An Initial Assessment of Radar Data Assimilation on Warm Season Rainfall Forecasts for Use in Hydrologic Models

Ben A. Moser
Iowa State University

William A. Gallus Jr.
Iowa State University, wgallus@iastate.edu

Ricardo Mantilla
University of Iowa

Follow this and additional works at: http://lib.dr.iastate.edu/ge_at_pubs
Part of the Atmospheric Sciences Commons, Climate Commons, and the Meteorology Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/ge_at_pubs/240. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
An Initial Assessment of Radar Data Assimilation on Warm Season Rainfall Forecasts for Use in Hydrologic Models

Abstract
The effect of introducing radar data assimilation into the WRF Model to improve high-resolution rainfall forecasts that are used for flash flood forecasting is analyzed. The authors selected 12 heavy rainfall events and performed two WRF 24-h simulations that produced quantitative precipitation forecasts (QPFs) for each, one using the standard configuration in forecast mode (QPF-Cold) and one using radar data assimilated at initialization (QPF-Hot). Simulation outputs are compared with NWS stage IV QPEs for storm placement, area over threshold coverage, and equitable threat scores. The two QPF products and stage IV data are used to force the distributed hydrological model CUENCAS for the same 800 km × 800 km domain centered over Iowa (and to calculate peak flows across the river network). The hydrological model responses to the three products are compared in terms of spatial location and flood intensity. In general, QPF-Hot outperformed QPF-Cold in replicating stage IV QPE statistics. However, QPF-Hot was too wet in the first 2 h of the event, and storms created by the radar-assimilation techniques dissipated quickly, with rainfall forecasts resembling QPF-Cold after 12 h. Flash flooding predicted by CUENCAS using QPF-Hot was more consistent with stage IV in terms of placement and intensity; however, results were not consistent for all events evaluated. The most encouraging result is that expected flash flooding was indeed predicted in all 12 cases using QPF-Hot and not QPF-Cold even though placement and intensity were not a perfect match. The initial results of this study indicate that radar assimilation improves WRF's ability to capture the character of storms, promising more accurate guidance for flash flood warnings.

Keywords
Forecasting, Forecast verification/skill, Numerical weather prediction/forecasting, Short-range prediction, Models and modeling, Data assimilation, Hydrologic models, Applications, Flood events

Disciplines
Atmospheric Sciences | Climate | Meteorology

Comments
An Initial Assessment of Radar Data Assimilation on Warm Season Rainfall Forecasts for Use in Hydrologic Models

BEN A. MOSER AND WILLIAM A. GALLUS JR.
Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

RICARDO MANTILLA
Department of Civil and Environmental Engineering, and Iowa Flood Center, The University of Iowa, Iowa City, Iowa

(Manuscript received 8 October 2014, in final form 23 May 2015)

ABSTRACT

The effect of introducing radar data assimilation into the WRF Model to improve high-resolution rainfall forecasts that are used for flash flood forecasting is analyzed. The authors selected 12 heavy rainfall events and performed two WRF 24-h simulations that produced quantitative precipitation forecasts (QPFs) for each, one using the standard configuration in forecast mode (QPF-Cold) and one using radar data assimilated at initialization (QPF-Hot). Simulation outputs are compared with NWS stage IV QPEs for storm placement, area over threshold coverage, and equitable threat scores. The two QPF products and stage IV data are used to force the distributed hydrological model CUENCAS for the same 800 km × 800 km domain centered over Iowa (and to calculate peak flows across the river network). The hydrological model responses to the three products are compared in terms of spatial location and flood intensity. In general, QPF-Hot outperformed QPF-Cold in replicating stage IV QPE statistics. However, QPF-Hot was too wet in the first 2 h of the event, and storms created by the radar-assimilation techniques dissipated quickly, with rainfall forecasts resembling QPF-Cold after 12 h. Flash flooding predicted by CUENCAS using QPF-Hot was more consistent with stage IV in terms of placement and intensity; however, results were not consistent for all events evaluated. The most encouraging result is that expected flash flooding was indeed predicted in all 12 cases using QPF-Hot and not QPF-Cold even though placement and intensity were not a perfect match. The initial results of this study indicate that radar assimilation improves WRF’s ability to capture the character of storms, promising more accurate guidance for flash flood warnings.

1. Introduction

Although recent improvements in computing, better analyses of the atmosphere, and more accurate understanding of atmospheric microphysics have fostered tremendous improvements in deterministic numerical weather prediction (NWP) efforts in recent decades, contemporary NWP models remain limited in their ability to predict rainfall amounts. Warm season convective rainfall, the main driver of flooding in the upper Midwest and other areas of the world, is especially poorly forecasted (Olson et al. 1995; Stensrud et al. 2000; Fritsch and Carbone 2004; Sun et al. 2014). This is a particularly pressing issue because annual flood losses (inflation adjusted) in the United States increased from about $1 billion in the 1940s to about $5 billion in the 1990s (Pielke and Downton 2000).

The idea of using high-resolution quantitative precipitation forecasts (QPFs) from mesoscale weather models in real-time flood forecasting is gaining popularity (e.g., Liu et al. 2013). Using QPFs instead of QPEs after rain has already fallen would significantly increase lead time for flood warnings. However, since QPF skill associated with warm season convective rainfall has been poor, QPF has not been considered for use in hydrologic modeling for streamflow. Consequently, nowcasting (short-range and case-specific forecasting for periods shorter than a few hours) has been used instead of model QPFs to provide short-lead-time operational flood forecasts. These nowcasting methods extrapolate radar echoes (e.g., Dixon and Wiener 1993; Mueller et al. 2003), and their success appears limited since they
do not allow for the development of new storms/rain areas and become less accurate as lead time increases (Ebert et al. 2004). In addition, so-called expert now-casting systems, which attempt to predict storm initiation, growth, and decay from the location of boundary layer convergence lines, have only shown an improvement in skill during the first hour when compared to more primitive extrapolation methods (Roberts et al. 2012). Therefore, high-resolution model QPFs may hold more promise since models can develop detailed precipitation fields with longer-lasting forecast reliability.

The accuracy of mesoscale NWP models in the first 3–6 h suffers from the spinup effect (Daley 1991), and the models can be less accurate than predictions based on advection of radar echoes (Austin et al. 1987). Additionally, model performance depends on the accuracy of the initial conditions and on model errors both in the discretization of equations and in the physical parameterizations. Several studies have shown that radar data assimilation improves precipitation forecasts, especially in the short term (first 6 h or so), in part by reducing the spinup effect and by reducing errors in initial and lateral boundary conditions. For instance, Macpherson (2001), along with Davolio and Buzzi (2004), found that assimilation of radar data via nudging techniques improved rainfall forecasts in the first 6 h. Sugimoto et al. (2009) focused on the use of three-dimensional variational data assimilation (3DVAR) techniques and found that the assimilation of radial velocity and reflectivity from Doppler radar results in the most accurate short-range precipitation forecasting. Radar data assimilation in the Center for the Analysis and Prediction of Storms (CAPS) ensemble has also been found to improve forecasts significantly, especially over the first 6 h or so (Kain et al. 2010; Berenguer et al. 2012). Sun et al. (2014) concluded that future advances in data assimilation have the potential to yield even greater improvements.

