The Use of a Modified Ebert–McBride Technique to Evaluate Mesoscale Model QPF as a Function of Convective System Morphology during IHOP 2002

JEREMY S. GRAMS* AND WILLAM A. GALLUS JR.

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

STEVEN E. KOCH, LINDA S. WHARTON, AND ANDREW LOUGHE

NOAA Research–Forecast Systems Laboratory, Boulder, Colorado

ELIZABETH E. EBERT

Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

(Manuscript received 29 July 2004, in final form 20 October 2005)

ABSTRACT

The Ebert–McBride technique (EMT) is an entity-oriented method useful for quantitative precipitation verification. The EMT was modified to optimize its ability to identify contiguous rain areas (CRAs) during the 2002 International H_2O Project (IHOP). This technique was then used to identify systematic sources of error as a function of observed convective system morphology in three 12-km model simulations run over the IHOP domain: Eta, the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5), and the Weather Research and Forecasting (WRF). The EMT was fine-tuned to optimize the pattern matching of forecasts to observations for the scales of precipitation systems observed during IHOP. To investigate several error measures provided by the EMT, a detailed morphological analysis of observed systems was performed using radar data for all CRAs identified in the IHOP domain. The modified EMT suggests that the Eta Model produced average rain rates, peak rainfall amounts, and total rain volumes that were lower than observed for almost all types of convective systems, likely because of its production of overly smoothed and low-variability quantitative precipitation forecasts. The MM5 and WRF typically produced average rain rates and peak rainfall amounts that were larger than observed in most linear convective systems. However, the rain volume for these models was too low for almost all types of convective systems, implying a sizeable underestimate in areal coverage. All three models forecast rainfall too far northwest for linear systems. The results for the WRF and MM5 are consistent with previous observations of mesoscale models run with explicit microphysics and no convective parameterization scheme, suggesting systematic problems with the prediction of mesoscale convective system cold pool dynamics.

1. Introduction

Summertime convective systems are among the most difficult weather events for operational meteorologists and numerical models to predict. Verification of a quantitative precipitation forecast (QPF) made by a fine-grid numerical model for these small-scale features can be just as difficult. Standard grid-based measures often result in scores that are not consistent with the subjective impression of the forecaster. Traditional verification statistics severely penalize a precipitation system that may have been forecast with a small positional error or incorrect shape, with resultant low correlation coefficients, high root-mean-square errors (rmse), and poor values of categorical statistics (Ebert and McBride 2000; Baldwin and Wandishin 2002). This type of forecast could still be useful to a forecaster or modeler if the model has known biases with its QPF. Fine-resolution models, which subjectively can often be of more value to an operational forecaster, are typically

* Current affiliation: NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma.

Corresponding author address: Jeremy Grams, NOAA/NWS/NCEP Storm Prediction Center, 1313 Halley Circle, Norman, OK 73069.
E-mail: jeremy.grams@noaa.gov

penalized more for spatial errors than coarser models (e.g., Gallus 2002; Mass et al. 2002; Kain et al. 2003). Common verification measures, such as the equitable threat score, reward smoothly varying forecast models over those with relatively high amplitude structures (Baldwin et al. 2001). Therefore, operational models have tended to be designed to produce smoothly varying QPFs, despite the preference of some human forecasters for more realistic-looking detail and the increasing simulation by research models of finer representations of QPF.

Numerous approaches have been applied to deal with the deficiencies of traditional verification methods (e.g., Du et al. 2000; Zepeda-Arce et al. 2000; Davis et al. 2006a). One such approach is the Ebert–McBride technique (EMT), which employs the concept of matching individual forecast and observed areas (Ebert and McBride 2000, hereafter EM2000). The technique utilizes contiguous rain areas (CRAs), defined as the areas of contiguous observed and forecast rainfall enclosed within a specified isohyet. A displacement is performed using an objective pattern-matching technique to optimally align the forecast with the observations. The EMT method was originally developed for application to synoptic-scale precipitation systems in the Australian region. The current study adapts the EMT to mesoscale convective systems (MCSs) characteristic of the central United States during the warm season.

The International H_2O Project (IHOP), which took place from 16 May to 26 June 2002, was designed to help improve the understanding and prediction of QPFs. High-resolution model datasets produced for this project offered the opportunity to investigate precipitation forecast accuracy as a function of convective system morphology. Around 175 MCSs were identified during the 4-week IHOP period for 12-km grid spacing runs of 1) the National Centers for Environmental Prediction (NCEP) Eta Model (Black 1994; Janjic 1994; Rogers et al. 1998) using the Betts–Miller–Janjic (BMJ) convective parameterization (Betts 1986; Betts and Miller 1986; Janjic 1994); 2) the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5, version 3.5; Grell et al. 1995); and 3) the Advanced Research Weather Research and Forecasting (WRF) model, version 1.3 (Skamarock et al. 2001). The Eta utilized its own continually cycling, three-dimensional variational data assimilation (3DVAR) scheme (Rogers et al. 1998). The MM5 and WRF were both initialized with the "Hot Start" procedure (McGinley and Smart 2001; Shaw 2004) developed for the National Oceanic and Atmospheric Administration (NOAA) Forecast Systems Laboratory (FSL) Local Analysis and Prediction System (Albers et al. 1996). The Hot Start is comprised of two parts: a cloud analysis and a dynamical balancing scheme. The cloud analysis diagnoses estimates vertical velocity profiles based on cloud type, depth, horizontal scale, and stability criteria. These estimates are then used as constraints in a 3DVAR system to develop model initial conditions that are in dynamic balance with the observed cloud field while maintaining consistency with the observations. Neither the MM5 nor the WRF employed a convective parameterization scheme. Jankov et al. (2005) found that in WRF simulations using 12-km grid spacing, the use of the Hot Start procedure resulted in rainfall forecasts in fully explicit simulations at least as skillful as those from runs using convective parameterizations. Each model was run every 6 h, and only the first 6-h forecast period was analyzed by the EMT.

All rain systems producing a volume of rainwater exceeding 3.13×10^8 m^3 (i.e., 1 in. of rain over a 1° by 1° box; 1 in. = 25.4 mm) in a 6-h forecast period were identified. Although traditional grid point–to–grid point objective skill measures were computed for the full domain, the focus of our analysis was on various skill measures applicable to the EMT-identified CRAs. An analysis of these parameters was performed as a function of the convective system morphology. This analysis should reveal whether certain types of systems, such as linear squall lines with trailing stratiform rain areas, had larger errors than other types within each model.

For every CRA identified, a corresponding observed system was classified using radar-based MCS characteristics. This radar-based morphology used 2-km composite base reflectivity radar imagery available from the Next-Generation Radar (NEXRAD) Information Dissemination Systems (NIDS; Baer 1991). The radar-based convective systems were divided into seven general types. These types were continuous linear (CL), continuous linear bowing (CLB), continuous nonlinear (CNL), discontinuous areal (DA), isolated cells (IC), orographically fixed (OF), or false alarms (FA). Linear systems were subclassified as having trailing stratiform (TS), leading stratiform (LS), parallel stratiform (PS), or combinations of the three types based on the classification system presented by Parker and Johnson (2000). The evolutionary characteristics of squall lines were further characterized as back building (BB), broken areal (BA), broken line (BL), or embedded areal (EA), following Bluestein and Jain (1985).

The overriding goal of this study was to use the EMT objective verification measures in concert with an observed morphological classification scheme to reveal systematic errors for certain types of MCSs. This paper

is organized as follows. Section 2 provides an overview of the EMT adapted from EM2000. Section 3 discusses the modifications that were made to optimize the performance of the EMT for central U.S. convective systems on the time scale of 6 h. Section 4 describes how the convective system classification schemes were developed and used. Section 5 presents the distribution of observed MCS types. Section 6 shows statistical results from the modified EMT applied to the Eta, MM5, and WRF models. A summary and conclusions follow in section 7.

2. Overview of EMT

The aim of the EMT approach is to verify to what extent the forecast entity has the same location, shape, and magnitude as the observed one (EM2000), with resulting error statistics based on the properties of each entity. Ebert and McBride (1998) first introduced the CRA as the area of contiguous observed and/or forecast rainfall enclosed within a specified isohyet. Typically, a CRA is initially identified as an overlap between an observed and forecast entity, especially as rainfall systems increase in size and magnitude. However, separated forecast and observed entities can be successfully matched if the forecast entity is within the specified search radius of the observed one. After matching, what started out as two rain areas then becomes a single CRA.

CRA verification utilizes pattern-matching techniques to horizontally translate the forecast entity over the observed one. The best match can be determined in a number of ways, usually by maximizing the correlation coefficient or by minimizing the total squared error. The forecast is permitted to shift within an expanded box enclosing the CRA (the maximum distance allowed between the forecast and observed areas, beyond which it is assumed the two areas are unrelated). Several user-defined parameters (discussed in section 3) can be adjusted to define the temporal and spatial scale of the CRA, the pattern-matching process, and how verification statistics are calculated.

Figures 1 and 2 show examples of CRA output from 0000 UTC 13 June 2002 for the Eta and WRF 6-h forecast of precipitation, respectively, in the top left, with the smoothed (section 3f) NCEP stage IV 6-h accumulated precipitation product (Baldwin and Mitchell 1997) in the bottom-left panel. A displacement vector (arrow with boldface outline) is determined by shifting the forecast entity to maximize the correlation coefficient between the forecast and observed entities. Various measures of error were determined before and after displacement (shown at the right in Figs. 1 and 2).

3. Modified EMT parameters

The EMT objectifies the intuitive process of pattern matching. It is therefore important to choose values of parameters that give the best agreement between the objective pattern matching and the investigator's visual interpretation. One of the advantages of this technique is that a variety of arbitrary parameters can be tuned based on the needs of the user. In our case, modifications were made for the purpose of gathering statistical information for central U.S. convective systems.

Some of the parameters determine how the CRA is chosen in order to make it most meaningful for the time and space scales of interest. These include the rainfall threshold, critical water mass threshold, and minimum area threshold. Other choices influence the way the verification is done. These include the pattern-matching criterion (maximizing the correlation coefficient versus minimizing the total squared error), search radius within which a forecast object can search for a corresponding observed object, tolerance for allowing a portion of the forecast object to shift off the domain grid, and filtering of observations. In addition, improvements were made in the code to redefine the areas over which some error measures and diagnostics were computed. Table 1 summarizes sensitivity tests performed raising or lowering the CRA rainfall, critical mass, and search radius thresholds. The following subsections provide greater detail on each modified EMT parameter.

a. CRA rainfall threshold

For 24-h QPF verification, EM2000 used a CRA critical rainfall threshold of 5 mm (~0.20 in.) per day for the minimum accumulation required for a grid point to be considered part of a CRA. For our purposes, a critical threshold of 0.25 inches for 6 h was found to work reasonably well at identifying an MCS, as might be expected for the higher rain rates accompanying springtime MCSs in the central United States compared with those found over a full year in Australia.

