

The Impact of Large-Scale Forcing on Skill of Simulated Convective Initiation and Upscale Evolution with Convection-Allowing Grid Spacings in the WRF*

JEFFREY D. DUDA⁺ AND WILLIAM A. GALLUS JR.

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

(Manuscript received 8 January 2013, in final form 11 March 2013)

ABSTRACT

A set of mesoscale convective systems (MCSs) was simulated using the Weather Research and Forecasting model with 3-km grid spacing to investigate the skill at predicting convective initiation and upscale evolution into an MCS. Precipitation was verified using equitable threat scores (ETs), the neighborhood-based fractions skill score (FSS), and the Method of Object-Based Diagnostic Evaluation. An illustrative case study more closely examines the strong influence that smaller-scale forcing features had on convective initiation.

Initiation errors for the 36 cases were in the south-southwest direction on average, with a mean absolute displacement error of 105 km. No systematic temporal error existed, as the errors were approximately normally distributed. Despite earlier findings that quantitative precipitation forecast skill in convection-parameterizing simulations is a function of the strength of large-scale forcing, this relationship was not present in the present study for convective initiation. However, upscale evolution was better predicted for more strongly forced events according to ETs and FSSs. For the upscale evolution, the relationship between ETs and object-based ratings was poor. There was also little correspondence between object-based ratings and the skill at convective initiation. The lack of a relationship between the strength of large-scale forcing and model skill at forecasting initiation is likely due to a combination of factors, including the strong role of small-scale features that exert an influence on initiation, and potential errors in the analyses used to represent observations. The limit of predictability of individual convective storms on a 3-km grid must also be considered.

1. Introduction

Mesoscale convective systems produce a significant portion of warm-season rainfall (Fritsch et al. 1986) and a large amount of severe weather (Doswell et al. 1996; Wheatley et al. 2006; Wakimoto et al. 2006; Gallus et al. 2008; Duda and Gallus 2010). Many mesoscale convective systems (MCSs) also spawn mesoscale convective

vortices (MCVs) that can serve later as focal points for the development of new convection that may not be tied to any other large-scale forcing and can produce heavy precipitation (Fritsch et al. 1994; Trier and Davis 2002). Therefore, accurate forecasting of the occurrence and location of MCSs is important.

Initiation of deep, moist convection is difficult to simulate well using numerical weather prediction models, so much so that large field projects such as the International H₂O Project of 2002 have focused on it (Weckwerth and Parsons 2006). The strength of forcing for initiation on the large scale and the clarity of definition of surface boundaries that force surface-based deep, moist convection have been shown to impact how well a numerical model is able to predict initiation (Jankov and Gallus 2004; Szoke et al. 2004; Wilson and Roberts 2006). Wilson and Roberts (2006) noted a 10-km grid spacing version of the Rapid Update Cycle (RUC) model predicted the initiation of storms better when the forcing mechanism was a synoptic front as opposed to a gust

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/WAF-D-13-00005.s1>.

⁺ Current affiliation: University of Oklahoma, Norman, Oklahoma.

Corresponding author address: Jeff Duda, Center for Analysis and Prediction of Storms, University of Oklahoma, National Weather Center, Rm. 2500, 120 David Boren Blvd., Norman, OK 73072-7309.
E-mail: jeffduda319@gmail.com

front. Szoke et al. (2004) concluded that the RUC and fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model simulated initiation better when the surface boundary was clearly defined.

As technology and computing resources have improved over the last several years, mesoscale models are now often run routinely using grid spacing sufficient to not require the use of convective parameterization (e.g., Fowle and Roebber 2003; Done et al. 2004, Kain et al. 2008; Weisman et al. 2008; Weiss et al. 2008; Xue et al. 2009). A range of grid-spacing values between $O(100\text{ m})$ and 4 km have been stated as being fine enough to eliminate the need for convective parameterization (Weisman et al. 1997; Bryan et al. 2003). At these fine grid spacings, it is not known if earlier findings on the role of larger-scale forcing or surface features and their relationship to convection initiation (CI) and quantitative precipitation forecast (QPF) skill are valid.

The goal of this study is to investigate the effect of large-scale forcing on model skill of initiation and upscale evolution of convection simulated on a 3-km horizontal grid. Jankov and Gallus (2004) discovered using traditional contingency table skill measures applied to 10-km Eta Model output that QPF skill for MCSs was better when the large-scale forcing was stronger. The present study examines if the same relationship holds for 3-km grid spacing using the Weather Research and Forecasting model (WRF), but focuses instead on initiation and upscale evolution using object-based and neighborhood-based measures of skill as well as the traditional skill measures equitable threat score (ETS) and frequency bias. Section 2 discusses the experimental setup, and section 3 discusses the results from the main portions of the experiment. Discussion and conclusions are presented in section 4.

2. Methodology

a. Case selection

Cases were chosen in the United States from the 2006–10 period based on the rapid, isolated initiation of convection of a cellular morphology (or short line segments), obvious upscale evolution to an MCS over the few hours after initiation, and formation of a mature MCS small enough to fit reasonably into an approximately $1000\text{ km} \times 1000\text{ km}$ domain. To minimize complexity, the sample was also restricted to cases in which no prior convection was present in the region where the convection of interest that developed into the MCS to be studied initiated. Only cases for which a full dataset for analysis existed were included, and only cases considered a “hit” in terms of capturing convective initiation

TABLE 1. List of cases simulated.

Case No.	Initialization time (UTC)	Initialization date
1	1800	7 May 2009
2	0600	19 Jul 2008
3	1200	22 May 2007
4	1200	22 Jun 2007
5	1200	5 May 2008
6	1200	13 May 2009
7	1200	1 Jun 2010
8	1200	2 Jun 2010
9	1200	7 Jun 2010
10	1200	25 Jun 2010
11	1200	22 May 2008
12	1200	11 Aug 2007
13	1200	11 Jun 2008
14	1200	11 Jul 2008
15	1200	10 Jun 2010
16	1200	13 Jun 2010
17	1200	17 Jun 2010
18	1200	20 Jun 2010
19	1200	30 Aug 2010
20	1200	15 Jun 2009
21	1200	1 Sep 2010
22	0000	18 Jun 2010
23	1200	2 Aug 2008
24	1200	1 Apr 2006
25	0600	29 May 2007
26	1200	23 May 2006
27	1200	26 May 2006
28	1200	10 Sep 2009
29	1200	10 Sep 2010
30	1200	14 Sep 2010
31	0600	12 Aug 2007
32	1200	27 Mar 2010
33	1200	8 May 2006
34	1200	10 Feb 2009
35	1200	31 May 2007
36	1200	23 May 2007

were included. A case was included if convection developed within roughly 300 km and 4 h of observed systems, stronger constraints than those used in a similar study by Fowle and Roebber (2003), but weaker constraints than those used in Weisman et al. (2008). In addition, a subjective analysis was performed to determine if the simulated system shared enough attributes such as general track and causation with the observed system to be regarded as the same event. The final sample contained 36 cases (Table 1). The terms “target MCS” and “target convection” will be used hereafter to refer to the MCS on which each case focused and the convection that initiated and evolved into that MCS, respectively.

b. Model configuration

Simulations were conducted using the WRF model, version 3.1.1, using the Advanced Research (ARW) dynamics core (Skamarock et al. 2008). A pair of one-way

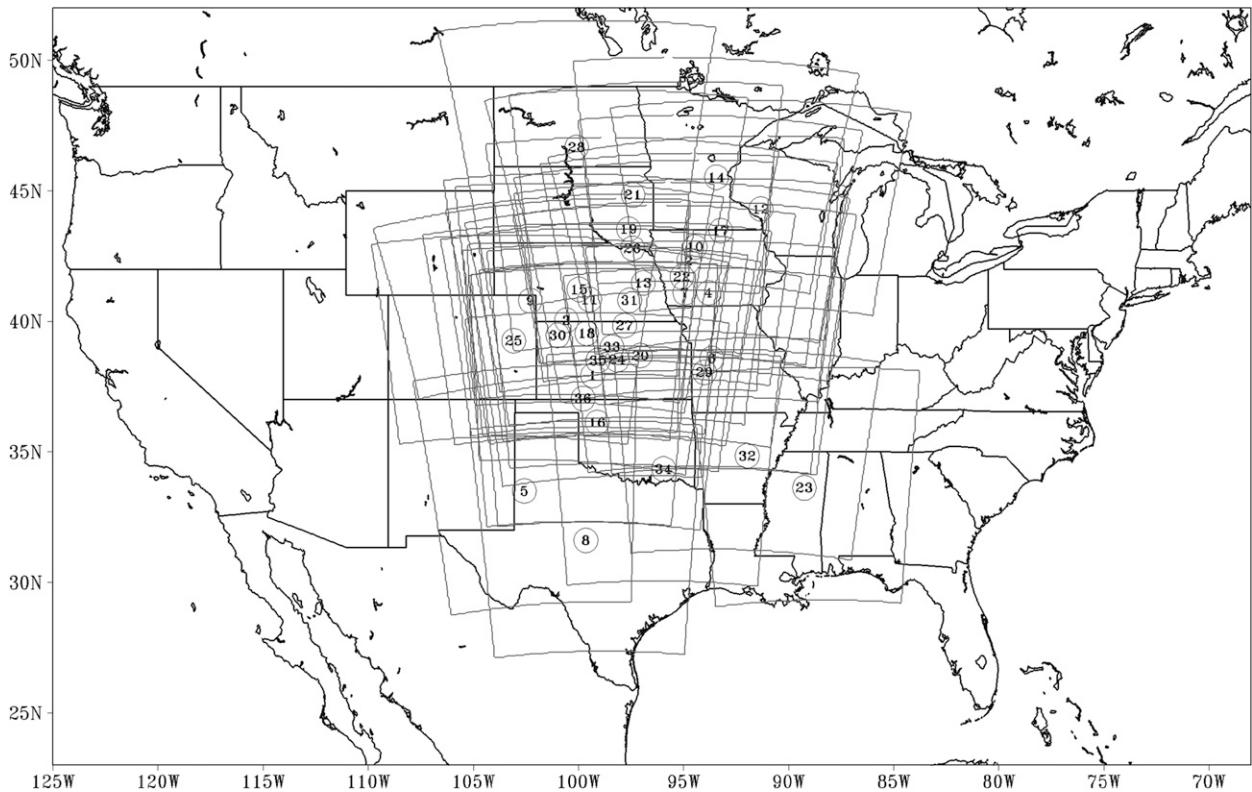


FIG. 1. Model domains for the 36 cases. The center of each domain is indicated by the encircled number for each case.

nested domains was used where the outer domain was $3000 \text{ km} \times 3000 \text{ km}$ with a horizontal grid spacing of 12 km, and the inner domain covered a roughly $1000 \text{ km} \times 1000 \text{ km}$ area with a horizontal grid spacing of 3 km. Figure 1 contains a map of the model domain for each case. The aspect ratio of the inner domain varied to best fit the initial and mature stages of the target MCS. Analysis for each case was performed on only the inner domain. The Thompson microphysics scheme was used (Thompson et al. 2008) with the Mellor–Yamada–Janjić planetary boundary layer scheme (Mellor and Yamada 1982; Janjić 2002) and the Monin–Obukhov (Janjić Eta) surface layer scheme (Janjić 2002). The model also used the Noah land surface model (Ek et al. 2003) and the Rapid Radiative Transfer Model shortwave (Mlawer et al. 1997) and Dudhia longwave (Dudhia 1989) radiation schemes. Initial and lateral boundary conditions came from 12-km grid spacing North American Mesoscale (NAM) model analyses. All cases were integrated for 24 h with 15-min output, and most were initialized at 1200 UTC.

c. Analysis procedures

It has been shown (e.g., Ahijevych et al. 2009) that traditional contingency table statistical measures such as

ETS may be inadequate for validating model output at high resolutions, primarily because they may give a much different impression of the skill of a forecast than a human would. Because of the high spatial variability of the high-resolution model output, a nontraditional method of verification was used in this study: the Method of Object-Based Diagnostic Evaluation (MODE; Davis et al. 2006). MODE creates objects within both forecast and observation datasets and then matches objects in the forecast field with those in the observation field according to various parameters and fuzzy logic. [Detailed instructions on the operation of MODE can be found in Davis et al. (2006).] It outputs several object-based measurements that can be used for nontraditional verification.

Convective initiation was defined as the first hour in which 3 mm of precipitation was produced from the target convection (or in the case of the model, the model's representation of the target convection). Stage IV multisensor quantitative precipitation estimates (ST4) from the National Centers for Environmental Prediction (NCEP) served as the observed precipitation. ST4 data have a spatial resolution of approximately 4.7 km, and are available in 1-, 6-, and 24-h accumulation intervals. Bilinear interpolation was used to regrid the ST4 data to the model grid for each case.

Differences in temporal resolution between the ST4 data and the model output created uncertainty in determining the timing error for convective initiation. Because ST4 data only cover hourly blocks ending at the top of each hour, Next Generation Weather Radar (NEXRAD) level III rain-rate estimates were used instead to more precisely determine the observed initiation time. Although these estimates, like ST4, cover 1-h periods, they are available approximately every 5 min. Thus, the use of these data in conjunction with 5-min radar reflectivity images can allow a fairly precise determination of initiation time. Because the model output had a temporal frequency of 15 min and thus initiation could only be specified with a precision of 15 min, 15-min bins were also assigned to the observed initiation. The accumulation threshold used was 0.10 in. This is slightly different than 3 mm, but its use was required since 0.10 in. is the contour value closest to 3 mm used to display NEXRAD level III 1-h precipitation in the National Climatic Data Center's Weather and Climate Toolkit. As an example, if 0.10 in. of precipitation was indicated by radar to occur in the 1900–1915 UTC period and in the model in the 1930–1945 UTC period, then the actual temporal error in initiation could range from 15 to 45 min. For this study the midpoint of the range of values was used so the error in the example would be 30 min. Spatial error in initiation was computed using 1-h ST4 data instead of radar-derived precipitation since the distance between centroids of precipitation objects was output from MODE, which required the gridded ST4 output.