The present study uses a 3DVAR data assimilation system (ADAS; Brewster 1996) to adjust the hydrometeor and cloud fields based off of reflectivity data. Simulations were run for 12 heavy rainfall cases that occurred in and around Iowa using the Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model. We performed two simulations for each case, one using radar data (hot start) to produce the initial conditions and one in which the initial conditions used NAM data that were interpolated onto the WRF grid (cold start). We evaluated the two quantitative precipitation forecasts (QPF-Hot and QPF-Cold for short) using a variety of methods including point to point, object based, and visual verification. Finally, the distributed hydrologic model CUENCAS was forced with WRF Model. We performed two simulations for each case, one using radar data (hot start) to produce the initial conditions and one in which the initial conditions used NAM data that were interpolated onto the WRF grid (cold start). We evaluated the two quantitative precipitation forecasts (QPF-Hot and QPF-Cold) to produce flash flood predictions in order to assess the impact that the improvement in QPF provided for the 12 events. Our strategy enables a fair comparison of the QPF field’s ability to reproduce the rainfall features that are conducive to flash flooding and eliminates the need to contrast the results with observations of actual flash flooding.

2. Methodology

We selected 12 cases of heavy rainfall in Iowa for this study. A case was selected if the 24-h rainfall total from the NWS 24-h Cooperative Observer reports exceeded 5 in. for at least two stations in Iowa. The events occurred between 1 May 2001 and 1 September 2011, a period when North American Model (NAM) analyses were available to initialize the WRF Model. The dates and initialization times of the WRF runs are shown in Table 1.

We used stage IV rainfall data to verify the events. These analyses are created using regional multisensor precipitation analyses produced at River Forecast Centers (RFCs) across the United States; these analyses are then mosaiced/quilted into a national product on a 4-km horizontal grid at NCEP. A combination of the WSR-88D network of radars and surface rain gauges at RFCs is used to produce the stage IV product. Stage IV data benefit both from surface gauge input and manual and automatic quality controls at each individual RFC before being interpolated onto the Hydrologic Rainfall

<table>
<thead>
<tr>
<th>Year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>0000 UTC 25 Jun</td>
<td>0000 UTC 2 Aug</td>
<td>1800 UTC 19 Aug</td>
<td>0000 UTC 30 May</td>
<td>1800 UTC 26 Aug</td>
<td>0000 UTC 5 Jun</td>
<td>0000 UTC 10 Jun</td>
</tr>
</tbody>
</table>

Table 1. Dates and WRF Model initialization times for the 12 heavy rainfall events.
Analysis Project (HRAP) grid to become the NCEP stage IV dataset (Fulton 2005).

The stage IV product is widely considered to be the best gridded rainfall analysis dataset over the contiguous United States and is often used as a benchmark when evaluating other remotely sensed precipitation products (Wu et al. 2012; Gourley et al. 2010; Lin and Hou 2012). A study by Smalley et al. (2014) showed the stage IV product struggles with frozen precipitation and light precipitation events when compared with observations from the 94-GHz CloudSat Cloud Profiling Radar. That study also showed that stage IV QPE performed best for areas and events like the ones in this study where frozen precipitation is not present and heavier rainfall occurs.

Simulations were run using WRF version 3.3.1 with the Advanced Research core of the WRF (ARW). The initial and lateral boundary conditions used 12-km horizontal grid spacing NAM output. The domain was roughly 800 km × 800 km centered over Iowa (Fig. 1). We also used convection-allowing 4-km horizontal grid spacing with 40 grid levels in the vertical and the Mellor–Yamada–Janjić (MYJ; Janjić 1994) planetary boundary layer (PBL) scheme and Thompson microphysics (Thompson et al. 2008). The simulations were integrated for 24 h and were initialized at the closest possible time to when NAM data were available and showers/thunderstorms associated with the heavy rain event either began to develop or entered the domain.

a. ARPS 3DVAR

We used the 3DVAR Analysis System developed as a part of ARPS along with a cloud analysis procedure that is a component of both ARPS 3DVAR and ADAS in this study to produce initial conditions (analysis) for the WRF Model simulations (hot-start runs). Essentially, nudging via ARPS 3DVAR was carried out for the velocity field in the simulations, while reanalysis through a cloud analysis procedure created three-dimensional cloud and precipitation fields. A single time assimilation step was used, allowing only one volume scan of data from each
radar. The ARPS data assimilation package was used instead of the WRF Data Assimilation (WRFDA) package because of familiarity with the package and its inclusion of analysis and assimilation methods. In addition, the ARPS cloud analysis procedure has proven it can effectively build storms in the initial conditions (e.g., Xue et al. 2003; Hu et al. 2006; Hu and Xue 2007a) and effectively initialize WRF forecasts (Hu and Xue 2007b).

The ARPS 3DVAR system performs multiple analysis iterations with different spatial scales in order to accurately represent the discontinuous or sporadic nature of convective storms (Gao et al. 2013). To carry out the data assimilation using the radial wind data, the ARPS 3DVAR uses an incremental form of a cost function that includes the background, observation, and equation constraint terms (Hu et al. 2006). The cloud analysis uses radar reflectivity data to construct three-dimensional cloud and precipitation fields. A latent heat adjustment to temperature based on added adiabatic liquid water content also occurs in order to make the in-cloud temperature consistent with the cloud fields (Hu et al. 2006). For this project, NAM analyses at the time of initialization (either 1800 or 0000 UTC depending on the case) were interpolated onto the ARPS grid and used as the background (or first guess) field for the 3DVAR data assimilation procedure for both wind and cloud analyses. To perform the data assimilation, the radar data (reflectivity and wind) are first interpolated onto the analysis grid using a local least squares procedure.

The level II radar data used for the radar data assimilation in this project came from nine NEXRAD WSR-88D sites that are located within the model’s domain (Fig. 1): Minneapolis, Minnesota (KMPX); Des Moines, Iowa (KDMX); Omaha, Nebraska (KOAX); Davenport, Iowa (KDVN); Sioux Falls, South Dakota (KFSD); Aberdeen, South Dakota (KABR); Lacrosse, Wisconsin (KARX); Kansas City, Missouri (KEAX); and St. Louis, Missouri (KILX). Two radar sites that are located near the boundary of the domain—Topeka, Kansas (KTWX) and Lincoln, Illinois (KILX)—were not used in the assimilation.

b. Hydrology model and flash flood forecasts

We used the CUENCAS hydrological model (Mantilla and Gupta 2005), which is a distributed rainfall–runoff hillslope model, in this study to forecast streamflow for rivers and streams within Iowa and in adjacent areas (Fig. 2). CUENCAS is a parsimonious model, which means it minimizes the computational resources needed for physically based models by capturing only the essential features in a watershed and uses as few parameters as possible to obtain acceptable results.