The CRA rainfall threshold was the most critical element for the inclusion or division of multiple objects in a CRA. Since pattern matching generally requires both a forecast and observed entity, CRAs are model dependent. This can make it difficult to individually compare statistical results from different models. Gallus (1999) found that the Eta using the BMJ convective scheme often depicted relatively large areas of contiguous low-to-moderate rainfall because of the design of the scheme. High diffusion (filtering) in NCEP's operational Eta runs may also contribute to this tendency (W. Skamarock 2004, personal communication). These

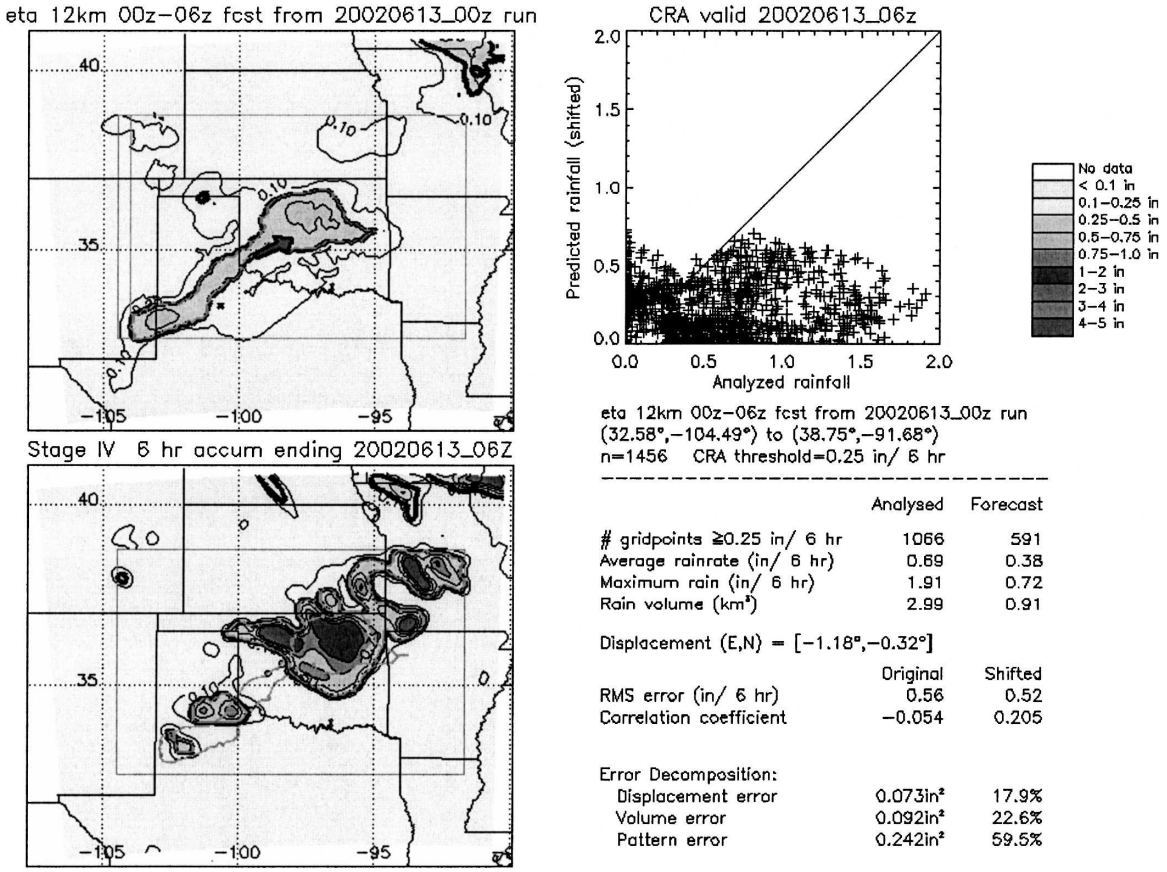

FIG. 1. Example of CRA output from the EMT for the 0000 UTC run of the 12-km Eta Model on 13 Jun 2002. (top left) The 6-h model forecast of rain above the 0.25-in. threshold is outlined in dark gray. Displacement vector (arrow outlined in boldface) shows computed displacement of forecast rain area to the northeast. (bottom left) Stage IV 6-h observed rainfall accumulation above the 0.25-in. threshold is outlined in drak gray, with the shifted forecast overlaid in light gray. Outer box shows the area over which CRA statistics (shown to the right) were calculated. (top right) Graph showing point-to-point verification of the shifted forecast rainfall vs observed rainfall. (bottom right) Table showing various statistical measures used in the study. The legend shows the thresholds for the 6-h rainfall accumulations.

smoothed patterns do not resemble typical observed rainfall patterns during warm season convective episodes. An example of this can be found by comparing Figs. 1 and 2. One can see the broad area of low-to-moderate rainfall forecast in the Eta versus the intense, but small, area of rainfall forecast in the WRF.

An overly broad forecast rainfall area can be responsible for two or more distinct observed systems getting combined into one large CRA. In this type of situation, morphological classification of the observed system can be difficult (section 4a). Fortunately, this problem was limited to a minority (12%) of the identified CRAs in the Eta. For comparison, the MM5 and WRF only had 3% of CRAs where more than one observed system was clearly identifiable. These models typically had smaller forecast rain areas above the CRA rainfall threshold and thus would have less of a chance to overlap two or more observed systems.

b. Critical mass threshold

The critical mass threshold defines a minimum volume of rainfall necessary for a system to be identified by the EMT. Since our study focused on the first 6 h of a model forecast, we chose a critical mass threshold ($\sim 3 \times 10^{11}$ kg) corresponding to a combined forecast and observed system producing a minimum of 0.25 in. of rain in 6 h over a 40 000-km^2 area.

In the complete absence of a forecast system, the threshold will allow the EMT to identify observed systems whose spatial scales and intensities match the minimum radar-based criteria for a MCS. Every observed system from the four primary MCS morphological types (CL, CLB, CNL, and DA) that composed a CRA was matched to a corresponding forecast. Systems with very little or no forecast rainfall but enough observed rainfall to meet the CRA critical mass thresh-

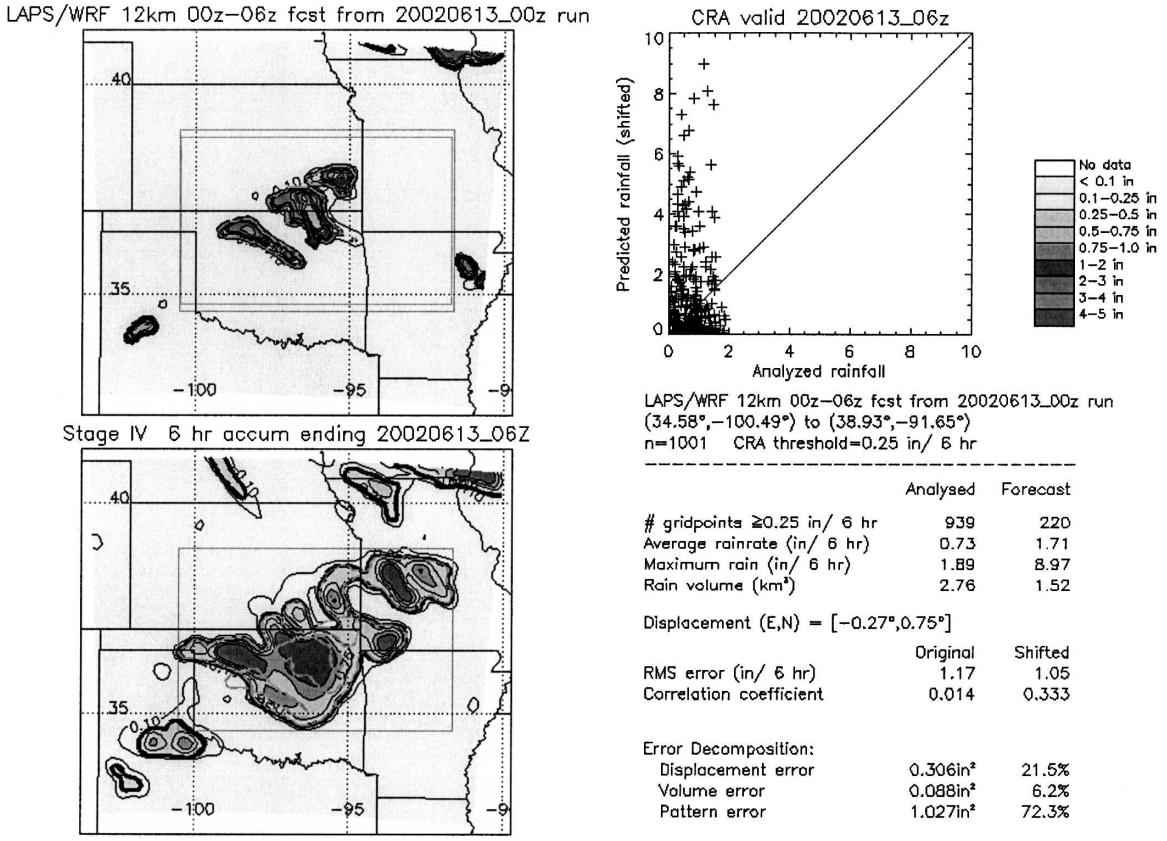

FIG. 2. Same as Fig. 1, but for the WRF model.

old were included in the statistical analysis and were classified based on radar morphology just like any other CRA. Thus, the full spectrum of model forecasts to observed events was represented.

c. Search radius

The search radius allows for initially separated forecast and observed entities to be matched provided they are located within the search radius limit. After matching, these two rain areas become a single contiguous area (section 2). We chose a search domain of 20 grid points (240 km) over which a forecast system could be shifted to match an observed one. This was roughly equal to the length scale used in defining the critical mass threshold.

d. Pattern-matching criterion

Hoffman et al. (1995) found that minimization of rmse and maximization of correlation coefficient were the best methods for determining the fit of a forecast spatial pattern to an observed one over a rectangular domain. In the EMT, the points in the verification domain are composed of the original CRA, plus any points that may have been added as a result of shifting the forecast. EM2000 found that minimizing the total squared error gave the best pattern matches for 24-h

TABLE 1. Summary of EMT threshold sensitivity tests.

