because BS rewards correct forecasts of "no rain," BSs improve at the heavier thresholds where relatively few grid points are forecast or observed to have rainfall exceeding the thresholds. Because the BS showed similar behavior in the comparison

TABLE 4. ETS and $p$ values for runs using both convective schemes, for rainfall exceeding five precipitation thresholds, and for different magnitudes of the larger-scale forcing. Adjusted ETS (equalized BIAS) and corresponding $p$ values are presented in parentheses.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ ETS | 0.297 (0.293) | 0.273 (0.276) | 0.243 (0.245) | 0.176 (0.179) | 0.112 (0.111) |
| Weak BMJ ETS | 0.154 (0.154) | 0.127 (0.127) | 0.112 (0.112) | 0.089 (0.089) | 0.012 (0.012) |
| BMJ $p$ values | 0.050 (0.050) | 0.014 (0.012) | 0.008 (0.005) | 0.142 (0.113) | 0.120 (0.104) |
| Strong KF ETS | 0.293 (0.301) | 0.264 (0.285) | 0.226 (0.279) | 0.136 (0.185) | 0.062 (0.059) |
| Weak KF ETS | 0.166 (0.182) | 0.128 (0.136) | 0.106 (0.111) | 0.043 (0.063) | 0.021 (0.043) |
| KF $p$ values | 0.028 (0.069) | 0.014 (0.010) | 0.042 (0.008) | 0.044 (0.011) | 0.104 (0.425) |

of strongly and weakly forced events to ETS, and ETS is the most frequently used skill measure operationally at the present time, all further analysis and discussion will emphasize ETS values alone and the difference in ETS between different model versions and different forcing magnitudes.

An investigation of MCS morphology associated with different magnitudes of larger-scale forcing (Table 1) designated TS as the most frequent morphology type (four out of five cases, with one SL) under strong larger-scale forcing while in the case of weak forcing NC dominates (four NCs versus one TS case).

A similar analysis was performed treating upper-level and lower-level dynamical forcing separately. For this purpose a differential absolute vorticity advection between 250 and 850 mb was chosen as an example of the upper-level forcing, and surface frontogenesis was used to represent the low-level forcing. It should be noted that a sensitivity test was performed to determine the impact of approximating the differential vorticity advection term by using only the 500-mb vorticity advection, a common approximation in operational meteorology. Although for most cases the two measures did not differ substantially, enough impact occurred in a few cases so as to affect the categorization of events and the significance of the ETS difference computed. The use of the difference between the 250- and 850-mb vorticity advection resulted in a more pronounced difference in ETSs between the strongly and weakly forced events than when the 500-mb advection alone was used.

An analysis of ETS for runs with both convective schemes with strong and weak differential vorticity advection indicates results similar to those obtained for net strong and weak larger-scale forcing. Cases characterized by a strong differential vorticity advection receive higher ETSs compared to those characterized by weak forcing (Table 7), and ETSs are usually slightly higher for the BMJ run. By interpreting related $p$ values it can be seen that differences are statistically significant with 90% or 95% confidence for the first two precipitation thresholds considering the BMJ runs and the first three precipitation thresholds considering the KF runs. The same trend was obtained for both adjusted and unadjusted ETSs.

Regarding MCS morphology for the five events characterized by strong upper-level forcing, the dominant type is TS (three out of five cases) with two events classified as NC. For events characterized by weak differential vorticity advection, the same two morphologies are noted, NC in three cases and TS in two.

An analysis of the low-level dynamical forcing (represented by surface frontogenesis) shows interesting results (Table 8). Once again, ETSs are higher for events under strong forcing, but this time, the BMJ runs are impacted much more strongly than the KF runs. Statistically significant differences occur in the BMJ model runs for thresholds of 0.1 in. or less (adjusted and unadjusted values), but are not present in the KF runs. This might indicate that the magnitude of the low-level forcing has a noticeably bigger impact on the BMJ runs than on the KF. Based on the fact that vertical motion is actually a part of the KF scheme's triggering function, this finding is counterintuitive.

