Ensemble-Based Forecast Uncertainty Analysis of Diverse Heavy Rainfall Events

RUSS S. SCHUMACHER

National Center for Atmospheric Research, Boulder, Colorado, and Department of Atmospheric Sciences, Texas A&M University, College Station, Texas*

CHRISTOPHER A. DAVIS

National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 3 November 2009, in final form 7 February 2010)

ABSTRACT

This study examines widespread heavy rainfall over 5-day periods in the central and eastern United States. First, a climatology is presented that identifies events in which more than 100 mm of precipitation fell over more than 800 000 km^2 in 5 days. This climatology shows that such events are most common in the cool season near the Gulf of Mexico coast and are rare in the warm season. Then, the focus turns to the years 2007 and 2008, when nine such events occurred in the United States, all of them leading to flooding. Three of these were associated with warm-season convection, three took place in the cool season, and three were caused by landfalling tropical cyclones. Global ensemble forecasts from the European Centre for Medium-Range Weather Forecasts Ensemble Prediction System are used to assess forecast skill and uncertainty for these nine events, and to identify the types of weather systems associated with their relative levels of skill and uncertainty. Objective verification metrics and subjective examination are used to determine how far in advance the ensemble identified the threat of widespread heavy rains. Specific conclusions depend on the rainfall threshold and the metric chosen, but, in general, predictive skill was highest for rainfall associated with tropical cyclones and lowest for the warm-season cases. In almost all cases, the ensemble provides very skillful 5-day forecasts when initialized at the beginning of the event. In some of the events—particularly the tropical cyclones and strong baroclinic cyclones—the ensemble still shows considerable skill in 96–216-h precipitation forecasts. In other cases, however, the skill drops off much more rapidly as lead time increases. In particular, forecast skill at long lead times was the lowest and spread was the largest in the two cases associated with meso-α-scale to synoptic-scale vortices that were cut off from the primary upper-level jet. In these cases, it appears that when the vortex is present in the initial conditions, the resulting precipitation forecasts are quite accurate and certain, but at longer lead times when the model is required to both develop and correctly evolve the vortex, forecast quality is low and uncertainty is large. These results motivate further investigation of the events that were poorly predicted.

1. Introduction

In the United States in 2007 and 2008, several extended periods of heavy rainfall took place, resulting in destructive flooding and record rainfall in many locations. Major floods occurred in parts of the southern plains and the Upper Midwest in 2007, the 2008 floods in Iowa and Wisconsin received national attention, and three tropical cyclones made landfall in 2008 that soaked large regions of the country. Considering the numerous impacts that heavy rainfall and flooding have on society, and the continuing challenges in predicting precipitation at all scales (e.g., Olson et al. 1995; Fritsch and Carbone 2004; Hamill et al. 2008), there is a need to understand the processes that lead to extreme precipitation events. Additionally, because numerical weather prediction (NWP) models provide much of the guidance for making precipitation forecasts, it is important to evaluate their performance in high-impact events so that forecasters and end users can have a better understanding of the skills and shortcomings of these forecasts.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. Russ Schumacher, Dept. of Atmospheric Sciences, Texas A&M University, College Station, TX 77843.
E-mail: russ.schumacher@tamu.edu

DOI: 10.1175/2010WAF2222378.1

The years 2007 and 2008 also represent the first two full years of an international initiative to create and archive global ensemble forecasts. This initiative is part of The Observing System Research and Predictability Experiment (THORPEX), and is known as the THORPEX Interactive Grand Global Ensemble (TIGGE; Bougeault et al. 2010). TIGGE provides ensemble forecasts from seven national centers and, among other benefits, allows access to previously unavailable data for investigating the predictability of, and potential improvements in, warnings for extreme events (Pappenberger et al. 2008).

Medium-range predictions of precipitation and other components of the hydrologic cycle are used for multiple purposes, including flood forecasting and warning, water resources management, and agriculture. Efforts are being made in many parts of the world to couple numerical weather prediction ensembles to hydrologic models for real-time flood prediction (e.g., Schaake et al. 2007; Pappenberger et al. 2008; Thielen et al. 2009; He et al. 2009). Thielen et al. (2009) ''test the limits of predictability'' for flood warnings with such a coupled forecast system. The goal of the present study is similar, but with a slightly different approach: we aim to estimate the predictability of widespread, extended periods of heavy rain in an NWP ensemble, and to identify how these limits change for events associated with different types of weather systems. Coupling the NWP forecasts to hydrologic models is not performed here, but this study provides information that may be useful for forecasters when they evaluate ensemble predictions of rainfall in different meteorological situations.

Theoretical limits to the predictability of different scales of atmospheric motion were first examined by Thompson (1957) and Lorenz (1963, 1969). The interaction between small and large scales also impacts predictability (e.g., Zhang et al. 2003; Rotunno and Snyder 2008; Jung and Leutbecher 2008). Precipitation systems can be especially difficult to predict because of the small-scale, chaotic nature of deep moist convection. As a result, warm-season precipitation, which often occurs with weak synoptic-scale forcing, presents a very difficult forecast challenge (e.g., Carbone et al. 2002) that can be partially overcome by using ensemble prediction. Past studies (e.g., Mullen and Buizza 2001; Hamill et al. 2008) have investigated the performance of quantitative precipitation forecasts (QPFs) in global ensembles and confirmed that they provide skillful forecasts at low precipitation thresholds and in the winter season; they perform much worse in the warm season and for heavy precipitation. The predictability of extreme rainfall has been examined in the United States by Mullen and Buizza (2001) and Zhang et al. (2006), and in Europe by Romero et al. (2006). However, these investigators focused primarily on 24–36-h accumulations. Predictions of 5-day rainfalls in Africa were analyzed by Knippertz and Fink (2009). Their approach is similar to that taken here, where we examine multiple-day, widespread rain events that occur on scales that should be well resolved by global NWP models, pose flooding threats over large areas, and have implications for seasonal climate predictions.

The atmospheric processes associated with heavy rainfall at various scales in the central and eastern United States have been established in numerous past studies (e.g., Maddox et al. 1979; Giordano and Fritsch 1991; Mo et al. 1997; Konrad 2001; Schumacher and Johnson 2005, among many others). In the cool season, U.S. heavy rainfall is typically produced by strong synoptic-scale weather systems, such as extratropical cyclones. In some cases, a front remains over a given area for several days, as in the ''synoptic'' type of flash flood discussed by Maddox et al. (1979). During the warm season, the processes leading to heavy rainfall are more varied: extratropical cyclones, mesoscale convective systems (MCSs; Houze 2004), and tropical cyclones are all important heavy rain producers (e.g., Schumacher and Johnson 2006). It is common for MCSs to occur within latitudinal ''corridors'' (Tuttle and Davis 2006), and when these corridors are more persistent than normal, periods of MCS development over the same area may last for a week or more. This type of situation led to the historic 1993 floods in the Midwest (e.g., Junker et al. 1999), as well as the 2008 Midwest floods, which will be discussed further later in this article.

In section 2, the data and methods used in the study are described. Section 3 presents a climatology of widespread heavy rain events in the United States, and brief descriptions of the events during 2007–08 are given in section 4. In section 5, the forecast skill and uncertainty in global ensemble forecasts of these events is presented, and the study concludes with section 6.

2. Data and methods

The primary precipitation dataset used in this study is the U.S. Daily Precipitation Analysis (Higgins et al. 2000), obtained from the National Oceanographic and Atmospheric Administration/Climate Prediction Center (NOAA/CPC) (information online at http://www.cpc.ncep.noaa.gov/products/precip/realtime/GIS/retro.shtml). This dataset is created from over 8000 rain gauge observations, gridded to a 0.25° latitude × 0.25° longitude grid. This dataset has relatively coarse resolution, so it does not faithfully represent local precipitation maxima, but it is adequate for analyzing the widespread, multiple-day rainfall events that will be discussed below. This dataset covers the period 1948–2008; however, the number and

TABLE 1. Climatological frequency bins used in the calculation of the BSS for different precipitation thresholds and seasons, as described in the text.

Threshold (mm)	Bin 1 (%)	Bin 2 (%)	Bin 3 (%)	Bin 4 (%)	Bin 5 (%)	Bin 6
50, JJA	<0.1	0.1–0.5	0.5–2.5	2.5–5	5–7.5	>7.5%
50, other	<0.25	0.25–1	1–2.5	2.5–5	5–9	>9%
100, JJA	<0.005	0.005–0.1	0.1–0.25	0.25–0.5	0.5–1	>1%
100, other	<0.005	0.005–0.1	0.1–0.5	0.5–1	1–1.5	>1.5%
150, JJA	<0.001	0.001–0.025	0.025–0.05	0.05–0.075	0.075–0.125	>0.125%
150, other	<0.001	0.001–0.025	0.025–0.075	0.075–0.15	0.15–0.25	>0.25%

locations of rain gauges used in creating the analysis have varied over time.

The model datasets were obtained from the TIGGE archive, which is available online (http://tigge-portal.ecmwf.int). The model that will be used is the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS; Buizza et al. 2007), which has 51 members with a spectral truncation of T399 (corresponding to approximately 50-km horizontal grid spacing) and 62 vertical levels through 240 forecast hours (10 days). The ECMWF EPS was chosen because of its large number of members and because of its superior performance compared to other EPSs based on several verification methods (e.g., Park et al. 2008). The ECMWF EPS comprises a control run and 50 members that are initially perturbed by singular vectors in pairs (i.e., a positive and negative perturbation), and by stochastic physics (ECMWF 2009a). The singular-vector perturbations are designed so that their impacts are maximized over Europe at 48 h into the forecast. Furthermore, diabatic singular vectors are used to create initial perturbations near tropical cyclones (ECMWF 2009a). The horizontal scale of the singular-vector perturbations is T42 with 62 vertical levels. QPFs for all 51 members were obtained at 12-h initialization intervals, on a 0.5° latitude × 0.5° longitude grid, for the events of interest. Other details about the ECMWF EPS can be found in Buizza et al. (2007) and references therein. Several upgrades to the ECMWF EPS took place in 2007 and 2008 (ECMWF 2009b), so the version of the model is not exactly the same for all events. Because a set of forecasts with an unchanging model was unavailable, the impacts of the model upgrades could not be examined. The TIGGE archive also includes forecasts from other EPSs; these are not considered in the present study, but investigating the forecasts of a combined grand ensemble is an intended route for future research.

For the calculation of verification statistics, the precipitation analysis was coarsened to a 0.5° × 0.5° grid using neighborhood averaging. The metrics used include the Brier skill score (BSS) and the area under the relative operating characteristic (ROC) curve [hereafter ROC area; e.g., Wilks (2005)]. The BSS calculation follows the method proposed by Hamill and Juras (2006), which accounts for the varying climatological probability of events by calculating the score for subsets of climatological probability. Specifically, we use their Eqs. (6), (8), and (9) to calculate the BSS as a sample-weighted average of the skill score for each distinct climatological regime. For the three precipitation thresholds used in this work (50, 100, and 150 mm in 120 h), the climatological frequency of occurrence of these thresholds was calculated both for the full year, as well as for the four meteorological seasons. Then, six climatological frequency subsets were defined to obtain bins that defined distinct climatological regimes while still containing a sufficiently large number of grid points. The climatological frequencies in March–May (MAM), September–November (SON), and December–February (DJF) were very similar to one another, so the same bins were used for these seasons; different bins were used for June–August (JJA) (Table 1). In general, because the climatological probabilities of these events are so small, the effects of accounting for spatially varying climatology on the BSS are minimal. For simplicity, the ROC area is calculated without accounting for the varying climatology; it is compared with a random reference forecast with area 0.5. The ROC area is calculated using Eqs. (10)–(12) of Hamill and Juras (2006).