Parameter	Chosen threshold	Lowering threshold	Raising threshold
CRA rainfall	0.25 in. $(6\ h)^{-1}$	Combining of individual observed systems into one large CRA	Statistics calculated over areas not representative of a typical observed MCS.
Critical mass	$\sim 3 \times 10^{11}$ kg	More CRAs; observed systems too small for MCS classification	Fewer CRAs; exclusion of small MCSs
Search radius	20 grid points (240 km)	Typically lowered correlation coefficient and raised rmse	Little change to error scores; sometimes a different match of forecast to the observed

QPFs, although they noted that maximizing the correlation coefficient generally gave similar results. In our study, we tested both measures of fit to determine what looked best to the eye of the human forecaster. For most CRAs, maximization of correlation coefficient and minimization of total squared error generally gave similar displacements, agreeing with EM2000's findings. However, our tests showed that maximization of correlation coefficient worked better overall near the edge of the IHOP domain. When using total squared error minimization, forecast rain areas would typically shift off the verification grid instead of matching up with nearby observed systems. The reason for this pathological behavior is described below.

In the original code, 25% of the grid points used for verification were allowed to shift off the domain if such a move would give the best pattern match. In our study, this threshold resulted in practically every near-boundary forecast system being shifted off of the verification grid, since this shifting resulted in the lowest total squared error calculation. If the forecast does not sufficiently resemble the observations (likely in fine-grid resolution rainfall simulations where detailed structures can be depicted), the total squared errors may be minimized by a no-rain forecast that eliminates half of the double penalty (rain in the wrong place, no rain in the right place). After testing, we found that if the best fit of the forecast rainfall pattern to the observed pattern was determined by minimizing total squared error, allowing almost none (0.1%) of the grid points to be shifted off of the domain resulted in more reasonable matches. However, this small threshold meant that all forecast systems near the verification domain corners (upper and lower Mississippi River valley), were very limited in how far they could be shifted toward an observed rainfall system.

Maximizing the correlation coefficient resulted in more reasonable matches and fewer problems of systems being shifted off of the domain. This result was also found in a CRA case study over Cyprus by Tartaglione et al. (2005). Since correlation coefficient maximization matches rainfall gradients, this matching strategy allowed the forecast rainfall maxima to be shifted to closely align with observed maxima. In addition, for most cases in our IHOP study, the use of correlation coefficient maximization resulted in little relative increase in total squared error. However, use of total squared error minimization resulted in much lower correlation coefficients for smaller CRAs. Thus, correlation coefficient maximization (and allowing up to 25% of the forecast grid points to shift off the domain) was the method of choice for the IHOP study domain.

e. Error decomposition

Forecast errors in rain events can be expressed in terms of errors in displacement, intensity, and pattern or variability of the rainfall (EM2000). The switch to correlation coefficient maximization instead of total squared error minimization was found to occasionally result in incorrectly negative rmse for the displacement portion. Therefore, a new error decomposition method was developed using correlation coefficient and mean square error (MSE) terms based on Murphy (1995). In that paper, MSE was represented as

$$\mathrm{MSE} = (\bar{g} - \bar{y})^2 + (s_g - r_o s_y)^2 + (1 - r_o^2)s_y^2, \quad (1)$$

where s represents the standard deviation, and r_o is the original correlation coefficient between the forecast (represented by $\bar{y}$, where the overbar indicates the mean) and observed (represented by $\bar{g}$) rain fields before the forecast is shifted by the EMT. Rearranging the second and third terms on the rhs gives

$$\mathrm{MSE} = (\bar{g} - \bar{y})^2 + (s_g - s_y)^2 + 2s_g s_y(1 - r_o). \quad (2)$$

The first term on the rhs is the unconditional bias, or volume error ($\mathrm{MSE}_{\mathrm{volume}}$). The second term compares the sample standard deviations of the forecast and observations and is a type of pattern error ($\mathrm{MSE}_{\mathrm{pattern}}$). The third term contains additional pattern error and the displacement error. These can be separated by adding and subtracting r (optimal correlation) in the third term:

$$\mathrm{MSE} = (\bar{g} - \bar{y})^2 + (s_g - s_y)^2 + 2s_g s_y(1 - r) + 2s_g s_y(r - r_o). \quad (3)$$

The third term on the rhs in (3) represents the shape, or finescale pattern error ($\mathrm{MSE}_{\mathrm{pattern}}$), as it includes the difference between a perfect correlation ($r = 1$) and the optimal correlation for the forecast, r. The fourth term in (3) represents the contribution of displacement error ($\mathrm{MSE}_{\mathrm{displacement}}$), as it includes the difference in covariances before and after shifting the forecast. Combining both the second and third terms in (3), the error decomposition (shown in Figs. 1 and 2) can be summarized in Eq. (4) as

$$\mathrm{MSE}_{\mathrm{total}} = \mathrm{MSE}_{\mathrm{volume}} + \mathrm{MSE}_{\mathrm{pattern}} + \mathrm{MSE}_{\mathrm{displacement}}. \quad (4)$$

The error decompositions based on total squared error minimization (EM2000) and correlation maximization [Eq. (3)] gave very similar results. CRA verification of several thousand 24-h QPFs over Australia using both approaches gave mean pattern errors that were virtually identical, and differences of only a few percent between methods for volume and displacement errors.

TABLE 2. General classification types.

Identification	Name	Criteria for classification
MCS	Mesoscale convective system	>30 dBZ in a >100 km × 100 km area >40 dBZ in a >50 km × 50 km area Both conditions for >3 h
CL	Continuous linear	CNL and MCS criteria Major axis of >40 dBZ and >100 km in length Major axis 3 times greater than minor axis in length
CLB	Continuous linear bowing	CNL, CL, and MCS criteria Bulging, convex shape. Angle of bow >30°
CNL	Continuous nonlinear	Contiguous region of echoes and MCS criteria
DA	Discontinuous areal	Discrete convective elements and MCS criteria
IC	Isolated cells	>30 dBZ in a 40 km × 40 km area >40 dBZ in a 20 km × 20 km area
OF	Orographically fixed	Nearly stationary with respect to Rocky Mountains/Black Hills
FA	False alarm	None of the above criteria were met

Error terms were originally shown as percentages relative to the other terms (EM2000). A problem with such a method is that relative errors for small rain systems are viewed no differently than those for large rain systems. In our study we also included absolute magnitudes for each error decomposition term in the CRA output (Figs. 1 and 2) to allow for comparison of the contribution each term makes to the total MSE (section 6e).

f. Filtering

It is well known that models cannot adequately predict the spatial structure of convective scales because of interpolation from finite-differencing schemes and parameterized horizontal diffusive processes. The minimum resolvable feature varies as a function of not only the grid spacing of models, but also the numerics and physics in each type of model. Most mesoscale models will generally be able to resolve only rainfall features of wavelength at roughly 5 times the grid spacing. Harris et al. (2001) showed that the 3-km Advanced Regional Prediction System model could not resolve less than five delta waves. Baldwin and Wandishin (2002) also found three–five delta waves to be the smallest resolvable wavelength in the 22-km Eta with the Kain–Fritsch parameterization and in 10- and 22-km versions of the WRF model. However, in the 12-km Eta with the BMJ parameterization, features less than 200 km were not resolved well, which might argue for filtering of 17 delta waves. In the present study, we decided that the stage IV observations should be filtered so that the observed rain areas resembled what the majority of the 10–12-km grid spacing models run by FSL during IHOP were able to show (Koch et al. 2004). Thus, the stage IV data were remapped to each native model grid and filtered using a low-pass Lanczos filter (Duchon 1979) to remove wavelengths less than six delta (72 km). This procedure does not remove any mismatch in the spatial variability of the model QPFs. As will be shown in section 6e, error measures reflect the Eta's low variability in QPFs compared with the MM5 and WRF.

g. Area over which verification statistics were computed

Usually categorical statistics are computed over entire model domains. For an entity-oriented technique, like the EMT, there is some uncertainty over which areas should be used for calculating various verification statistics. In this study, the definitions for four verification quantities were adjusted. Rain volume, maximum rainfall, average rain rate, and number of grid points exceeding the user-defined threshold were previously calculated over the union of the observed, original forecast, and shifted forecast regions (EM2000). To better describe the characteristics of each individual entity, these parameters were computed exclusively over the observed and original forecast portions of the CRA before any displacement occurs. Only grid points at or above the CRA rainfall threshold were included in the analysis area for each portion (the areas enclosed by the dark gray isohyet in Figs. 1 and 2).

4. Classification of convective systems

The EMT allows a user to focus on rainfall systems of a specified temporal or spatial scale. For our purposes, the EMT was used to examine MCS forecasts valid for fixed 6-h periods to note errors as a function of the observed system morphology. A classification scheme was necessary for exploring this aspect. Tables 2, 3, and 4 provide a summary of the various classes and defini-

TABLE 3. Stratiform classification types for linear systems. (Adapted from Parker and Johnson 2000.)

Identification	Name
TS	Trailing stratiform
LS	Leading stratiform
PS	Parallel stratiform
TS/PS	Trailing and parallel stratiform
LS/PS	Leading and parallel stratiform

tions, with additional information provided in the following paragraphs on how the scheme was defined.

In the first stage of classification, the observed system highlighted in the stage IV 6-h accumulated precipitation product was cross-referenced with an observed system indicated in radar observations. The radar-based morphology used 2-km NIDS composite base reflectivity radar imagery with a temporal resolution of 30 min.

The radar-based system classification was determined mainly by previous literature dealing with MCS classification (e.g., Houze 1993; Geerts 1998; Parker and Johnson 2000). We defined a radar-based MCS as a convective system containing continuous or discontinuous convective echoes that propagated and/or organized in nearly the same manner as other convective echoes within the system. We required the minimum MCS criteria to have at least 30 dBZ of base radar reflectivity over at least a 10 000-km^2 (i.e., 100 km $\times$ 100 km) area and at least 40 dBZ in a 2500-km^2 (i.e., 50 km $\times$ 50 km) area. Both dBZ conditions had to exhibit temporal continuity for at least 3 h. Using the Z–R relationships (where Z is reflectivity and R is rain rate) of $Z = 200 \times R^{1.6}$ for stratiform and $Z = 300 \times R^{1.4}$ for convective rain, 30 dBZ corresponds to a rain rate of around 0.10 in. h^{-1} with 40 dBZ corresponding to a rain rate of nearly 0.50 in. h^{-1}.

a. General classification

The first series of classifications distinguished between linear and nonlinear systems for those meeting the MCS criteria. Since not all observed systems identified by the EMT met our radar-based MCS criteria, separate categories had to be made for these smaller or shorter-lived systems. The classification scheme included seven general types of systems (as shown in Table 2) and is summarized as follows.