To better understand this finding, further investigation of the parameters presented in Tables 1–3 is necessary. All cases with strong frontogenesis also have moderate or strong values of 850-mb temperature advection and differential vorticity advection. Although the 200-mb divergence is weak in two cases, all events are characterized by moderate or strong ascent, so that the dynamical measures in general seem to support the development of rainfall. However, examination of thermodynamic measures, and in particular surface-based CAPE, suggests a possible cause for the more modest difference in ETSs in the KF runs. In four (out of five) cases characterized by strong surface frontogenesis,

TABLE 5. Average BIAS values for runs using both schemes, for rainfall exceeding five precipitation thresholds with different magnitudes of the larger-scale forcing.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ BIAS | 1.061 | 1.168 | 1.241 | 1.233 | 1.445 |
| Weak BMJ BIAS | 1.327 | 1.515 | 1.533 | 1.393 | 0.808 |
| Strong KF BIAS | 0.733 | 0.785 | 0.812 | 0.869 | 1.312 |
| Weak KF BIAS | 1.063 | 1.090 | 1.009 | 0.776 | 0.548 |

TABLE 6. As in Table 5 but for the Brier score.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ BS | 0.229 | 0.207 | 0.118 | 0.141 | 0.084 |
| Weak BMJ BS | 0.313 | 0.270 | 0.230 | 0.140 | 0.066 |
| Strong KF BS | 0.205 | 0.180 | 0.163 | 0.133 | 0.087 |
| Weak KF BS | 0.293 | 0.247 | 0.204 | 0.130 | 0.069 |

TABLE 7. As in Table 4 except for the differential absolute vorticity advection (250–850 mb).

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ ETS | 0.329 (0.329) | 0.278 (0.278) | 0.224 (0.224) | 0.151 (0.151) | 0.077 (0.077) |
| Weak BMJ ETS | 0.158 (0.149) | 0.149 (0.140) | 0.141 (0.142) | 0.088 (0.103) | 0.040 (0.052) |
| BMJ *p* values | 0.038 (0.032) | 0.055 (0.044) | 0.268 (0.270) | 0.410 (0.532) | 0.528 (0.660) |
| Strong KF ETS | 0.285 (0.289) | 0.242 (0.250) | 0.195 (0.217) | 0.118 (0.129) | 0.078 (0.087) |
| Weak KF ETS | 0.070 (0.081) | 0.064 (0.042) | 0.065 (0.071) | 0.069 (0.082) | 0.043 (0.054) |
| KF *p* values | 0.056 (0.059) | 0.038 (0.040) | 0.062 (0.042) | 0.646 (0.71) | 0.489 (0.631) |

CAPE is less than average. As will be shown later, the KF runs are sensitive to CAPE with large differences in ETSs between high and low CAPE events. In the fifth case, the KF run is significantly later with its initiation during the first 6 h of the forecast (JG), and this possible model spinup problem results in low ETSs. On the other hand, of the cases characterized by weak surface frontogenesis, two out of the five have higher than average CAPE. This findings demonstrates a difficulty with isolating the impact of a particular forcing measure with a limited sample of cases. Ideally, an investigation of the impact of surface frontogenesis on ETSs would require that cases be found where all measures were similar except for frontogenesis. Such an investigation would require vastly more cases than are available in the present study. Nonetheless, the comparison of the behavior of ETSs between the BMJ and KF runs should provide some information that could assist operational forecasters.

Concerning MCS morphology, strong fronotogenesis is associated with four different types (TS in two cases PS in one case, NC in one, and SL in one). For those cases characterized by weak low-level forcing, NC is the most common morphology (four out of five) with one TS. The dominance of the NC type in the weak forcing cases has been consistent among all of the dynamic measures investigated.

To obtain further insight into the impact of thermodynamical forcing on ETSs when different covective schemes are used, the same analysis was performed for events characterized as having a high/low deviation from average in surface-based CAPE (Table 9). It can be seen that in the case of the KF runs events characterized by high values of CAPE on average have noticeably higher ETSs compared to those characterized by low CAPE. In contrast, in the BMJ runs that difference is often negligible. Calculated *p* values for the KF runs indicate statistically significant differences for both unadjusted and adjusted ETSs for all thresholds except the heaviest one. The *p* values for the lighter three thresholds are the lowest obtained for any parameter, with confidence exceeding 99%. This clearly reveals a big positive impact of high CAPE on the KF runs, apparently related to the scheme's design. An examination of 6-h rainfall plots (not shown) indicates that the higher ETSs in high CAPE events than in low CAPE ones are associated with better forecasts of both the location and amount of rain. The improvement in forecast amount of rain is more pronounced than that of location, which implies that the sensitivity to CAPE is related to the KF scheme's vertical reorganization of mass. Only at the heaviest thresholds does CAPE seem to affect the BMJ-run ETSs.