In addition, a new statistic is proposed that represents the spread in the precipitation forecasts in relation to the typical predicted spatial extent of the event. This will be referred to as area spread (AS). This statistic aims to quantify the spread that one sees in displays of precipitation forecasts on a ''spaghetti chart''—a single plot that shows forecast contours from all of the ensemble members. It is calculated by dividing the total area enclosed by all predicted rainfall contours by the average area enclosed by an ensemble member at a specified threshold (Fig. 1). Let $\mathbf{p}_{e,j} = [p_{1,j}, \ldots, p_{n,j}]$ be an n-member ensemble of precipitation forecasts at the jth of m grid points. This ensemble is converted to binary notation so that $p_{i,j} = 1$ if the forecast precipitation meets or exceeds the threshold, and $p_{i,j} = 0$ if the forecast precipitation

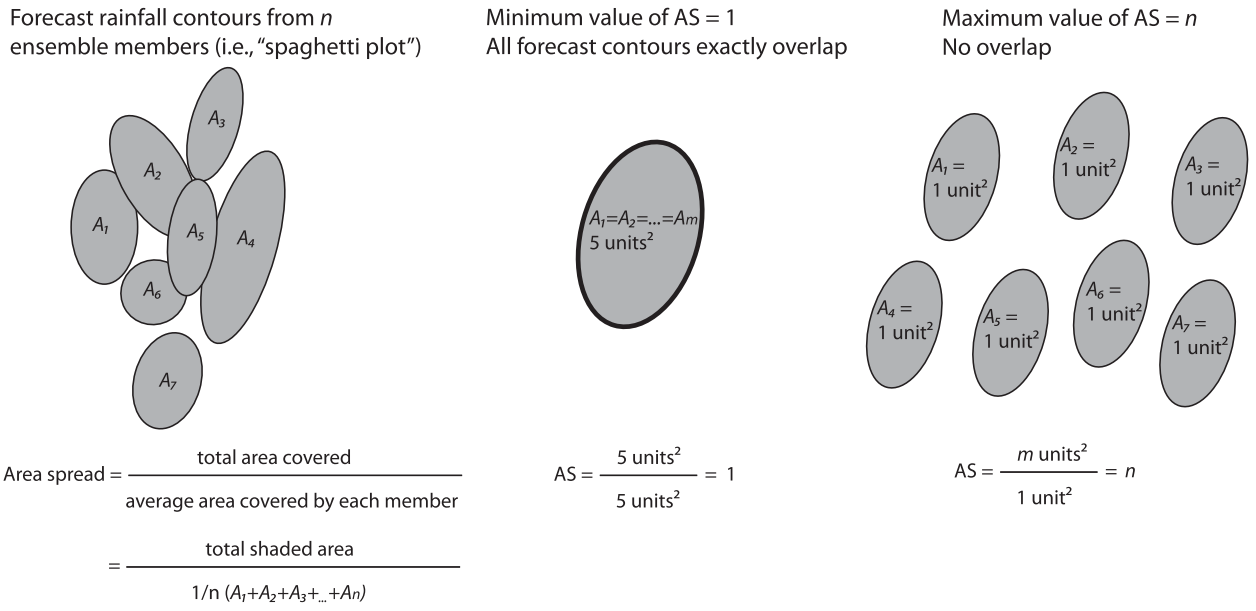

FIG. 1. Schematic diagram illustrating the method of calculation of the AS statistic. (left) An idealized plot of the predicted rainfall contours at a specified threshold (say, 100 mm) from each of n ensemble members. The area covered by each member is denoted by A_1 through A_n. The AS statistic is determined by calculating the total area covered by all of the ensemble members (i.e., the total area shaded in gray), and dividing by the average area covered by each member. If not all members predict rainfall over the specified threshold, the denominator is reduced, and the resulting value of AS increases. (middle) The minimum possible value of AS, which is unity. Here, every ensemble member predicts that the specified contour will be in exactly the same place (i.e., there is no spread). (right) The maximum possible value of AS is equal to n, the number of ensemble members. This occurs when all members predict a rainfall contour of the same size, but there is no overlap among the contours.

is less than the threshold. The AS statistic is then expressed as

$$\mathrm{AS} = \frac{\sum_{j=1}^{m} P_j}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,j}}, \quad \text{where} \quad P_j = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} p_{i,j} \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

In other words, the value of P at each grid point is be either one (if any ensemble members predicted precipitation greater than the threshold at that point), or zero (if no members did). These binary values are then added over all grid points, and divided by the average area enclosed by the predicted precipitation contours at the threshold (Fig. 1). Higher values indicate greater spread. The AS is similar to the spread ratio (SR) introduced by Stensrud and Wandishin (2000) and Wandishin et al. (2001); both have the area of the union of the ensemble precipitation forecasts as their numerator. However, the SR's use of the area of the intersection of the forecasts as the denominator is problematic for the present study. At high precipitation thresholds and with a large ensemble, it is common for a few ensemble members to have no grid points where the predicted precipitation exceeds the threshold. In these instances, the intersection is zero and the SR is undefined. In contrast, the AS still provides a measure of the ensemble's spread in these situations. In a situation such as that shown in the middle of Fig. 1, AS and SR are equivalent.

In addition to the precipitation analyses and forecasts, other atmospheric data will be presented where applicable. The primary source for upper-level atmospheric information is the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis dataset (Kalnay et al. 1996). Six-hourly and daily average analyses of pressure-level geopotential height and temperature are used to illustrate some of the large-scale atmospheric processes during the heavy-rain events. Additionally, locations of other important weather phenomena (such as tropical cyclone tracks, fronts, etc.) were obtained from NOAA surface analyses archived at the National Climatic Data Center.

3. Selection of events and climatology

To select the cases that will be discussed in this study, the gridded precipitation data were examined to find

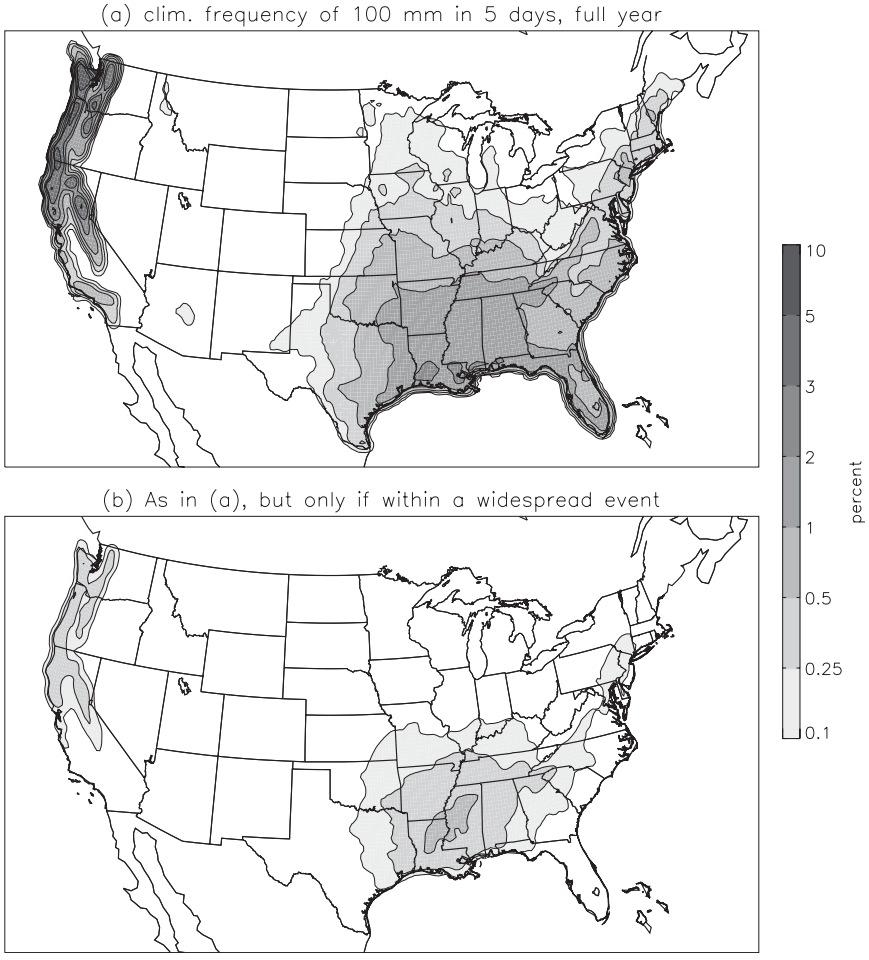

FIG. 2. Climatological frequency (based on 1948–2006) of (a) 100 mm of rainfall in 5 days and (b) as in (a) but only if it occurred as part of a widespread rain event as defined in the text.

instances of widespread, multiday heavy rainfall. All 5-day periods in which 100 mm of rain fell over more than 350 grid points (approximately 800 000 km^2) were identified; these will be referred to as widespread rain events. The 350 grid points were not required to be contiguous. Five days was chosen as the accumulation period to separate localized, short-term rainfall events from those that persist for multiple days and are likely to have a substantial impact over a large area. This choice is also consistent with past research that has analyzed ''pentad'' rainfall (e.g., Knippertz and Fink 2009), and with the issuance of 5- and 6–10-day precipitation forecasts by the U.S. Hydrometeorological Prediction Center and Climate Prediction Center, respectively.

Before discussing and assessing the predictions of the cases from 2007 and 2008 in more detail, some climatological context for this type of long-lived, widespread rainfall is warranted. The climatological frequency of 100 mm of rainfall in 5 days at a given grid point was calculated (Fig. 2a), as was the frequency of receiving 100 mm in 5 days *as part of a widespread rain event* (Fig. 2b). This climatology was calculated using data from the years 1948–2006, and the percentages reflect the probability that, if one were to randomly choose a 5-day period, the total precipitation *in those 5 days* would exceed the threshold.[1] As with the general precipitation climatology in the United States (e.g., Higgins et al. 1997, their Fig. 2), the climatology of 100 mm (5 day^{-1}) rainfall amounts shows a maximum in the Pacific Northwest, with a secondary maximum near the Gulf of Mexico coast and frequencies decreasing poleward from there (Fig. 2a).

[1] The climatology is calculated in this way for use in the computation of the Brier skill score (see section 2). This method contrasts with other possible ways of calculating the climatological percentages, such as the chance that a given *day* would fall within a 5-day rainy period.

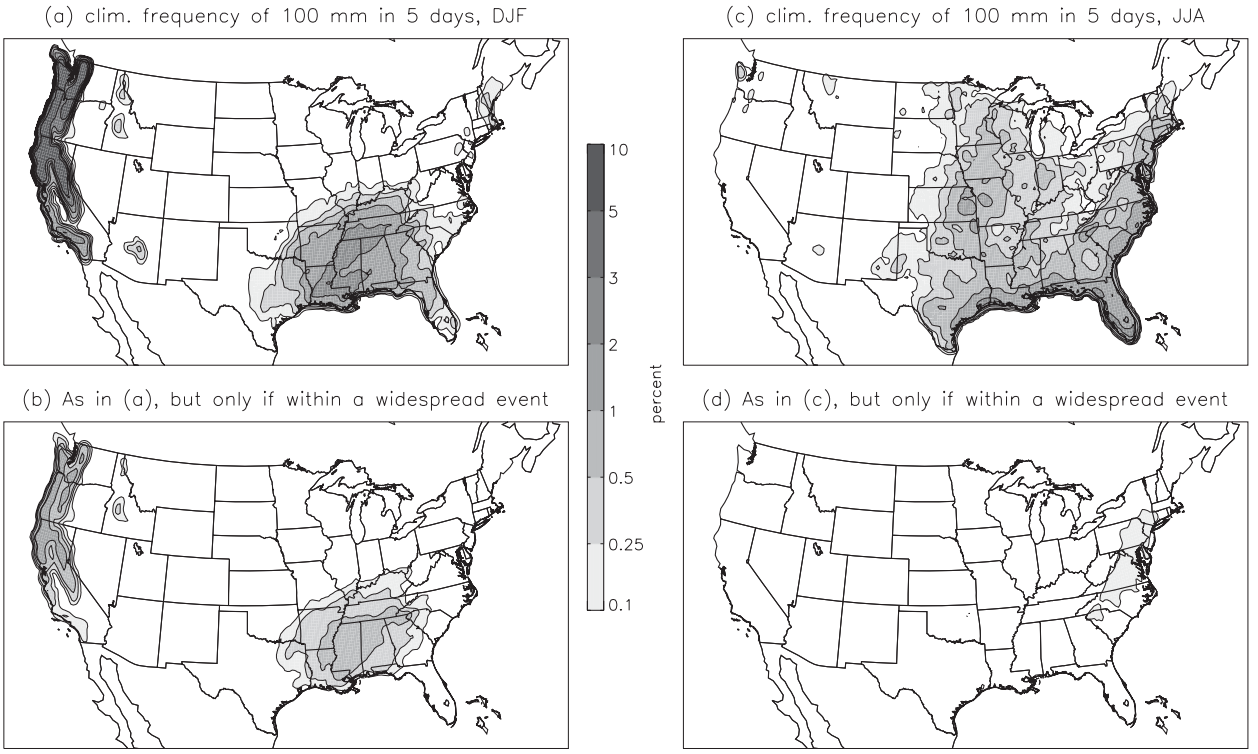

FIG. 3. As in Fig. 2, but for the months of (a),(b) DJF and (c),(d) JJA.