TABLE 4. Development classification types for linear systems. (Adapted from Bluestein and Jain 1985.)

Identification	Name
BA	Broken areal
BB	Back building
BL	Broken line
EA	Embedded areal

The CL systems contained a continuous major axis of at least 40-dBZ convective echoes of at least 100-km length that shared a common leading edge and moved in tandem. In addition, the major axis had to be at least three times as long as the minor axis. CLB systems not only met the CL criteria above, but also had to contain a bulging, convex shape (angle greater than 30°) of continuous convective cells with a tight reflectivity gradient on the front edge of the convective region. This shape had to exist for at least 1.5 h. If the minimum MCS requirements were met in a contiguous area but did not meet the linear requirements of CL or CLB, then the system was identified as CNL. If the above minimum MCS requirements were not met in a continuous area, but were met in an area of discrete convective elements in which no element was separated by more than 200 km from another, then the system was identified as DA.

If a system remained nearly stationary with respect to the western edges of the IHOP domain (the Rocky Mountains and Black Hills), then the system was classified as OF because the mesoscale processes influencing these mountain systems may differ from systems over the plains. If cells were too small, isolated, or lacked temporal continuity to meet any of the above classifications, but had at least 40 dBZ in a 400-km^2 (i.e., 20 km $\times$ 20 km) area and at least 30 dBZ in a 1600-km^2 (i.e., 40 km $\times$ 40 km) region, then an IC classification was assigned. It is well understood that a 12-km model cannot fully resolve isolated cell events, but for completeness these events were included in the classification scheme. If none of the above criteria were met, then the observed system in the CRA was classified as an FA, a falsely predicted event.

In the 6-h period over which CRAs were defined from accumulated rainfall data, multiple radar-based systems might be observed within one larger CRA. When this was the case, the system with the greater temporal, spatial, and/or rain volume was used to define the morphology of the CRA. In other cases, when the morphology of a single system changed over time, the morphology that occurred over the majority of the 6-h period was used to classify the CRA. It is understood that defining a single convective morphology for multiple radar-based systems will increase the amount of statistical uncertainty. However, it is pertinent to include these CRAs in the statistical analysis, since a clearly dominating type occurred in the vast majority of these cases.

b. Additional linear classifications

For every linear system (CL or CLB), additional subclassification was performed by using the taxonomy

proposed by Parker and Johnson (2000) and Bluestein and Jain (1985). First, the arrangement of stratiform rainfall with respect to the intense convection was classified according to the definitions TS, LS, and PS as given by Parker and Johnson (2000; Fig. 3). Combinations of these types were noted when both were seen for at least 1.5 h. Second, a classification was made based on Bluestein and Jain's (1985) categories for squall line development (Fig. 4): BB, BA, BL, and EA.

Figure 5 shows a four-panel plot of radar reflectivity corresponding to the system observed in the stage IV product of Figs. 1 and 2. The first image (Fig. 5a) is valid 1 h after convective initiation in south-central Kansas and north-central Oklahoma. This image represents conditions 1 h before the start time of the 6-h period evaluated in Figs. 1 and 2, and helps to show the BA development assigned to this system. Figures 5b and 5c show the development into a CL system. Figure 5d shows the system at maturity as a TS area expands. A PS area is also noted at this time, but this did not last the required 1.5 h. The TS region was largest around 0700 UTC (not shown) as the convective line decayed and then ended around 1000 UTC.

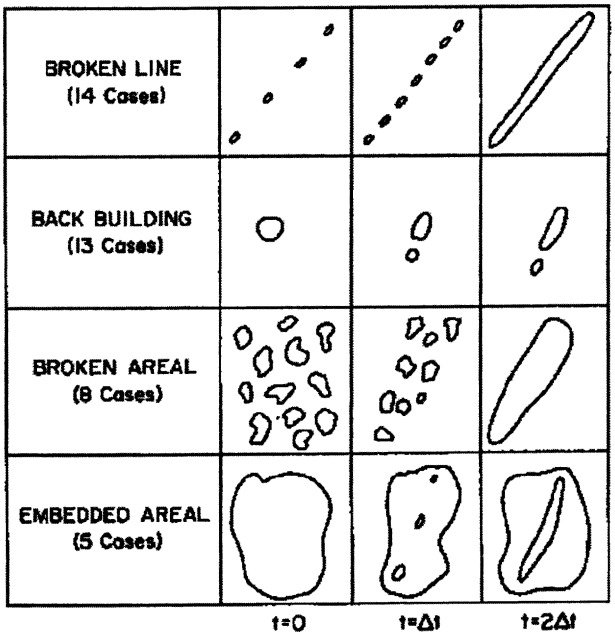

FIG. 4. Development types for linear MCSs; from Bluestein and Jain (1985).

5. Observed MCS morphology distributions

A total of 190 CRAs were identified for the Eta, 164 for the MM5, and 163 for the WRF during the 4-week IHOP period. Of the CRAs identified, 7% of Eta systems and 2% of MM5 and WRF systems were classified as FA (little or no observed rain). A small number of the observed systems (4%–5% in the three models) were classified as OF to the Rockies and Black Hills at the western edge of the IHOP domain. The IC systems accounted for 12% of the CRAs in the Eta, 6% in the MM5, and 5% in the WRF. Other than to note the number of occurrences, we exclude IC, FA, and OF systems (22% in the Eta, 13% in the MM5, and 12% in the WRF) from further analysis in this study. The IC and FA systems were only identified because of forecasted rainfall; observations did not show enough rain volume to meet the CRA critical mass threshold (section 3b).

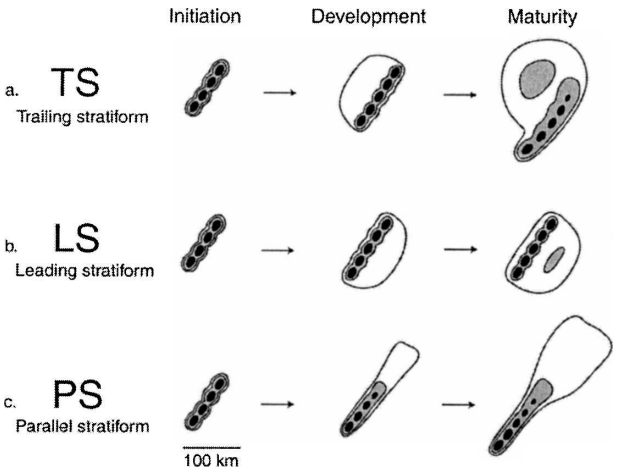

FIG. 3. Stratiform types for linear MCSs; from Parker and Johnson (2000).

These results show the Eta produced more occurrences of higher rain volumes for non-MCS cases than the MM5 or WRF. The higher percentage of non-MCS occurrences in the Eta may be due to its lack of differentiation in rain rate and volume between convective system types. As will be shown in section 6, the Eta produced nearly the same average rain rate and volume across the various convective system categories. Thus, a higher percentage of cases on the non-MCS side of the spectrum should result. The focus of subsequent evaluation is on model performance as a function of the observed system morphology of the 148 remaining events for the Eta, 144 events for the MM5, and 143 events for the WRF.

Fifty-five (37%) observed cases were classified as linear in the Eta, 62 (43%) in the MM5, and 60 (42%) in the WRF. Ninety-three (63%) observed cases were classified as nonlinear in the Eta, 82 (57%) in the MM5, and 83 (58%) in the WRF. Figure 6 shows a histogram of general, squall, and development types for every