The skill of the model for CI was obtained using a formula that has a 3D bell-curve shape:

$$\text{score} = e^{-(At^2 + Bl^2)}, \quad (1)$$

where t represents the temporal error, l the spatial error, and A and B are scaling parameters so that equal weight is given to both types of error. Both A and B are scaled such that a 1-h error is equivalent to the average distance covered by the first storm in the first hour after initiation from all 36 cases, which was approximately 44 km. This resulted in values for A of $4.514 \times 10^{-5} \text{ min}^{-2}$ and for B of $8.492 \times 10^{-5} \text{ km}^{-2}$. The best score of 1.0 is achieved when a forecast has a temporal error of 0 min and a spatial error of 0 km.

Although MODE was run using both 1- and 6-h precipitation accumulation intervals (with the accumulation interval starting at initiation in both datasets), only the object-based statistics obtained from the 6-h accumulation interval were used in the present study. Statistics from the 1-h MODE runs were not used because large errors in object shape, size, and magnitude (of 1-h

precipitation) were often judged to not represent the true nature of the skill of the model regarding CI. The 6-h accumulated precipitation was used to evaluate the upscale evolution of convection toward an MCS. Since the mesoscale observations needed to evaluate upscale evolution from the standpoint of airflows (rear-inflow jet, overturning updraft, etc.) and thermodynamics (i.e., latent heating) do not exist for the majority of cases, precipitation accumulated over a timespan that contains the upscale evolution was regarded to be the best alternative.

The present study followed the method used in Jankov and Gallus (2004) to determine the large-scale forcing present at initiation. RUC 20-km analyses¹ (on NCEP grid 252) were used to represent observations. Time- and area-averaged forcing values were computed for three different scenarios: 1) in the RUC analyses using the 6 h prior to the time and location of initiation of the observed convection (scenario A), 2) in the WRF output using the 6 h prior to the time and location of observed initiation regardless of whether or not the WRF correctly simulated the target convection (scenario B), and 3) in the WRF output using the 6 h prior to the time and location of initiation in the WRF (scenario C). In scenarios A and B the centroid of observed 1-h precipitation at initiation was used as the center point of the $4^\circ \times 4^\circ$ latitude–longitude box over which the forcing terms were averaged, whereas in scenario C the centroid of the 1-h precipitation at initiation in the forecast was used instead as the center of the box. Five measures of large-scale forcing were used: 700-hPa omega, surface frontogenesis, 200-hPa divergence, 250–850-hPa differential vorticity advection (DVA), and 850-hPa temperature advection. Forcing fields were filtered using the General Meteorological Package (GEMPAK) Gaussian weighted smoothing function (GWFS) to remove features smaller than $13\Delta x$ (39 km), thus restricting the scale of features resolved on the WRF grid roughly to those that are resolved on the RUC analysis grid. Consistent with quasigeostrophic theory, the forcing fields (except for 700-hPa omega and 200-hPa divergence) were also coupled, by division, to the stability term $\sigma = -T(d\ln\theta/dp)(R/p)$, where T and θ are temperature and potential temperature, respectively; R is the dry-air gas constant; and p is pressure. To keep the height above ground used in the quasigeostrophic forcing measures roughly consistent between cases, the 800-hPa level was used in place of the 850-hPa level when the center of the

¹ Although 13-km analyses were available for some of the cases, they were not available for all cases. Therefore, for consistency, only 20-km analyses were used to represent observations.

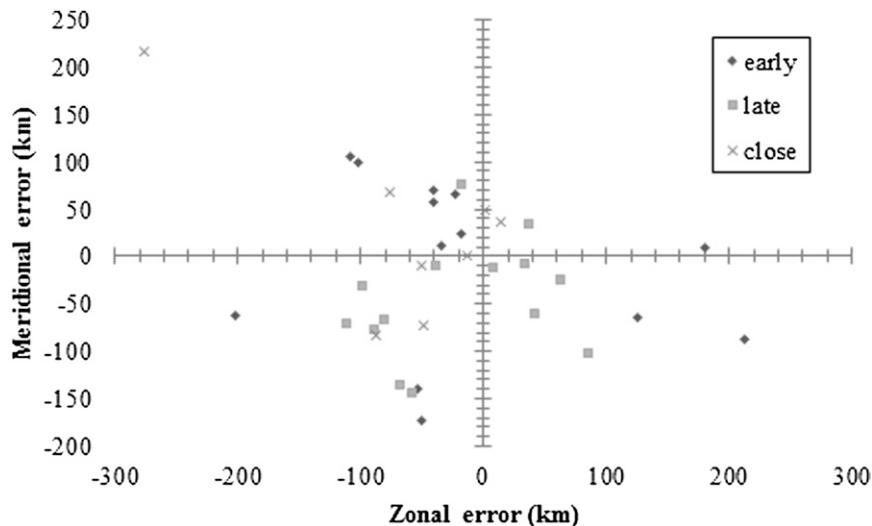


FIG. 2. Initiation location error classified by timing error (symbols in inset). Early initiation errors had a timing error of -30 min or less, while late initiation errors had a timing error of $+30$ min or more. Close initiation errors had a timing error of less than 30 min. Negative values of the zonal and meridional error imply a westerly and southerly error, respectively.

averaging box was over higher terrain west of 100°W in the central plains.

3. Results

a. Initiation

Temporal and distance errors for WRF forecasts of initiation were computed for each of the 36 cases (Fig. 2; Table 2). In all but two of the cases where the WRF initiated late, a southerly component was present in the location error. Also, the error had a westward component in most cases (24, or 65%). The average west–east and north–south components of the location error were 14 km west and 27 km south, respectively, indicating a preference for WRF initiation to occur south-southwest of the observed location. The mean absolute error of the displacement was 105 km. The average timing error for initiation was only -0.56 min, as timing errors were approximately normally distributed with early and late errors canceling each other. Given the model configuration and the initial and lateral boundary data used to feed the model, these errors are reasonably consistent with the limit of predictability for a single intense, convective storm, which is poorly resolved on the model grid (Zhang et al. 2003; Skamarock 2004; Hohenegger 2007; Hohenegger and Schär 2007).

The overall strength of large-scale forcing was computed by summing the ranks of the forcing values of each measure (Fig. 3). It is clear that no relationship exists between the strength of large-scale forcing and the skill

in CI, nor between the strength of forcing and the location error alone (not shown). There was a weak relationship between the strength of forcing and the timing error, however (not shown). While there were cases that were both strongly forced and well forecast, as well as cases that were weakly forced and poorly forecast, there were just as many cases for which the strength of forcing was not commensurate with the skill at forecasting CI. Additionally, no relationship was found between the strength of any one particular measure of forcing to the skill in forecasting CI (not shown). It was also found that there was no correlation between the magnitude of the error in the large-scale forcing and the initiation skill score (not shown). The large-scale forcing was reasonably well forecast (Figs. 4–7), although the presence of a few outlier cases caused lower correlation coefficients than might otherwise be expected. Some outliers were a result of relatively small values of the forcing term at a small number of points influencing the quotient and spatial average. However, other outliers were not related to the stability. Regarding outliers, if a case contained a forcing measure that was an outlier, then that was the only forcing measure for that case for which the forcing was an outlier. There was no commonality among the cases for which a forcing term was an outlier in terms of skill at forecasting CI; the outliers appeared to be randomly scattered throughout the cases. The lack of relationship between the strength of forcing and skill at forecasting CI likely has several contributions. One contribution that was tested herein is that convective initiation is sensitive enough to smaller-scale features so

TABLE 2. Spatial and temporal errors for CI for each case, triggering mechanism for initiation, and computed skill score and its rank. Positive (negative) temporal errors indicate the model initiated convection late (early) compared to the observations.

Case No.	Initialization date	Triggering mechanism	Spatial error (km)	Temporal error (min)	Skill score	Rank of skill score
1	7 May 2009	North of stationary front	132	+90	0.16	24
2	19 Jul 2008	Warm front	105	+75	0.30	20
3	22 May 2007	Stationary front near surface low and dryline	70	-60	0.56	11
4	22 Jun 2007	Warm or stationary front	69	-105	0.41	17
5	05 May 2008	Orography	13	+30	0.95	1
6	13 May 2009	Cold front	150	-105	0.09	28
7	1 Jun 2010	Outflow boundary (OFB)	350	0	<0.01	36
8	2 Jun 2010	Random warm sector (MCV)	230	-180	<0.01	35
9	7 Jun 2010	Dryline-pressure trough	72	+90	0.44	16
10	25 Jun 2010	Warm front-OFB-wind-shift-line triple point	103	0	0.40	18
11	22 May 2008	Warm front-dryline triple point	15	+75	0.76	5
12	11 Aug 2007	Cold front	35	-210	0.12	26
13	11 Jun 2008	Prefrontal trough or cold front	181	-90	0.04	30
14	11 Jul 2008	Warm and cold fronts near surface low	142	-45	0.16	23
15	10 Jun 2010	Pressure trough	117	+45	0.28	21
16	13 Jun 2010	OFB-dryline triple point	133	+240	0.02	33
17	17 Jun 2010	Prefrontal trough	121	-15	0.28	22
18	20 Jun 2010	Triple point between stationary fronts	50	+60	0.69	6
19	30 Aug 2010	Quasi-stationary front-random warm sector	78	+60	0.50	14
20	15 Jun 2009	Dryline bulge and stationary front triple point	212	-75	0.02	31
21	1 Sep 2010	Surface low and associated fronts and boundaries	87	-30	0.51	13
22	18 Jun 2010	Low-level jet nose	180	-75	0.05	29
23	2 Aug 2008	Cold front	152	-210	0.02	31
24	1 Apr 2006	Dryline	40	+90	0.61	9
25	29 May 2007	Wind-shift line (+ orography?)	38	0	0.88	2
26	23 May 2006	Cold front	67	+45	0.63	8
27	26 May 2006	Weak dryline or cold front and warm front near triple point	51	+30	0.77	4
28	10 Sep 2009	Cold front	151	+195	0.03	31
29	10 Sep 2010	Surface low, warm front, and cold front	34	-90	0.63	7
30	14 Sep 2010	Orography (Palmer Divide, CO)	34	+120	0.47	15
31	12 Aug 2007	Elevated front	104	+75	0.31	19
32	27 Mar 2010	Cold front near triple point	49	-15	0.81	3
33	8 May 2006	Cold front	140	-75	0.15	25
34	10 Feb 2009	Dryline	155	+90	0.09	27
35	31 May 2007	Low-level jet or random warm sector	80	-45	0.53	12
36	23 May 2007	Stationary front	31	-105	0.56	10

as to significantly impact the influence of the large-scale forcing. The influence of smaller-scale forcing on initiation is explored more in section 3c. Another contribution is from errors in the data used to represent the observations; namely, the ST4 and RUC data. Finally, model errors from physics parameterizations (i.e., microphysics, boundary layer, land surface, and radiation schemes) may have contributed to the lack of a relationship. We will focus the rest of this section on the impact of large-scale forcing in this study.

For each scenario defined in section 2c, a composite of large-scale forcing was computed for each of the large-scale forcing fields by translating the forcing field for each case so that the location of initiation in each case lay at the same point. Due to sample size and large

spreads among forcing values at a given grid point, the median value was judged to be the most representative. The resulting composited fields for scenarios A, B, and C are shown in Figs. 5–7. The forcing terms from the RUC analyses (scenario A) showed many features supportive of the initiation of convection on the synoptic scale. A region of upward motion was centered over the region of initiation with mostly weak, downward motion elsewhere (Fig. 5a). Positive DVA generally occurred west and northwest of the region of initiation (Fig. 5b) with weaker DVA elsewhere, a pattern consistent with a mid-to upper-level trough that commonly spawns organized convective events. A strong signal was also seen in the 850-hPa temperature advection (Fig. 5c). Warm-air advection (WAA) was located to the east of the location of

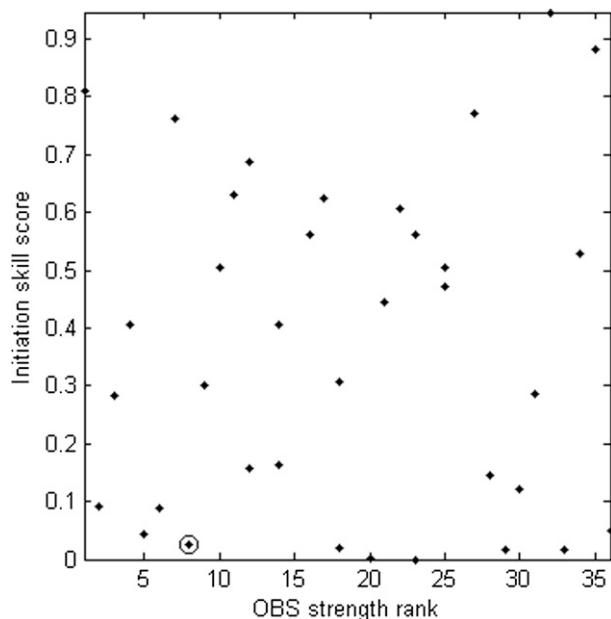


FIG. 3. Skill score as a function of large-scale forcing strength using a $4^{\circ} \times 4^{\circ}$ box for all five measures defined in the text. The circled dot near the bottom represents the data for case 28, which is examined in the case study in section 3c.

initiation with cold-air advection to the west. A gradient in advection ran through the location of initiation from SSW to NNE. This pattern is consistent with a low-level height or pressure trough oriented SSW–NNE, which is commonly associated with the synoptic forcing that supports convection. No strong, clear signal existed in the frontogenetical forcing, although some stronger forcing was found near the region of initiation (Fig. 5d). A noticeable pattern of divergence at 200 hPa, locally strong, was found at and north of the region of initiation (Fig. 5e). However, comparable values were also seen in areas well removed from the location of CI. The spread of the distribution of forcing values around the location of CI was one to two orders of magnitude larger than the median values. This high spread reflects the high variability of the forcing among the cases.