The chance that a point will receive 100 mm of rainfall in 5 days *as part of a widespread rain event* (i.e., with areal coverage greater than 350 grid points) is substantially lower (Fig. 2b). The highest probabilities are in parts of Louisiana, Arkansas, and Mississippi. After removing overlapping periods, there were 163 total events in the 59 yr from 1948 to 2006. The most widespread events occurred in 1957, which had 10. There were seven years that had no events, and the median number of events per year is two. The two years of interest in this study (2007 and 2008) had three and six events, respectively.

The climatology of multiday heavy precipitation in the summer contrasts strongly with that during the rest of the year (Fig. 3). The probabilities for DJF (Figs. 3a and 3b) resemble the overall climatology (cf. Fig. 2), as do the probabilities in MAM and SON (not shown). In JJA, higher probabilities of 100 mm of rainfall in 5 days expand poleward (Fig. 3c); however, very few widespread events occur during these months (Fig. 3d). This is the case because summer is generally characterized by weak synoptic-scale flow over the United States, which is conducive to MCSs that are more likely to produce heavy rain over small areas than over widespread regions (e.g., Fritsch et al. 1986; Schumacher and Johnson 2006). In most of the Upper Midwest, no widespread rain events occurred in the 59 yr between 1948 and 2006. In total, there were only 12 widespread rain events in the months of June, July, and August during the period of record, with 6 of these occurring in the last 6 yr of the record (2003–08). The summers of 2007 and 2008 were particular outliers in relation to this climatology: two summer events occurred in each of these years, including two that affected the Upper Midwest.

The events occurring in 2007 and 2008, which are listed in Table 2, will be the focus of the remainder of this study. In addition to widespread coverage of 100 mm of rain, each of these events had point accumulations exceeding 200 mm, and all resulted in flooding. The nine events conveniently fall evenly into three categories: three were associated with series of warm-season organized convective systems; three with cool-season, synoptic-scale weather systems; and three were landfalling tropical cyclones. All nine events occurred east of the Rockies. Widespread rain events are also relatively common along the Pacific coast (Fig. 2b), but none took place in the two years analyzed here.

4. Description of the events

a. 25–30 June 2007

The large-scale atmospheric conditions during this extended period of heavy precipitation in the southern

TABLE 2. Listing of the nine widespread 5-day rain events in 2007 and 2008, selected as described in the text. Each event had an areal coverage of 100 mm of precipitation greater than 350 precipitation grid points (approx. 800 000 km^2). Each event is considered to have started at 1200 UTC on the first date given and ended at 1200 UTC on the second date given. The events have been classified as either cool season (CS), warm season (WS), or tropical cyclone (TC). The number of flood-related fatalities and the amount of flood damage for each event are also shown, based on data from the Storm Events database. Asterisks are shown for two of the landfalling tropical cyclones that had damage both from flooding and from other causes; more complete descriptions are given in sections 4g and 4h of the text.

Dates	Location	Type	Process	No. of fatalities	Damage ($M)
25–30 Jun 2007	Southern Plains	WS	Midlevel vortex	4	350
18–23 Aug 2007	Upper Midwest	WS	Stationary front	15	600
22–27 Oct 2007	Southeast	CS	Cutoff synoptic cyclone	0	0
15–20 Mar 2008	Mississippi Valley	CS	Synoptic-scale front	9	235
4–9 Jun 2008	Upper Midwest	WS	Stationary front	11	2000
22–27 Aug 2008	Southeast	TC	Tropical Storm Fay	0	560
1–6 Sep 2008	South	TC	Hurricane Gustav	*	*
10–15 Sep 2008	South, Midwest	TC	Hurricane Ike; frontal system	*	*
8–13 Dec 2008	Southeast	CS	Synoptic cyclone	0	1.2

plains consisted of an anomalously strong upper-level ridge over the central United States, with a slow-moving midlevel circulation equatorward of the ridge (Figs. 4a and 5a). This circulation arose from a mesoscale convective vortex (MCV; Bartels and Maddox 1991) that grew upscale into a much larger warm-core vortex (Goebbert et al. 2008). The rainfall during this 5-day period exceeded 300 mm in much of southeastern Kansas.

b. 18–23 August 2007

The conditions during this period were characterized by stronger-than-normal zonal upper-level flow and an associated low-level baroclinic zone in the northern United States (Figs. 4b and 5b). Simultaneously, Tropical Storm Erin (Arndt et al. 2009) made landfall along the Texas coast and recurved poleward. Erin itself was responsible for extreme rainfall and flooding, and it also transported tropical moisture poleward. Galarneau et al. (2010) show that this tropical moisture substantially enhanced the rainfall totals over the Midwest. Because the largest swath of rain was associated with convection along the baroclinic zone in the Midwest, this event was classified as a warm-season case rather than a tropical cyclone, although both tropical and extratropical influences were important.

The most destructive flooding during this period resulted from an MCS on 18–19 August. In Hokah, Minnesota, a new statewide, all-time record for 24-h rainfall accumulation was set [384 mm or 15.1 in; Minnesota State Climatology Office (2009)].

c. 22–27 October 2007

Two large areas in the eastern United States experienced rainfall accumulations greater than 100 mm during the period 23–27 October. The heavy precipitation in this case was associated with an upper-level cyclone that developed over the Midwest on 21–22 October and then cut off from the upper-level jet on 23–24 October (Figs. 4c and 5c). There were numerous reports of flooding and flash flooding resulting from this rainfall; however, no fatalities, injuries, or economic costs of flood damage are given in the Storm Events database.

d. 15–20 March 2008

An extratropical cyclone that intensified in the lee of the Rocky Mountains and moved eastward was responsible for heavy rains during 15–20 March 2008 (Fig. 4d). A strong upper-level trough developed over the western United States on 15 March and strengthened on 16–17 March over northern Mexico. A surface cyclone developed over Texas on 18 March, and heavy rains developed along several slow-moving surface boundaries positioned along a baroclinic zone northeast of its center (Fig. 5d). Within a large swath of 100-mm accumulations in the Mississippi River valley were local amounts exceeding 300 mm in Missouri and Illinois.

e. 4–9 June 2008

The rainfall during 4–9 June 2008 (Figs. 4e and 5e) was associated with an upper-level jet that was displaced southward relative to its climatological position, as well as with anomalously high values of atmospheric water vapor (e.g., Dirmeyer and Kinter 2009). On 4–5 June, an MCS in Nebraska produced local rainfall maxima greater than 150 mm, and a long-lived squall line on 5–6 June produced another large swath of precipitation. A backward-propagating (e.g., Corfidi 2003) MCS developed in Illinois and Indiana on 6–7 June, which produced local totals over 250 mm. Multiple MCSs then occurred between 7 and 9 June. The combination of heavy rains and already wet soil conditions led to devastating

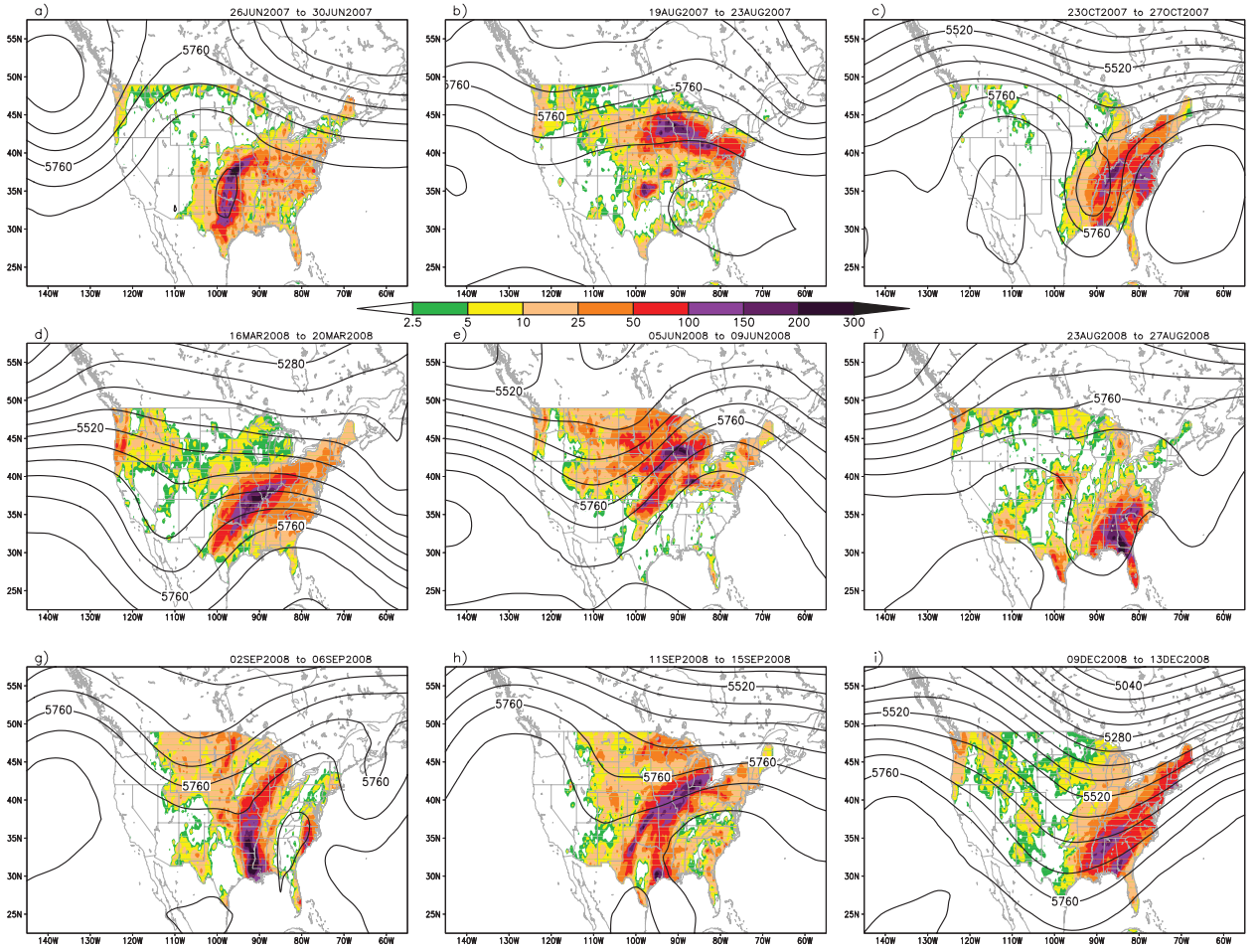

FIG. 4. Five-day accumulated precipitation (color shading) and 5-day-average 500-hPa geopotential height (solid contours every 60 m) for each of the nine widespread rain events during 2007–08 (see also Table 1): (a) 25–30 Jun 2007, (b) 18–23 Aug 2007, (c) 22–27 Oct 2007, (d) 15–20 Mar 2008, (e) 4–9 Jun 2008, (f) 22–27 Aug 2008 (TS Fay), (g) 1–6 Sep 2008 (Hurricane Gustav), (h) 10–15 Sep 2008 (Hurricane Ike), and (i) 8–13 Dec 2008.

floods. Many rivers in Iowa crested at record levels in the subsequent days, causing unprecedented damage according to the U.S. Geological Survey (USGS 2009).

f. 22–27 August 2008

Slow-moving Tropical Storm Fay (Figs. 4f and 5f) became a depression and a tropical storm just east of the island of Hispaniola on 15 August, tracked westward over Hispaniola on 15–16 August, and then moved westward along the southern coast of Cuba on 17–18 August (Stewart and Beven 2009). It then curved poleward and moved slowly across Florida from 19 to 23 August, continued into Alabama and Mississippi from 23 to 25 August, and curved northeastward on 25 August. The slow forward speed of Fay allowed it to produce extreme rainfall accumulations along its path. Fay produced heavy rains in Florida on 19–21 August, and a total of 698 mm of rain was reported at Thomasville, Georgia, on 22–24 August.

g. 1–6 September 2008

Just over a week later, Hurricane Gustav made landfall along the Louisiana coast and moved inland, producing heavy rainfall along a south-to-north corridor (Figs. 4g and 5g). Gustav caused substantial damage in Haiti, the Dominican Republic, Jamaica, and Cuba on its path toward the Gulf coast (Beven and Kimberlain 2009). After making U.S. landfall, Gustav's forward motion slowed on 2–4 September, before it moved poleward on 4–5 September. The maximum point rainfall accumulation was 533 mm at Lake Larto, Louisiana (Beven and Kimberlain 2009). Gustav caused over $4 billion in damage in the United States overall, with approximately $30 million of that resulting from flooding.