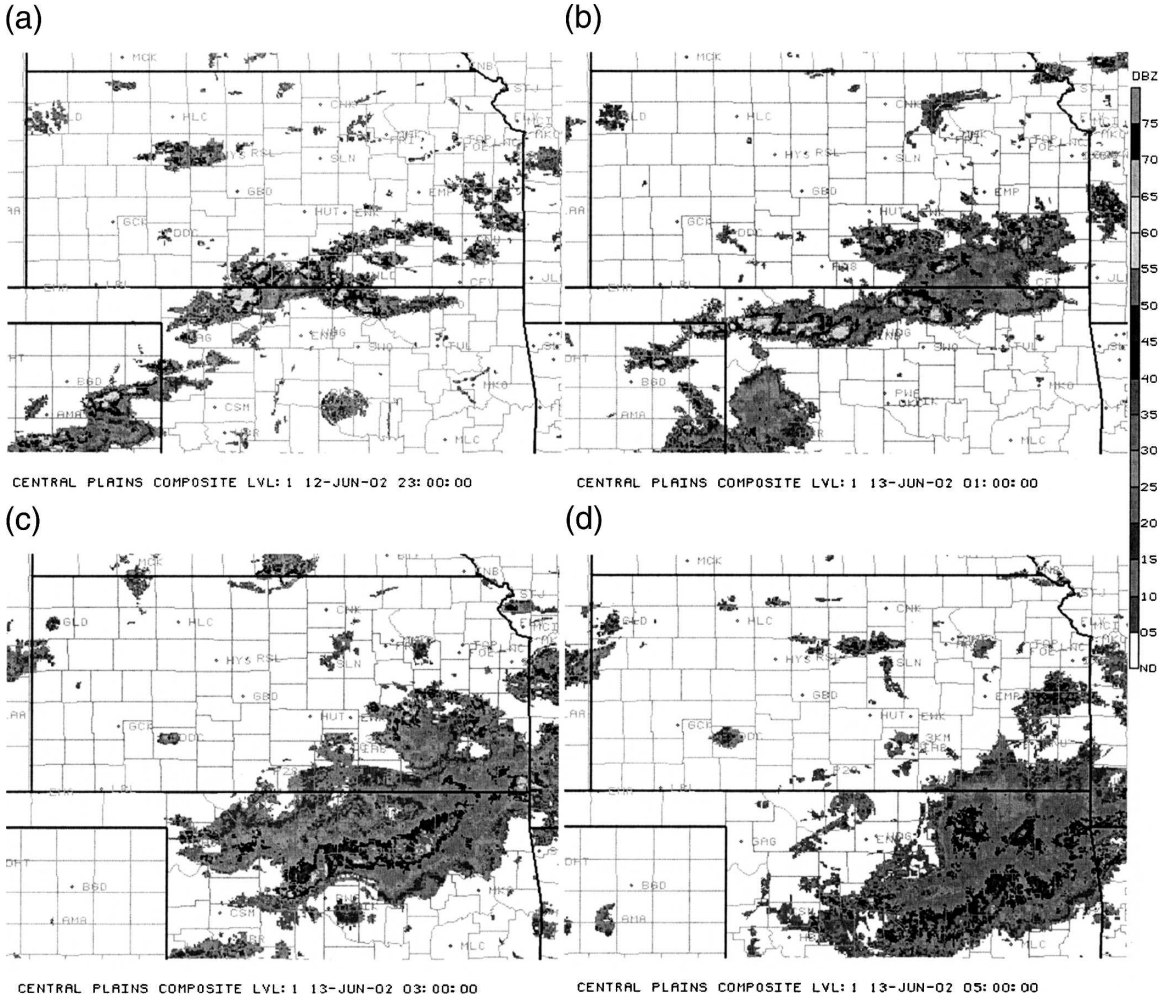

FIG. 5. Example of 2-km NIDS radar reflectivity corresponding to CRA output from Figs. 1 and 2. Radar images are at 2-h intervals [(a) 2300, (b) 0100, (c) 0300, and (d) 0500 UTC] beginning at 2300 UTC 12 Jun 2002 and ending at 0500 UTC 13 Jun 2002. The legend in the upper right shows the dBZ thresholds for instantaneous reflectivity values.

identified CRA that met MCS criteria (see Tables 2, 3, and 4 for abbreviations used for the various classification types in the graphs). Of the linear systems, 86% were classified as CL, with 14% as CLB. Because of the low sample sizes associated with the CLB category, these systems have been lumped into the CL category for the statistical analysis in section 6. Nonlinear systems were led by the CNL category with 61%, followed by DA with 39%.

For the stratiform rain area classification, TS dominated with 67% of the linear systems. The TS/PS type (generally large systems since stratiform rain occurred in both regions) garnered the second highest total with 16%. The categories of LS, PS, and LS/PS (substantial areas of each) all had five or fewer occurrences in each model. Little statistical significance of LS, PS, and LS/PS classifications was found, likely owing to the small sample size in each of these categories. These results were fairly similar to the Parker and Johnson (2000) survey of central U.S. linear MCS. They found TS was the dominant mode, though only accounting for 40% of the cases. The TS/PS type was second highest with 18%. In our study, the other categories of stratiform had slightly less of a representation than in the Parker and Johnson (2000) study, due to a greater domination of the TS type.

Among the development types, BA was the most common with 44%, followed by BL with 32%, and then BB with 23%; EA had only one (1%) occurrence for each model CRA and was, therefore, excluded from further study. The results for BA differed greatest from Bluestein and Jain (1985) who found these events only 20% of the time for severe squall-line cases in Oklahoma.

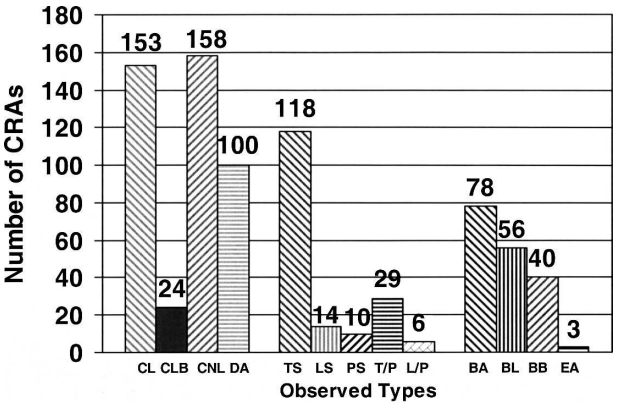

FIG. 6. Histogram of observed systems (general types, stratiform types for linear systems, and development types for linear systems) for all CRAs identified by the EMT applied to the Eta, MM5, and WRF models. Tables 2, 3, and 4 give definitions used for the abbreviation of types.

Since the classification scheme is conditioned on the observed system, forecasts that miss the event (i.e., no or very little precipitation is simulated) are included in the dataset. In totaling the number of MCS cases where no forecast rain volume existed above the 0.25-in. rainfall threshold there were 19 (13%) cases in the Eta, 31 (22%) cases in the MM5, and 20 (14%) cases in the WRF. Thus, the MM5 had a larger number of missed events than the Eta or WRF.

6. CRA statistical analysis

Statistics were calculated for the following parameters: rain volume, rain rate, maximum gridpoint rainfall, phase displacement, and MSE decomposition. This analysis was performed for all of the observed systems over the plains meeting minimum MCS criteria. Errors were then examined as a function of the observed system morphology. Individual comparisons of forecasts with observations are not provided since the definition of a CRA can occasionally cause different statistical results for the observations, a limitation of the EMT (section 3a).

All statistical results discussed in this section were formally evaluated by a Student's *t* test, a multiple comparison analysis of variance (ANOVA) utilizing Tukey's honestly significant difference (HSD) procedure (see Tukey 1993) and Levene's (1960) test for homogeneity of population variances. These tests determined statistical significance at the 0.05 alpha level. For the *t* test, basic assumptions were made regarding adequate sample sizes, approximate normality, and that the data comprising each sample were randomly selected from their larger population. In comparing the general-type categories, sample sizes range from around 30 to 60 cases per type while development-type categories have sample sizes ranging from around 15 to 30 cases per type. These sample sizes should be sufficient for statistical testing, as long as the sampling distribution is approximately normal. For Tukey's HSD, assumptions of approximate normality and nearly equal variances in samples and populations were made in order to accurately perform this test, with deviations noted. Thus, categories with extreme skewness or many outliers and vastly different sample sizes are excluded from discussion below. Levene's test was used to determine if population variances were not all equal for multiple comparisons. However, the assumption of the equality of population variances is less critical when the sample sizes are nearly equal. The variances can be significantly different, but the *p* values for any analysis of variance procedure will only be mildly distorted (Ott and Longnecker 2001).

The Student's *t* test determined whether errors between forecast and observed values were biased for each type (e.g., is a mean wet bias in the DA category for the Eta truly statistically significant?). Tukey's HSD determined statistical significance of differences in mean errors between types in a given model (e.g., if both DA and CNL have a statistically significant mean wet bias for the WRF model, does one type have a greater mean bias versus the other?). This conservative test was performed to protect the true alpha level of 0.05 during multiple comparisons from the effects of multiplicity (e.g., Wilks 1995; Ott and Longnecker 2001).

All graphical results are presented by using box plots showing the following: medians, 75% and 25% quartiles, and minimum/maximum values within 1.5 times the interquartile range; mean diamonds show means and 95% confidence intervals for the mean. Additional differences that did not pass formal statistical significant tests can be gleaned from the graphs, but they are generally not discussed in this section.

a. Rain volume

The Eta showed a mean dry bias (forecast compared with observations) with the CL category (as noted in section 5, for the CRA statistical analysis this category represents the combination of CL and CLB systems) and a mean wet bias for the DA category (Fig. 7a). The mean bias of the CL category was significantly drier in comparison with the CNL and DA categories. This confirms that the Eta produces too little rain volume for linear systems, and this behavior differs from its performance with nonlinear systems. We speculate that the mean dry bias with linear systems reflects the lack of

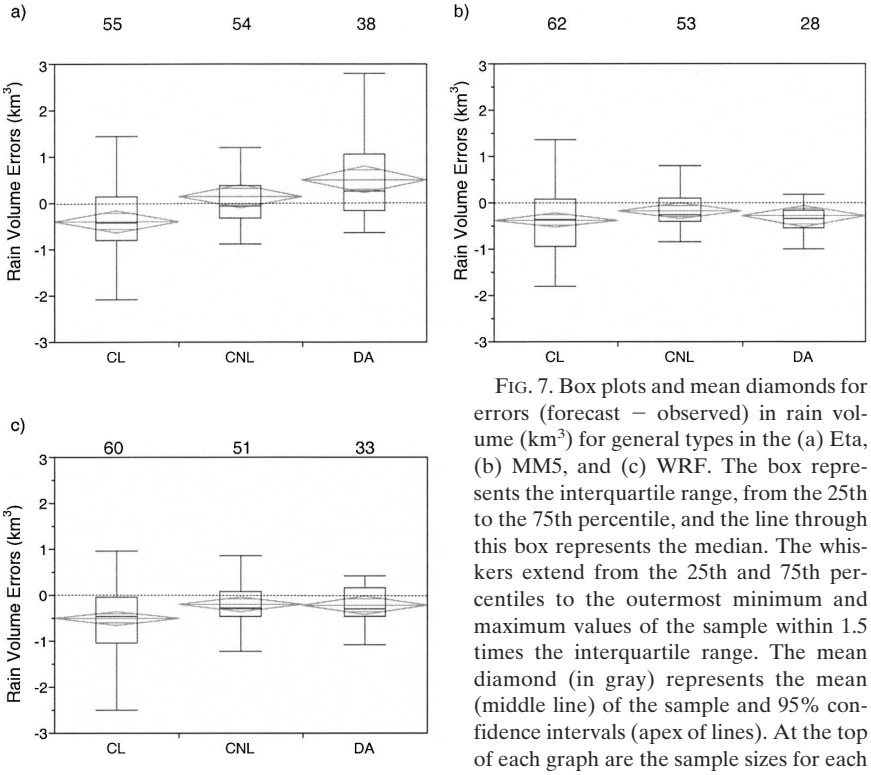

FIG. 7. Box plots and mean diamonds for errors (forecast − observed) in rain volume (km^3) for general types in the (a) Eta, (b) MM5, and (c) WRF. The box represents the interquartile range, from the 25th to the 75th percentile, and the line through this box represents the median. The whiskers extend from the 25th and 75th percentiles to the outermost minimum and maximum values of the sample within 1.5 times the interquartile range. The mean diamond (in gray) represents the mean (middle line) of the sample and 95% confidence intervals (apex of lines). At the top of each graph are the sample sizes for each category.

transport of condensate away from more intense convective cells (which is not included in the BMJ convective scheme), a process known to be very important in the upscale growth of organized linear systems (e.g., Rutledge 1986).