Some of the composite forcing fields from the WRF in scenario B agreed reasonably well with those of the RUC analyses from scenario A, whereas other fields disagreed. One field for which the WRF in scenario B agreed with the RUC in scenario A was the 700-hPa omega field. There was agreement both in the magnitudes and in the distribution of upward and downward motion (Fig. 6a, where the entire panel fits within the dashed box in Fig. 5a). The other field for which there was agreement was the 850-hPa temperature advection field. The spatial distribution agreed fairly strongly with that in the RUC analyses, although the peak magnitudes

of WAA in the WRF were slightly higher than they were in the RUC analyses (Fig. 6c). As in the composited surface frontogenesis field in scenario A, there was not a definitive signal in the surface frontogenesis field in scenario B (Fig. 6d). Although a front and associated narrow band of strong frontogenesis was present near the region of initiation in many, but not all, cases, this frontogenesis band did not have a preferred orientation, so compositing washed out the feature. No attempt was made to rotate the frontogenesis field for each case to align the frontogenesis band since the band was not always linear (thus precluding defining an angle for the band) or was not always present. Therefore, the lack of a definitive signal is not believed to represent a lack of influence of surface frontogenesis on convective initiation. On the contrary, detailed examination of the cases revealed a strong tendency for CI to be intimately linked to areas of strong surface frontogenesis. No such localized bands of intense forcing were commonly found near the region of initiation in the composited DVA and 200-hPa divergence fields (Figs. 6b and 6e). Much as with scenario A, spread in the composited forcing fields for scenario B was very large compared to the median values. Nonetheless, the lack of pattern in the DVA and 200-hPa divergence fields may indicate a lack of dependence of CI on these forcing terms. This lends credence to the notion that while the large-scale support had some influence over the location of initiation, smaller-scale forcings potentially played a larger role in pinpointing where CI would occur.

Due to timing errors of CI, the forcing terms were also composited for scenario C (Fig. 7). It is expected that these composited fields may agree better with those from the RUC analyses in scenario A than those from the WRF in scenario B since each composited field from scenario C represents the forcing from the 6-h period leading up to initiation in the model. Also, convection that initiates early in the model may contaminate the fields used to composite the data in scenario B, which ignored timing errors in CI. However, some errors in the forcing terms in scenario C are expected to remain simply due to the timing mismatch between the RUC analyses and the model output. Better agreement was seen between the 200-hPa divergence from scenario C and that in scenario A (cf. Figs. 5e and 7e) than was seen between scenarios B and A, as stronger values of divergence were located closer to the region of initiation in scenario C. A somewhat more coherent signal was present in the surface frontogenesis (Fig. 7d) than was seen in scenario B, but the magnitudes were not much different from those in scenario B. A general lack of signal was found in the DVA field (Fig. 7b). The patterns of 700-hPa omega and 850-hPa temperature advection

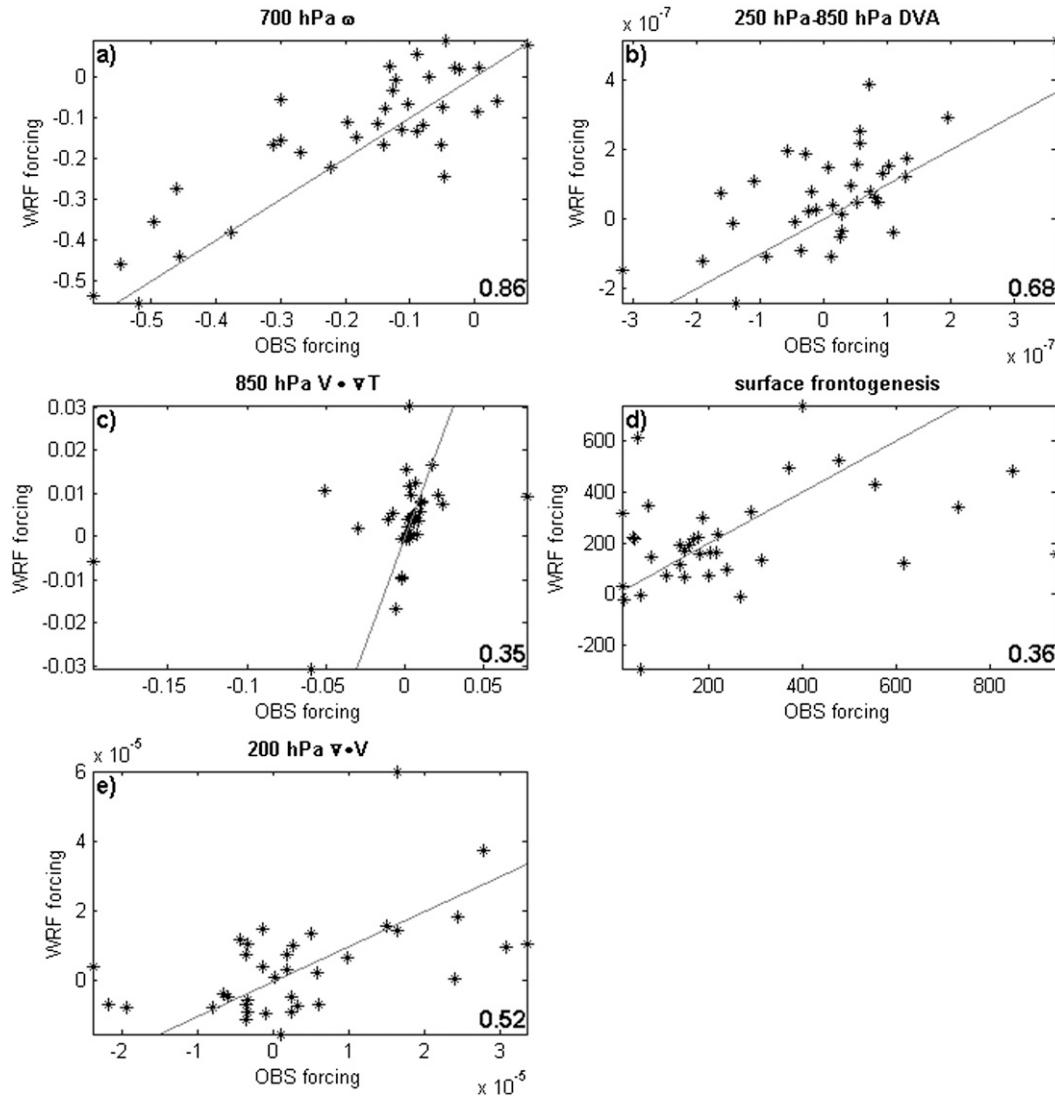


FIG. 4. Scatterplots of large-scale forcing from the observations and the model at the time and location of the observed initiation: (a) 700-hPa omega ($\mu\text{b s}^{-1}$), (b) 250 – 850 hPa differential vorticity advection (10^{-9} s^{-2}), (c) 850-hPa temperature advection (10^5 K s^{-1}), (d) surface frontogenesis [$^{\circ}\text{C (100 km)}^{-1} (3 \text{ h})^{-1}$], and (e) 200-hPa divergence (10^5 s^{-1}). A perfect-fit line is included in each panel. Values were scaled for graphical purposes by the indicated magnitude. Correlations between the observed and modeled large-scale forcing values are indicated in the bottom right of each panel. Frontogenesis was not coupled to the stability term in (d). Some values in (b),(d) were computed using the 800-hPa level instead of the 850-hPa level.

were very similar to those from scenarios A and B. The consistency of the signals from these two forcing terms may indicate a stronger influence of them on CI than that of the other forcings. Overall, it is clear that these cases contained synoptic-scale forcings supportive of CI, and the model forecasts of the synoptic forcings generally agreed with the RUC analyses (as shown in Fig. 4). However, the specific location of CI was not usually where the magnitude of the synoptic forcing was greatest, implying smaller-scale forcings played a bigger role in exactly when and where CI occurred.

The specific surface or near-surface triggering mechanism responsible for initiation in each of the observed cases and in the simulations was classified (e.g., cold front, warm front, dryline, orography, outflow boundary; Table 2). The mechanism was the same in 22 (61%) of the cases, different in 11 (31%) of the cases, and undetermined in the remaining 3 (8%). Of the nine cases with the strongest forcing, the triggering mechanism was a surface low or triple point associated with two fronts (warm and cold, warm and dryline, cold and dryline, etc.) in five of those cases. No common triggering mechanism

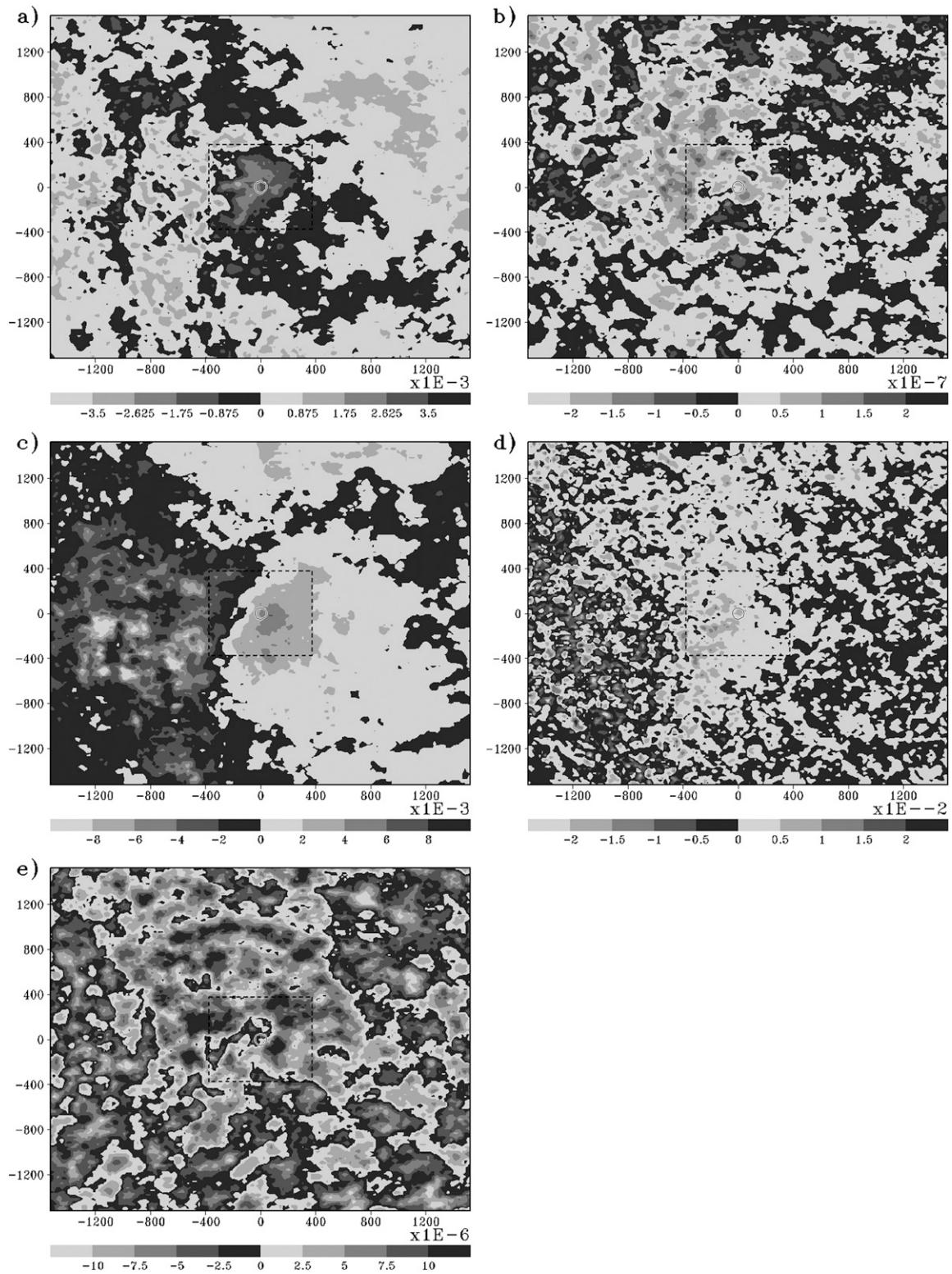


FIG. 5. Composed forcing fields for scenario A: (a) 700-hPa omega ($1000 \mu\text{b s}^{-1}$), (b) 250 – 850 hPa differential vorticity advection [$\text{s}^{-2} (\text{kg}^2 \text{m}^{-4} \text{s}^{-2})$], (c) 850-hPa temperature advection [$\text{K s}^{-1} (\text{kg}^2 \text{m}^{-4} \text{s}^{-2})$], (d) surface frontogenesis [$^{\circ}\text{C} (100 \text{ km})^{-1} (3 \text{ h})^{-1} (\text{kg}^2 \text{m}^{-4} \text{s}^{-2})$], and (e) 200-hPa divergence (s^{-1}). Values along the horizontal and vertical axes represent the approximate horizontal distance (km) from the point of initiation. The dashed box in each panel represents the size of the area shown in each panel of Figs. 6 and 7. The circle at the origin represents the location of initiation. These values include data obtained from the 800-hPa level for some cases, as explained in section 2c.