The model consists of a large number of river links (the portion of a river channel between two junctions of a river network) and hillslopes (adjacent areas that drain into the links), with each link and hillslope having a system of differential equations assigned to it in order to solve for water fluxes and storages (Mantilla and Gupta 2005). This rainfall–runoff model accounts for the routing of water through the river network’s channels, hillslope runoff generation due to surface ponding (i.e., rainfall rate overcoming infiltration rate), and soil water storage dynamics (Small et al. 2013). Figure 3 depicts a schematic diagram of the model’s components for which there is a corresponding system of differential equations that govern rainfall runoff, infiltration, and movement of infiltrated water:

\[
\frac{ds_u}{dt} = p(t) - q_{pc} - q_{pl} - e_p, \tag{1a}
\]

\[
\frac{ds_l}{dt} = q_{pl} - q_{ls} - e_l, \tag{1b}
\]

\[
\frac{ds_s}{dt} = q_{ls} - q_{sc} - e_s. \tag{1c}
\]
The variables that compose the system of differential equations represent the water storage in the hillslope surface $s_p$, top soil $s_u$, and deep soil $s_d$. Fluxes in, across, and out of the hillslope include precipitation $p(t)$, overland runoff $q_{pc} = k_3 s_p$, infiltration into the top soil $q_p = k_{DRY} (1 - (s_u/p_t))^3 s_p$, percolation from the top soil into the groundwater $q_h = k_2 s_u$, and baseflow into the channel $q_w = k_3 s_d$, and finally evaporation from the ponded, top soil, and deep soil layers ($e_p$, $e_u$, and $e_d$, respectively). The hillslope area $a_h$ for the elements in the distributed model is on average 0.052 km$^2$ and link lengths $l_i$ are on average 400 m. Note that $a_h/(2l_i)$ is the hillslope length. The CUENCAS software (Mantilla and Gupta 2005) was used to analyze a 30-m digital elevation model (DEM) and to decompose the river network into links and landscape into hillslope units. The model parameters are kept constant in time and in space, and they are set to $k_2 = 0.01 \times 2 \times l_i/a_h$, $k_{DRY} = 99k_2$, $k_1 = 0.02k_2$, and $k_3 = 0.003k_2$, as well as $T_i = 0.1$ m. We show in Fig. 4 an example of the model performance for a few locations in Iowa for simulated events in 2008. In this paper, the model is used in diagnostic mode, and therefore model parameters that fall in the range of values typically observed in Iowa are used (Tatard et al. 2008; FAO 1990). We conceptualize the hydrologic model as a filter of the rainfall fields, one that captures many of the aspects of rainfall fields that are relevant in flood prediction. The parameter values are not optimized to fit observed hydrographs in specific locations.

Water transport through the river network is nonlinear and governs how channel links propagate flows through the river network. It is formulated within the context of a mass conservation equation developed by Gupta and Waymire (1998), and it uses the parameterization given by Mantilla (2007) as

$$\frac{dq_c}{dt} = \frac{u_p q_{pc} A_1}{(1-A_2)l} \left[ a_h (k_2 s_p + k_3 s_d) - q_c + q_1 + q_2 \right], \quad (2)$$

where $q_c$ is the discharge from the link at time $t$, $a_h$ is the total hillslope area draining to the link, $q_1(t)$ and $q_2(t)$ are the incoming flows of the two upstream tributaries, $A$ is the upstream basin area, and $\lambda_1, \lambda_2,$ and $\nu_0$ are global parameters of the model and are set to 0.2, $-0.1$, and 0.3, respectively. (See schematic in Fig. 3.)

To provide a scale-independent reference value for each link in the river network, the hydrological model was run for an 11-yr period from 2002 to 2012 using stage IV rainfall as forcing. This simulation allowed us to calculate a model-based mean annual flood (MAF) value for each link in the river network. The MAF calculated for each link allows us to define a flood severity index (FSI), which is calculated as the ratio between streamflow and MAF. There is a significant body of literature relating MAF to bankfull streamflow (e.g., Dury 1961); therefore, values above 1 will typically indicate that water levels are outside the river banks. For Iowa, regionalization equations (Eash 2001) show that the 10-yr flood follows the relation $728DA^{0.465}$, where DA is the catchment area in square miles, while the mean annual flood (2-yr flood) follows the relation $182DA^{0.540}$. Dividing these two equations, one can see that the ratio of the 10-yr flood to MAF is at least 2 for the range of DA values considered in this study ($0.01$–$15,000$ mi$^2$). A simple interpretation of the flood severity index is that values below 1 indicate no flooding, a value from 2 to 3 indicates minor flooding, and values above 3 indicate major flooding.

Three hydrology model runs were implemented for the case dates: one run was forced with the stage IV rainfall (the benchmark), a second was forced with the hot-start forecasted rainfall, and a third was forced with the cold-start forecasted rainfall. The initial conditions

---

**Fig. 3.** Schematic diagram showing the hydrological model components. Three control volumes ($s_p$, $s_u$, and $s_d$) are defined for each hillslope to represent vertical flows into the soil and lateral flows to the channel link. A channel link is defined as the portion of the river reach in between two consecutive river network junctions. See text for definition of the flux terms ($q_{pc}$, $q_h$, and $q_w$).
at the beginning of each simulated event for the three hydrologic-model runs were the same. The initial conditions for each event (storage in hillslope surface, top soil layer, subsurface, and channel storage) were determined from a simulation of the hydrologic model that starts on 1 April of the year of the event using as input the stage IV rainfall. Note that the initial conditions for the simulations are spatially variable fields of the model-independent variables.

To determine the locations in the river network where the hydrology model would predict flash flooding when forced by the different products, the FSI was calculated for the peak streamflow at each link during the 24-h window when rainfall forecasts are given. The spatial distribution of the FSI allows us to map in space the extent of the domain where flooding (minor or major) occurs.

c. Analysis methods

For all cases, we visually compared the QPF (1 and 6 h) and postprocessed reflectivity of the hot-start and cold-start model simulations with corresponding stage IV rainfall analyses and NEXRAD reflectivity. It should be noted that the postprocessed reflectivity refers to a simulated composite radar reflectivity product computed from the forecast mixing ratios of grid-resolved hydrometeor species. We used equitable threat scores
(ETTs; Schaefer 1990) and bias to more objectively evaluate the forecast performance, where

\[
\text{ETT} = \frac{(\text{CFA} - \text{CHA})}{(F + O - \text{CFA} - \text{CHA})},
\]

(3)

\[
\text{CHA} = \frac{\text{F}}{\text{V}}, \quad \text{and}
\]

(4)

\[
\text{bias} = \frac{\text{F}}{\text{O}}.
\]

(5)

In the above equations, it is determined at each grid point whether (i) rainfall was correctly forecasted to exceed the specified threshold (CFA), (ii) rainfall was forecasted to exceed the threshold (F), (iii) rainfall was observed to exceed the threshold (O), and (iv) a correct forecast would occur by chance (CHA), where V is the total number of evaluated grid points. A perfect forecast would result in an ETT score of 1, with lower values indicating a less accurate forecast. Bias values greater (less) than 1 indicate that the model overpredicted (underpredicted) the areal coverage. For this study, it is advantageous to use ETT and bias since they provide a simple method to score and compare the QPFs for each case. However, it should be noted that ETT is known to penalize forecasts more as the horizontal grid interval decreases (Mass et al. 2002). Nonetheless, these indices are assumed adequate for this study as all simulations use the same horizontal grid spacing. In order for the stage IV data to be used in the calculation of ETT (and other measures), the data were regrided to match the WRF Model grid. Finally, it should be noted that high biases often help yield higher ETTs because displacement errors are common in rainfall forecasts (Hamill 1999).

We calculated the ETT and bias for five different 6-h rainfall thresholds and three different 1-h thresholds. The thresholds chosen for 6-h periods were 0.01 (0.254), 0.10 (2.54), 0.50 (12.7), 1.0 (25.4), 2.0 (50.8), and 3.0 in. (76.2 mm), while rainfall thresholds of 0.01 (0.254), 0.25 (6.35), and 0.50 in. (12.7 mm) were used for the 1-h periods. In cases where fewer than 10 grid points exceeded the rainfall threshold for the stage IV rainfall, the bias was not used in the case-average calculations because these cases can potentially produce extremely high bias values that significantly affect the full sample average (where each case was given equal weight).