Both the MM5 and WRF showed a mean dry bias for all three general types (Figs. 7b and 7c). The MM5 had no categories that were significantly different from the other categories. In WRF, the CL category was significantly drier than the CNL and DA categories. It should be noted that Levene's test for the assumption of equal variances was not passed for the WRF comparison, because of the larger spread in the CL category versus the CNL and DA categories. However, as in the Eta, the WRF appears to have larger dry biases for linear systems than nonlinear ones.

As might be expected because all three models had a mean dry bias with the linear category, a dry bias was also apparent with the dominant stratiform type, TS. Statistical significance was not found for any model in a comparison among stratiform types, likely because the small sample size for all types except TS. For development types, the Eta had a mean dry bias with both the BB and BL categories (Fig. 8a). The mean for the BB category was significantly higher than the BA category. However, it is easily seen through the spread of the box plots that the assumption of equal variances was not valid for this comparison. Both the BB and BL categories also suffered from modest skewness and relatively low sample sizes, further lowering confidence in the above results. Dry biases were present in the MM5 and WRF for both BA and BL categories (Figs. 8a and 8b). The dry biases present with development types likely reflect the dry bias already noted for linear systems. Differences in biases were not significant among development types for the MM5 and WRF.

b. Rain rate

The Eta's forecast average rain rate (for all CRA points above the 0.25-in. threshold) was significantly lower than observed for both the CL and CNL general categories. It was also significantly lower for the TS and TS/PS stratiform categories and the BB, BA, and BL development categories. In addition, the Eta produced nearly the same average rain rate for practically all general types (Fig. 9a), unlike observations (Fig. 9b), implying the model may not have the capability to differentiate its rate of rainfall for highly efficient precipitation systems from those with lower efficiency. Gallus (1999) showed that the Eta with the BMJ convective scheme was fairly insensitive to changes in horizontal grid resolution. He speculated that the BMJ scheme

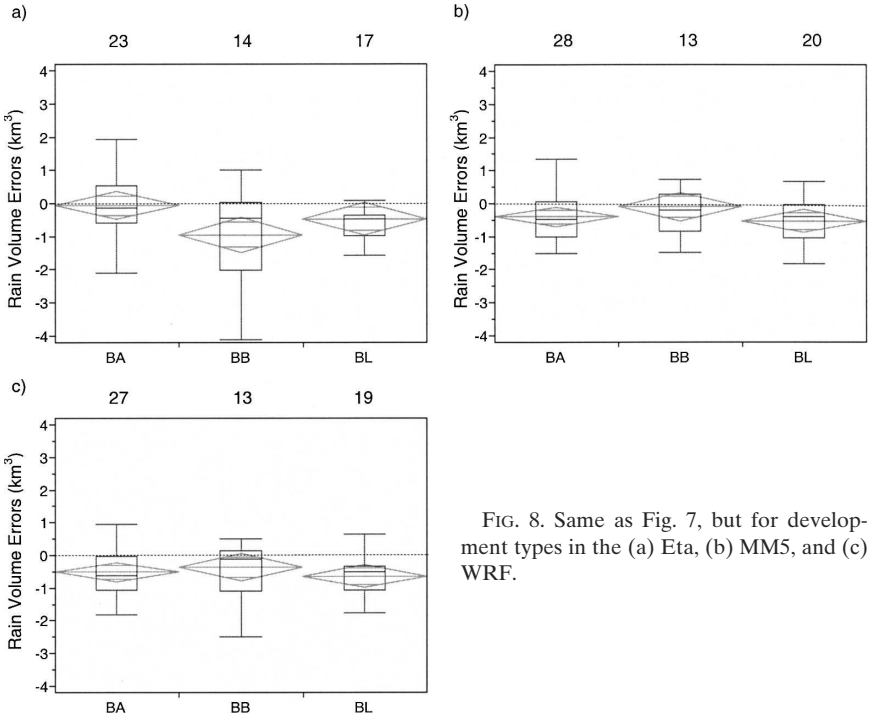

FIG. 8. Same as Fig. 7, but for development types in the (a) Eta, (b) MM5, and (c) WRF.

was so aggressive at drying the atmosphere that small-scale structures more likely to be produced in the grid-resolved component of the rainfall were often eliminated. Operational forecasters have long noted that the rainfall forecasts from the Eta appear to be overly smooth and lack finescale structure. The current analysis agrees with those observations.

However, for both the MM5 and WRF, the CL category had a significantly higher forecast average rain rate than that observed (Figs. 10a and 10b). The mean errors of the CL category were also significantly higher than the DA category. For stratiform types, these same trends were noted. Both models had significantly higher average rain rates than observed in the TS category, a result consistent with a failure to develop larger areas of lighter stratiform rain (such that the heavier convective rates dominated these systems). For development types, the MM5 and WRF were both significantly higher than observed with the BB category. The MM5 and WRF results are in contrast to the much lower average rain rates of the Eta forecasts.

These results suggest a systematic rainfall distribution and amount error arising from problems with prediction of cold pool dynamics. Weisman et al. (1997) showed from three-dimensional midlatitude squall-line simulations performed at a variety of grid resolutions that a delayed strengthening of the cold pool occurs with explicit models run at resolutions coarser than

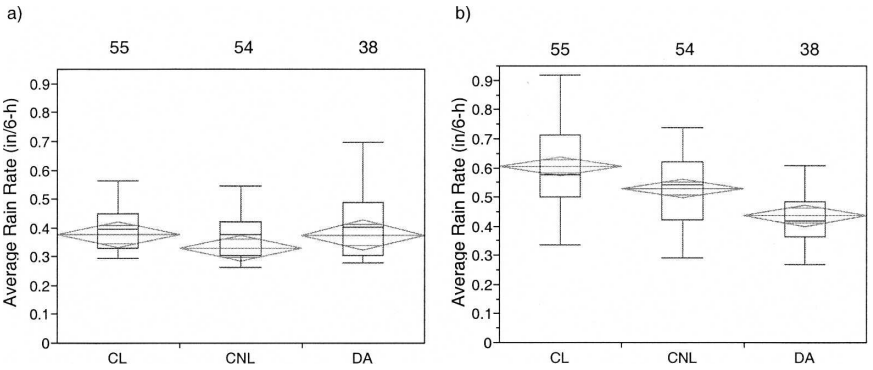

FIG. 9. Same as Fig. 7, but for (a) forecast and (b) observed average rain rate [in. (6 h)$^{-1}$] for general types in the Eta.

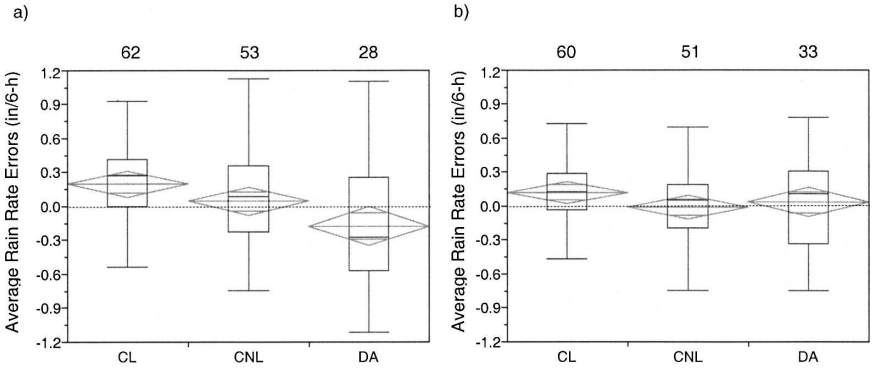

FIG. 10. Same as Fig. 7, but for errors in average rain rate [in. (6 h)$^{-1}$] for general types in the (a) MM5 and (b) WRF.

4 km. Since the cold pool is crucial to the evolution of an MCS into an upshear-tilted mature system, such models can be expected to underestimate the TS precipitation region commonly produced by the upshear-tilted front-to-rear flow, while overpredicting the precipitation in the convective leading line. Both characteristics are observed with the 12-km MM5 and WRF models examined in the present study.

c. Maximum rainfall

Maximum rainfall was defined as the highest observed amount of precipitation in the model's 12-km grid and in the filtered stage IV observed accumulation grid. The Eta significantly underpredicted average rainfall maxima overall, for all general and development types, and TS and TS/PS stratiform types. Both the CL and CNL categories had greater mean dry biases than the DA category (Fig. 11a). For the development types, the BL category had a greater mean dry bias compared with the BA category.

As with the average rain-rate category, the Eta was very uniform in its distribution of average maximum rain rate for each system type. The tendency of the Eta to have far smaller average maximum rain rates than observed agrees with Gallus (1999), who showed that

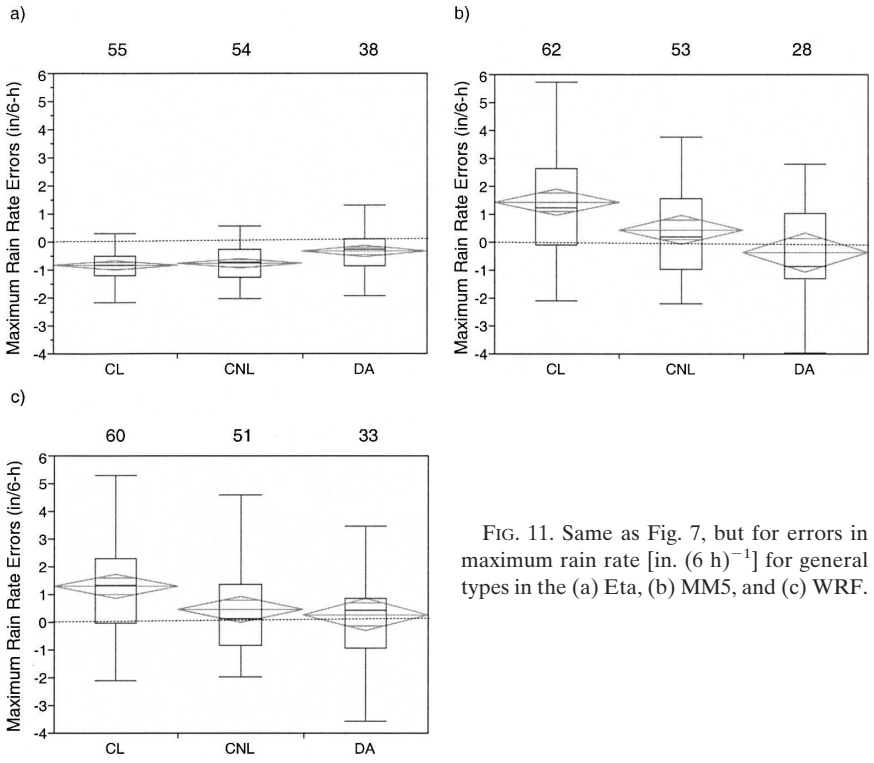

FIG. 11. Same as Fig. 7, but for errors in maximum rain rate [in. (6 h)$^{-1}$] for general types in the (a) Eta, (b) MM5, and (c) WRF.