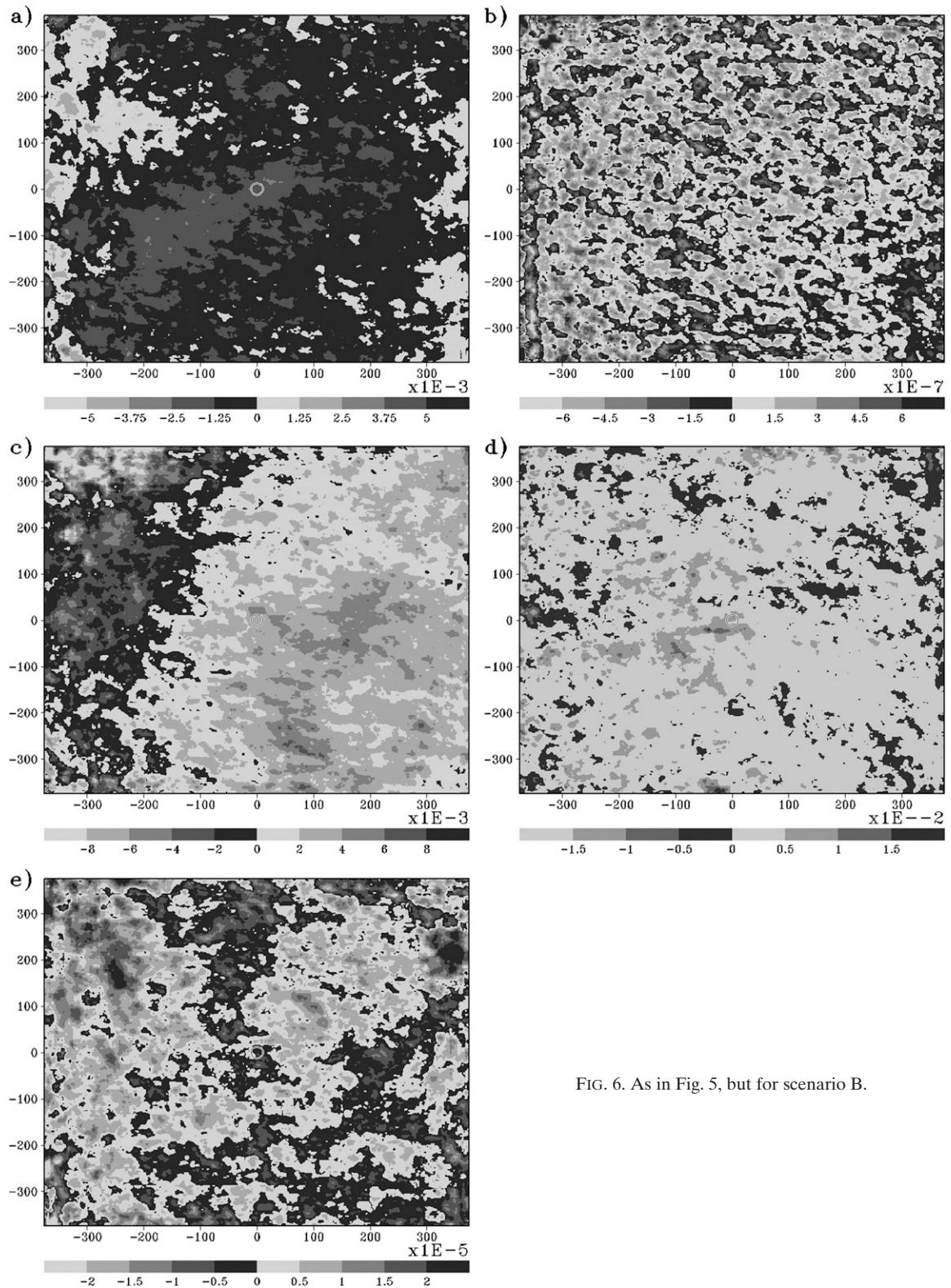


FIG. 6. As in Fig. 5, but for scenario B.

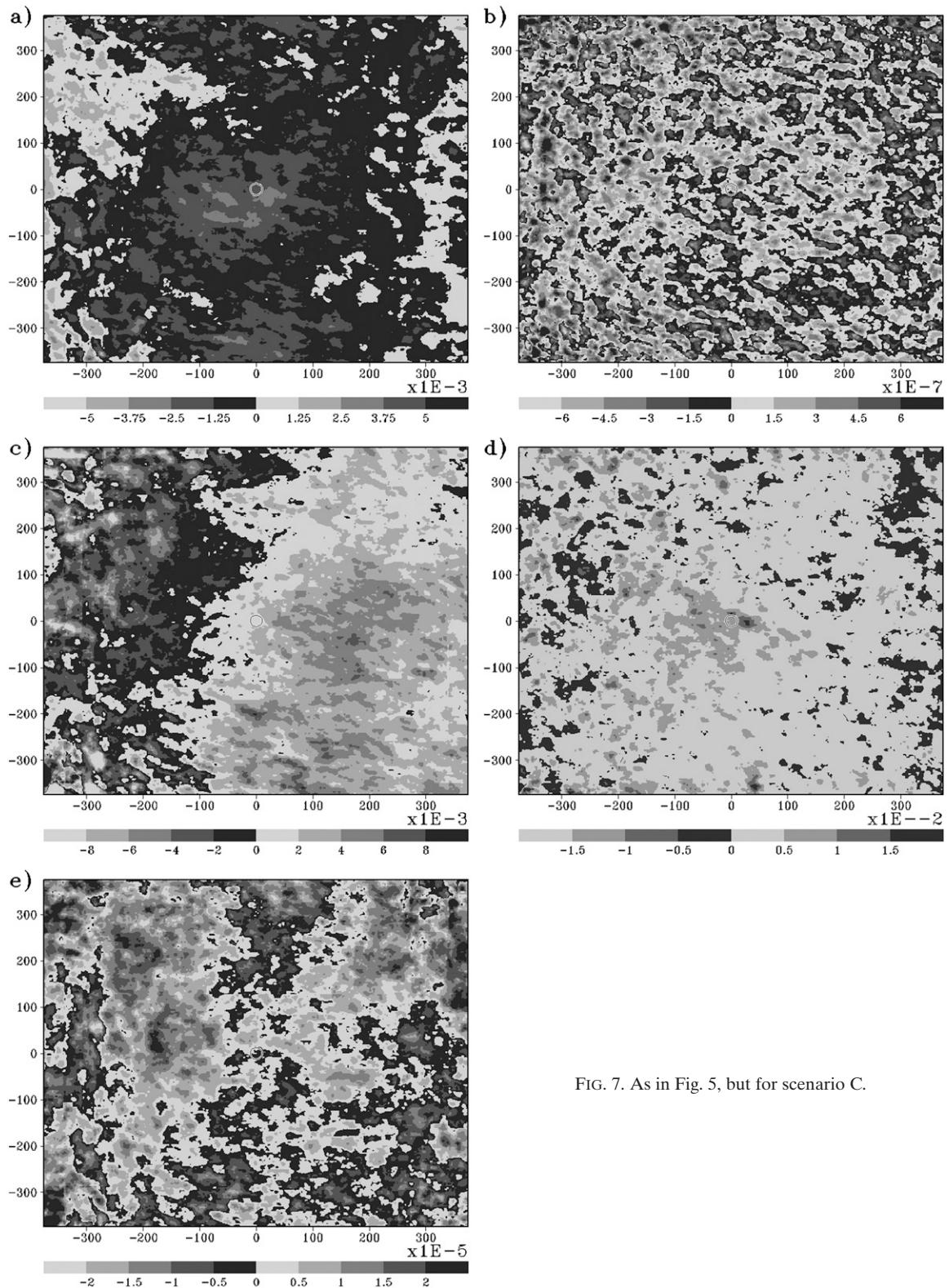


FIG. 7. As in Fig. 5, but for scenario C.

existed among the weakest forced nine cases. Of the nine cases where WRF had the best skill at forecasting initiation, a surface low or triple point was also common, present in four of the cases. No common triggering mechanism existed among the nine least skillful cases. Despite the number of cases for which triggering mechanisms were the same, the spatiotemporal error in CI further suggests that smaller-scale variability within these triggering mechanisms played a role in differences in the timing and location of CI in the WRF runs compared to the observations.

b. Upscale evolution

The upscale evolution of convection into an MCS was evaluated using several methods applied to rainfall exceeding a threshold, including the traditional objective skill measures of ETS and frequency bias, a neighborhood verification metric called the fractions skill score (FSS; Schwartz et al. 2010), and MODE object-based measures using 6-h precipitation accumulation starting at initiation. The traditional skill measures and FSS are given by

$$\text{ETS} = \frac{\text{CFA} - \text{CHA}}{F + O - \text{CFA} - \text{CHA}}, \quad (2)$$

$$\text{CHA} = \frac{O \times F}{N}, \quad (3)$$

$$\text{bias} = \frac{F}{O}, \quad (4)$$

$$\text{FSS} = 1 - \frac{\text{FBS}}{\text{FBS}_{\text{worst}}}, \quad (5a)$$

$$\text{FBS} = \frac{1}{N} \sum_{i=1}^N (\text{NP}_i^f - \text{NP}_i^o)^2, \quad \text{and} \quad (5b)$$

$$\text{FSS}_{\text{worst}} = \frac{1}{N} \left[\sum_{i=1}^N (\text{NP}_i^f)^2 + \sum_{i=1}^N (\text{NP}_i^o)^2 \right], \quad (5c)$$

where F is the number of grid points at which the event was forecast, O is the number of grid points at which the event occurred, N is the total number of grid points, CFA is the number of grid points at which the event was correctly forecast, and CHA is the number of points where a correct “yes” forecast could be made by chance. For the fractions skill score, FBS is the fractions Brier score, which is the same as the traditional Brier score but with forecast and observed probabilities of precipitation based on a spatial neighborhood [NP in Eqs. (5b) and (5c)] rather than being binary (0 or 1). Subscripts f and o in Eqs. (5b) and (5c) refer to forecast and observed

quantities, respectively. ETS and FSS were computed for thresholds of 0.01, 0.05, 0.10, 0.25, and 0.50 in. Specific MODE object-based attributes examined include the area of the object representing the target convection, the value of the 50th percentile of precipitation intensity within that object, and the total precipitation volume of that object. A threshold amount of 2.54 mm (0.10 in.) and a convolution radius of three grid squares were used in MODE to define objects (Davis et al. 2006). The object-based scores from MODE focused only on the target convection, an important distinction due to the implications of the representativeness of the results.

Because the simulated convection frequently did not initiate at the same time as the observed convection, two different time periods were used for evaluation. One period, hereafter the free time forecast, refers to the time period corresponding to the first 6 h after the target convection initiated in the model regardless of whether that period matches the first 6 h after convection initiated in the ST4 analyses. The other period, hereafter the fixed time forecast, refers to the time period corresponding to the first 6 h after initiation of convection in the ST4 data regardless of whether or not the target convection had developed in the model during that period. For 10 cases MODE was unable to completely separate either the forecast or observed precipitation from other nontarget precipitation when creating objects; these cases were excluded from the evaluation. The thresholds used in MODE resulted in an average area of the precipitation objects of approximately 60 000 km². This is about one-half the value used in Grams et al. (2006), but is not unreasonable since only the first 6 h after initiation—covering only the development of the target MCS and only the early stages of the mature MCS—were considered. The object-based measures were ranked by the average relative error to obtain one score for each case.