For the events characterized by high CAPE, TS and NC are equally represented among MCS morphologies. In contrast, the dominant MCS structure for events characterized by low CAPE is NC (three out of four), with one PS case.

An analysis of surface-based CIN indicates that lower CIN generally leads to higher ETSs (both adjusted and unadjusted) for runs with both convective schemes, at all precipitation thresholds except for the two heaviest in the case of the BMJ and the heaviest one in the case of the KF (Table 10). However, none of the differences in ETS are statistically significant. Events characterized by low CIN have mostly the NC morphology (three out of four) with one SL event. Interestingly, NC appears to be the dominant morphology for cases characterized by high CIN too (three out of four), with one PS event.

In addition to the investigation of the impacts of these thermodynamic measures of forcing on ETS, the same type of analysis was performed for 1000–500-mb average RH and PW. For RH, more humid cases received higher unadjusted and adjusted ETSs for runs with both

TABLE 8. As in Table 4 except for surface frontogenesis. An asterisk next to a *p* value indicates a better ETS for weak forcing than for strong forcing.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ ETS | 0.290 (0.290) | 0.262 (0.264) | 0.224 (0.239) | 0.150 (0.156) | 0.090 (0.077) |
| Weak BMJ ETS | 0.160 (0.160) | 0.156 (0.156) | 0.130 (0.130) | 0.073 (0.073) | 0.042 (0.042) |
| BMJ *p* values | 0.009 (0.004) | 0.026 (0.022) | 0.089 (0.044) | 0.192 (0.166) | 0.854 (0.629) |
| Strong KF ETS | 0.235 (0.244) | 0.201 (0.197) | 0.165 (0.167) | 0.094 (0.126) | 0.039 (0.036) |
| Weak KF ETS | 0.186 (0.189) | 0.153 (0.171) | 0.120 (0.138) | 0.066 (0.080) | 0.048 (0.048) |
| KF *p* values | 0.324 (0.238) | 0.338 (0.626) | 0.410 (0.586) | 0.565 (0.266) | 0.914* (0.784) |

TABLE 9. As in Table 8 except for surface-based CAPE.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| High BMJ ETS | 0.233 (0.234) | 0.208 (0.211) | 0.180 (0.185) | 0.121 (0.138) | 0.062 (0.061) |
| Low BMJ ETS | 0.225 (0.225) | 0.214 (0.214) | 0.183 (0.183) | 0.066 (0.066) | 0.015 (0.015) |
| BMJ *p* values | 0.926 (0.332) | 0.948* (0.999) | 0.999* (1.0) | 0.492 (0.258) | 0.718 (0.488) |
| High KF ETS | 0.271 (0.293) | 0.250 (0.273) | 0.215 (0.267) | 0.132 (0.181) | 0.076 (0.069) |
| Low KF ETS | 0.130 (0.134) | 0.103 (0.114) | 0.085 (0.098) | 0.064 (0.069) | 0.049 (0.049) |
| KF *p* values | 0.002 (0.004) | 0.004 (0.006) | 0.002 (0.006) | 0.244 (0.054) | 0.646 (0.828) |

schemes at all precipitation thresholds (Table 11). By analyzing *p* values, it can be seen that differences in ETS between cases with high and low values of RH are statistically significant with 95% confidence for two moderate precipitation thresholds (0.05 and 0.1 in.). The *p* values are relatively low for most other thresholds as well. The morphology of events is similar to that present for high and low CIN, with all high RH events having the same NC structure and low RH events all being NC except for one TS case.

Due to the fact that RH is a function of temperature, it is not unusual to have events with high RH associated with a low amount of PW and vice versa. Thus, PW has also been analyzed (Table 12). Interestingly, differences in ETS for different amounts of PW are rather small for both schemes at all precipitation thresholds. Adjusted ETS values for the BMJ runs show that at some thresholds high PW cases are simulated better while at other thresholds, low PW events are better simulated. The differences in ETS are not statically significant. This implies a negligible impact of PW amount on the behavior of runs using the two convective parameterizations. Events characterized by high PW are equally divided between the NC and TS morphologies, while those with low PW are consistently NC.