To determine whether differences in ETTs between QPF-Cold and QPF-Hot were statistically significant, a bootstrap hypothesis test that uses a resampling methodology described in Hamill (1999) was performed on the 6-h period ETTs of the 0.254-, 12.7-, and 25.4-mm rainfall thresholds. The chosen bootstrap test is a good method for small sample sizes since many other statistical significance tests require the population of data to be normally distributed. Differences in ETTs statistically significant with 95% confidence will be noted in the discussion of results.

In addition to ETT and bias, the number of grid points with 6-h simulated rainfall exceeding the five thresholds (areal coverage) and the 6-h rainfall volume and rain rate for those points were also compared with stage IV observations for each case. The same was done for hourly periods using the thresholds applied for ETT and bias. The hourly statistics were analyzed to more precisely understand the QPF behavior.

In addition, an object-based verification method known as the Method for Object-Based Diagnostic Evaluation (MODE; Davis et al. 2006a,b) was applied. The MODE verification is objective (whereas visual verification is subjective) and provides more information on QPF skill than do ETTs. MODE identifies “objects” in the forecast and observed fields, describes them geometrically, and allows for attributes of the forecast and observed objects to be compared. These attributes include location, shape, orientation, and size. Furthermore, MODE computes properties of the objects such as intensity, area, centroid, axis angle, aspect ratio, and curvature for comparison between forecast and observed fields (Davis et al. 2006a).

Also, MODE can determine which areas in the forecast field have matching counterparts in the observations and provide error statistics on the match such as centroid distance, angle difference (agreement in orientation of system), area, intersecting region area, intensity, and more. An interest parameter [Total Interest (TI)] that is a weighted function of the other parameters is calculated and can be used to gauge the overall quality of the match between objects in the forecast and observed fields. The default weighting for calculation of TI is used in our study. The assigned weights are 4.0 to boundary distance (minimum distance between the boundaries of two objects), 2.0 each to centroid distance and intersection area, and 1.0 each to orientation angle difference and area ratio. The parameter TI quantifies the overall degree of accuracy between two objects with a fuzzy value between 0.0 and 1.0.

In this study, MODE is applied to QPF-Hot and QPF-Cold for the first two 6-h forecast periods of two representative cases examined in greater detail later. A threshold of 12.7 mm is chosen for the MODE analysis in this study; thus, only areas with rainfall at or exceeding 12.7 mm are identified as objects in the forecast and observed fields. It should be noted that very little smoothing was done on the precipitation fields for the mode verification with the convolution
radius set to 4 km. Since smaller-scale convective features are of interest to this study, it makes sense to limit smoothing. As a result many objects are identified by MODE in most of the forecasts and emphasis is given to the object cluster to determine how well the system or cluster was forecast.

3. Results

WRF simulations for the 12 warm season heavy rainfall events generally exhibited improved placement and areal coverage of the precipitating regions in the hot-start runs when compared with the cold-start runs. However, deficiencies were still present in the hot-start runs. First, the initial 6-h simulations were generally too wet. Second, in some of the cases the improvements decreased significantly from the first 6-h period to the second. When compared to the cold-start runs, postprocessed composite reflectivity from the hot-start runs showed that thunderstorm/rain areas created from the radar assimilation that were not present in the cold-start runs tended to dissipate too soon in several of the cases. A good example occurred in the 25 June 2005 event, where thunderstorms that were present over northern Iowa in the hot-start run shortly after the 0000 UTC time of initialization (0100 UTC) dissipated

Fig. 5. NEXRAD 2-km resolution of (left) base reflectivity for 25 Jun 2005 with (center) the hot-start run and (right) the cold-start run simulated composite reflectivities (4-km resolution) at (top) 0100, (middle) 0300, and (bottom) 0600 UTC.
rapidly by 0300 UTC (Fig. 5). The dissipation of the thunderstorms over northern Iowa in the hot-start simulation coincided with thunderstorms developing farther south in central Iowa in the vicinity of a surface cold front (not shown). In reality, thunderstorm activity on 25 June 2005 (Fig. 5) remained focused over northern Iowa at 0300 and 0600 UTC.

The impact of radar data assimilation varied substantially among events, with large improvements in the forecasts for 26 August 2009, but very limited impact for 23 July 2010.
The simulations for the 26 August 2009 and 23 July 2010 heavy rain events were chosen as examples to be emphasized in the following subsections, because they reflect the range of improvement from the cold-start to the hot-start runs, with most other cases falling between these two.

a. Case 1 (26 August 2009)

At 1200 UTC 26 August 2009, a weak (1014 mb; 1 mb = 1 hPa) surface low pressure system was over north-central Kansas with a stationary front extending northeastward from southwest to east-central Iowa (Fig. 6a). At 500 mb (Fig. 6e), a broad short-wave trough was approaching Iowa from the west, providing large-scale (synoptic scale) forcing for the rainfall event. Showers and thunderstorms over northern Kansas and southern Nebraska in the vicinity of the low pressure system moved east-northeastward into Iowa and northern Missouri through the morning and early afternoon hours. Thunderstorms developed ahead of the system in the vicinity of the frontal boundary as daytime surface heating increased instability and upper-level forcing moved in. A midlevel closed low pressure system (noted at the 700-mb level) developed north of the surface low over central Iowa during the day, and the circulation around the feature was evident on NEXRAD reflectivity loops. The surface and midlevel lows moved eastward throughout the day and into the overnight hours of 26 August and focused the rainfall in a swath from east-central Iowa southwestward along a trailing cold front through northwest Missouri and east-central Kansas during the overnight hours (Fig. 6b). The primary forcing mechanism throughout the event appeared to be the aforementioned upper-level short-wave trough as low-level winds in the vicinity of the frontal boundary were generally weak; thus, the low-level convergence near the frontal boundary was not very strong.

The 6-h QPF for the 26 August 2009 case (Fig. 7) was placed better and had better spatial coverage in the hot-start run for the first 6-h period (although the rainfall was too intense, especially over eastern Iowa), while the cold-start run was rather dry over west-central Iowa where the stage IV analysis had the heaviest rainfall amounts. The QPF-Hot simulation continued to outperform QPF-Cold throughout the last three 6-h periods as well, with rainfall placement, coverage, and intensity all being closer to the observed conditions.

The ETS and bias (Table 2) for the 26 August 2009 case agree with the subjective analysis of the 6-h rainfall
plots. The 26 August 2009 QPF-Hot simulations yielded higher ETSs than did QPF-Cold for all four 6-h periods for all rainfall thresholds (Table 2a), and the bias showed that the rainfall was overpredicted for all thresholds during the first 6-h period (Table 2b).

ETS, bias, and domain rainfall volume for the 0.254-mm rainfall threshold for the 26 August 2009 case are given in Fig. 8. The ETS for QPF-Hot was around 0.6 for the first hour, gradually fell to just under 0.3 by hour 5, and then subsequently stayed between 0.25 and 0.40. The QPF-Hot ETS remained higher than the ETS for the QPF-Cold runs throughout the 24-h simulations.