Box-and-whiskers plots (Fig. 8) hint at a relationship between the strength of large-scale forcing and ETS. In general, the strongest forced cases had the highest ETSs, but there was much overlap between the distributions for strongest and weakest forced cases. Correlation coefficients between ETS and strength of large-scale forcing range between -0.3 and -0.55 , depending on threshold. Correlations strengthened with an increasing threshold except for the highest threshold. A similar conclusion is reached by examination of FSSs (not shown). This is not surprising given the very strong correlation between ETS and FSS for the cases (Fig. 9). There is also a slight dependence of bias on the strength of forcing (Fig. 10). Biases tended to be concentrated more tightly around 1.0 for the stronger forced cases. A frequency bias of 1.0 is a necessary, but not sufficient, condition for a forecast

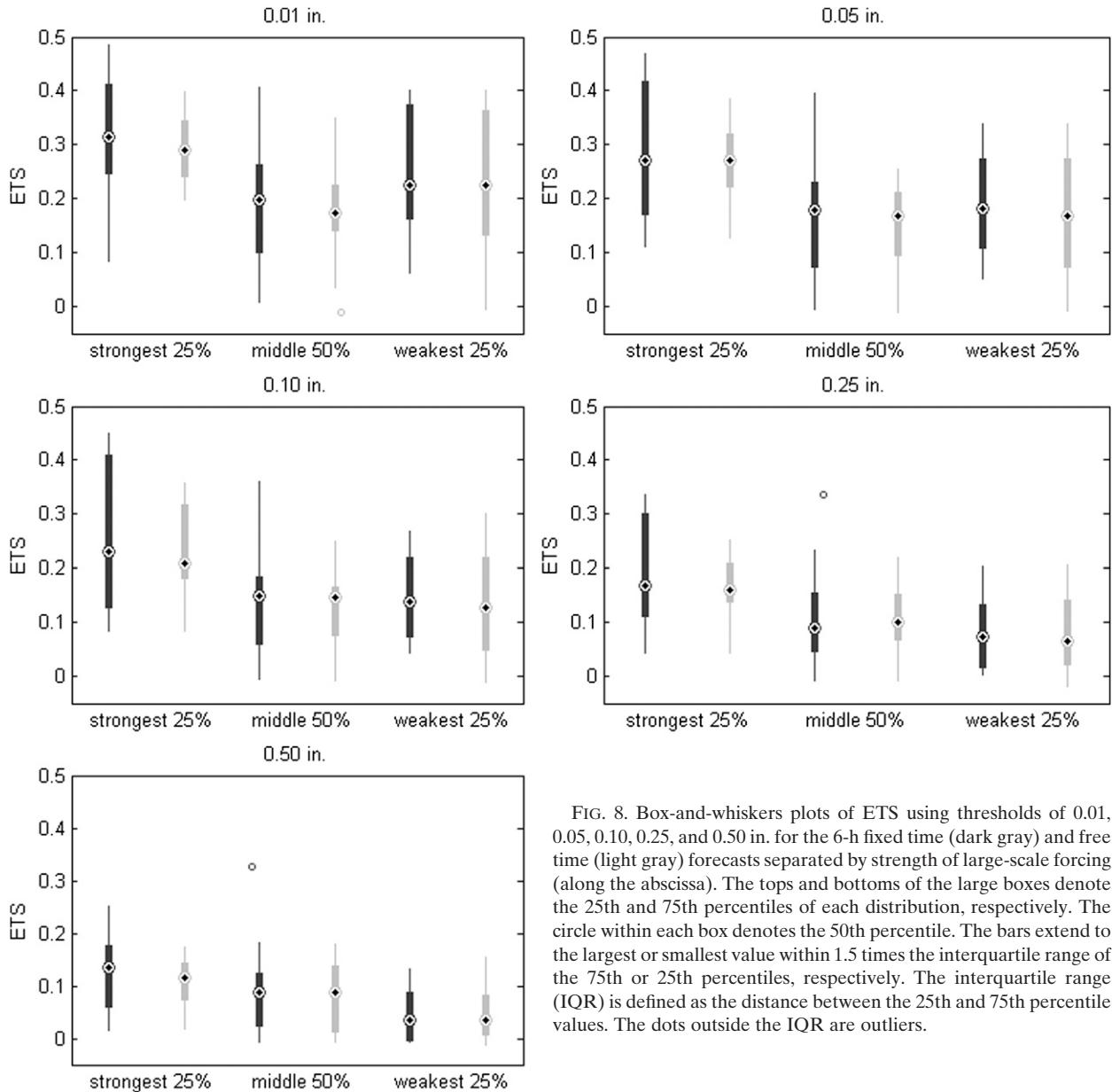


FIG. 8. Box-and-whiskers plots of ETS using thresholds of 0.01, 0.05, 0.10, 0.25, and 0.50 in. for the 6-h fixed time (dark gray) and free time (light gray) forecasts separated by strength of large-scale forcing (along the abscissa). The tops and bottoms of the large boxes denote the 25th and 75th percentiles of each distribution, respectively. The circle within each box denotes the 50th percentile. The bars extend to the largest or smallest value within 1.5 times the interquartile range of the 75th or 25th percentiles, respectively. The interquartile range (IQR) is defined as the distance between the 25th and 75th percentile values. The dots outside the IQR are outliers.

to be perfect, but when examined in conjunction with the ETSS, the notion that stronger forced events were more accurately forecast is further supported. Also, with the exception of a few outliers, biases are not very high, indicating that bias was not inflating ETSS to mask a signal representing a dependence of ETS on large-scale forcing (Hamill 1999). While these metrics do not indicate a strong relationship between the strength of large-scale forcing and QPF skill, they are domain-wide computations and may not reflect the skill of the QPF representing the target MCS since other significant precipitation occurred within the model domain during the fixed and free time forecasts for a significant number

of cases. Therefore, object-based verification from MODE is crucial for a representative evaluation in this study. However, only a poor correlation was found between the ranking of the scores from MODE and the strength of the large-scale forcing (Fig. 11). The overall conclusion is that a relationship between the strength of the large-scale forcing and the QPF skill representing the upscale evolution of the target convection toward an MCS exists but it is not strong.

c. Case study

To illustrate the influence that smaller-scale features had on the location and timing of convective initiation

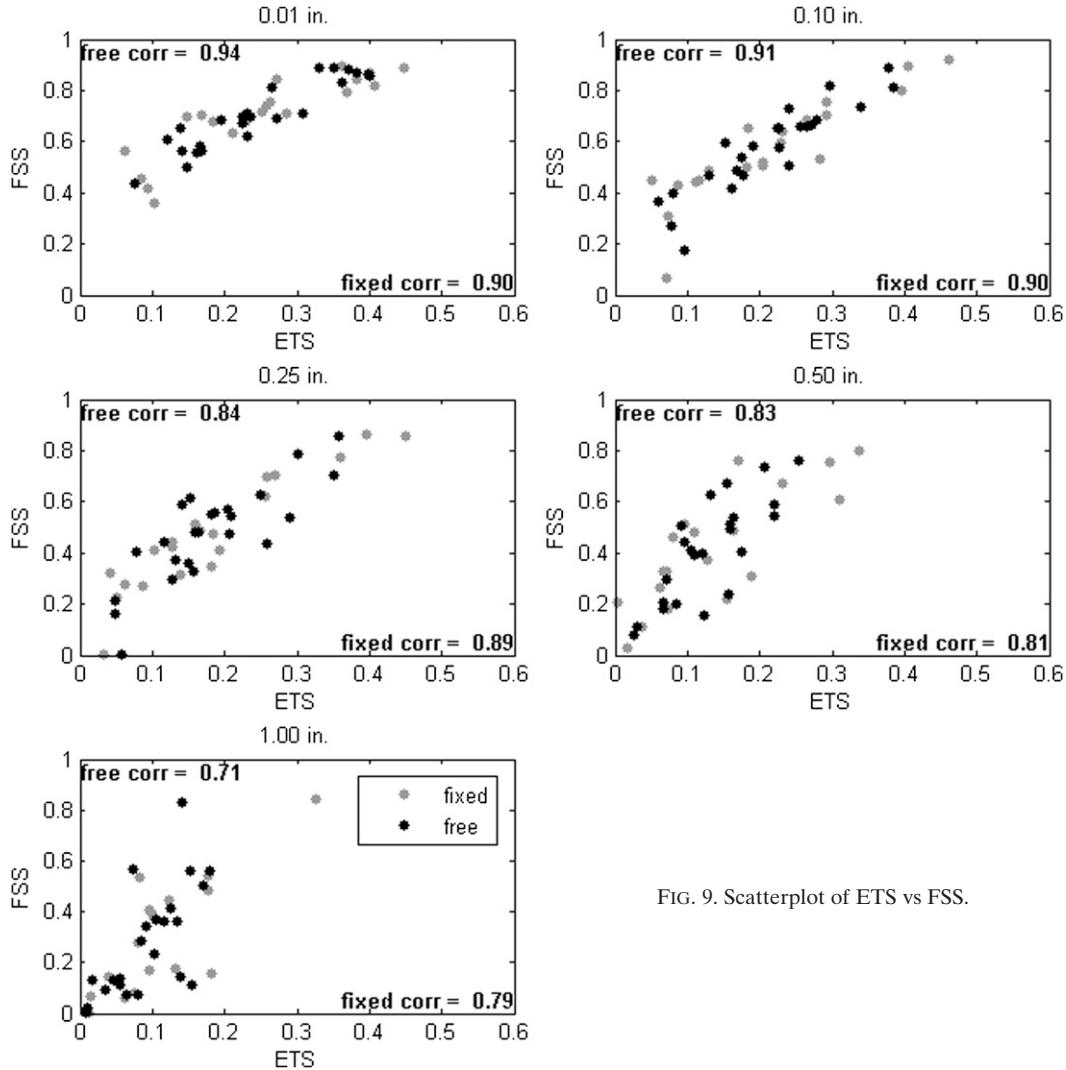


FIG. 9. Scatterplot of ETS vs FSS.

in the model, a more detailed analysis of an illustrative case from the present study, case 28 initialized at 1200 UTC 10 September 2009, follows. This case was selected since the large-scale forcing was among the strongest of all cases (8 of 36), as well-defined synoptic-scale features were present, but the skill at CI was among the worst of all cases (31 of 36). This case featured a broken line of cells that developed across southwestern and central North Dakota late in the afternoon of 10 September and evolved into a trailing-stratiform precipitation squall line as it moved east across North Dakota. The convective portion of the squall line lengthened as it matured and eventually encompassed much of the central and eastern Dakotas as well as portions of southern Manitoba during its most organized state. About one dozen large hail and damaging-wind reports came from the early stages of

this MCS. Animations of observed and simulated radar reflectivity are included in the supplemental material accompanying this work (file WAF-D-13-00005.1s1 for the observed reflectivity and file WAF-D-13-00005.1s2 for the model reflectivity).

There was a short-wave trough in the middle and upper troposphere that approached the region of interest from the northwest during the preconvective² hours of this case. The curved flow was associated with a large amount of relative vorticity at 500 hPa (Figs. 12a and 12b), and a large amount of stability-coupled DVA accompanied this trough (Fig. 13b). In fact, the box-averaged DVA for this case was the fourth strongest among all

²The term “preconvective” will be used in this section to denote the 6-h time period leading up to initiation of the target convection.

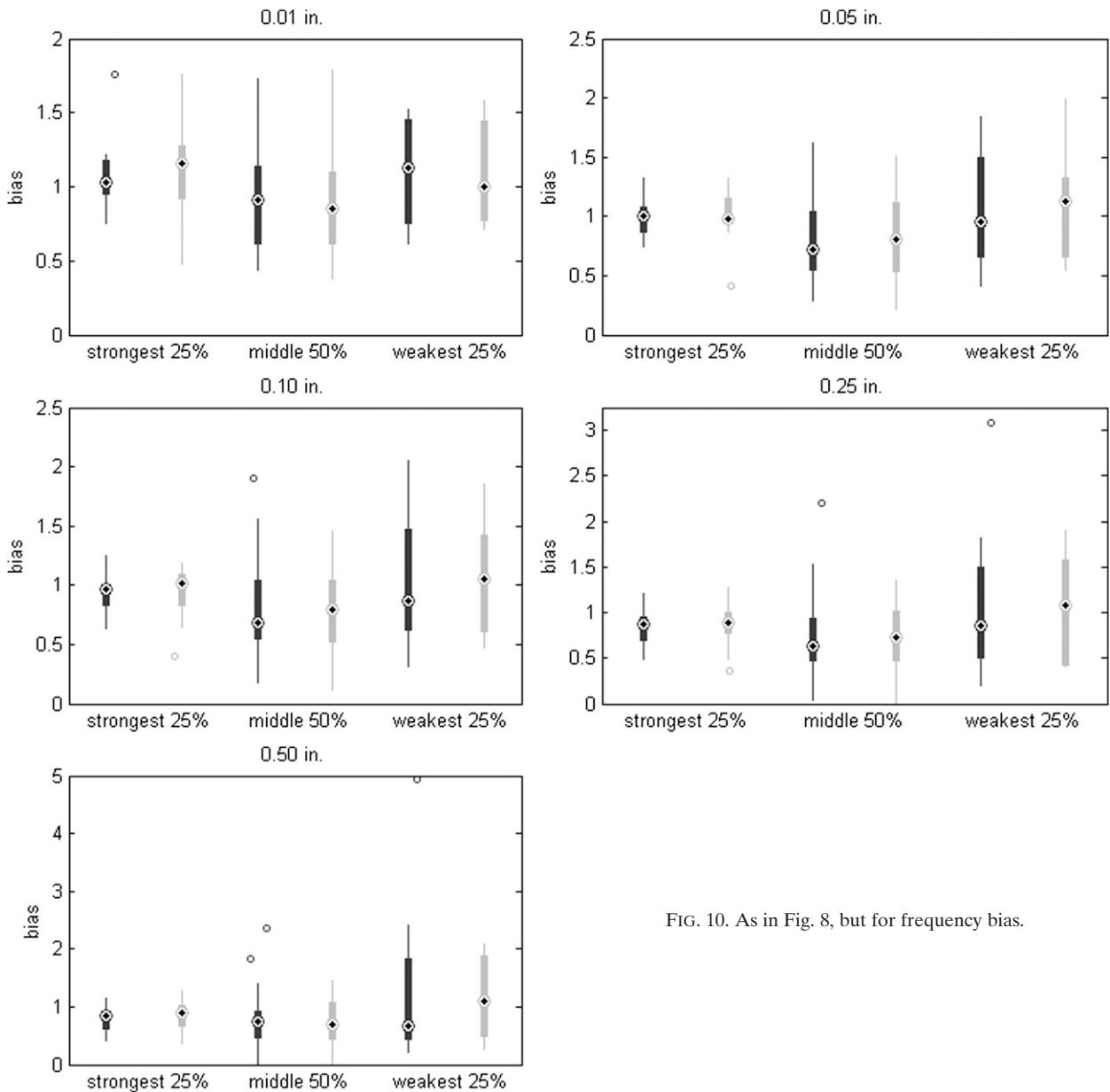


FIG. 10. As in Fig. 8, but for frequency bias.