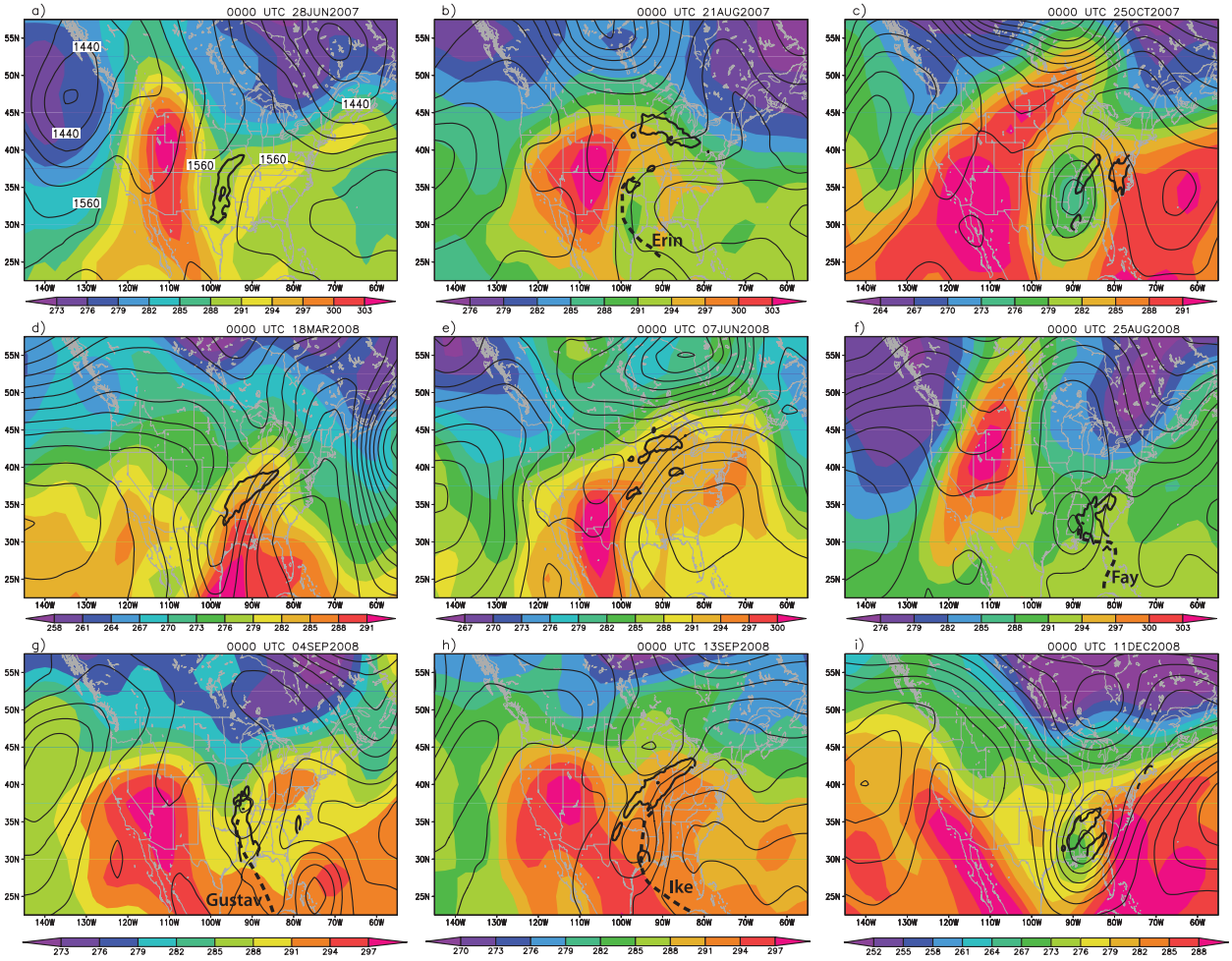

FIG. 5. Instantaneous 850-hPa air temperature (K, color shading) and geopotential height (solid contours every 30 m) in the middle of each of the nine widespread rain events. At 0000 UTC on (a) 28 Jun 2007, (b) 21 Aug 2007, (c) 25 Oct 2007, (d) 18 Mar 2008, (e) 7 Jun 2008, (f) 25 Aug 2008, (g) 4 Sep 2008, (h) 13 Sep 2008, and (i) 11 Dec 2008. The 100-mm rainfall contours are shown by the thick solid black lines, and tropical cyclone tracks are shown by thick dashed lines, where appropriate.

h. 10–15 September 2008

Hurricane Ike made landfall along the Texas coast in mid-September 2008 (Berg 2009) and went on to produce heavy rains as it turned poleward. The situation with Ike was more complex than that with Gustav, as rain fell both from the tropical cyclone itself as well as from a predecessor rain event (PRE; Galarneau et al. 2010) along a southwest–northeast-oriented baroclinic zone that extended from the Texas Panhandle to Michigan (Figs. 4h and 5h). In addition to tropical moisture brought onshore by Ike, another source of moisture was eastern Pacific Tropical Cyclone Lowell, whose remnants were carried eastward by an upper-level trough over the western United States. The combined rainfall from Ike, the remnants of Lowell, and the front across the Midwest broke records and led to major flooding. Ike was an extremely damaging storm, with the cost estimated at over $19 billion in the United States (Berg 2009). Furthermore, there were millions of dollars of additional damage from flooding across the Midwest that was directly and indirectly a result of Ike.

i. 8–13 December 2008

The last widespread rain event to be addressed in this study occurred in December 2008 in the southeast (Figs. 4i and 5i). This rain fell ahead of an extratropical cyclone that moved eastward across the United States. This strong synoptic-scale system initiated numerous convective systems over the 5-day period 8–13 December, with the largest rainfall amounts being reported in Mississippi, Alabama, Georgia, and Florida. The location of this event was very near the climatological maximum for such events (Figs. 3a and 3b).

5. Forecast skill and uncertainty

a. Overview

The ECMWF forecasts will now be used to assess skill and uncertainty associated with the weather systems responsible for the nine widespread rain events that occurred during 2007–08. Since these events had impacts on local and regional scales, and on temporal scales ranging from subdaily to seasonal, an understanding of how well a current state-of-the-art ensemble predicts them may be useful to a variety of users. This analysis can also potentially provide useful information about where predictions are poorest, which can pinpoint the most important topics for future research. In this section, basic verification statistics and some of their notable aspects will be presented, and the following subsection will delve deeper into the reasons for the results presented here. The goal of this analysis is not to perform a systematic verification of the ECMWF EPS, but instead to use basic verification statistics to quantify the uncertainties in predictions of heavy precipitation and to learn about how the meteorological conditions affect the quality of the predictions.

We begin our discussion of this analysis at a relatively low precipitation threshold for these events: 50 mm (5 day^{-1}). The predictions at this threshold provide a baseline for whether the ensemble correctly identifies that a particular large-scale region will receive a moderate amount of rain, without considering the details of the heaviest precipitation. For almost all of the events, both the BSS and ROC area metrics indicate skillful predictions at this threshold, especially for forecasts initialized at the start of the rainy 5-day period (Figs. 6a and 6d). In most of the events, the skill then gradually decreases with increasing lead time, with the ensemble still showing considerable skill in 96–216-h precipitation forecasts. This suggests that, in general, the ECMWF ensemble QPFs provide high-quality guidance many days in advance as to the locations of widespread rainfall. However, two of the events are notable outliers at the 50-mm threshold: the June 2008 event (the purple line in Figs. 6a and 6d) and the June 2007 event (blue line). There is minimal skill in predicting the June 2008 event at short lead times; the skill is actually somewhat higher at longer lead times. In the June 2007 event, the forecasts are comparable to the other events at short lead times, but the level of skill drops off sharply at longer lead times, such that the BSS is negative at lead times longer than 84–204 h (Fig. 6a) and the area under the ROC curve is substantially smaller than for any of the other events at the longest lead times considered here (Fig. 6d).

At the 100-mm threshold, the level of skill is generally lower than at 50 mm, and the forecasts for several of the events have considerable forecast-to-forecast variability in skill (Figs. 6b and 6e). For 120-h precipitation forecasts initialized at the starting times of the events, the ensemble again shows skill in almost all of the cases. Two of the tropical cyclone cases—Fay and Gustav—have the best forecasts at short lead times, whereas there is essentially zero skill for the June 2008 event at short lead times. At long lead times, the June 2007 and October 2007 events have the least skillful forecasts.

The verification statistics using a precipitation threshold of 150 mm appear to depend in part on the areal coverage of 150 mm of rain in each event. The ensemble has considerable skill in predicting the events with widespread 150-mm accumulations, such as Tropical Cyclones Fay and Gustav (Figs. 6c and 6f), whereas it has no skill for several of the events with smaller 150-mm areas. There is again large run-to-run variability in the forecast quality for some of the events, particularly the March 2008 and Hurricane Ike cases. At the longest lead times, the forecasts for the rainfall from Fay have by far the highest skill, with there being near-zero BSS and small ROC areas for 96–216-h forecasts for all of the other events.

The area spread (AS) statistic indicates that, as expected, the spread in the area covered by specific precipitation amounts increases with increasing lead time (Fig. 7). At the 50-mm threshold (Fig. 7a), the June 2007 event stands out as having the greatest spread at all lead times. At long lead times, the October 2007 case approaches it. At the 100-mm threshold (Fig. 7b), there is a clear separation between the three tropical cyclone events and the other six cases. The AS is similar for the three TC cases, and it remains lower for these cases at all lead times than for any of the other events. This result holds even when further normalizing AS to account for differences in total areal coverage between different events (not shown). In general, there is large (small) spread in the forecasts with the lowest (highest) skill, suggesting an inverse relationship between spread and skill (e.g., Whitaker and Loughe 1998).

b. Discussion

As with all single-number forecast verification statistics, the results discussed above and shown in Figs. 6 and 7 do not paint a complete picture of the quality and usefulness of the ensemble forecasts for these cases. This section will delve deeper into the reasons that the spread and skill vary among the nine cases, and how the weather systems responsible for producing the heavy rainfall determine the quality of the forecasts of each event.