Bias (Fig. 8b) indicated an overprediction of areal coverage in the hot-start run through the first 8 h of the simulation. After that, the biases remained fairly constant between 1.2 and 0.7. Early on (especially for the first hour), the 1-h domain rainfall volumes (Fig. 8c) were too high in the hot-start runs. The heaviest observed rainfall occurred between model hours 9 and 16. The hot-start underforecasted the rainfall during this peak period, although the forecasted rainfall amounts were closer to the stage IV rainfall amounts during the peak rainfall period for this case than for the 23 July 2010 case to be discussed next.

MODE verification of the first two 6-h QPFs for 26 August 2009 agrees with the visual and point-to-point verification showing the hot-start run outperformed the cold-start simulations in most aspects. The location and shape of matched cluster object 2 (first 6-h forecast period) and matched cluster object 1 (second 6-h period) in QPF-Hot (Figs. 9a and 10a) during the first 6-h period more closely matches the corresponding cluster objects in the stage IV rainfall (Figs. 9b and 10b) for both 6-h periods than any of the matched cluster objects found in QPF-Cold (Figs. 9c and 10c). However, the intensity of the forecast rainfall in cluster pair 2 found in the stage IV and QPF-Hot simulations for the first 6-h forecast period was too high for QPF-Hot. This was especially true for the most intense rain areas as the average rainfall for areas with amounts exceeding the 90th percentile (INTP90) was 51.65 mm for QPF-Hot compared to 38.03 mm for the stage IV rainfall. Also the rain area (in grid boxes) shown in green was much too large and widespread for QPF-Hot (Fig. 9a) compared to stage IV (Fig. 9b). The QPF-Cold simulation had little rainfall over the 12.7-mm threshold and thus only a very small area of rainfall in southeast Iowa (Fig. 9c) exhibited a significant match with any cluster object in the stage IV rainfall for the first 6-h forecast period.

The MODE analysis for the second 6-h period indicated cluster pair 1 in the QPF-Hot and stage IV fields was a much stronger match than any of the cluster object matches between the QPF-Cold and stage IV fields for the same period with a TI of 0.9964. For comparison cluster pair 2 exhibited the strongest match between the QPF-Cold and stage IV fields with a TI of 0.8752. For cluster pair 1 (in the QPF-Hot and stage IV fields) INTP90 was lower for QPF-Hot (45.35 mm) than for stage IV (60.93 mm). Similarly, for cluster pair 2 (in the QPF-Cold and stage IV fields) INTP90 was lower for QPF-Cold (45.35 mm) than for stage IV (60.35 mm). Thus, the heaviest rainfall regions were less intense than observed for the aforementioned cluster objects in QPF-Hot and QPF-Cold. A secondary smaller but significant

### Table 2. (a) ETS and (b) bias values (for 6-h QPF) at six different rainfall thresholds (mm) for WRF simulations of the 26 Aug 2009 case for both hot- and cold-start runs initialized at 1800 UTC.

<table>
<thead>
<tr>
<th>(a) ETS</th>
<th>Period (h)</th>
<th>0.254</th>
<th>2.54</th>
<th>12.7</th>
<th>25.4</th>
<th>50.8</th>
<th>76.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold start</td>
<td>0–6</td>
<td>0.388</td>
<td>0.202</td>
<td>0.003</td>
<td>−0.001</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Hot start</td>
<td>0.586</td>
<td>0.469</td>
<td>0.141</td>
<td>0.036</td>
<td>0.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>6–12</td>
<td>0.339</td>
<td>0.219</td>
<td>0.05</td>
<td>0.004</td>
<td>−0.003</td>
<td>−0.001</td>
</tr>
<tr>
<td>Hot start</td>
<td>0.491</td>
<td>0.372</td>
<td>0.171</td>
<td>0.06</td>
<td>0.006</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>12–18</td>
<td>0.353</td>
<td>0.254</td>
<td>0.04</td>
<td>0.002</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Hot start</td>
<td>0.482</td>
<td>0.467</td>
<td>0.274</td>
<td>0.137</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>18–24</td>
<td>0.23</td>
<td>0.098</td>
<td>−0.005</td>
<td>−0.003</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Hot start</td>
<td>0.374</td>
<td>0.344</td>
<td>0.17</td>
<td>−0.002</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Bias</th>
<th>Period (h)</th>
<th>0.254</th>
<th>2.54</th>
<th>12.7</th>
<th>25.4</th>
<th>50.8</th>
<th>76.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold start</td>
<td>0–6</td>
<td>0.939</td>
<td>0.501</td>
<td>0.195</td>
<td>0.093</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td>Hot start</td>
<td>1.273</td>
<td>1.578</td>
<td>3.841</td>
<td>7.045</td>
<td>10.196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>6–12</td>
<td>0.812</td>
<td>0.706</td>
<td>0.707</td>
<td>0.552</td>
<td>0.43</td>
<td>0.264</td>
</tr>
<tr>
<td>Hot start</td>
<td>1.014</td>
<td>1.023</td>
<td>0.852</td>
<td>0.456</td>
<td>0.289</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>12–18</td>
<td>0.939</td>
<td>0.597</td>
<td>0.415</td>
<td>0.616</td>
<td>3.167</td>
<td></td>
</tr>
<tr>
<td>Hot start</td>
<td>0.934</td>
<td>0.686</td>
<td>0.521</td>
<td>0.785</td>
<td>6.278</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>18–24</td>
<td>1.644</td>
<td>0.94</td>
<td>0.483</td>
<td>0.389</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Hot start</td>
<td>1.49</td>
<td>0.923</td>
<td>0.518</td>
<td>0.332</td>
<td>0.048</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
matched cluster object (cluster pair 1 in red) was located over east-central Iowa and northwest Illinois (Figs. 10c,d) in QPF-Cold, and INTP90 was more intense than observed for this cluster.

In the first 6-h period, MODE found 12 individual objects in the QPF-Hot, 10 in the QPF-Cold, and 12 in the stage IV rainfall. In general for the object pairs (Table 3), many of the individual object matches were not very strong. In QPF-Cold only 2 out of 119 object pairs had an interest value exceeding the default 0.70 threshold (this threshold determines whether an object pair is considered a match) with 11 out of 144 object pairs exceeding the threshold in QPF-Hot. The maximum interests for object matches in QPF-Hot and QPF-Cold were 0.8734 and 0.8617 respectively. In the second 6-h period MODE found 11 objects in the QPF-Hot, 27 in the QPF-Cold, and 11 in the stage IV rainfall. Statistics indicate most of the matches were again poor with only 5 of 120 and 11 of 297 object pairs having a total interest exceeding the threshold. Maximum interest was 0.91 for QPF-Hot and 0.87 for QPF-Cold.

b. Case 2 (23 July 2010)

On 23 July 2010 (Fig. 11) little upper-level forcing was present over Iowa, with positive vorticity advection at 500 mb (Figs. 11e,f) primarily to the northwest and
north of the state ahead of a trough swinging through the northern plains. At 300 mb (not shown), the jet stream was well north of Iowa over northern North Dakota, Minnesota, and southern Canada. At the surface, a 1007-mb low pressure system was located over southeastern South Dakota at 1200 UTC 22 July (Fig. 11c). The surface low moved northeastward into southern Canada by 1200 UTC 23 July (Fig. 11d) and strengthened to 1002 mb, while a stationary boundary extending to the east of the low pressure over northern
Iowa remained stationary throughout the day and into the evening hours. The frontal boundary was a focal point for thunderstorm development throughout the evening and overnight hours of 23 July 2010 (Figs. 11a,b), as strong south-to-southwesterly flow across the southern and central plains and much of the state of Missouri resulted in ample heat and moisture transport into the region. The strong southerly flow was accompanied by strong low-level convergence and overrunning of the front, which forced the thunderstorm activity. Though upper-level forcing was weak over Iowa for this event, some forcing was present on the mesoscale and convective scales associated with convergence near the frontal boundary. As a result, the convection associated with the event was smaller in scale during the time of model initialization, with a thin line of