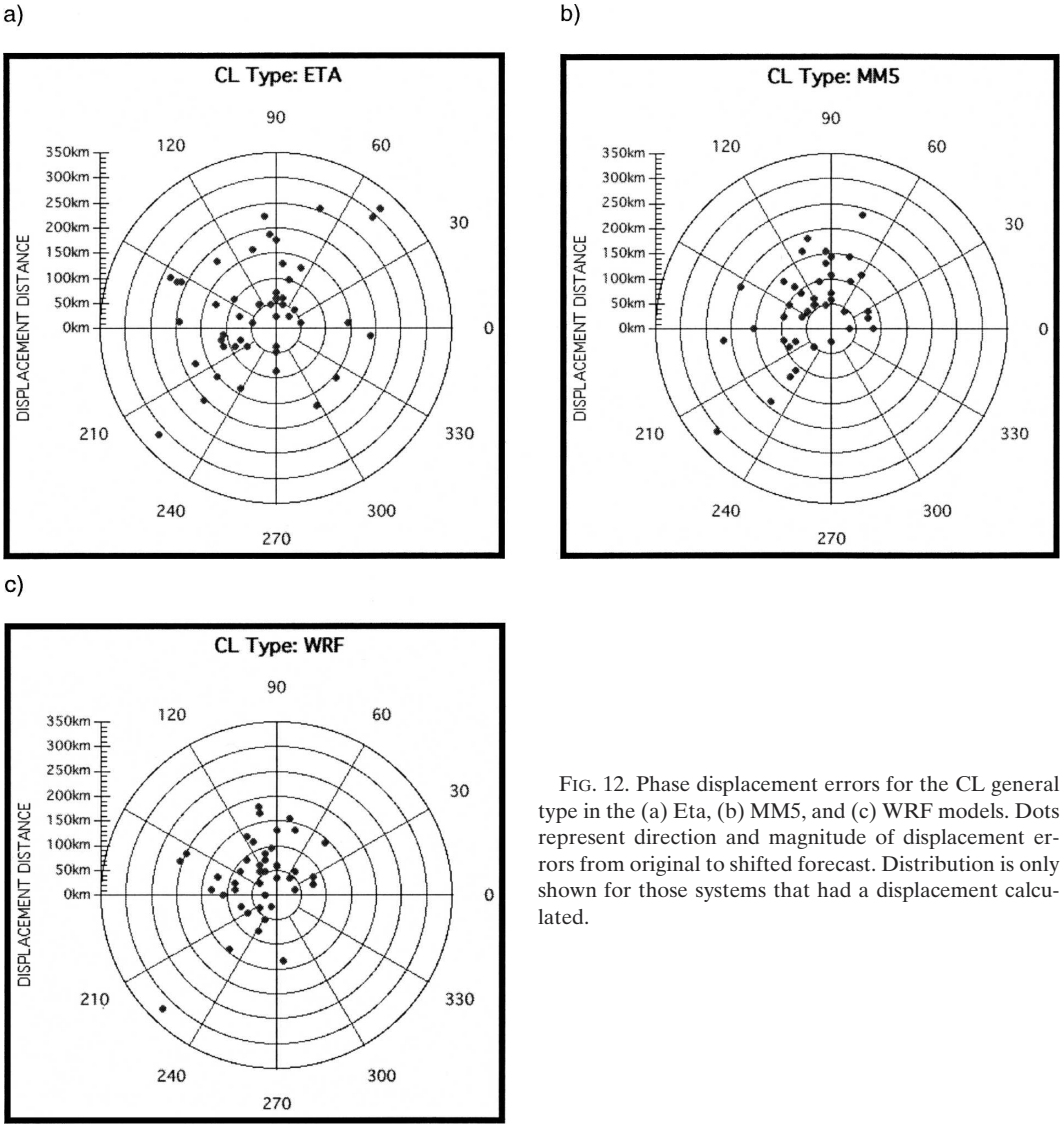

FIG. 12. Phase displacement errors for the CL general type in the (a) Eta, (b) MM5, and (c) WRF models. Dots represent direction and magnitude of displacement errors from original to shifted forecast. Distribution is only shown for those systems that had a displacement calculated.

the use of the BMJ scheme prevented large rainfall amounts from occurring with fine-grid resolution. When the Kain–Fritsch scheme (Kain and Fritsch 1993) was used instead, Gallus (1999) noted that much larger rain rates resulted. He showed the maximum rain rates in simulated convective systems occurred in regions with large grid-resolved rainfall components.

Once again the MM5 and WRF results were in stark contrast to the Eta (Figs. 11b and 11c). Both the MM5 and WRF had significantly larger maximum rainfall rates on average, than observed for the CL and CNL types. For both models, CL systems had significantly larger wet biases than both the CNL and DA types. This trend continued into the stratiform categories with TS forecasts being significantly wetter than observed in both the MM5 and WRF. For all three development types, the MM5 and WRF had a mean wet bias. The MM5 and WRF also exhibited much more variability with the spread of the interquartile range (from the 0.25 to the 0.75 percentile), usually double that of the Eta for most types. Since the MM5 and WRF typically underestimated rain volume, the greater rainfall intensities indicate much smaller rainfall areas than observed.

d. Phase displacement errors

None of the models displayed a strongly preferred direction and magnitude of displacement error vectors for any particular MCS classification, except for the CL type. All three models exhibited a majority of displacements from the northwest for this type (Fig. 12). These systems were likely forecast too slowly by the three models (especially MM5 and WRF). This may suggest

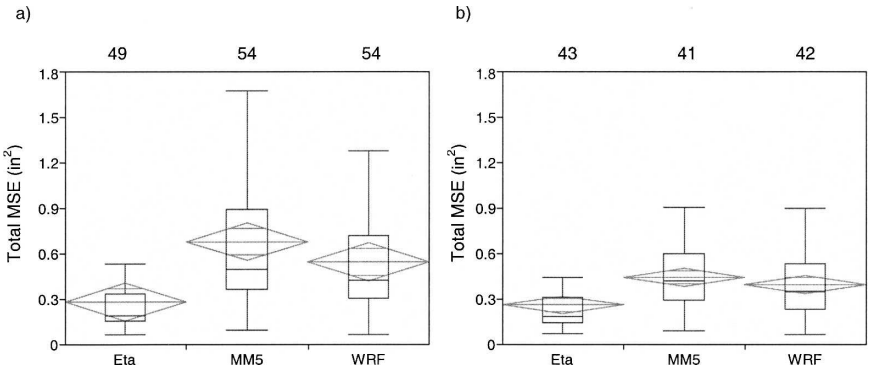

FIG. 13. Same as Fig. 7, but for total MSE (in.^{2}) for (a) the CL type and (b) the CNL type in the Eta, MM5, and WRF.

that MM5- and WRF-simulated cold pools for squall-line systems were too weak or delayed, a hypothesis fully consistent with the rainfall rate bias problems (underprediction of the stratiform rain region, overprediction of the rain rates in the convective leading lines) discussed in section 6b. In the Eta, the BMJ convective scheme does not directly affect the model environment below the subcloud layer. This makes the scheme's behavior difficult to correlate to specific observed physical processes (Kain et al. 2003). Consequently, linear MCS cold pools are not realistically simulated.

e. MSE decomposition

For all three models, the CL category had the highest average total MSE, while the CNL category had the lowest average out of the general types. Differences in MSE magnitudes between classification categories likely reflect the complexity (e.g., scale, spatial variability, intensity, and longevity) of the system being forecast. A system with greater spatial variability and more intense precipitation maxima will tend to be more difficult to forecast and will have the potential for greater MSE scores (Murphy 1993). Normalizing the MSE by average rainfall is one approach to account for differences in complexity (in particular, intensity and longevity) that affect predictability. The only significant difference found for the general types in any of the models was that for the Eta, the DA category had higher normalized average total MSE values than the CL and CNL categories.

Given similarly observed average rainfall volumes between the models' CRAs, one can test whether a certain model had significantly lower or higher average MSE than the others for a specific type. The average total MSE for the Eta was significantly lower than both the MM5 and WRF for the CL and CNL general types (Figs. 13a and 13b). The TS was the only stratiform type to be significantly lower in the Eta versus MM5 and WRF. However, the test for assumption of equal variances in both the CL and TS types failed, because both the MM5 and WRF clearly had much larger variances than the Eta (Fig. 13a). There were no statistically significant differences between the models for the development types.

For the CL and CNL types, both the MM5 and WRF had their largest source of total MSE from pattern errors, followed by displacement errors, and then volume errors. The Eta was similar in this distribution for the CL type. But for the CNL type, larger errors for pattern were followed by volume errors, and then displacement errors. Figure 14a shows the combined results from all three models for the CL type. The DA type did not display the same distribution as the other three general types; there were no significant differences between the types of errors for any of the models. The magnitude of errors were nearly equally distributed among all decomposition terms for this type (Fig. 14b).