cases. At 200 hPa, a moderately strong jet streak extended across the northern plains of the United States along the leeward side of the upper-level trough. Associated with the accelerated flow was a region of divergence (Fig. 13e) concentrated at and just east of the trough axis. In the lower troposphere, there was a localized area of intense temperature advection (Fig. 13c) associated with isentropic flow over a steeply sloped 310-K isentropic surface that intersected the ground in southwest North Dakota (Fig. 14a). Despite the strength of upper-air forcing, only a diffuse boundary with a corresponding diffuse wind shift existed at the surface,

although a cold front was only a few hundred kilometers to the west. The boundary was oriented SW–NE and was marked by an inverted mean sea level pressure trough extending across eastern Wyoming, northwestern South Dakota, and western and central North Dakota (Fig. 15a). While no temperature gradient was present across this boundary, a moisture gradient was generally coincident with the wind shift in the Dakotas. This boundary was manifest as local bursts of convergence and frontogenesis throughout the preconvective hours. There was one particular band of intense forcing located in far southwestern North Dakota, which served to aid in CI in

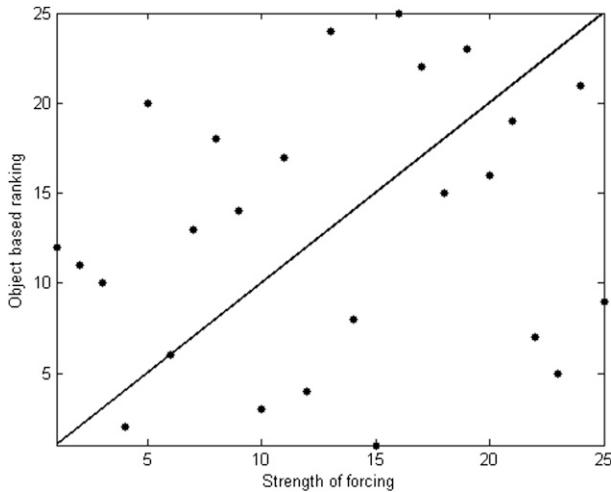


FIG. 11. Scatterplot of rank of large-scale forcing strength vs rank of object-based scores. A perfect-fit line is included.

that region. The local maximum in convergence was collocated with the most intense part of the isentropic lift on the 310-K surface. These two forcings combined to result in a local maximum of 700 hPa omega (Fig. 13a) over southwestern North Dakota. As surface-based convective available potential energy (SBCAPE) increased from around 1000 to 2500 J kg⁻¹ (Fig. 16a) and what little there was of surface-based convective inhibition (SBCIN) eroded during the late afternoon, the forcing became strong enough to cause CI. Despite the location of the strongest forcing, however, *the first intense cell did not develop in the area of strongest forcing*. Instead, the first cell that marked observed CI developed along a weaker part of that convergence band, just to the northeast of where the most intense convergence was. Cells eventually formed over the area of strongest forcing about 45–60 min after the first cell developed, however. It should be noted that this analysis is predicated on the accuracy of the RUC analyses; the accuracy in this particular area may be questionable. There were few surface observing stations in North Dakota online at the time of this case, especially in the region where the boundary was analyzed (Fig. 15a).

The model erred on initiation by about 150 km to the southwest and about 3 h late. Because of the model domain size, it is difficult to get a full picture of the forecast synoptic conditions. Nonetheless, a 500-hPa chart is presented for the model forecast in Fig. 12c for comparison with the corresponding subset from a RUC analysis. The model did a reasonably good job with the upper-level trough, although it appears to have missed the more southwesterly 500-hPa flow in southern Canada, forecasting instead mostly WSW or westerly flow.

Significant in this case was that the large-scale forcing in the model was weaker than in the RUC analyses, especially for temperature advection, surface frontogenesis, and 700-hPa vertical motion. The isentropic surfaces in the lower troposphere were less sloped than in the RUC analyses (Fig. 14b), so with weaker winds at those levels, there was less temperature advection and less lift than in the RUC analyses. This difference in strength of the large-scale forcing did not appear to significantly harm the forecast for organized convection, however, since the simulated MCS looked similar to the observed MCS (see animations in the supplementary material previously cited). This is consistent with the finding presented earlier that no correlation existed between the magnitude of the error and the skill in forecasting CI.

The model captured the surface boundary, the moisture gradient, and the lack of a temperature gradient. However, the WRF winds near the boundary were weaker, and the wind shift was very gradual and diffuse compared to the RUC analyses (Fig. 15b), as was the boundary in general. At times there were two separate wind shift lines. The moisture gradient was gradual across the boundary as well. Also, the surface winds in the model were generally weaker over a larger area near the boundary than in the RUC analyses. Therefore, convergence along the boundary was generally less organized and weaker. This weaker flow was related to the pressure trough being shallower than in the RUC analyses.

The reasons for the model error in CI are as follows. In southwestern North Dakota, where CI occurred in the observations, northwest surface winds dominated early in the preconvective period. These winds pushed dry air southeast, which initially limited the instability. As time passed, these northwest winds veered to the east, which served to 1) bring the moisture back to the west and 2) redefine the convergence and forcing for CI. Earlier in the period, the northwest winds converged with southerly winds along a wind shift farther to the east across south-central North Dakota and northern South Dakota. After the winds veered, the convergence was redefined as this flow converged with westerly and northwesterly winds behind the advancing cold front in eastern Montana and western North Dakota (Fig. 17). There was significant SBCIN and insufficient SBCAPE in the model at 2300 UTC for initiation to occur as it did in the observations (Fig. 16b). Given the lack of agreement in location between the various forcing mechanisms and the insufficient thermodynamic environment, initiation was delayed. Initiation finally began in the model after 0300 UTC. This occurred despite the presence of 50–75 J kg⁻¹ of SBCIN along the convergence band, and generally less than 1000 J kg⁻¹ SBCAPE, since sufficient forcing for CI

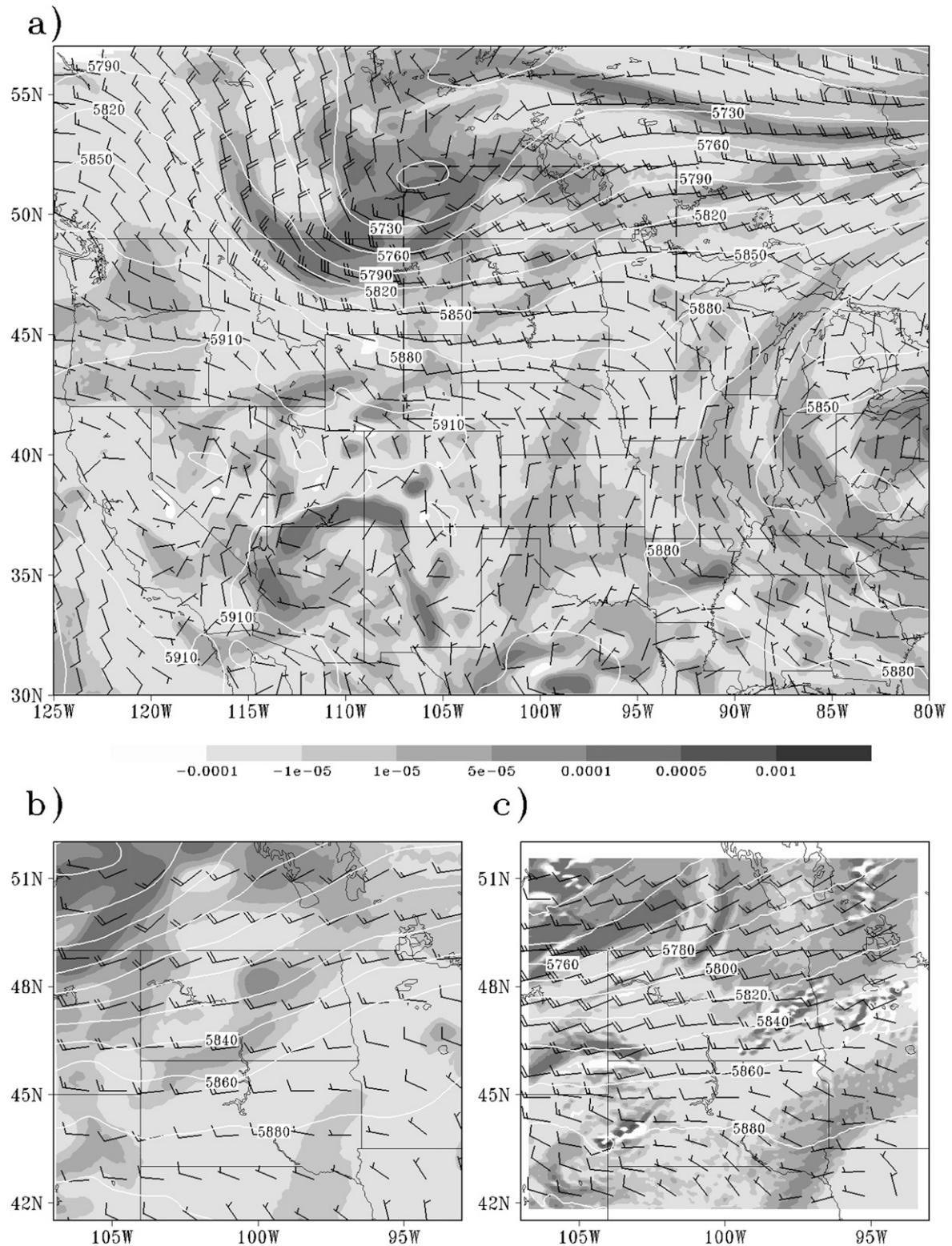


FIG. 12. The 500-hPa analysis valid 2300 UTC 10 Sep 2009 for (a) RUC analysis and (b) RUC analysis restricted to a subset matching that of the WRF domain. (c) The 11-h forecast using filtered WRF output. Relative vertical vorticity (s^{-1}) is shaded. Geopotential height (m) is contoured. Winds are indicated by the barbs, with a half-barb representing 5 m s^{-1} , a full barb representing 10 m s^{-1} , and a flag representing 50 m s^{-1} . The dashed box in (a) shows the area covered in (b) and (c), which is meant to cover the area the size of the WRF domain.

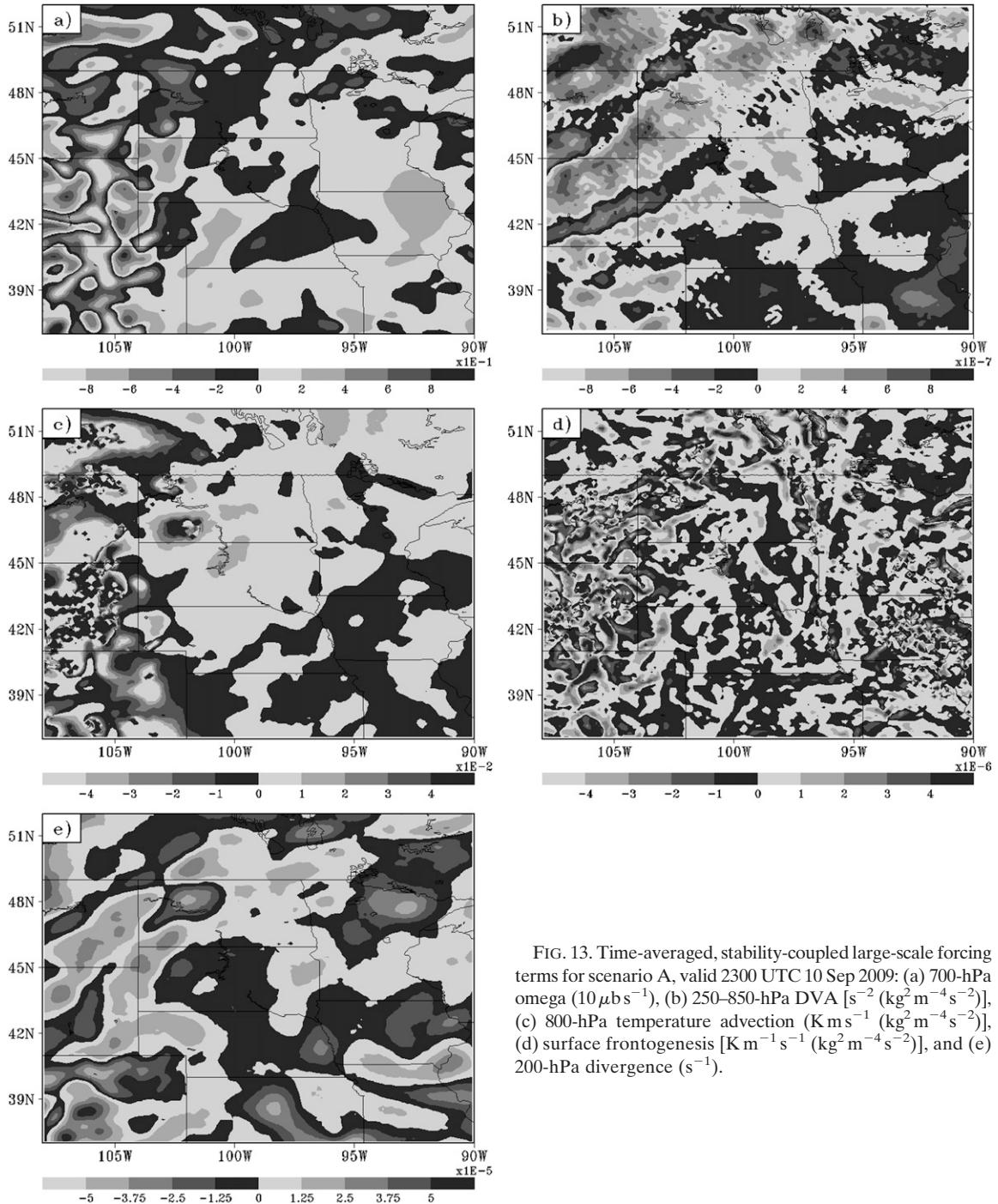
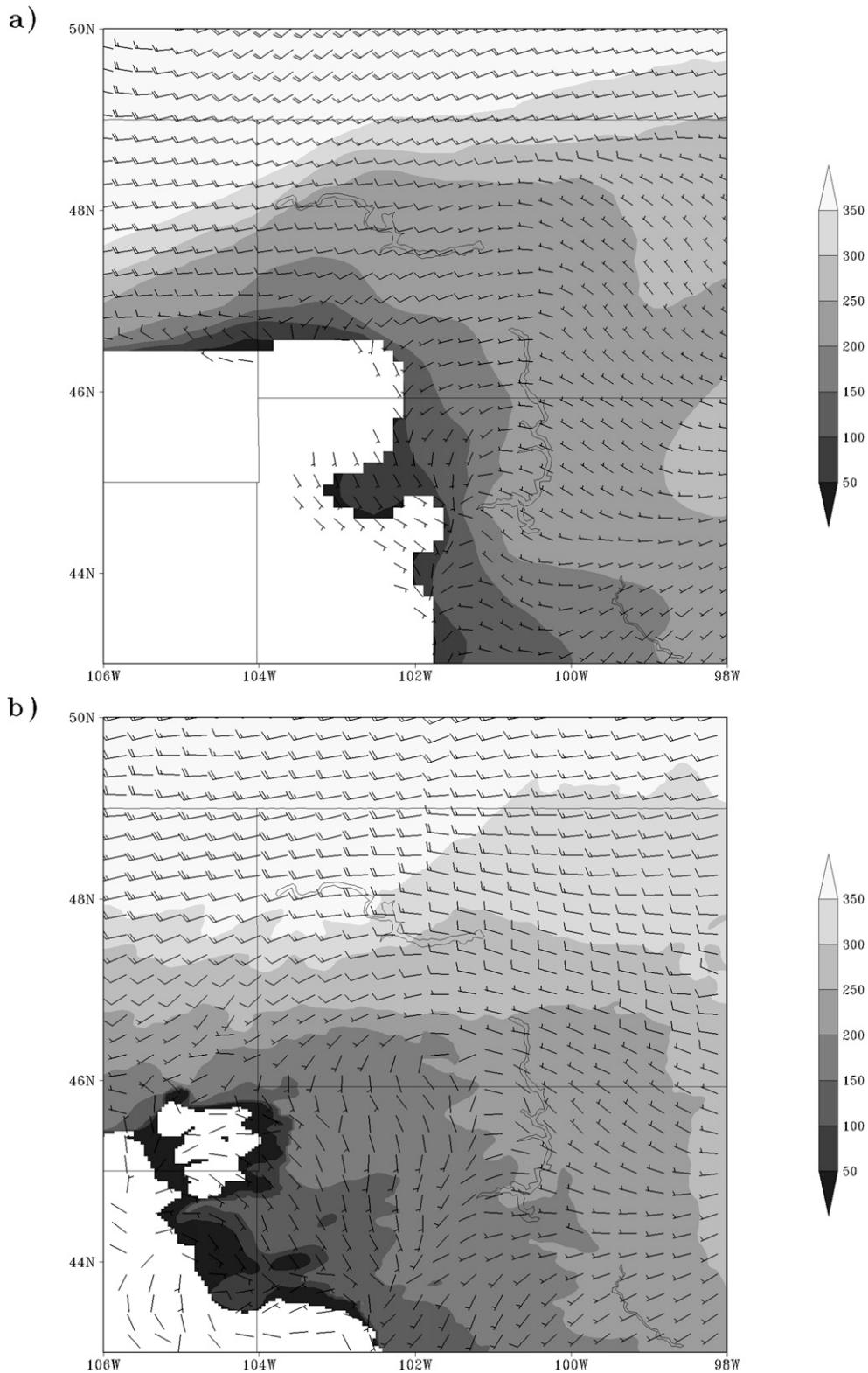