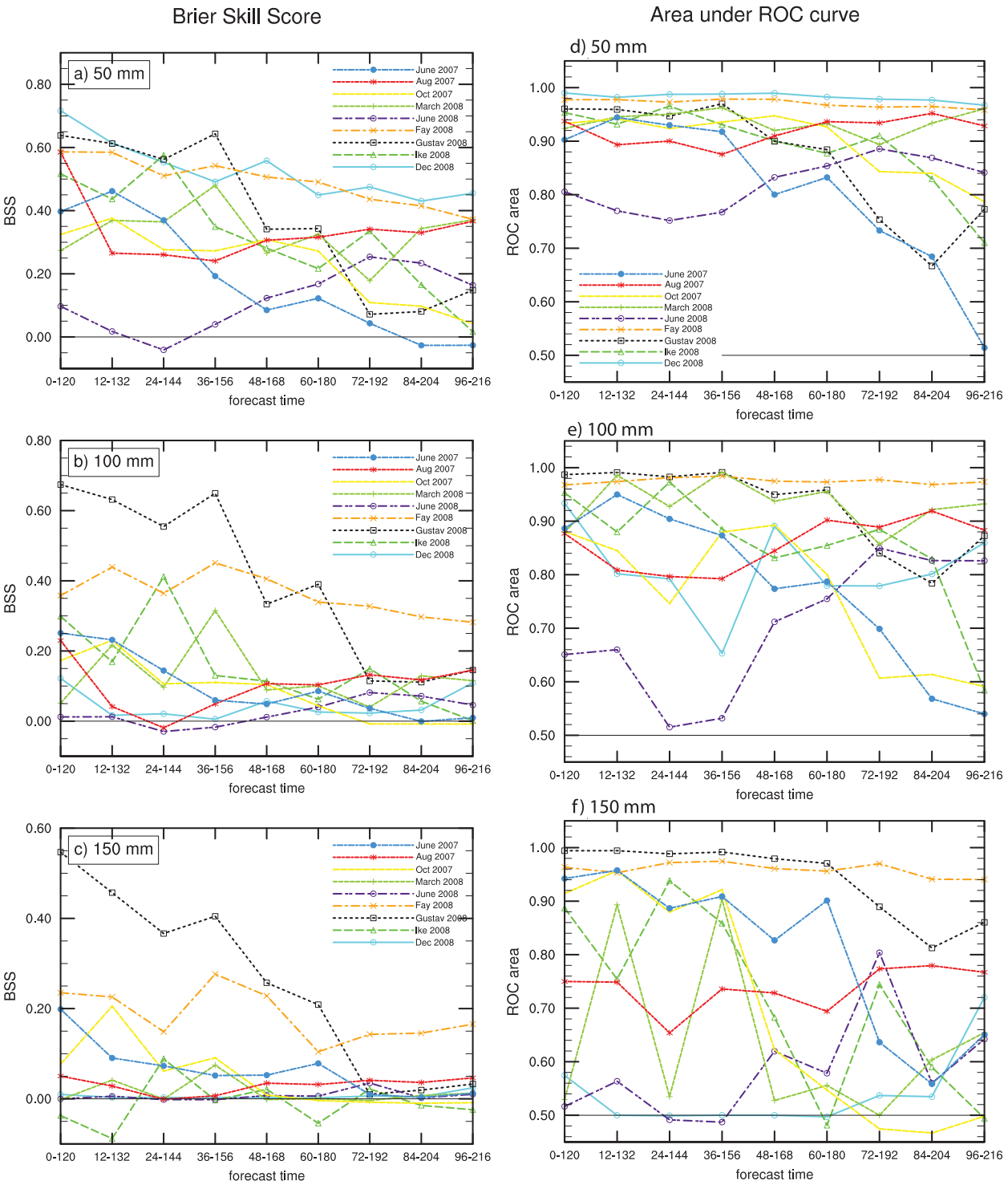

FIG. 6. (left) BSS and (right) area under the ROC curve for the ECMWF ensemble for the nine widespread 5-day rain events. Shown are 120-h precipitation accumulation thresholds of (a),(d) 50, (b),(e) 100, and (c),(f) 150 mm. The BSS is calculated relative to the seasonal climatology, as described in the text. A BSS of 1 is a perfect forecast; 0 indicates no skill. The ROC area does not account for seasonal climatology. A perfect forecast has a ROC area of 1; a random reference forecast has an area of 0.5. Note that the values shown on the ordinate are scaled differently in each panel.

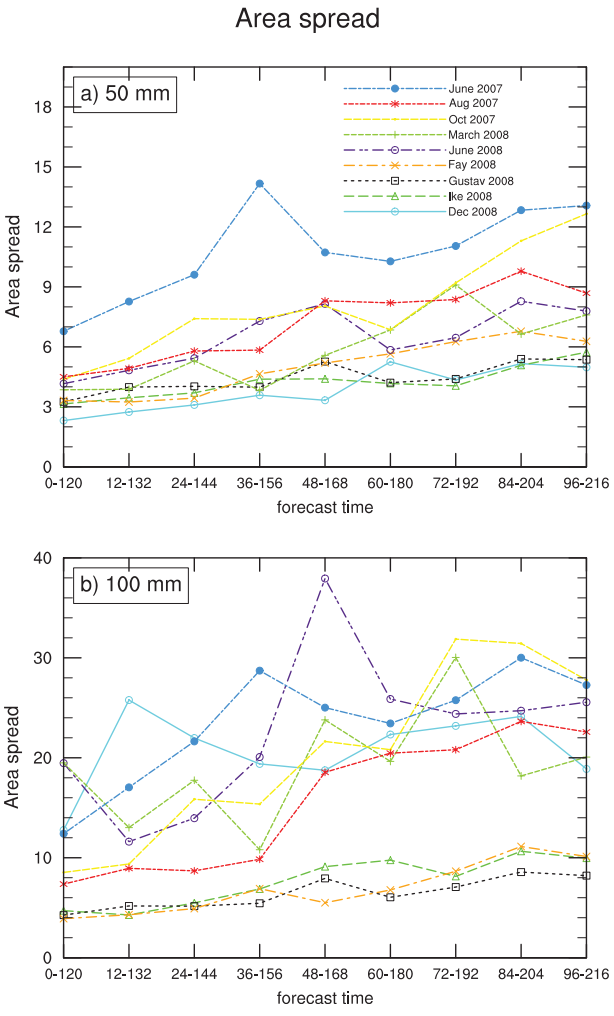

FIG. 7. As in Fig. 6, but for the area spread statistic at thresholds of (a) 50 and (b) 100 mm. Higher values of AS indicate greater spread. Note that the values shown on the ordinate are scaled differently in each panel. The maximum value of AS is equal to the number of ensemble members, which in this case is 51.

1) EVENTS WITH LOW FORECAST SKILL AND HIGH UNCERTAINTY

One of the most striking features of both the BSS and the ROC area (Fig. 6) statistics is the poor performance at short lead times for the June 2008 event (shown by the dark purple lines in the figures) in the Midwest. Given the severe impacts of the rainfall during that time period, the low skill scores for this event are a reason for concern. A summary of the raw forecast probabilities (Fig. 8a) shows that there are several reasons for the low verification scores. At the shortest lead times considered here (Figs. 8a and 8b), probabilities of 50 mm are very high throughout the Upper Midwest, particularly in the northern halves of Minnesota and Wisconsin, and also farther west in South Dakota. However, the observed heavy precipitation mainly occurred farther south. The ensemble did not provide any indication of heavy rainfall in Indiana—this was caused by a single quasi-stationary MCS, and global models at coarse resolutions are generally unable to predict an event of this type. Finally, the swath of observed precipitation in Missouri, Kansas, and Oklahoma, which occurred near the end of the 120-h forecast period, was not well forecast. At longer lead times (Figs. 8c and 8e), probabilities were lower overall (which is consistent with the larger spread) but they correctly identified the possibility for widespread rainfall in the general area where it would occur. As a result, the skill scores are higher at longer lead times for this event.

The other notable outlier at the 50-mm threshold in Figs. 6a and 6d is the lack of forecast skill at long lead times for the June 2007 event (shown by the blue lines in the figures) in the southern plains. While forecasts for all of the other events are shown to be skillful relative to climatology even at the 96–216-h forecast time, the June 2007 case shows a negative BSS for lead times of 84 and 96 h (Fig. 6a) and has a ROC area consistently smaller than that for the other events beyond 48 h (Fig. 6d). The AS for this event is also consistently larger than for the other events (Fig. 7a). At short lead times, the ensemble forecasts of this event appear subjectively to be quite good (Figs. 9a and 9b), consistent with the objective metrics. With increasing lead time, however, the forecasts degrade substantially (Figs. 9c and 9d), such that in the 96–216-h forecast, there was zero probability of 50 mm of rain in much of the region that received greater than that amount (Fig. 9e). Additionally, the highest probabilities of 50 mm at long lead times were in the northern plains (Figs. 9d and 9e), which generally had no precipitation at all during this time period (Fig. 4a).

An initial understanding of the large uncertainty in this event can be gained by examining the atmospheric processes responsible for the heavy rains in this case. Convective heating was primarily responsible for the spinup of a midlevel vortex, and also for its development into a warm-core circulation. The vortex then led to additional convection and rainfall, with this sequence continuing for several days. By 28 June, a lower-tropospheric vorticity center was apparent (Fig. 9k), and this vortex became better defined over the coming days (not shown). Ensemble forecasts of this vortex showed rapidly increasing spread and decreasing quality at increasing lead times (Figs. 9k–o). At shorter lead times (Figs. 9k and 9l), numerous ensemble members predicted low-level vorticity maxima in the southern plains; the precipitation forecasts from these same ensemble initializations were also quite accurate (Figs. 9a, 9b, 9f, and 9g). At longer lead times (Figs. 9m–o), there was much greater spread

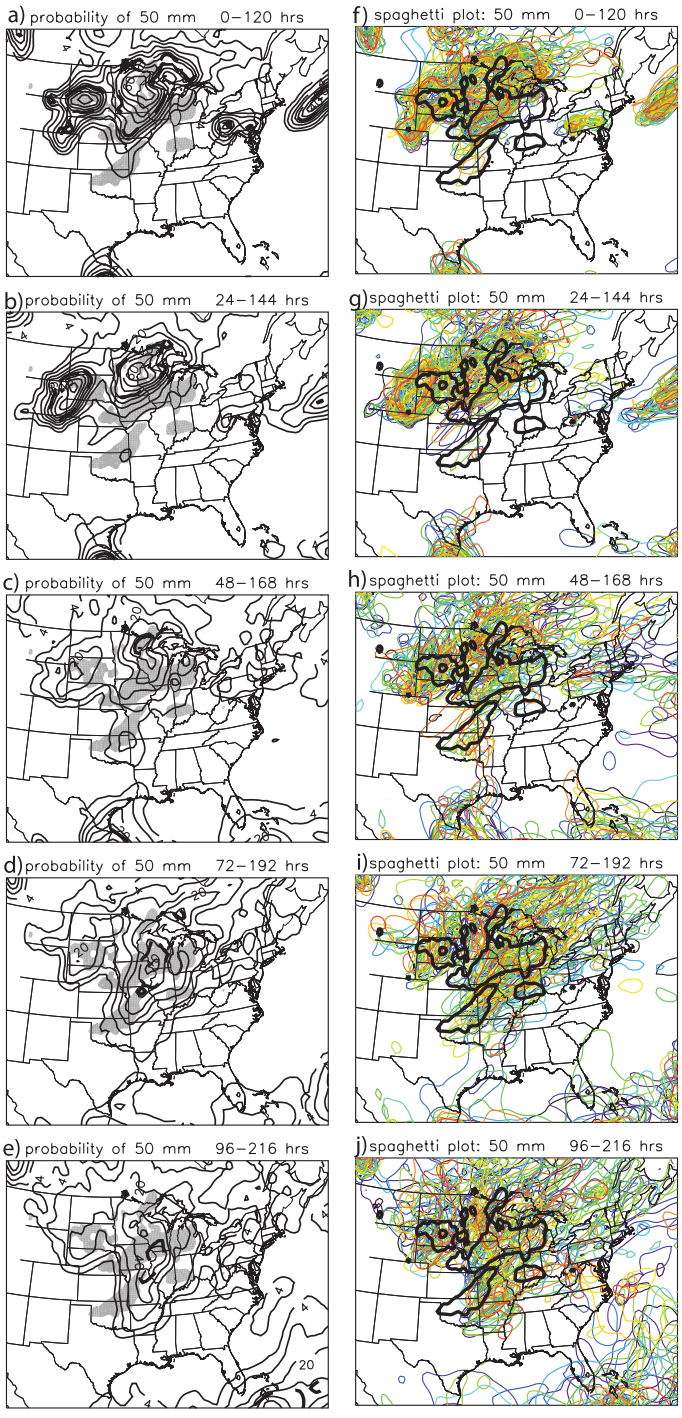

FIG. 8. (left) Raw ensemble probabilities, at increasing lead times, of 50 mm of precipitation in the 120-h period between 1200 UTC 4 Jun and 1200 UTC 9 Jun 2008. Probabilities are contoured at 4% (i.e., two ensemble members), 10%, and every 10% above that. The ensemble mean is shown by the thick black dashed line. Areas where 50 mm of precipitation was observed are shaded. (right) "Spaghetti" plot showing the predicted 50-mm rainfall contour from each ensemble member in a different color. The observed 50-mm contour is shown in thick black. Model initialization times shown are at 1200 UTC on (a),(f) 4 Jun, (b),(g) 3 Jun, (c),(h) 2 Jun, (d),(i) 1 Jun, and (e),(j) 31 May 2008.