---

**Fig. 10.** As in Fig. 9, but for precipitation occurring during 0000–0600 UTC 27 Aug 2009.
convection extending from northeast Iowa into southern Wisconsin. That thin line of convection showed up well initially in the hot-start run but then dissipated rapidly. The 6-h QPF for this event (Fig. 12) was placed well for the first forecast period, with rainfall being in the same general location and of similar intensity for the hot-start and cold-start runs and stage IV analysis; however, the QPF for the next 6-h period was much less accurate. For this period, the hot-start run failed to produce much heavy rainfall (maximum around 38.5 mm), while the stage IV analysis showed heavy rainfall (101.6–127+ mm over northeast Iowa) occurring in an arching band across northern Iowa and into northern Illinois. The cold-start run may have performed slightly better in this second 6-h period because it had somewhat heavy rainfall in a band from northeast Iowa into northern Illinois.

The simulated reflectivity (Fig. 13) shows that, initially, the hot-start run accurately located the intense reflectivity echoes (0100 UTC), but the intense reflectivity echoes had moved too far south by 0600 UTC, while the cold-start had them positioned farther north more in line with NEXRAD reflectivity at the time. Neither the cold-start nor hot-start runs were able to accurately simulate the observed development of thunderstorms farther west (over western Iowa and eastern Nebraska) starting around 0600 UTC (Fig. 13). As a result, neither model simulation produced the widespread thunderstorm activity over Iowa that was shown by the NEXRAD reflectivity around 0900 UTC (Fig. 13), but the hot-start simulation at least showed activity over the eastern third of Iowa.

As expected, QPF-Hot yielded higher ETSs than did QPF-Cold for the first 6-h period but then had lower ETSs for the second 6-h period for the 0.5-in. (12.7 mm) and 1-in. (25.4 mm) rainfall thresholds (Table 4a). Bias values (Table 4b) indicated the hot-start run overforecasted the rainfall coverage for all thresholds during the first 6-h forecast period, underforecasted the coverage for all rainfall thresholds during the second and third 6-h periods, and then overforecasted the rainfall coverage for the final 6-h period for all thresholds. ETS, bias, and domain rainfall volume for the 0.254-mm rainfall threshold of 1-h QPF for 23 July 2010 are shown in Fig. 14. Forecast skill fell rapidly for this case with the ETS for QPF-Hot beginning around 0.55 for the first hour and then falling to around 0.3 by the third hour and then to around 0.15 by hour 7. After that time, it fell to nearly zero by hour 13, a time when little rain was observed or simulated. The QPF-Cold ETS exceeded the QPF-Hot ETS around simulation hour 5, when there was heavy rainfall occurring within the model domain (with around 0.4 mm of rainfall per grid point over the entire domain; Fig. 14c), which made the higher ETS of the cold-start more troubling. Bias values (Fig. 14b) indicated an overprediction of areal coverage in the hot-start run through the first 6 h of the simulation. After that, the biases dropped rapidly with an underprediction of areal coverage.

Early on (especially for the first hour), the 1-h domain rainfall volumes were too high in the hot-start run. The 1-h domain rainfall volume peaked between forecast hours 7 and 13 and neither the hot-start nor the cold-start simulations were able to reproduce the rainfall observed during this period. For example, at its heaviest (around model hour 10), the stage IV domain-averaged rainfall was around 0.71 mm per grid point, while the QPF-Hot value was around 0.14 mm and for QPF-Cold it was around 0.06 mm. Domain rainfall was also underforecast for the peak rainfall period in the 26 August 2009 case; however, the forecasted rainfall amounts during the peak rainfall period were closer to the stage IV rainfall amounts in that case.

MODE analysis indicated that both QPF-Hot and QPF-Cold were much more accurate for the first 6-h forecast period with cluster pair 1 (Fig. 15a,c; in red), each having an interest value of 0.9926 and 1.00, respectively. Angle difference (difference between the axis angle of an object in forecast versus observation) and intensity appear to have been the two largest discrepancies for the first 6-h period. The angle difference was 10.44° between cluster object 1 in the QPF-Hot versus stage IV rainfall and 11.87° comparing the cluster object 1 in QPF-Cold and stage IV. INTP90 was too high for cluster object 1 in QPF-Hot.
at 73.59 mm compared with 64.85 mm for the stage IV rainfall, and INTP90 for cluster object 1 in QPF-Cold was too low at 57.86 mm compared with 64.84 mm for stage IV.

MODE statistics for the second 6-h forecast decreased dramatically for both sets of QPFs with interest values for cluster pair 1 in QPF-Hot at 0.8778 and for cluster pair 1 in QPF-Cold a slightly higher value of 0.8850. The area of rainfall exceeding 12.7 mm for both QPF-Hot and QPF-Cold was far too small (Fig. 16). The QPF-Cold result was closer to the stage IV rainfall than that of QPF-Hot for many of the MODE statistics pertaining to cluster pair 1, although QPF-Hot was notably closer in angle difference and centroid.

FIG. 11. As in Fig. 6, but for (a),(c),(e) 1200 UTC 22 Jul and (b),(d),(f) 1200 UTC 23 Jul 2010.
distance. It appears that the object (in red) shown in QPF-Cold (Fig. 16c) is somewhat similar to the eastern portion of the object (in red) found in the stage IV rainfall (Fig. 16d) in terms of orientation and rainfall intensity. The orientation of the observed rainfall area in that region is west-northwest to east-southeast as opposed to west-southwest to east-northeast in stage IV cluster object 1.

Many of the individual object matches for the 23 July 2010 case were poor, like in the 26 August 2009 case. However, during the first 6-h period there were strong individual object matches found in both the QPF-Hot and QPF-Cold simulations as indicated by higher interest values (Table 3). The match between object 1 in the QPF-Hot and object 1 in the stage IV rainfall had an interest of 0.9927, and the match between object 1 in the QPF-Cold and object 1 in the stage IV rainfall had an interest of 1.00. None of the other object matches in QPF-Hot or QPF-Cold had an interest above 0.88, although 4 of 12 and 3 of 12 object pairs exceeded the 0.70 interest threshold in QPF-Hot and QPF-Cold, respectively. During the second 6-h period many of the individual object matches were poor and none of the object matches (in either QPF-Hot or QPF-Cold) had an interest above 0.89.

c. Summary of all cases

Average (Table 5a) and case-specific ETSs (Tables 2a and 4a) for WRF 6-h QPFs showed a noticeable improvement in forecast accuracy in QPF-Hot compared to QPF-Cold during the first two 6-h periods. This was especially true for rainfall thresholds up to 1 in. (25.4 mm); for heavier thresholds, the ETSs were generally low and little if any increase in score occurred in the hot-start runs. The improvement in average ETS and the magnitudes of the ETSs themselves decreased during the third and fourth 6-h periods. One would expect the forecast accuracy to decrease since model error grows with time (Sugimoto et al. 2009).