Finally, for a broader view, a general analysis of synoptic features present in the five most and least accurate forecasts is discussed. For this purpose, the five events with the highest and lowest ETSs are used. The five cases with the highest ETS are 28 June 1998, 22 July 1998, 8 July 1999, 10 June 1999, and 11 June 2000, and the five with the lowest ETS are 1 June 1999, 4 June 1999, 18 July 1999, 20 July 1999, and 12 August 1999. Of note, all five cases with the highest ETSs were associated with a cold front moving from west to east into the domain at the time of initialization. Considering different forcing measures, higher values of vertical velocity and differential vorticity advection, as compared to cases with low ETSs are found for four of the five events. Similarly, surface frontogenesis is higher in three of the events. Thus, cold front events often appear to be associated with relatively large values of the dynamic forcing measures discussed earlier. On the other hand, all events with low ETS values were cases with elevated convection occurring to the north of a stationary or warm front. In these events, the magnitude of the dynamic forcing measures was generally less than that with the high ETS cases. Also, the cases were characterized by generally low CAPE and two of them by high CIN. These poorly forecast events appear to be affected both by unfavorable thermodynamic conditions and less than ideal dynamic ones. These cases did have much higher PW amounts than well-simulated events, but as discussed earlier, PW amount was not found to significantly influence the ETSs.

#### 2) COMPARISON OF DIFFERENT INITIALIZATION ADJUSTMENTS

The impact of different modifications to the initial conditions on skill scores also was examined for both strongly and weakly forced events. [A similar analysis of these events but without differentiation by synoptic forcing magnitude was performed in Gallus and Segal (2001).] Adjusted and unadjusted changes in ETS (compared to the appropriate BMJ or KF control run) for the different modifications (cold pool initialization, assimilation of mesoscale observations, and relative humidity increase) and for the two groups of events (strongly and weakly forced) are shown in Tables 13–15. It can be seen that the cold pool adjustment to the initial conditions for cases characterized as strongly forced does not on average produce improvement in the forecast (Table 13). In fact, differences in both adjusted and unadjusted ETSs for both schemes at all thresholds are negative but not statistically significant. The same analysis but

TABLE 10. As in Table 8 except for surface-based CIN.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Low BMJ ETS | 0.215 (0.218) | 0.185 (0.184) | 0.158 (0.160) | 0.068 (0.094) | 0.031 (0.032) |
| High BMJ ETS | 0.203 (0.203) | 0.151 (0.151) | 0.138 (0.138) | 0.072 (0.072) | 0.037 (0.037) |
| BMJ *p* values | 0.672 (0.536) | 0.209 (0.207) | 0.398 (0.389) | 0.942* (0.538) | 0.844* (0.799*) |
| Low KF ETS | 0.174 (0.165) | 0.146 (0.154) | 0.124 (0.139) | 0.078 (0.098) | 0.028 (0.039) |
| High KF ETS | 0.150 (0.150) | 0.119 (0.119) | 0.096 (0.110) | 0.072 (0.075) | 0.037 (0.054) |
| KF *p* values | 0.398 (0.536) | 0.289 (0.152) | 0.290 (0.285) | 0.891 (0.654) | 0.672* (0.536*) |

TABLE 11. As in Table 8 except for mean relative humidity in the 1000–500-mb layer.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| High BMJ ETS | 0.314 (0.297) | 0.290 (0.283) | 0.263 (0.260) | 0.207 (0.198) | 0.149 (0.152) |
| Low BMJ ETS | 0.236 (0.236) | 0.186 (0.186) | 0.159 (0.159) | 0.109 (0.109) | 0.025 (0.025) |
| BMJ *p* values | 0.148 (0.146) | 0.034 (0.028) | 0.051 (0.024) | 0.226 (0.242) | 0.120 (0.104) |
| High KF ETS | 0.316 (0.293) | 0.292 (0.269) | 0.253 (0.245) | 0.177 (0.198) | 0.085 (0.089) |
| Low KF ETS | 0.207 (0.208) | 0.157 (0.163) | 0.134 (0.144) | 0.069 (0.095) | 0.042 (0.048) |
| KF *p* values | 0.118 (0.159) | 0.044 (0.016) | 0.101 (0.008) | 0.426 (0.192) | 0.610 (0.932) |

for cases defined as weakly forced indicates that the cold pool modification does increase both adjusted and unadjusted ETSs for runs with both schemes at lighter thresholds (0.1 in. or less). Still, this enhancement is not large enough to be statistically significant.