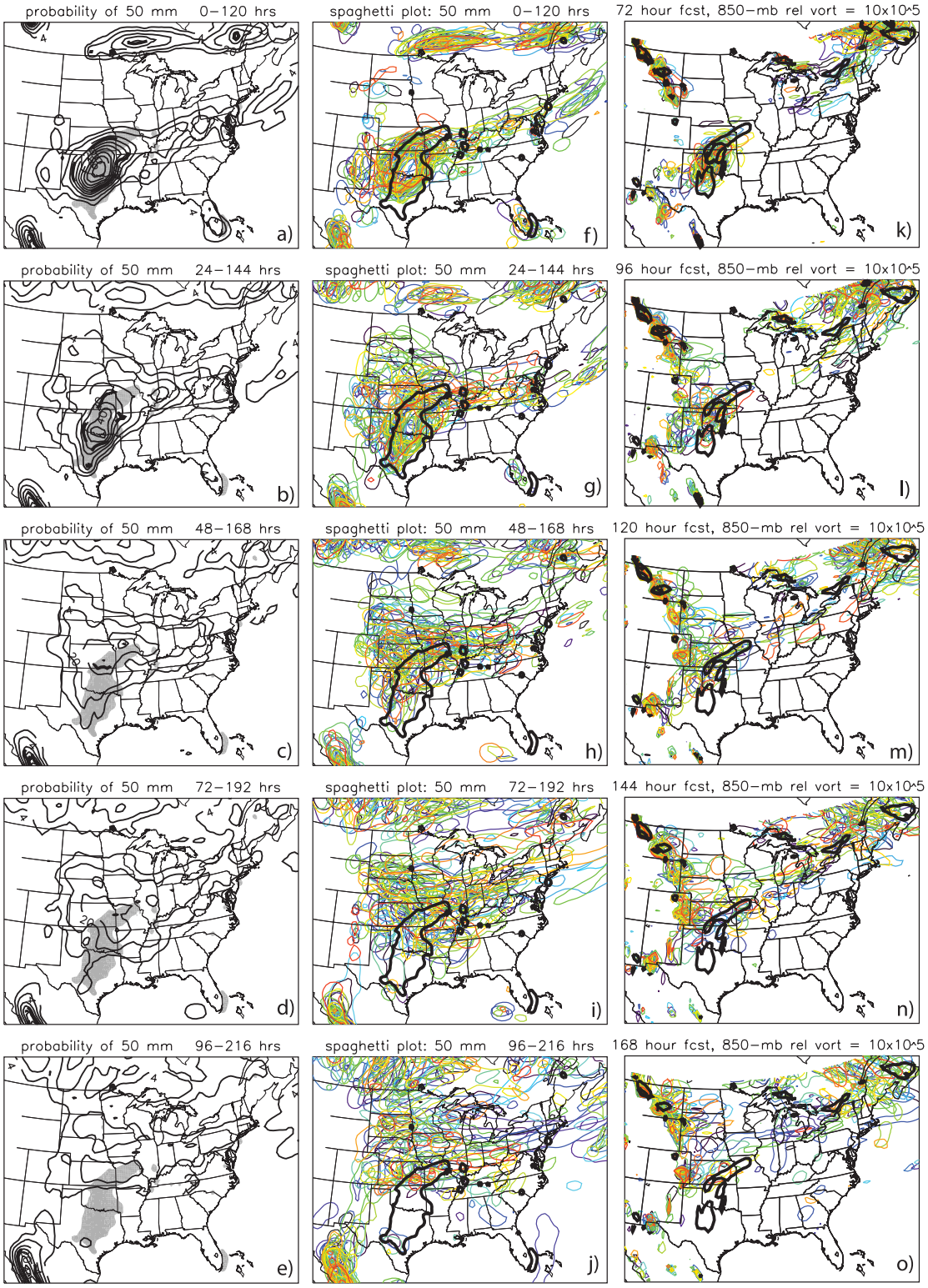

FIG. 9. As in Fig. 8, but for the June 2007 event. Model initialization times shown are at 1200 UTC on (a),(f) 25 Jun, (b),(g) 24 Jun, (c),(h) 23 Jun, (d),(i) 22 Jun, and (e),(j) 21 Jun 2007. (right) Spaghetti plot showing the locations of the 10×10^{-5} s^{-1} 850-hPa relative vorticity contours for each of the ensemble members (for members that have vorticity values exceeding this threshold). All panels are forecasts valid at 1200 UTC 28 Jun 2007, for ensemble forecasts initialized at 1200 UTC on (a) 25 Jun, (b) 24 Jun, (c) 23 Jun, (d) 22 Jun, and (e) 21 Jun 2007. The thick black contour shows the 10×10^{-5} s^{-1} 850-hPa relative vorticity contour from the ECMWF control initial analysis at 1200 UTC 28 Jun 2007.

in the locations of the predicted vortices, and very few of these vortices were in the correct location over the southern plains. For the longest lead time shown (Fig. 9o), the southern plains region was devoid of predicted low-level vorticity maxima, and as a result, there was also no suggestion of heavy rainfall in that area in the QPFs (Figs. 9e and 9j).

Inspecting the evolution of the vortices in the ensemble forecasts suggests a possible connection between the presence (and strength) of a midlevel vortex in the model initialization and the resulting forecasts of the vortex (Fig. 10). For example, in the forecast initialized at 1200 UTC 25 June (Fig. 10a), the vortex was already present in its early stages and was captured in the initial analysis, as illustrated by the closed 500-hPa height contour and broad region of midlevel vorticity in north Texas. In these runs, the predicted vortices followed similar paths: the spread in the locations of these vortices was small, and the resulting precipitation forecasts were good (Figs. 9a and 9k). At earlier initialization times, accurate precipitation forecasts only resulted if the model was successful in capturing the timing and location of the deep convection that initially formed the vortex, then the intensification of the vortex, and then the resulting deep convection and the vortex's maintenance. These results are similar to those of Davis et al. (2002) for MCVs, who found that Rapid Update Cycle (RUC; Benjamin et al. 2004) forecasts were often unable to predict the *formation* of MCVs but were able to evolve it once it was present in the initialization.

A somewhat similar pattern was in place during the October 2007 event (section 4c), which also showed a rapid decline in forecast quality with increasing lead time and relatively large spread.[2] At short lead times, when the cutoff low was already present in the model initialization, the resulting precipitation forecasts were accurate (Figs. 11a–c). At longer lead times, the likelihood of widespread heavy precipitation was much less certain, and many of the ensemble members predicted heavy rainfall in locations far removed from where it was observed (Figs. 11d,e). The similarities in the quality of the predictions of the June and October 2007 events suggest that the evolution of a balanced atmospheric circulation that is cut off from the upper-level jet may be fairly predictable, but *predicting the development* of that circulation may be considerably more difficult. On the other hand, the large spread in the ensemble may provide guidance to forecasters that there is the potential for low

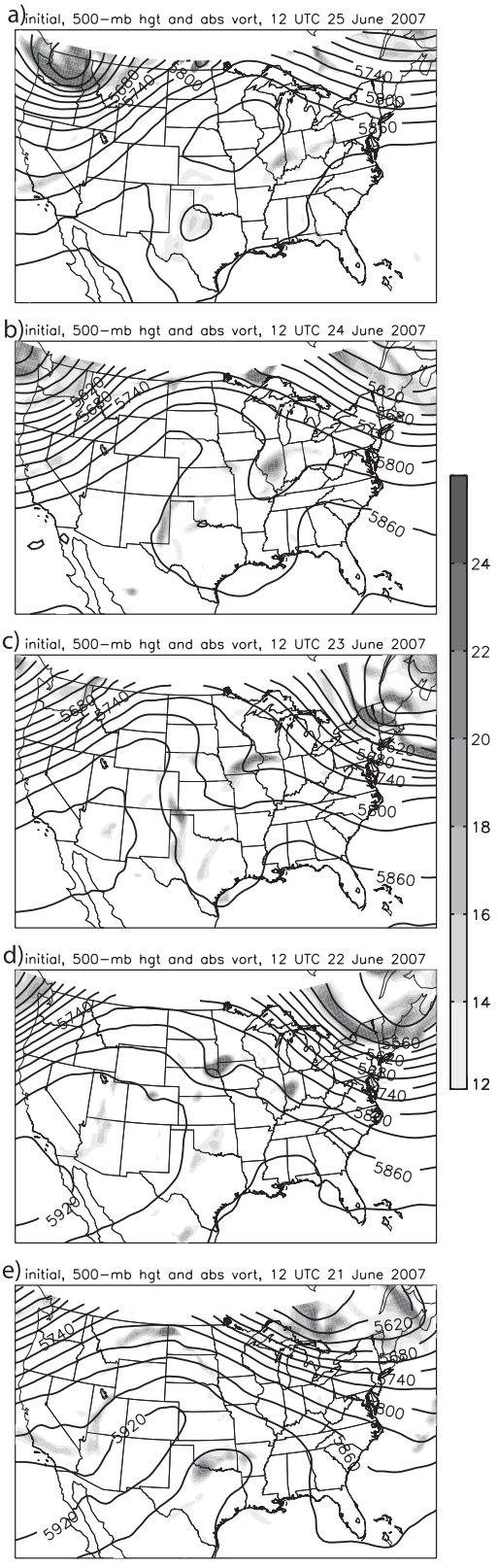

FIG. 10. Analysis of 500-hPa geopotential height (m, solid contours) and absolute vorticity ($\times 10^{-5}$ s^{-1}, shading) from the initialization of ECMWF control forecast at 1200 UTC on (a) 25 Jun, (b) 24 Jun, (c) 23 Jun, (d) 22 Jun, and (e) 21 Jun 2007.

[2] This time period was identified as a skill "dropout" (i.e., a period with particularly low skill) for the Global Forecast System (GFS) model by Alpert et al. (2009).

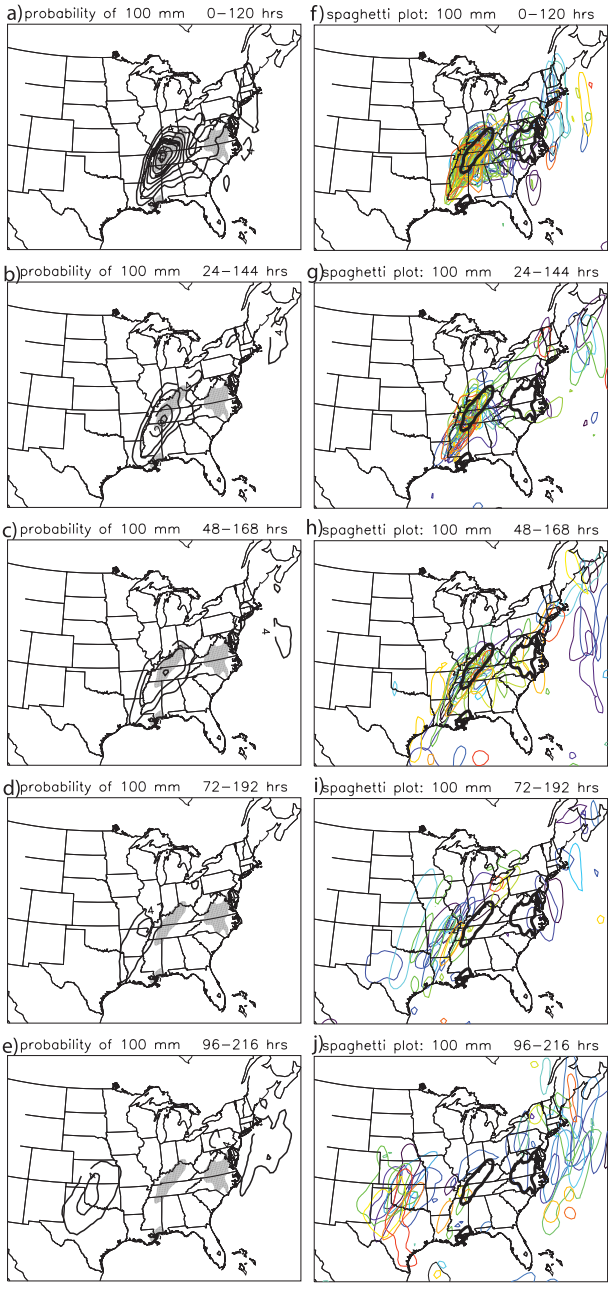

FIG. 11. As in Fig. 9, but for the October 2007 event and for a 100-mm precipitation threshold. Model initialization times shown are 1200 UTC on (a),(f) 22 Oct, (b),(g) 21 Oct, (c),(h) 20 Oct, (d),(i) 19 Oct, and (e),(j) 18 Oct 2007.

forecast skill at the medium range, based on the apparent inverse spread–skill relationship seen here.