Case-average ETS could be influenced by one or two cases that had a large jump in ETS between QPF-Cold and QPF-Hot; however, the increase in ETS was found to be widespread between the cases for the first two 6-h periods. QPF-Hot had a higher ETS than QPF-Cold for the 0.254-mm rainfall threshold (entire precipitating region) in all 12 simulated cases for the first 6-h period and in 11 of the 12 cases for the second 6-h period. In addition, the
bootstrap test yielded a $P$ value of 0 for the first 6-h forecast period and 0.019 for the second 6-h period, indicating that for both periods the increase in ETS for QPF-Hot compared to QPF-Cold was statistically significant.

Isolating the heavier rain areas (threshold of 12.7 mm) QPF-Hot had a higher ETS in 11 of 12 cases for the first 6-h forecast period and in 10 of 12 cases for the second 6-h period. When applied to this threshold, the bootstrap method gave a $P$ value of 0.001 for the first 6-h forecast period and 0.013 for the second period, again indicating statistically significant improvement of ETSs for the hot start. For the 25.4-mm rainfall threshold, the increase in ETS for QPF-Hot was statistically significant only during the first 6-h forecast period.

Case-average bias values (Table 5b) generally indicated the hot-start runs overforecasted the rainfall coverage for all thresholds (albeit just slightly for the lower thresholds) during the first 6-h forecast period, underforecasted the rainfall coverage for all rainfall thresholds except for the 2-in. (50.8 mm) and 3-in. (76.2 mm) thresholds during the second 6-h period, and then overforecasted the rainfall coverage for the final two 6-h periods for all thresholds.

Regarding case-average 6-h rainfall characteristics (Fig. 17), QPF-Hot generally overpredicted areal coverage, especially for the higher rainfall thresholds (>25.4 mm) and during the first 6-h model period. Also, QPF-Hot was too intense for the higher thresholds,
with the system rain rates generally exceeding the observed system rain rates, especially during that first 6-h period. The hot-start runs performed best for rainfall characteristics during the second 6-h period in the simulations, as average areal coverage, rain rate, and rain volume tended to be closer to that observed for this period when compared with others. This agrees with the finding that QPF-Hot in the first 6-h period was too wet for most of the cases, and QPF (both QPF-Hot and QPF-Cold) amounts in the second 6-h period were more in line with observed rainfall amounts.

To better pinpoint the times at which the model runs had performance difficulties, ETSs (Fig. 18) and biases (Fig. 19) for three rainfall thresholds [0.01 in. (0.254 mm), 0.25 in. (6.35 mm), and 0.50 in. (12.7 mm)] were calculated using 1-h QPF-Hot and QPF-Cold. Rainfall volume, intensity, and areal coverage were also calculated using the 1-h rainfall for the three rainfall thresholds. ETSs for all three thresholds and for all cases (not shown) indicated little accuracy for QPF-Cold during the first few hours of the simulations, while QPF-Hot showed a significant increase in accuracy during those first few hours. This appears to be largely attributable to spinup issues in the cold-start runs as they are far too dry in the first two to three simulation hours (Fig. 20) with absolute errors of around −0.44 (0.254-mm threshold), −0.28 (6.35-mm threshold), and −0.19 (12.7-mm threshold) mm per grid point for the first hour in the cold-start runs. The hot-start runs started out with fairly high ETSs in the first one or two simulation hours and then decreased in the 3–6-h range before rising again and achieving a second peak in the 7–12-h range. The decrease in ETSs after the first hour or two would seem to support the findings from the postprocessed composite radar plots (shown in Figs. 4 and 12 for two cases), which indicated that thunderstorms/rain areas produced from the radar assimilation and that are not present in the cold-start simulations tend to dissipate rapidly.

The average domain rainfall volume (Fig. 20) is far too wet in the first hour of the hot-start simulations with average absolute errors of around 0.5 (0.254- and 6.35-mm thresholds) and 0.38 (12.7-mm threshold) mm per grid point of rainfall. After the first hour, the case-average absolute errors decreased substantially to around 0.12 mm per grid point (0.254- and 6.35-mm thresholds) and 0.10 mm per grid point (12.7-mm threshold) in the second hour of the hot-start runs. Therefore, the overly intense rainfall in the first hour of the runs appears to be the greatest contributor to the overly wet conditions that were noted in the 6-h performance measurements.

Case-average areal coverage (Fig. 21) for QPF-Hot was much closer to the stage IV analysis during the first few hours of the simulations than QPF-Cold (especially for the 0.254-mm rainfall threshold) as a result of the cold-start runs having very little rainfall coverage early on. The hot-start runs begin with too large of a precipitating region (0.254-mm threshold), but they approached that of the stage IV rainfall analysis over the first few hours and then became too small from hour 4 until around hour 12. This period also coincides fairly well with the stage IV analysis becoming slightly wetter than QPF-Hot, as shown by the 1-h domain rainfall volumes for the 0.254-mm rainfall threshold (Fig. 20). However, the 1-h stage IV rainfall volume does not exceed the hot-start result until around hour 7.

<table>
<thead>
<tr>
<th>Period (h)</th>
<th>0.254</th>
<th>2.54</th>
<th>12.7</th>
<th>25.4</th>
<th>50.8</th>
<th>76.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) ETS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold start</td>
<td>0–6</td>
<td>0.306</td>
<td>0.384</td>
<td>0.387</td>
<td>0.21</td>
<td>0.103</td>
</tr>
<tr>
<td>Hot start</td>
<td>0–6</td>
<td>0.314</td>
<td>0.469</td>
<td>0.503</td>
<td>0.384</td>
<td>0.141</td>
</tr>
<tr>
<td>Cold start</td>
<td>6–12</td>
<td>0.166</td>
<td>0.147</td>
<td>0.171</td>
<td>0.134</td>
<td>0.029</td>
</tr>
<tr>
<td>Hot start</td>
<td>6–12</td>
<td>0.245</td>
<td>0.212</td>
<td>0.119</td>
<td>0.046</td>
<td>0</td>
</tr>
<tr>
<td>Cold start</td>
<td>12–18</td>
<td>0.162</td>
<td>0.086</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hot start</td>
<td>12–18</td>
<td>0.03</td>
<td>−0.015</td>
<td>−0.002</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cold start</td>
<td>18–24</td>
<td>−0.01</td>
<td>−0.019</td>
<td>−0.008</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hot start</td>
<td>18–24</td>
<td>−0.014</td>
<td>−0.023</td>
<td>−0.008</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| (b) Bias  |       |      |      |      |      |      |
| Cold start | 0–6   | 1.493| 1.275| 1.015| 0.817| 0.482| 0.172 |
| Hot start  | 0–6   | 1.513| 1.214| 1.349| 1.442| 1.236| 1.931 |
| Cold start | 6–12  | 0.435| 0.252| 0.266| 0.266| 0.215| 0.031 |
| Hot start  | 6–12  | 0.668| 0.539| 0.273| 0.08 | 0      | 0      |
| Cold start | 12–18 | 0.56 | 0.397| 0.174| 0     | 0      | 0      |
| Hot start  | 12–18 | 0.433| 0.221| 0.06 | 0     | 0      | 0      |
| Cold start | 18–24 | 2.124| 2.423| 7.783| 72.571|      |      |
| Hot start  | 18–24 | 2.303| 2.777| 6.788| 54.714|      |      |
d. Hydrology model flash flood predictions

The previous evaluation of QPF accuracy shows that there is improvement in the hot-start runs over the cold start. However, the hot-start forecasts were far from perfect and still exhibited errors in timing, placement, and intensity of rainfall, all of which would be better sensed by a hydrology model. Hydrologic model–predicted flash flooding for all 12 cases (Fig. 22 shows the FSI for three cases) based on QPF-Hot was fairly accurate in terms of general placement and intensity for a few of the better-forecasted cases, while placement, intensity, and timing discrepancies existed for many of the cases when compared with the flash flooding produced via the stage IV rainfall. The QPF-Cold simulation failed to produce substantial flash flooding in multiple cases where flash flooding was produced using the stage IV rainfall and also under-predicted the coverage and intensity of flash flooding in nearly all of the other cases.