The mesoscale observation adjustment in both the BMJ and KF runs has a larger impact in the strongly forced cases than in the weakly forced ones, with the impacts being greater at heavier rainfall thresholds (Table 14). It must be noted, though, that the adjusted and unadjusted ETSs occasionally differ substantially. For example, unadjusted ETS values for BMJ runs under strong forcing that used the mesoscale observation adjustment show an improvement for all thresholds, while adjusted ETS values decrease for light thresholds. These differences suggest that BIAS changes significantly when the mesoscale observations are assimilated. Very high *p* values calculated for all of the ETS differences indicate that the mesoscale observation adjustment does not impact in a statistically significant way simulations performed with the two different convective parameterizations under different forcing magnitudes. The nonsignificant improvements that do occur are greater in the strongly forced events. This result is somewhat discouraging since it would seem that weakly forced events would be the ones where an improved depiction of mesoscale structures would result in relatively large improvements in forecast accuracy.

Finally, when the relative humidity modification is applied to both strongly and weakly forced events in runs using the BMJ scheme, adjusted and unadjusted ETSs improve (Table 15) but not substantially enough to be statistically significant. For both groups of cases and for all thresholds, *p* values are very high. In the KF runs, the impacts of the modification are not as favorable. The humidity modification generally has a negative impact on the adjusted ETSs. However, the unadjusted ETSs typically improve when the humidity is modified in strongly forced cases, again implying a large increase in BIAS when the humidity adjustment is used. Although the changes in ETSs are often the largest occurring with any of the initialization modifications investigated, the impacts are not statistically significant.

## 4. Summary and discussion

Simulations of 20 warm season MCS events performed with a 10-km grid spacing and 32 vertical layers in the Eta Model were used to investigate variations in rainfall forecast accuracy as a function of larger-scale forcing and thermodynamic conditions. In addition to control runs performed using two different convective parameterizations (BMJ and KF) for all 20 events, runs with three different adjustments to the initial conditions were also performed for each event with each convective scheme. These adjustments included (i) the use of a cold pool initialization scheme (Stensrud et al. 1999), (ii) the inclusion of mesoscale surface observations of temperature and specific humidity using the model's own vertical eddy diffusion to assimilate data into a deeper layer, and (iii) elimination of dry layers by setting a minimum relative humidity threshold of 80% for all levels warmer than $-10°C$ in locations where a radar echo was present at the initialization time. All model variations were included in an investigation of differences in rainfall forecast accuracy among the cases. In addition, all cases were classified based on their morphology to examine any relationship between particular morphology types and the magnitude of larger-scale forcing and thermodynamics.

First, all available cases were classified into three groups: strongly, moderately, and weakly forced, based on the magnitude of the larger-scale forcing (e.g., differential vorticity advection, frontogenesis, etc.). Further analysis concentrated on the two ''extreme'' groups of cases, strongly and weakly forced. Variations in the

TABLE 12. As in Table 8 except for PW.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| High BMJ ETS | 0.164 (0.164) | 0.148 (0.148) | 0.126 (0.126) | 0.074 (0.074) | 0.032 (0.032) |
| Low BMJ ETS | 0.143 (0.147) | 0.145 (0.142) | 0.133 (0.133) | 0.050 (0.104) | 0.010 (0.049) |
| BMJ *p* values | 0.568 (0.692) | 1.0 (0.830) | 0.866* (0.854*) | 0.549 (0.468*) | 0.479 (0.484*) |
| High KF ETS | 0.195 (0.196) | 0.165 (0.187) | 0.148 (0.167) | 0.101 (0.116) | 0.048 (0.049) |
| Low KF ETS | 0.189 (0.188) | 0.164 (0.166) | 0.132 (0.143) | 0.074 (0.098) | 0.046 (0.063) |
| KF *p* values | 0.979 (0.925) | 1.0 (0.738) | 0.804 (0.679) | 0.642 (0.760) | 0.978 (0.800*) |