FIG. 13. Time-averaged, stability-coupled large-scale forcing terms for scenario A, valid 2300 UTC 10 Sep 2009: (a) 700-hPa omega ($10 \mu b s^{-1}$), (b) 250–850-hPa DVA [$s^{-2} (kg^2 m^{-4} s^{-2})$], (c) 800-hPa temperature advection [$K m s^{-1} (kg^2 m^{-4} s^{-2})$], (d) surface frontogenesis [$K m^{-1} s^{-1} (kg^2 m^{-4} s^{-2})$], and (e) 200-hPa divergence (s^{-1}).

along this line was provided by a region of strong upward motion at 700 hPa (Fig. 18). This forcing had remained anchored over far eastern Montana and far western North Dakota during the preconvective hours, only moving slightly east during that time. This upward motion likely influenced the location of CI, causing it to be farther west than in the RUC analyses.

In summary, the lack of a strong surface forcing mechanism along with a thermodynamic environment insufficient for CI at the time when observed initiation occurred explains the delayed initiation. The along-boundary variability was on the mesoscale, so these subsynoptic-scale errors and forcings significantly contributed to the error in initiation timing and location for this case.



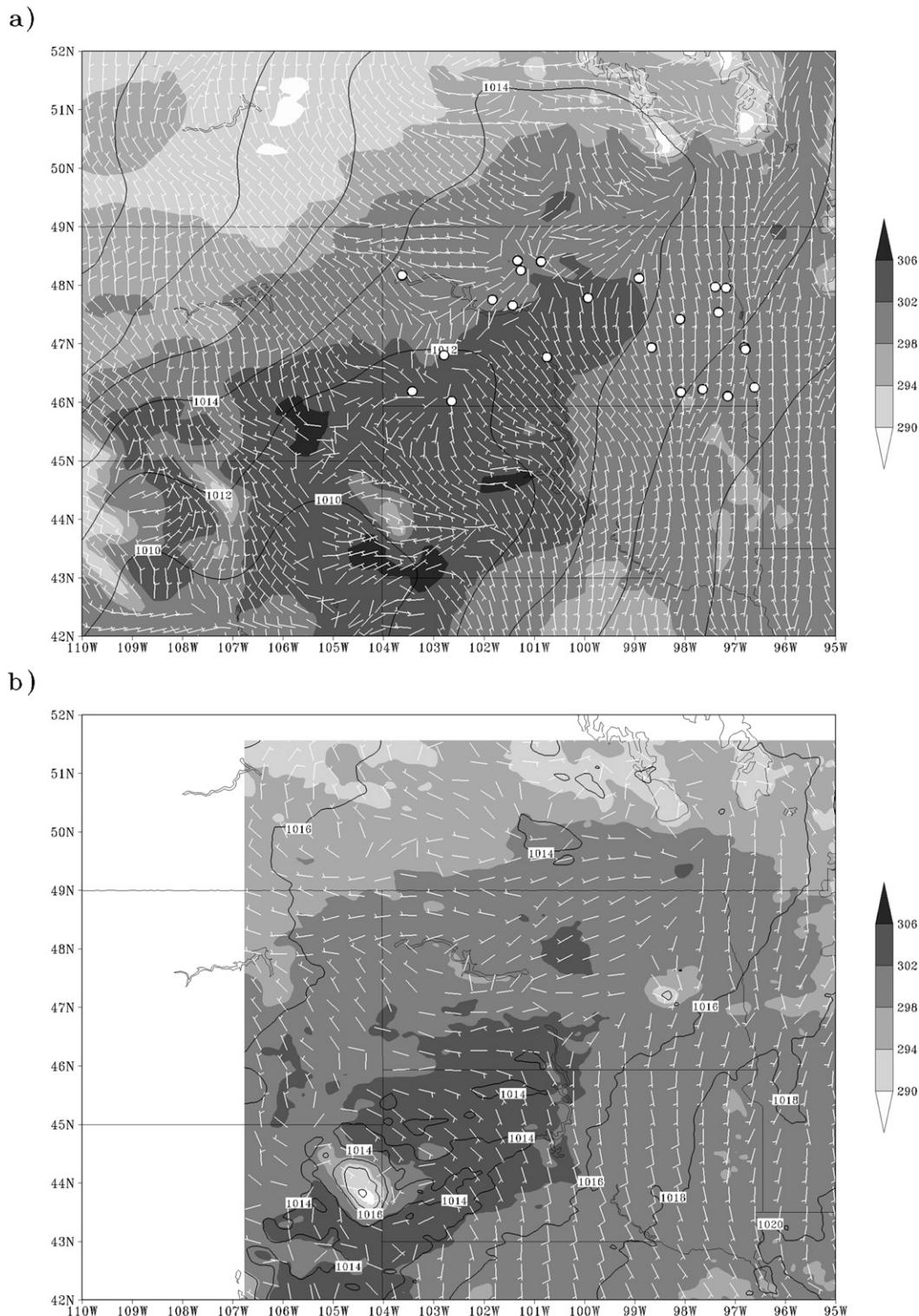
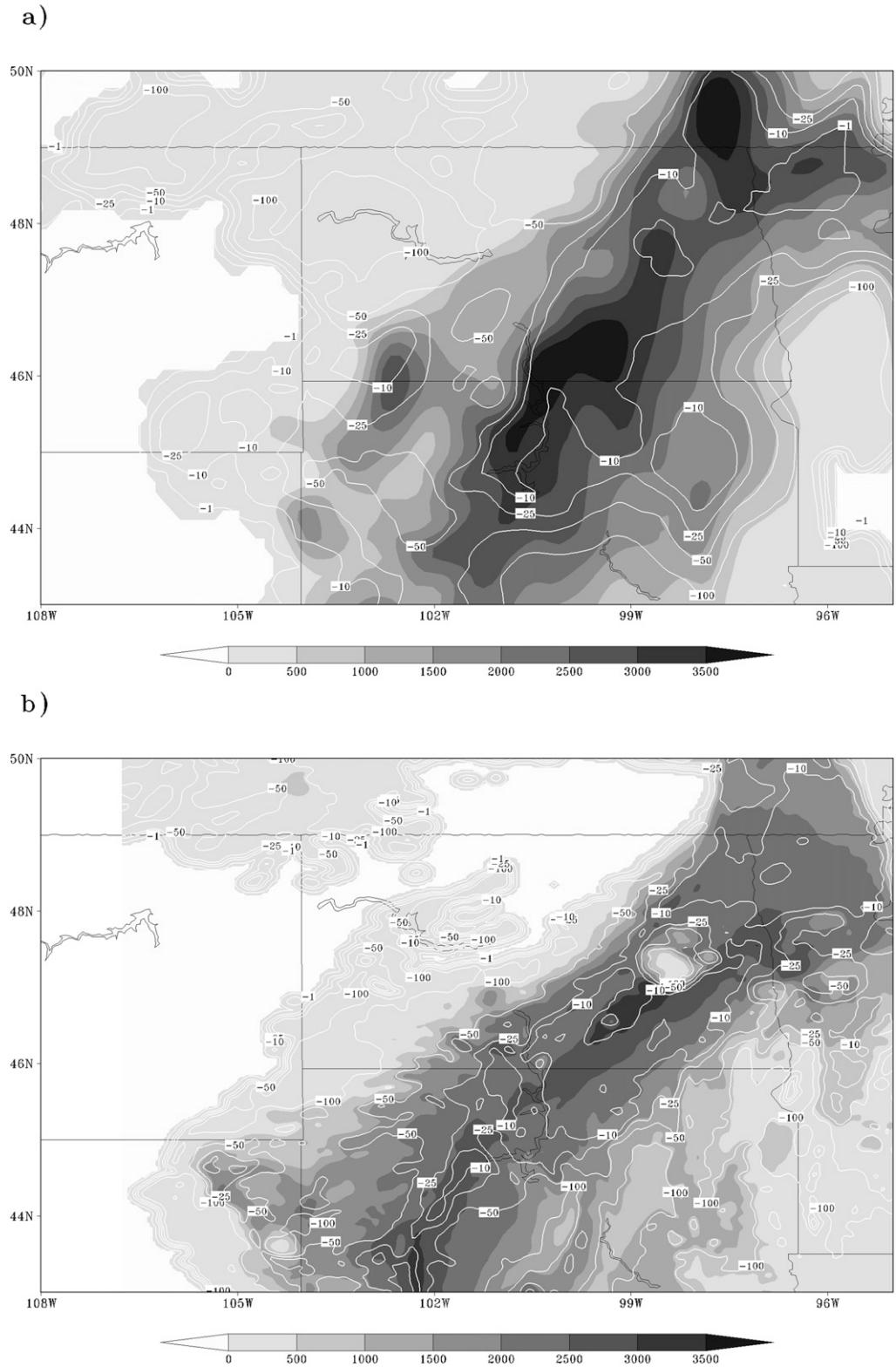


FIG. 15. Surface analysis valid at 2300 UTC 10 Sep 2009 for the (a) RUC analysis and (b) 11-h forecast using filtered WRF output. The 2-m temperature is shaded (K). Wind barbs are as in Fig. 12. Mean sea level pressure is contoured every 2 hPa. The dots in (a) represent the locations of surface observing stations in North Dakota online at the time.