Pattern error was the largest source of average error in the MSE decomposition for all three models for CL and CNL types. The Eta was also significantly lower in average total MSE for these categories in comparison with the MM5 and WRF. As mentioned in section 3f, pattern error is strongly influenced by spatial variability, which is a function of the effective model resolution. Because the Eta produces QPF on scales 3 to 4 times larger than the MM5 and WRF, the Eta has lower spatial variability. All other factors being equal, a model with lower spatial variability will have lower MSE. The magnitude of pattern errors are around twice as large for the MM5 and WRF compared with the Eta in the CL type, similar to the magnitude of total MSE (Fig. 13a). Despite forecaster and modeler interest in models simulating realistic-looking detail, commonly used error measures such as MSE suggest mod-

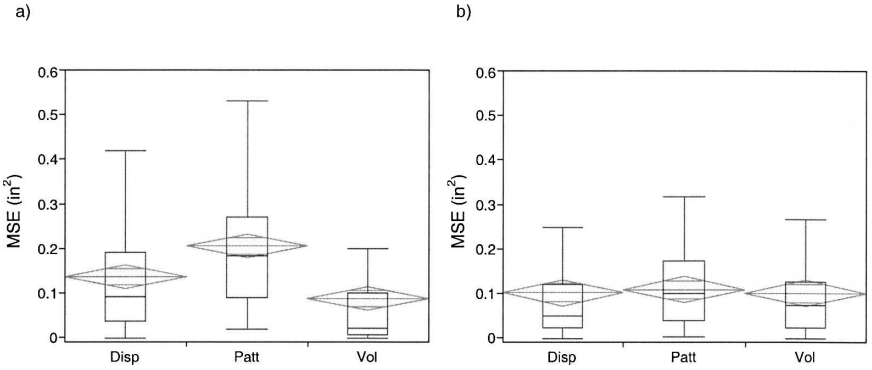

FIG. 14. Same as Fig. 7, but for displacement, pattern, and volume errors (in.^{2}) for (a) the CL type and (b) the DA type with combined results from the Eta, MM5, and WRF.

els with less spatial variability are more accurate. This result is consistent with many previous works (e.g., Baldwin et al. 2001; Gallus 2002; Mass et al. 2002; Kain et al. 2003).

7. Conclusions

The EMT was modified to optimize detection of MCSs occurring over the central United States and applied to forecasts of convective system rainfall from the 12-km Eta, MM5, and WRF models during IHOP 2002. This technique allowed for the objective determination of errors as a function of observed convective system morphology, a procedure not possible with typical grid point–to–grid point domainwide verification. No attempts were made to use the EMT for determining errors as a function of forecast convective system morphology, because of an inability of 12-km models to properly simulate detailed convective system characteristics.

Systematic deficiencies were found in these models for various types of convective systems when using the error measures supplied by the EMT. While almost all of the differences found in comparing the Eta and the MM5/WRF were not surprising, a goal of object-oriented verification is to produce nearly the same overall conclusions about model performance obtained by subjective evaluation (Davis et al. 2006b). The results as a function of the observed convective system morphology provide additional insight into the spectrum of MCS errors in each model. These results should help modelers in their assessments and may have some relevance to forecasters as well. For modelers, the error metrics can point out certain morphological types where the model has a systematic bias or relatively inaccurate forecast compared with other observed types. For forecasters, the utility of these results depends ultimately on an a priori knowledge of likely convective system morphological evolution, based on conceptual/numerical models and experience. Knowing what the numerical model QPFs typically depict for a certain type of system, a forecaster can further confirm or reject their forecast formulated on the environmental wind/thermodynamic fields and other observations. However, forecasting warm season convective system morphology is itself a problematic and uncertain process.

The modified EMT suggested that the Eta underestimated rain volume for linear systems and overestimated it for discontinuous nonlinear ones, while both the MM5 and WRF underestimated volume for all systems. The Eta also produced average rain rates and peak rainfall amounts that were much too light for almost all systems, likely because of its typically low variability and overly smoothed QPFs. On the other hand, the MM5 and WRF both produced average rain rates and peak rainfall amounts that were higher than observed for most linear classifications. These two models were dry-biased with rain volume reflecting a large underestimate of areal coverage compared with observations for linear systems. All three models forecast rainfall too far northwest for linear systems. These results suggest a systematic rainfall distribution and amount error arising from problems with prediction of cold pool dynamics, following Weisman et al. (1997). The Eta had smaller total mean square errors than the MM5 and WRF for both CL and CNL systems, and for TS types. The smaller errors again likely reflect its tendency to produce smoother rainfall fields than the WRF and MM5. For all general MCS types (except DA), the largest contributors to total MSE were pattern errors, typically followed by displacement, and then volume errors.

Overall, the modified EMT suggested various sys-

tematic errors are dependent on the convective system type and model. No one general type or model was consistently better or worse than the other. Out of the stratiform types, TS systems typically had the same biases as those of CL systems. Because of small sample sizes it was not possible to determine with certainty whether the other stratiform types had significant biases. Error measures did not consistently differ among the development types. It is plausible that processes occurring during development operate on scales too small for a 12-km model to differentiate.

In future work, this technique and observed morphology classification scheme could be used to evaluate other models or different versions of the same model. A goal of these comparisons would be to not only utilize the statistical results provided by the object-oriented technique, but to also evaluate the ability of the technique to match forecast and observed objects (Davis et al. 2006b). As model effective resolution becomes further refined in the future, a classification scheme based on the forecast MCS type could be used in lieu of an observed morphology. The model forecast MCS morphology and corresponding statistics generated by the EMT may provide a greater direct impact to forecasting users.

The EMT could also potentially be applied to verifying human gridded forecasts, in addition to those of numerical models. Since the National Weather Service has moved into the digital forecast era with the National Digital Forecast Database (Glahn and Ruth 2003), an object-oriented gridded verification could occur between human gridded forecasts and numerical models. Questions such as how does overall rainfall volume and rate differ from the human versus model forecast and do human forecasts exhibit the same northwest bias for linear MCSs could be addressed.

The EMT's flexibility for user-defined parameters in object-oriented verification, along with its production of several error metrics at once, makes the technique a valuable tool in the assessment of forecasts. By developing a classification scheme based upon the observed morphology, the technique can further differentiate its error measures and provide modelers with error information for specific types of observed systems. Such information may be useful in pinpointing specific shortcomings in model physics or dynamics, allowing for more potential improvement in numerical forecasts.

Acknowledgments. This research was supported in part by NSF Grants ATM0226059 and ATM0537043 and by NOAA support to the FSL provided through the U.S. Weather Research Program.

REFERENCES

Albers, S., J. McGinley, D. Birkenheuer, and J. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analysis of clouds, precipitation, and temperature. *Wea. Forecasting,* **11,** 273–287.

Baer, V. E., 1991: The transition from the present radar dissemination system to the NEXRAD Information Dissemination Service (NIDS). *Bull. Amer. Meteor. Soc.,* **72,** 29–33.

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.

——, and M. Wandishin, 2002: Determining the resolved spatial scales of Eta Model precipitation forecasts. Preprints, *15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 85–88.

——, S. Lakshmivarahan, and J. S. Kain, 2001: Verification of mesoscale features in NWP models. Preprints, *Ninth Conf. on Mesoscale Processes,* Fort Lauderdale, FL, Amer. Meteor. Soc., 255–258.

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, ATEX, and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.,* **112,** 693–709.

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

Bluestein, H. B., and M. H. Jain, 1985: Formation of mesoscale lines of precipitation: Severe squall lines in Oklahoma during the spring. *J. Atmos. Sci.,* **42,** 1711–1732.

Davis, C. A., B. G. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.,* in press.

——, ——, and ——, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.,* in press.

Du, J., S. L. Mullen, and F. Sanders, 2000: Removal of distortion error from an ensemble forecast. *Mon. Wea. Rev.,* **128,** 3347–3351.

Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.,* **18,** 1016–1022.

Ebert, E. E., and J. L. McBride, 1998: Routine verification of NWP quantitative precipitation forecasts for weather systems. Preprints, *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., J119–J122.

——, and ——, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.,* **239,** 179–202.

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.

——, 2002: Impact of verification grid-box size on warm-season QPF skill measures. *Wea. Forecasting,* **17,** 1296–1302.

Geerts, B., 1998: Mesoscale convective systems in the southeast United States during 1994–95: A survey. *Wea. Forecasting,* **13,** 860–869.

Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.,* **84,** 195–201.

Grell, G. A., J. Dudhia, and D. S. Stauffer, 1995: A description of the fifth-generation Penn State–NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398 + STR, 122 pp.

Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit, 2001: Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydrometeor.,* **2,** 406–418.

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

Houze, R. A., Jr., 1993: *Cloud Dynamics*. Academic Press, 573 pp.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

Jankov, I., W. A. Gallus Jr., B. Shaw, and S. E. Koch, 2005: The impact of different physical parameterizations and their interactions on warm season MCS rainfall. *Wea. Forecasting,* **20,** 1048–1060.

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,* K. A. Emanuel and D. J. Raymond, Eds., Amer. Meteor. Soc., 165–170.

——, 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.

Koch, S. E., and Coauthors, 2004: Real-time applications of the WRF model at the Forecast Systems Laboratory. Preprints, *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, 12.1.

Levene, H., 1960: Robust tests for the equality of variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling,* I. Olkinet et al., Eds., Stanford University Press, 278–292.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.,* **83,** 407–430.

McGinley, J. A., and J. R. Smart, 2001: On providing a cloud-balanced initial condition for diabatic initialization. Preprints, *18th Conf. on Weather Analysis and Forecasting/14th Conf. on Numerical Weather Prediction,* Fort Lauderdale, FL, Amer. Meteor. Soc., 40–44.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting,* **8,** 281–293.

——, 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting,* **10,** 681–688.

Ott, R. L., and M. Longnecker, 2001: *An Introduction to Statistical Methods and Data Analysis.* Duxbury, 1152 pp.

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 Technical Procedures Bulletin 477, National Oceanic and Atmospheric Administration/National Weather Service, 14 pp. [Available from Office of Meteorology, National Weather Service, 1325 East–West Highway, Silver Spring, MD 20910.]

Rutledge, S. A., 1986: A diagnostic numerical study of the stratiform region associated with a tropical squall line. *J. Atmos. Sci.,* **43,** 1337–1358.

Shaw, B. L., 2004: An objective inter-comparison of WRF, MM5, and NCEP Eta short-range quantitative precipitation forecasts for the International H_2O Project (IHOP) domain. *Extended Abstracts, Fifth WRF and 14th MM5 Users' Workshop,* Boulder, CO, NCAR, 3.6. [Available online at http://www.mmm.ucar.edu/mm5/workshop/workshop-papers_ws04.html.]

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,* Fort Lauderdale, FL, Amer. Meteor. Soc., J11–J15.

Tartaglione, N., S. Mariani, C. Accadia, A. Speranza, and M. Casaioli, 2005: Comparison of rain gauge observations with modeled precipitation over Cyprus using contiguous rain area analysis. *Atmos. Chem. Phys. Discuss.,* **5,** 2355–2376.

Tukey, J. W., 1993: The problem of multiple comparisons. *The Collected Works of John W. Tukey,* H. I. Braun, Ed., Vol. VIII, *Multiple Comparisons: 1948–1983,* Chapman Hall, 1–300.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.,* **125,** 527–548.

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

Zepeda-Arce, J., E. Foufoula-Georgiou, and K. K. Droegemeier, 2000: Space–time rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res.,* **105,** 129–146.