TABLE 13. Change in ETS from the appropriate control run due to application of the cold pool adjustment to the initial conditions, and $p$ values for model runs with both convective schemes for different magnitudes of larger-scale forcing at five precipitation thresholds. Changes in adjusted ETS and corresponding $p$ values are presented in parentheses. Asterisk notation is as in Table 8.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ | −0.002 (−0.010) | −0.011 (−0.009) | −0.011 (−0.009) | −0.009 (−0.008) | −0.008 (−0.007) |
| Strong $p$ values | 1.0* (0.882*) | 0.926* (0.910*) | 0.876* (0.908*) | 0.932* (0.936*) | 0.869* (0.934*) |
| Weak BMJ | +0.030 (+0.029) | +0.027 (+0.039) | +0.014 (+0.034) | −0.016 (−0.014) | −0.007 (−0.006) |
| Weak $p$ values | 0.646 (0.699) | 0.554 (0.334) | 0.684 (0.354) | 0.624* (0.718*) | 0.584* (0.552*) |
| Strong KF ETS | −0.011 (−0.008) | −0.009 (−0.010) | −0.008 (−0.012) | −0.001 (−0.007) | −0.001 (−0.001) |
| Strong $p$ values | 0.780* (0.924) | 0.944* (0.842*) | 0.956* (0.856*) | 0.999* (0.926*) | 0.984* (1.0*) |
| Weak KF | +0.013 (+0.002) | +0.012 (+0.005) | −0.004 (+0.015) | −0.019 (−0.008) | −0.004 (−0.016) |
| Weak $p$ values | 0.859 (1.0) | 0.868 (1.0) | 1.0* (0.788) | 0.692* (0.849*) | 1.0* (0.610*) |

accuracy of rainfall simulations were then studied to determine if certain larger-scale environments are more predictable than others, and if some favor a specific model configuration. For the basic analysis of variations in model performance under different magnitudes of larger-scale forcing, three different accuracy and skill measures (ETS, BIAS, and Brier score) were used. All of these measures led toward the same conclusion that the skill for runs using both convective schemes is significantly higher for the strongly forced than for the weakly forced events. This result is consistent with Stensrud et al. (2000) but valid for a much larger sample of cases.

A statistical significance testing resampling method suggested by Hamill (1999) was performed on the results. The testing was performed using both unadjusted ETS values and values adjusted to equalize BIAS, as suggested by Hamill. Although differences in BIASs resulted in some differences between adjusted and unadjusted ETSs, in general the changes were small and did not impact conclusions for most of the analysis undertaken. A similar analysis was performed separately for upper-level dynamic forcing, low-level dynamic forcing, and high/low values of CAPE, CIN, RH, and PW.

To represent upper-level dynamic forcing, the difference between the 250- and 850-mb absolute vorticity advection (differential absolute vorticity advection) was used. Generally, events characterized by strong differential absolute vorticity advection receive higher skill scores than those characterized by weak vorticity advection. Corresponding $p$ values indicate that differences in ETSs are statistically significant with 95% confidence for lighter precipitation thresholds for both schemes. The same results were obtained for both adjusted and unadjusted ETSs.

To represent a low-level forcing mechanism, surface frontogenesis was examined. ETSs were again higher for events under strong low-level forcing than weak forcing, but the difference between the magnitudes of the forcing was noticeably larger for the BMJ runs than for the KF runs. Statistically significant differences were present in the BMJ runs at light and moderate thresholds for both adjusted and unadjusted values, but were not present for the KF runs. This result implies that the magnitude of the low-level forcing has a noticeably larger impact on the BMJ scheme than on the KF scheme. It was shown, however, that a sensitivity to CAPE amount may have influenced the KF results. In general, though, the model using both convective schemes does a much better job for strong larger-scale dynamical forcing at both levels.

A similar analysis for different CAPE amounts showed that CAPE has a significant positive impact on runs using the KF scheme, but little or no impact on the BMJ run. The influence on the KF runs could be expected based on the scheme design. Although differences were not statistically significant, low CIN generally led toward higher ETSs for both schemes at all thresholds except for the heaviest one.

An analysis of events with high RH showed higher ETSs for runs using both schemes than in low RH events, with statistically significant differences at the moderate thresholds. Finally, it was found that variations in PW do not have a substantial impact on ETSs

TABLE 14. As in Table 13 except for the mesoscale observation adjustment to the initial conditions.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ | +0.014 (−0.002) | +0.022 (−0.008) | +0.031 (+0.016) | +0.028 (+0.026) | +0.029 (+0.008) |
| Strong $p$ values | 0.846 (0.992*) | 0.704 (0.918*) | 0.592 (0.789) | 0.704 (0.640) | 0.718 (0.976) |
| Weak ETS | +0.012 (−0.005) | +0.003 (+0.011) | −0.002 (+0.011) | −0.005 (+0.019) | +0.004 (−0.0003) |
| Weak $p$ values | 0.924 (0.958*) | 1.0 (0.792) | 0.896* (0.784) | 0.892* (0.718) | 0.988 (−0.990*) |
| Strong KF | −0.006 (−0.012) | −0.007 (−0.005) | +0.009 (−0.032) | +0.036 (+0.011) | +0.036 (+0.017) |
| Strong $p$ values | 0.934* (0.789*) | 0.946* (0.284*) | 0.938* (0.458*) | 0.502 (0.884) | 0.540 (0.776) |
| Weak KF | +0.012 (−0.001) | +0.004 (−0.003) | +0.002 (−0.006) | −0.007 (−0.004) | +0.012 (−0.004) |
| Weak $p$ values | 0.896 (0.992*) | 1.0 (1.0*) | 1.0 (0.906*) | 0.928* (1.0*) | 0.748 (1.0*) |