2) EVENTS WITH HIGH FORECAST SKILL AND LOW UNCERTAINTY

In contrast to the events with relatively poor forecasts discussed above, other events were notable for their very skillful forecasts at long lead times. One of these events was Tropical Storm Fay (section 4f), which was an outlier in terms of its high skill, even at the 150-mm precipitation threshold and at long lead times (the orange lines in Figs. 6c and 6f). The ensemble forecasts correctly identified the high probability of 50 mm of precipitation in 5 days over a large swath of the southeast (Figs. 12a–e), as well as the smaller area exceeding 150 mm of rainfall (Figs. 12f–j). In fact, at the 150-mm threshold, the ensemble shows a positive BSS all the way out to the 180–320-h precipitation forecast (not shown). The spread of the predicted heavy precipitation was also small for this event (Fig. 7).

The high level of skill in the heavy rain forecasts in this case can be related to the timing and evolution of the tropical cyclone. Fay initially developed in the Caribbean Sea on 15 August, nearly a week before it would affect the southeast. As a result, the focusing mechanism for the heavy rains was present in the model's initial conditions long before the event took place, and the model succeeded in predicting the approximate track of the storm such that it would make landfall in the United States, move slowly, and produce widespread heavy rains. The processes here can be contrasted with those in the June 2007 event: in that case the ''genesis'' of the lifting mechanisim (a mesoscale vortex) took place only a day or two before the heavy rains fell, and the ensemble predictions of precipitation at lead times beyond a few days were poor. The ensemble's success with Fay (and the other TC events discussed previously) should not be generalized too broadly, however. It is likely that medium-range forecasts for TCs that form near the U.S. coast and quickly make landfall would be much less skillful. In fact, the forecasts for Hurricane Gustav's rainfall, which were the best of all the events at short lead times and high rainfall thresholds, dropped off in skill rapidly with increasing lead time. At longer lead times, the storm was still in the Caribbean Sea, far from the coast, but as the storm neared the coast the track of the storm became more certain and the corresponding rainfall forecast became much more accurate.

The December 2008 event in the east (section 4i) also had relatively good predictions at the 50-mm threshold. At the longest lead times, the forecasts for this event had the highest BSS (the cyan line in Fig. 6a) and the third-highest ROC area (Fig. 6d). The forecast skill dropped off quickly for higher rainfall amounts, but the ensemble did give high probabilities for 50 mm of rain in the correct region far in advance of the event (Fig. 13a), and correctly identified the locations where over 100 mm would occur (Fig. 13b). The probabilities of 100 mm were quite low, however, hence the low skill scores at that threshold. Nonetheless, even these forecasts at long lead times

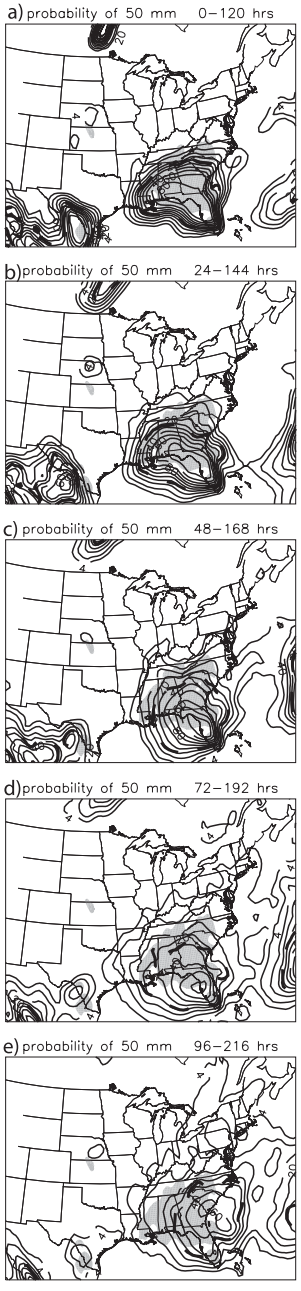

FIG. 12. As in Fig. 9, but for the Tropical Storm Fay rainfall. Probabilities of (left) 50 and (right) 150 mm of precipitation. Model initialization times shown are at 1200 UTC on (a),(f) 22 Aug, (b),(g) 21 Aug, (c),(h) 20 Aug, (d),(i) 19 Aug, and (e),(j) 18 Aug 2008.

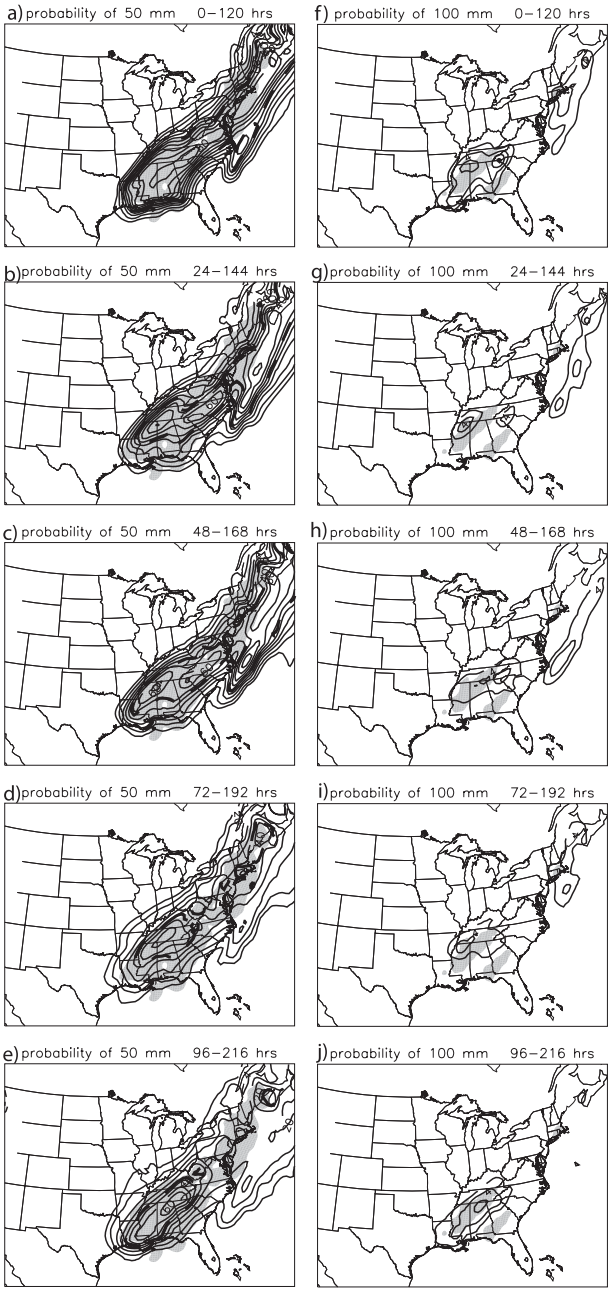

FIG. 13. As in Fig. 9, but for the December 2008 event. Probabilities of (left) 50 and (right) 100 mm of precipitation. Model initialization times shown are at 1200 UTC on (a),(f) 8 Dec, (b),(g) 7 Dec, (c),(h) 6 Dec, (d),(i) 5 Dec, and (e),(j) 4 Dec 2008.

could potentially be useful, as they correctly identified the precipitation pattern, showing a high probability of a smaller amount with smaller probabilities of a large amount, all in approximately the correct location. With a correct forecast of the precipitation *pattern*, proper ensemble calibration (e.g., Hamill et al. 2008) may improve on the predicted amounts.

3) TROPICAL CYCLONE TRACK AND PRECIPITATION UNCERTAINTY

As discussed above, the results of this study suggest an inverse relationship between spread and skill for heavy rain events: in addition to having the largest spread, the June 2007 case also had the least skillful forecasts at long

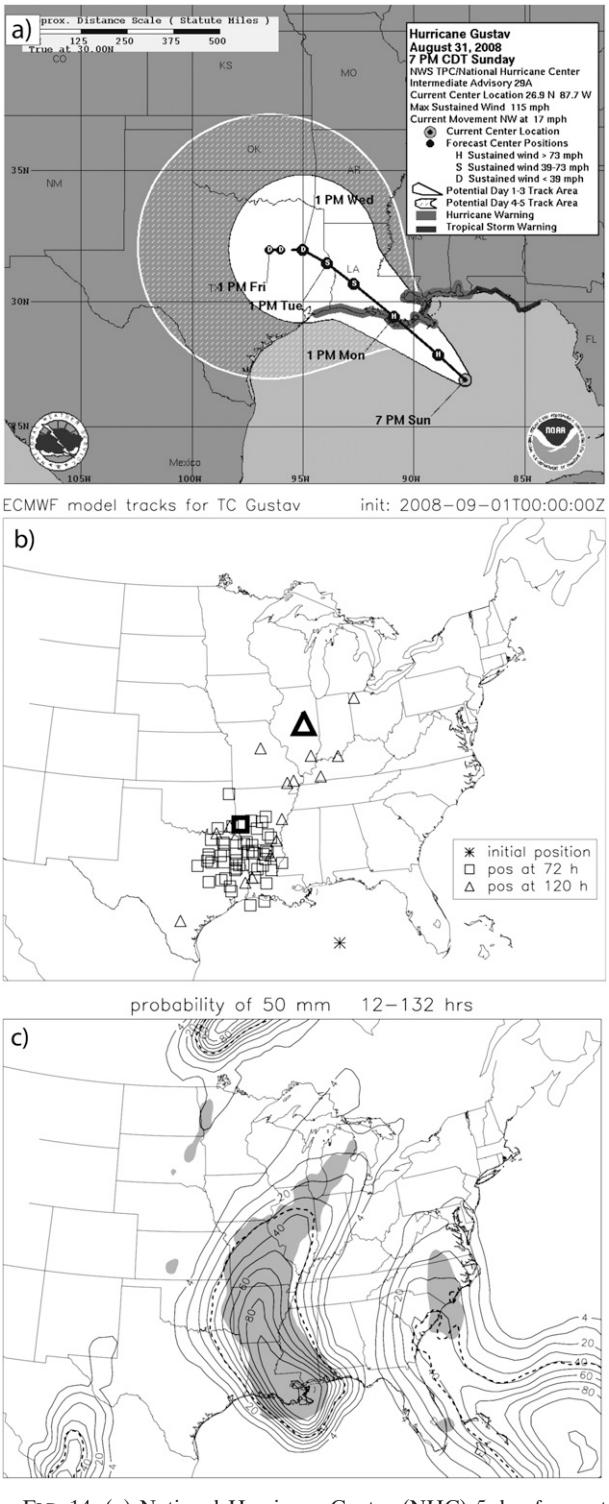

FIG. 14. (a) National Hurricane Center (NHC) 5-day forecast track of Hurricane Gustav, issued at 2300 UTC 31 Aug 2008. (b) Track forecasts for Gustav from the ECMWF ensemble for the forecast initialized at 0000 UTC 1 Sep 2008 [the same time as the NHC forecast shown in (a)]. The initial position of Gustav at this time is shown by the asterisk. Track locations from the different ensemble members at 72 h (0000 UTC 4 Sep) are shown by the squares, and at 120 h (0000 UTC 6 Sep) by the triangles. Observed track locations at these times are shown by the large bold symbols. (c) Raw ensemble probabilities of 50 mm of precipitation, plotted as in Fig. 9 for the forecast initialized at 0000 UTC 1 Sep, valid for the period 1200 UTC 1 Sep–1200 UTC 6 Sep 2008. The image in (a) was obtained from the NHC.

lead times, whereas the TC cases had low spread and high skill. Figure 7 also shows a large separation in spread between the TCs and all of the other events. The low spread and high skill in the TC cases suggests either that rainfall from TCs is indeed more predictable than that from other weather systems or that the ensemble may be underdispersive when it comes to predicting TC tracks.