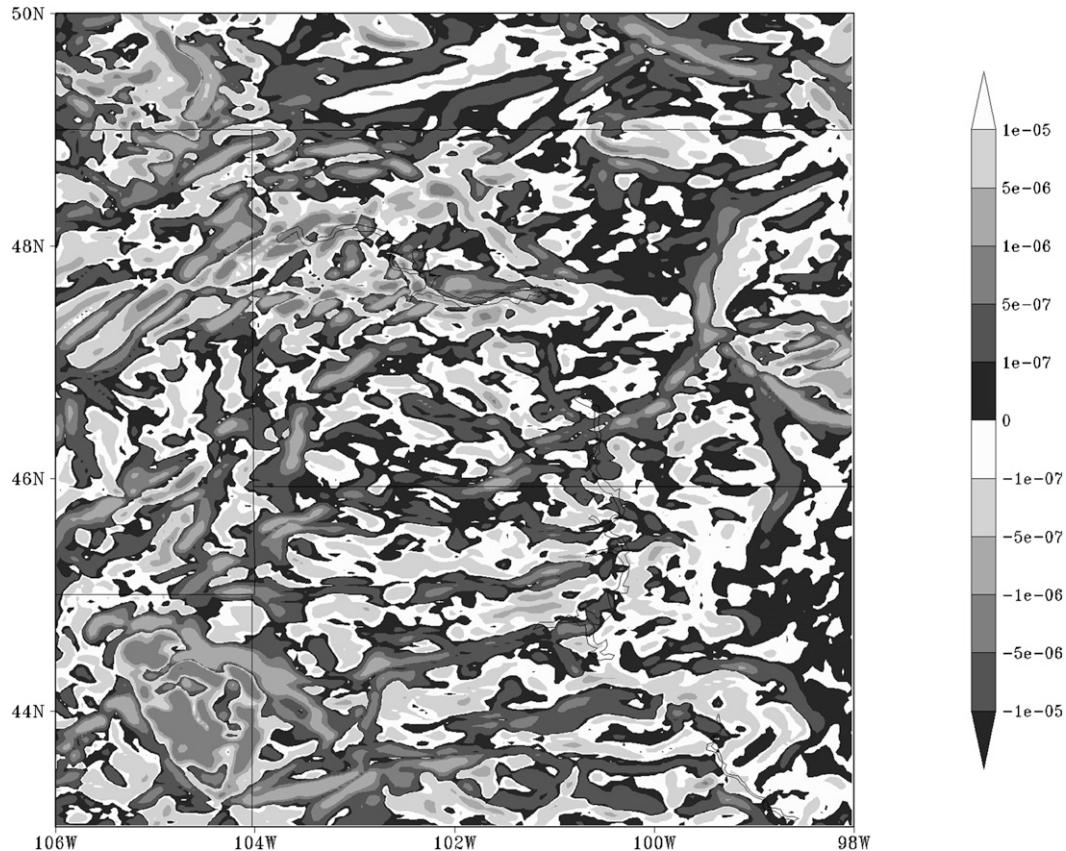


FIG. 17. Surface frontogenesis ($\text{K m}^{-1} \text{s}^{-1}$) valid at 0200 UTC 11 Sep 2009 from a 14-h forecast using filtered WRF output.

4. Discussion and conclusions

Convection-allowing 3-km grid-spacing WRF-ARW simulations were conducted for 36 cases of convection across the United States to explore the relationship between large-scale forcing and initiation and upscale evolution into an MCS. Knowledge of this relationship could improve forecasts of MCSs. Traditional (ETS and bias), neighborhood-based (FSS), and object-based techniques were used to determine model skill for the upscale evolution for each case. The different scoring techniques were also compared.

It was found that initiation erred in the southwest direction on average, but the average timing error was around zero due to a nearly normal distribution of temporal errors. No relationship existed between the skill of the model at forecasting initiation and the strength of large-scale forcing. Since the large-scale forcing itself was reasonably well simulated, we suspect subsynoptic-scale features such as storm-scale outflow boundaries, horizontal convective rolls in the boundary layer, and orographically forced circulations—all of which can provide for locally very strong forcing—exert a significant influence

on the timing and location of CI. These features are not evaluated by quasigeostrophic measures, and the two measures that might show the impacts of these smaller scale features, 700-hPa omega and 200-hPa divergence, did not show these features when averaged over space and time and after filtering in this study. Even though the RUC analyses were on a rather refined 20-km grid, it appears this may still be too coarse to accurately represent some of these small-scale features. Therefore, errors in the analysis values also likely influenced this finding of a lack of a relationship. The limits of predictability for a single convective cell at 3 km also may have contributed to the poor relationship, since the data used to feed the model were not fine enough in time or space to adequately resolve individual convective cells and since the model is generally only able to fully resolve features larger than about 20 km (Skamarock 2004). A correlation was found between the strength of large-scale forcing and model QPF skill in the 6-h forecast that represented upscale evolution according to ETSS, agreeing with Jankov and Gallus (2004), who used coarser-resolution models with convective parameterization. This relationship was not particularly strong, however. No particular verification

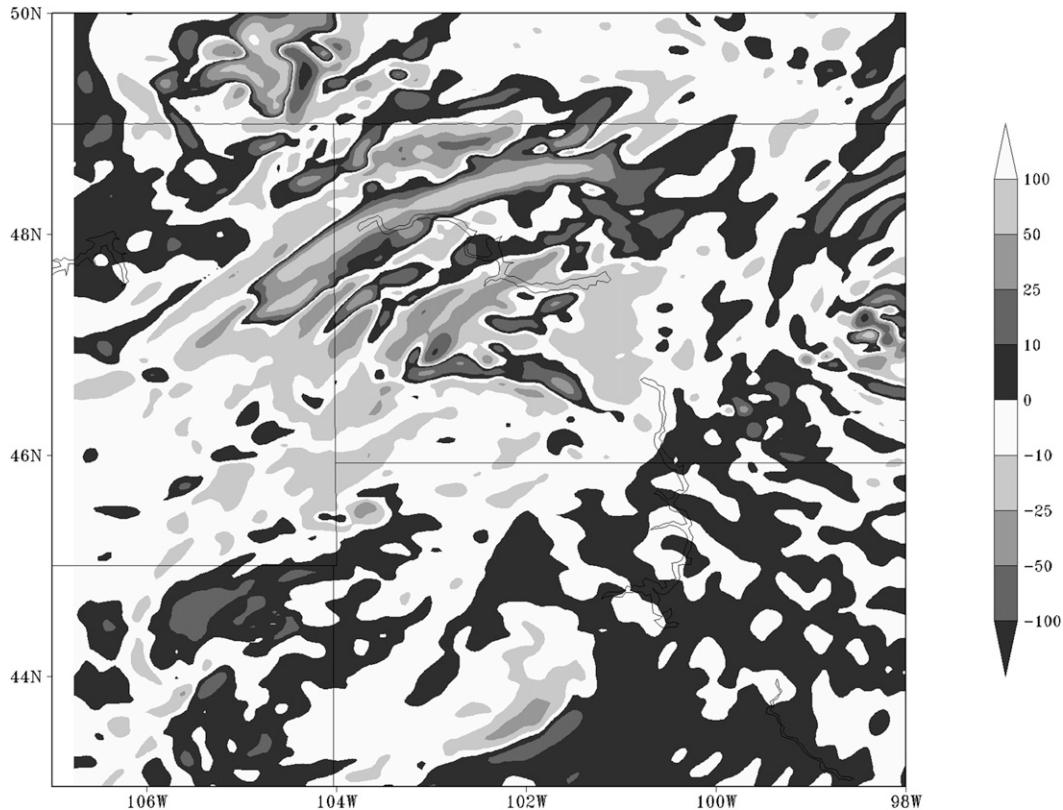


FIG. 18. As in Fig. 17, but for 700-hPa omega ($\mu\text{b s}^{-1}$).

method provided a strong relationship in this study. Object-based verification, which specifically focused on the target system, did not give a better result than the domain-averaged verifications of the ETS and FSS. ETSs and FSSs were found to be strongly correlated.

A detailed examination of one case where initiation was forecast poorly despite strong large-scale forcing was provided to gain more understanding of causes of the good and bad aspects of the forecast. In that case, large-scale features such as a short-wave trough, upper-level jet stream, and low-level warm advection were well represented, but the specific smaller-scale forcing features in the region of CI were not as accurately simulated. Because of the small-scale variability in the region near the pressure trough and moisture gradient, features driven by large-scale forcing in this case study, initiation erred in both its timing and location. It must also be acknowledged that errors possibly present in the RUC analyses used to represent the observations may have influenced the results.

Although a detailed case study was presented here for only one case, each case was evaluated with a similar level of detail. In each of the 36 cases, smaller-scale features such as storm-scale outflow boundaries, horizontal convective rolls in the boundary layer, and small-scale orographically forced circulations associated with

large bodies of water or complex terrain were found to be resolved, at least partially, in the model domain. In some of these cases, these smaller-scale features were either primarily responsible for initiation, or complicated the determination of specific causes for initiation.

High-resolution model output can be valuable in the investigation of small-scale features, such as storm-scale outflow boundaries and horizontal convective rolls in the boundary layer, features that models such as the one used in this study can resolve [although horizontal convective rolls are poorly resolved; D. Stensrud (2012, personal communication)] and that may impact the simulation of deep, moist convection. Fields such as surface dewpoint, temperature, divergence, and low-level vertical motion, when unfiltered, indicate the heterogeneity of the lower atmosphere and also show how easily the forecast for the initiation of convection can err, especially given how sensitive convection is to small-scale details that may be only partially resolved in higher-resolution model simulations. Given the spatial density of the observations used to feed high-resolution models, as we approach the limits of predictability on this scale, temporal and spatial errors in convective initiation are understandable, and a decrease in errors would be unlikely without significant

improvements in the initial and lateral boundary condition data. Further improvements could also come from better surface and boundary layer parameterizations. Although refinements in grid spacing will likely continue in the coming years as computer resources continue to increase, it is important to keep in mind that improvements in forecasting convective initiation and evolution may not be automatic as the limit of predictability will continue to be approached. Future studies should continue to explore the predictability of convection at these fine scales using an ensemble of high-resolution model simulations.

Acknowledgments. The authors thank Tara Jensen and John Halley Gotway of NCAR for their help with MODE, and Eric Aligo for help with WRF coding issues. The constructive feedback from three anonymous reviewers, and from Mike Baldwin, was invaluable. Stage IV data were provided by NCAR/EOL under sponsorship of the National Science Foundation (<http://data.eol.ucar.edu/>). Some RUC data were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. This project took advantage of netCDF software developed by UCAR/Unidata. This work was supported by NSF Grants ATM-0848200, with funds from the American Recovery and Reinvestment Act of 2009, and ATM-1222383.

REFERENCES

- Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of spatial verification methods to idealized and NWP-gridded model precipitation forecasts. *Wea. Forecasting*, **24**, 1485–1497.
- Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416.
- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methods and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117.
- Doswell, C. A., III, H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581.
- Duda, J. D., and W. A. Gallus, 2010: Spring and summer midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206.
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Fowle, M. A., and P. J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794.
- 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.
- , J. D. Murphy, and J. S. Kain, 1994: Warm core vortex amplification over land. *J. Atmos. Sci.*, **51**, 1780–1807.
- Gallus, W. A., Jr., N. A. Snook, and E. V. Johnson, 2008: Spring and summer severe weather reports over the Midwest as a function of convective mode: A preliminary study. *Wea. Forecasting*, **23**, 101–113.
- Grams, J. S., W. A. Gallus, S. E. Koch, L. S. Wharton, A. Loughe, and E. E. Ebert, 2006: The use of a modified Ebert–McBride technique to evaluate mesoscale model QPF as a function of convective system morphology during IHOP 2002. *Wea. Forecasting*, **21**, 288–306.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167.
- Hohenegger, C., 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478.
- , and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793.
- Janjić, Z., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp. [Available online at <http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf>.]
- Jankov, I., and W. A. Gallus, 2004: MCS rainfall forecast accuracy as a function of large-scale forcing. *Wea. Forecasting*, **19**, 428–439.
- Kain, J. S., and Coauthors, 2008: Severe-weather forecast guidance from the first generation of large-domain convection-allowing models: Challenges and opportunities. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.1. [Available online at <https://ams.confex.com/ams/pdfpapers/141723.pdf>.]
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102** (D14), 16 663–16 682.
- Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.
- , and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]
- Szoke, E. J., J. Brown, and B. Shaw, 2004: Examination of the performance of several mesoscale models for convective forecasting during IHOP. Preprints, *20th Conf. on Weather Analysis and Forecasting*, Seattle, WA, Amer. Meteor. Soc., J13.6. [Available online at <https://ams.confex.com/ams/pdfpapers/68930.pdf>.]
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved

- bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115.
- Trier, S. B., and C. A. Davis, 2002: Influence of balanced motions on heavy precipitation within a long-lived convectively generated vortex. *Mon. Wea. Rev.*, **130**, 877–899.
- Wakimoto, R. M., H. V. Murphey, A. Nester, D. P. Jorgensen, and N. T. Atkins, 2006: High winds generated by bow echoes. Part I: Overview of the Omaha bow echo 5 July 2003 storm during BAMEX. *Mon. Wea. Rev.*, **134**, 2793–2812.
- Weckwerth, T. M., and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.*, **134**, 5–22.
- 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.
- , C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecast with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437.
- Weiss, S. J., M. E. Pyle, Z. Janjić, D. R. Bright, and G. J. DiMego, 2008: The operational high resolution window WRF model runs at NCEP: Advantages of multiple model runs for severe convective weather forecasting. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., P10.8. [Available online at <https://ams.confex.com/ams/pdfpapers/142192.pdf>.]
- Wheatley, D. M., R. J. Trapp, and N. T. Atkins, 2006: Radar and damage analysis of severe bow echoes observed during BAMEX. *Mon. Wea. Rev.*, **134**, 791–806.
- Wilson, J. W., and R. D. Roberts, 2006: Summary of convective storm initiation and evolution during IHOP: Observational and modeling perspective. *Mon. Wea. Rev.*, **134**, 23–47.
- Xue, M., and Coauthors, 2009: CAPS real-time 4 km multi-model convection-allowing ensemble and 1 km convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2009 Spring Experiment. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 16A.2. [Available online at <https://ams.confex.com/ams/pdfpapers/154323.pdf>.]
- Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185.