TABLE 15. As in Table 13 except for the RH adjustment to the initial conditions.

| Threshold (in.) | 0.01 | 0.05 | 0.1 | 0.25 | 0.5 |
|---|---|---|---|---|---|
| Strong BMJ | +0.017 (+0.004) | +0.017 (+0.008) | +0.020 (+0.015) | +0.032 (+0.020) | +0.038 (+0.041) |
| Strong $p$ values | 0.797 (1.0) | 0.816 (0.928) | 0.772 (0.764) | 0.597 (0.734) | 0.414 (0.430) |
| Weak BMJ | +0.031 (+0.030) | +0.022 (+0.022) | +0.018 (+0.018) | +0.009 (+0.009) | +0.016 (+0.011) |
| Weak $p$ values | 0.694 (0.694) | 0.582 (0.566) | 0.450 (0.477) | 0.818 (0.811) | 0.322 (0.322) |
| Strong KF | +0.023 (−0.008) | +0.030 (−0.015) | +0.030 (−0.034) | +0.040 (+0.014) | +0.023 (+0.088) |
| Strong $p$ values | 0.618 (0.934*) | 0.552 (0.724*) | 0.582 (0.500*) | 0.496 (0.910) | 0.718 (0.002) |
| Weak KF | +0.007 (−0.007) | −0.007 (−0.010) | −0.009 (−0.003) | −0.020 (−0.015) | −0.003 (−0.018) |
| Weak $p$ values | 0.982 (0.908*) | 0.938* (0.818*) | 0.944* (1.0*) | 0.634* (0.708*) | 0.988* (0.660*) |

for runs using both schemes. Of all parameters examined, the amount of PW had the least influence on ETS.

A broad analysis of synoptic situations related to accurately and inaccurately simulated events was also performed. The analysis was based on the five events with the highest and lowest ETS values. All five well-simulated events included a cold front moving from west to east into the domain at the time of initialization. These cases were characterized by generally higher values of vertical velocity, differential vorticity advection, and surface frontogenesis compared to the poorly simulated events. On the other hand, all events with low ETS values appeared to be cases with elevated convection north of a stationary or warm front. These events also were characterized by generally low CAPE, which may have contributed to the low ETSs.

Regarding MCS morphology, the most frequent types observed were TS and NC. The TS type dominated for events characterized by strong dynamic forcing and often by high CAPE. Therefore, these systems were generally simulated better, at least in terms of rainfall accuracy, as measured by ETS and Brier score. On the other hand, events characterized by weak dynamic forcing and high CIN were almost always NC.

The impact on ETS of all three different adjustments to the initial conditions was also examined for both extremes of forcing. Despite some increases in ETSs for the mesoscale observation and relative humidity adjustments, statistically significant improvement was not obtained by using any of those adjustments, a result differing from Gallus and Segal (2001), who examined the full set of 20 cases using paired $t$ tests and Wilcoxon rank tests.

Knowledge that certain larger-scale environments might be better simulated than others, or might favor a specific model configuration, can be valuable for operational forecasting. Our results suggest that model forecasts will perform better when larger-scale forcing is strong. When forcing is weak, which may be the case in midsummer situations north of a warm or stationary front, both versions of the model will likely perform poorly. Runs with the BMJ scheme seem to be more affected by surface frontogenesis than the KF runs, and forecasters might emphasize the BMJ predictions more than the KF in a case of strong frontogenesis, and place less emphasis on the BMJ run in a weak frontogenesis case. Runs with the KF scheme may be more sensitive to CAPE than the BMJ runs and this information could also influence a forecast.

Clearly, a larger database should be investigated to extend and strengthen these findings. Because there was a fairly well-defined split in morphology, with TS likely in scenarios associated with higher ETSs and NC strongly dominating the more poorly forecast weakly forced events, future work should address in more detail the role of morphological evolution on model performance.