This issue is investigated further for Hurricane Gustav. The ensemble had considerable spread in the predicted Gustav tracks after landfall, yet still provided excellent probabilistic precipitation forecasts (Fig. 14). For the forecast initialized at 0000 UTC 1 September 2008, there were ensemble members with tracks leading everywhere from south Texas to Ohio by hour 120 (Fig. 14b). This uncertainty is even greater than that suggested by the National Hurricane Center's cone of uncertainty in the 3–5-day forecast (Fig. 14a), which is based on historical errors in track forecasts. Despite this uncertainty in the actual track of Gustav postlandfall, the precipitation probabilities (Fig. 14c) indicate a wide swath of heavy rainfall extending through the Mississippi Valley, associated with the scenario of Gustav becoming extratropical and curving poleward, which is in fact what occurred.

The high skill and low spread (suggesting high predictability) for TC rainfall makes some intuitive sense: given the prediction that a TC will make landfall, it is very likely that it will produce a wide swath of heavy rainfall: this is a fairly predictable result. It is also possible that the ensemble's forecasts for the three TCs considered here were particularly good compared with other cases; this could be a subject for future research. Nonetheless, the results of the analysis of both spread and skill suggest that heavy rainfall from TCs can be quite certain well in advance of landfall.

6. Conclusions

This study used an ensemble of global numerical weather forecasts to analyze their skill in predicting widespread, multiple-day rain events in the United States; related the performance of the ensembles to the relevant weather systems in each event; and came to conclusions about the relative predictability of such events. Nine

events in 2007–08 were examined in which widespread accumulations of 100 mm in 5 days occurred, comprising three warm-season cases, three cool-season cases, and three tropical cyclone cases. A summary of the primary findings is as follows:

- In general, the ECMWF ensemble provides skillful predictions of widespread heavy rain at relatively short lead times (e.g., 0–120- to 24–144-h forecasts). This is particularly true for rainfall amounts exceeding 50 and 100 mm in 120 h; the results were varied at higher thresholds.
- In a few of the events, the ensemble showed considerable skill at longer lead times, out to the 96–216-h precipitation forecasts.
- The ensemble performed best overall for the tropical cyclone events, particularly for the rainfall from Tropical Storm Fay and Hurricane Gustav. Two of the cool-season events associated with strong synoptic-scale forcing were also well predicted at longer lead times.
- Two of the warm-season events showed particularly poor forecasts: the June 2008 event in the Upper Midwest had very low verification scores at short lead times, as did the June 2007 event in the southern plains at long lead times. Both were associated with organized mesoscale convection and interactions between mesoscale features.
- The October 2007 event in the eastern United States, which was associated with a cutoff cyclone, also had poor forecasts at long lead times. The results from this case and the June 2007 event suggest that once a cutoff circulation is present in the model initial conditions, the forecasts tend to be good, but prior to this development, the forecast skill is low.
- In general, there was an inverse relationship between spread and skill in the precipitation forecasts: the events with the highest spread also had the lowest skill, and vice versa.

Only the ECMWF ensemble was considered in this study, but combining additional ensemble prediction systems into a grand ensemble is an intended avenue for future work, made possible by the availability of TIGGE data. Also, this study has not considered ''false alarm'' forecasts of heavy-rain events in the ECMWF ensemble, nor has it attempted to calibrate the precipitation forecasts, which could also be considered in future research. Understanding the situations in which these ensembles perform well (and not so well) may be a way that human knowledge of synoptic meteorology and numerical weather prediction can be combined to make improved forecasts. Although the limitations to long-range weather prediction owing to chaos will remain, ensembles will continue to be an important tool in providing enhanced forecasts, and better information about forecast uncertainty, for the benefit of the many users of this information.

Acknowledgments. The authors thank Joshua Hacker, Thomas Hamill, Richard Rotunno, and David Dowell for helpful suggestions and discussions relating to this research. The authors also thank Steve Mullen and two anonymous reviewers for their helpful and constructive suggestions. Flood reports presented in section 4 were obtained from the National Climatic Data Center's online Storm Events database unless stated otherwise.

REFERENCES

Alpert, J. C., D. L. Carlis, B. A. Ballish, and V. K. Kumar, 2009: Using pseudo RAOB observations to study GFS skill score dropouts. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction,* Omaha, NE, Amer. Meteor. Soc., 5A.6. [Available online at http://ams.confex.com/ams/pdfpapers/154268.pdf.]

Arndt, D. S., J. B. Basara, R. A. McPherson, B. G. Illston, G. D. McManus, and D. B. Demko, 2009: Observations of the overland reintensification of Tropical Storm Erin (2007). *Bull. Amer. Meteor. Soc.,* **90,** 1079–1093.

Bartels, D. L., and R. A. Maddox, 1991: Midlevel cyclonic vortices generated by mesoscale convective systems. *Mon. Wea. Rev.,* **119,** 104–118.

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.,* **132,** 495–518.

Berg, R., cited 2009: Tropical Storm Ike. National Hurricane Center Tropical Cyclone Rep. AL092008, 55 pp. [Available online at http://www.nhc.noaa.gov/pdf/TCR-AL092008_Ike.pdf.]

Beven, J. L., II, and T. B. Kimberlain, cited 2009: Tropical Storm Gustav. National Hurricane Center Tropical Cyclone Rep. AL072008, 36 pp. [Available online at http://www.nhc.noaa.gov/pdf/TCR-AL072008_Gustav.pdf.]

Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble (TIGGE). *Bull. Amer. Meteor. Soc.,* in press.

Buizza, R., J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart, 2007: The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quart. J. Roy. Meteor. Soc.,* **133,** 681–695.

Carbone, R. E., J. D. Tuttle, D. A. Ahijevych, and S. B. Trier, 2002: Inferences of predictability associated with warm season precipitation episodes. *J. Atmos. Sci.,* **59,** 2033–2056.

Corfidi, S. F., 2003: Cold pools and MCS propagation: Forecasting the motion of downwind-developing MCSs. *Wea. Forecasting,* **18,** 997–1017.

Davis, C. A., D. A. Ahijevych, and S. B. Trier, 2002: Detection and prediction of warm season midtropospheric vortices by the Rapid Update Cycle. *Mon. Wea. Rev.,* **130,** 24–42.

Dirmeyer, P. A., and J. L. Kinter III, 2009: The ''Maya Express'': Floods in the U.S. Midwest. *Eos, Trans. Amer. Geophys. Union,* **90,** 101–102.

ECMWF, cited 2009a: The creation of perturbed analyses. [Available online at http://www.ecmwf.int/products/forecasts/guide/The_creation_of_perturbed_analyses.html.]

——, cited 2009b: TIGGE model upgrades. [Available online at http://tigge.ucar.edu/documentation.htm.]

Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.,* **85,** 955–965.

——, 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. Appl. Meteor.,* **25,** 1333–1345.

Galarneau, T. J., Jr., L. F. Bosart, and R. S. Schumacher, 2010: Predecessor rain events ahead of tropical cyclones. *Mon. Wea. Rev.,* in press.

Giordano, L. A., and J. M. Fritsch, 1991: Strong tornadoes and flash-flood-producing rainstorms during the warm season in the mid-Atlantic region. *Wea. Forecasting,* **6,** 437–455.

Goebbert, K. H., A. D. Schenkman, C. M. Shafer, and N. A. Snook, 2008: An overview of the summer 2007 excessive rain event in the Southern Plains. Preprints, *22nd Conf. on Hydrology,* New Orleans, LA, Amer. Meteor. Soc., P1.2. [Available online at http://ams.confex.com/ams/pdfpapers/135330.pdf.]

Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.,* **132,** 2905–2923.

——, R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.,* **136,** 2620–2632.

He, Y., F. Wetterhall, H. Cloke, F. Pappenberger, M. Wilson, J. Freer, and G. McGregor, 2009: Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteor. Appl.,* **16,** 91–101.

Higgins, R. W., Y. Yao, E. S. Yarosh, J. E. Janowiak, and K. C. Mo, 1997: Influence of the Great Plains low-level jet on summertime precipitation and moisture transport over the central United States. *J. Climate,* **10,** 481–507.

——, W. Shi, E. Yarosh, and R. Joce, 2000: Improved United States precipitation quality control system and analysis. NCEP/Climate Prediction Center Atlas No. 7. [Available online at http://www.cpc.ncep.noaa.gov/research_papers/ncep_cpc_atlas/7/index.html.]

Houze, R. A., Jr., 2004: Mesoscale convective systems. *Rev. Geophys.,* **42,** RG4003, doi:10.1029/2004RG000150.

Jung, T., and M. Leutbecher, 2008: Scale-dependent verification of ensemble forecasts. *Quart. J. Roy. Meteor. Soc.,* **134,** 973–984.

Junker, N. W., R. S. Schneider, and S. L. Fauver, 1999: A study of heavy rainfall events during the Great Midwest Flood of 1993. *Wea. Forecasting,* **14,** 701–712.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Knippertz, P., and A. H. Fink, 2009: Prediction of dry-season precipitation in tropical West Africa and its relation to forcing from the extratropics. *Wea. Forecasting,* **24,** 1064–1084.

Konrad, C. E., II, 2001: The most extreme precipitation events over the eastern United States from 1950 to 1996: Considerations of scale. *J. Hydrometeor.,* **2,** 309–325.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307.

Maddox, R. A., C. F. Chappell, and L. R. Hoxit, 1979: Synoptic and meso-α scale aspects of flash flood events. *Bull. Amer. Meteor. Soc.,* **60,** 115–123.

Minnesota State Climatology Office, cited 2009: Heavy rains fall on southeastern Minnesota: August 18–20, 2007. [Available online at http://climate.umn.edu/doc/journal/flash_floods/ff070820.htm.]

Mo, K. C., J. N. Paegle, and R. W. Higgins, 1997: Atmospheric processes associated with summer floods and droughts in the central United States. *J. Climate,* **10,** 3028–3046.

Mullen, S. L., and R. Buizza, 2001: Quantitative precipitation forecasts over the United States by the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.,* **129,** 638–663.

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

Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and A. de Roo, 2008: New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophys. Res. Lett.,* **35,** L10404, doi:10.1029/2008GL033837.

Park, Y. Y., R. Buizza, and M. Leutbecher, 2008: TIGGE: Preliminary results on comparing and combining ensembles. *Quart. J. Roy. Meteor. Soc.,* **134,** 2029–2050.

Romero, R., A. Martin, V. Homar, S. Alonso, and C. Ramis, 2006: Predictability of prototype flash flood events in the Western Mediterranean under uncertainties of the precursor upper-level disturbance. *Adv. Geosci.,* **7,** 55–63.

Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.,* **65,** 1063–1076.

Schaake, J. C., T. M. Hamill, R. Buizza, and M. Clark, 2007: HEPEX: The Hydrological Ensemble Prediction Experiment. *Bull. Amer. Meteor. Soc.,* **88,** 1541–1547.

Schumacher, R. S., and R. H. Johnson, 2005: Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Mon. Wea. Rev.,* **133,** 961–976.

——, and ——, 2006: Characteristics of United States extreme rain events during 1999–2003. *Wea. Forecasting,* **21,** 69–85.

Stensrud, D. J., and M. S. Wandishin, 2000: The correspondence ratio in forecast evaluation. *Wea. Forecasting,* **15,** 593–602.

Stewart, S. R., and J. L. Beven II, cited 2009: Tropical Storm Fay. National Hurricane Center Tropical Cyclone Rep. AL062008, 29 pp. [Available online at http://www.nhc.noaa.gov/pdf/TCR-AL062008_Fay.pdf.]

Thielen, J., K. Bogner, F. Pappenberger, M. Kalas, M. del Medico, and A. de Roo, 2009: Monthly-, medium-, and short-range flood warning: Testing the limits of predictability. *Meteor. Appl.,* **16,** 77–90.

Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large-scale atmospheric flow patterns. *Tellus,* **9,** 275–295.

Tuttle, J. D., and C. A. Davis, 2006: Corridors of warm season precipitation in the central United States. *Mon. Wea. Rev.,* **134,** 2297–2317.

USGS, cited 2009: 2008 Iowa flood information page. [Available online at http://ia.water.usgs.gov/flood08/index.html.]

Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.,* **129,** 729–747.

Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.,* **126,** 3292–3302.

Wilks, D. S., 2005: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.,* **60,** 1173–1185.

——, A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting,* **21,** 149–166.