*Acknowledgments.* The authors would like to thank Dr. William Gutowski, Dr. Raymond Arritt, and Chris Anderson for useful suggestions and Aaron Todd for assisting in the MCS morphology classification. This work was supported by NSF Grants ATM-9908932 and ATM-0226059, along with NSF-REU Grant ATM-9908932/REU.

## REFERENCES

Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology,* Long Beach, CA, Amer. Meteor. Soc., 54–55.

——, J. S. Kain, and M. P. Kay, 2002: Properties of the convection scheme in NCEP's Eta Model that affect forecast sounding interpretation. *Wea. Forecasting,* **17,** 1063–1079.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather maps. *J. Appl. Meteor.,* **3,** 396–409.

——, 1973: Mesoscale objective map analysis using weighted time series observations. NOAA Tech. Memo. ERL NSSL-62, 60 pp. [NTIS COM-73-10781.]

Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.,* **112,** 677–692.

——, 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.

Black, T. M., 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Doswell, C. A., III, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting,* **11,** 560–581.

Fritsch, J. M., R. J. Kane, and C. R. Chelius, 1986: The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. *J. Climate Appl. Meteor.,* **25,** 1333–1345.

Fujita, T. T., 1955: Results of detailed synoptic studies of squall lines. *Tellus,* **7,** 405–436.

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, 2000: Sensitivity of forecast rainfall in a Texas convective system to soil moisture and convective parameterization. *Wea. Forecasting,* **15,** 509–526.

——, and ——, 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., 1999: Hypothesis test for evaluating numerical precipitation forecasts. *Wea. Forecasting,* **14,** 155–167.

Janjić, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

Jankov, I., and W. A. Gallus Jr., 2004: Contrast between good and bad forecasts of warm season MCS rainfall. *J. Hydrol.,* in press.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.,* **43,** 170–181.

——, and J. M. Fritsch, 1993: The role of convective ''trigger function'' in numerical forecasts of mesoscale convective systems. *Meteor. Atmos. Phys.,* **49,** 93–106.

——, M. E. Baldwin, P. R. Janish, S. J. Weiss, M. P. Kay, and G. W. Carbin, 2003: Subjective verification of numerical models as a component of a broader interaction between research and operations. *Wea. Forecasting,* **18,** 847–860.

Leary, C. A., and E. N. Rappaport, 1987: The life cycle and internal structure of a mesoscale convective complex. *Mon. Wea. Rev.,* **115,** 1503–1527.

Mesinger, F., 1998: Comparison of quantitative precipitation forecasts by the 48- and 29-km Eta Model: An update and possible implications. Preprints, *12th Conf. on Numerical Weather Prediction,* Phoenix, AZ, Amer. Meteor. Soc., J22–J23.

Newton, C. W., 1950: Structure and mechanism of the prefrontal squall line. *J. Meteor.,* **7,** 210–222.

——, 1966: Circulation in large cumulonimbus. *Tellus,* **18,** 699–712.

——, and J. C. Fankhauser, 1964: On the movements of convective storms, with emphasis on size discrimination in relation to water-budget requirements. *J. Appl. Meteor.,* **3,** 651–668.

Ogura, Y., and M. T. Liou, 1980: The structure of a mid-latitude squall line: A case study. *J. Atmos. Sci.,* **37,** 553–567.

Olson, D. A., N. W. Junker, and B. Korty, 1995: Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Wea. Forecasting,* **10,** 498–511.

Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 3413–3436.

Rogers, E., and Coauthors, 1998: Changes to the NCEP operational ''early'' Eta Analysis/Forecast System. NWS Tech. Procedures Bull. 477, National Oceanic and Atmospheric Administration/National Weather Service, 14 pp. [Available form Office of Meteorology, National Weather Service, 1325 East–West Highway, Silver Spring, MD 20910.]

Rogers, R. F., J. M. Fritsch, and W. C. Lambert, 2000: A simple technique for using radar data in the dynamic initialization of a mesoscale model. *Mon. Wea. Rev.,* **128,** 2560–2574.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting,* **5,** 570–575.

Stensrud, D. J., and J. M. Fritsch, 1994: Mesoscale convective systems in weakly forced large-scale environments. Part III: Numerical simulations and implications for operational forecasting. *Mon. Wea. Rev.,* **122,** 2084–2104.

——, G. S. Manikin, E. Rogers, and K. E. Mitchell, 1999: Importance of cold pools to NCEP mesoscale Eta Model forecasts. *Wea. Forecasting,* **14,** 650–670.

——, J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensembles of mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 2077–2107.

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.