The case of 26 August 2009 is one in which substantial improvement in the hydrology model–predicted flash flooding occurred using QPF-Hot. In this event, the hot-start run is much more accurate than the cold start in predicting the placement and areal coverage of QPF, especially during the first 18 h of the simulation (Fig. 6). For this case, the most significant problem

---

**Fig. 14.** As in Fig. 8, but for the 23 Jul 2010 case.
with the hot-start run was rainfall being too intense during the first hour or so of the simulation. When comparing the degrees of flash flood prediction accuracy for this event that were achieved through hydrology model runs forced with hot-start, cold-start, and stage IV rainfall (Fig. 21a), one can see that QPF-Hot would result in mostly minor flash flooding (FSI < 3) over the same general area of east-central Iowa as the stage IV rainfall, while QPF-Cold yielded no significant flash flooding. However, even though the hydrologic model run forced with QPF-Hot does well with the general placement of flash flooding (FSI) for this event, there are smaller-scale discrepancies in intensity within the general area of flash flooding.
On the other end of the performance spectrum are cases like 5 June 2008 (Fig. 21b), where QPF-Hot significantly overpredicted flash flooding in terms of coverage and intensity. Large areas of west-central Iowa would be predicted to experience major flash flooding (FSI > 3) with QPF-Hot, while the stage IV rainfall would produce only minor flooding (FSI between 1 and 3). QPF-Cold does not result in as large of an area of overly intense flooding as QPF-Hot, but it does result in flash flooding that is too intense in northeast Nebraska and fails to
predict minor flash flooding farther east into west-central and central Iowa that the stage IV rainfall is able to predict. Unlike in the 5 June 2008 case, the 23 July 2010 case (Fig. 21c) had stage IV rainfall that resulted in the prediction of major flash flooding over southern Fayette County in northeast Iowa, while QPF-Hot failed to result in the prediction of any major flash flooding and yielded only scattered areas of minor flash flooding. QPF-Cold failed to produce any significant flash flooding for this case when input into the hydrology model.

Despite obvious inconsistencies in performance in terms of intensity, location, and coverage of flash flooding predicted using QPF-Hot, it did correctly yield at least some flash flooding in each of the 12 cases. By contrast, QPF-Cold failed to predict any flash flooding in some of the cases, whereas flash flooding was predicted using the stage IV rainfall analysis in all cases. Although the location, intensity, and exact timing of the flash flooding may not be predictable from QPF-Hot, the fact that some flash flooding is predicted nearby may be of value to forecasters.

4. Discussion

Verification of model forecasts during the Hazardous Weather Testbed Spring Experiment in 2008 and 2009 (SE2008 and SE2009) showed the use of radar data assimilation provided greater improvement to larger-scale systems than to smaller-scale convective elements, with the predictability time scales for individual convective cells introduced through the assimilation procedure being rather short (Kain et al. 2010). It is also intuitive that NWP models should better forecast strongly forced large-scale events than small-scale weakly forced events. This results from the fact that large-scale forcing features are present in model analyses that are used to create model initial conditions, while smaller-scale forcing features may be absent. Xue et al. (2008) showed that a strongly forced convective event was well forecasted in all model simulations (including one that did not use radar data assimilation), while a weakly forced case was poorly forecasted by the model that did not use radar data. All model runs that used radar data gave a significantly better forecast for the weakly forced case, which demonstrated that radar data assimilation had a greater impact on the weakly forced event as compared with the strongly forced convective system. The Xue et al. (2008) study concluded that whether the convective system is controlled by large-scale features or by smaller-scale internal dynamics is the main factor in determining the effectiveness of the data assimilation on the model forecast.

To better understand the variability in improvement that was achieved as a result of the radar data assimilation for the present study, we conducted a synoptic analysis for each of the heavy rainfall events. In this section we will compare the 26 August 2009 and 23 July 2010 cases in detail and discuss the other events in a more general sense.

The large-scale nature of the 26 August 2009 event with broad upper-level forcing and corresponding large precipitating region would fit with cases in Kain et al. (2010) that were found to benefit more from radar data assimilation than small-scale convective elements. The 23 July 2010 event would fit more with the cases that had small convective elements in the Kain et al. (2010) study.
may explain to an extent why more improvement occurred in QPF-Hot for the 26 August 2009 case than the 23 July 2010 case. It should be noted, however, that the 23 July 2010 case is not particularly similar to the weakly forced case presented in Xue et al. (2008). Although the event has weak upper-level forcing, strong external forcing is present in the low levels near the frontal boundary helping to create the convection.

It is clear through the model data and skill scores, along with visual examination of QPF and simulated radar plots, that the cold-start and hot-start forecasts begin to look more alike as the runs progress through time. This was also noted in the Kain et al. (2010) study, where they suggest that this is due to the fact that both runs share similar forecasts of forcing mechanisms in the mesoscale environment. Thus, after the predictability of the small-scale convective features introduced by the radar data is lost, the evolution of the convective activity is strongly controlled by the mesoscale forcing mechanisms (Kain et al. 2010).
Analysis of all of the cases suggests that those that experienced the greatest impact from radar data assimilation were the ones in which the forcing was not overly strong and a large area of precipitation was present at the time of model initialization. In general, cases in which only small convective elements were present at the time of initialization tended to experience less of an impact from the radar data assimilation, which could be due to the short predictability time scale for the small convective elements (Kain et al. 2010) in conjunction with less “important” data being assimilated than with a larger convective system.

5. Conclusions

The present study examined the ability of radar data to improve high-resolution model QPFs for heavy rainfall events that often result in significant flash flooding. It also examined whether the improvements would be substantial enough to improve a hydrology model’s flash flood forecasts to a point where they might be used as a flood forecasting tool. In general, QPF-Hot had a higher level of precipitation forecasting accuracy than QPF-Cold, especially during the first 12 h of a simulation. The greatest increase in accuracy was noted for the lower rainfall thresholds, but thresholds up to 25.4 mm also saw an increase in accuracy. However, QPF-Hot still exhibited errors in placement, coverage, and timing. Also, the QPF-Hot runs were too wet in the first hour or two, and thunderstorms/rain areas created from the radar data assimilation and not present in the cold-start runs tended to dissipate too quickly.

The streamlined data assimilation strategy (single time step) and relatively coarse model grid (4 km) likely exacerbated the small-scale convective elements’ tendency to dissipate too soon and the problem of the runs...
being too wet over the first hour of integration (Kain et al. 2010). There appears to be some dynamic adjustment occurring early on in the model runs, which may also partly account for the high precipitation bias. Sun et al. (2014) noted that radar data assimilation via 3DVAR methods can lead to dynamically inconsistent initial conditions and, consequently, a dynamic adjustment. Sun and Wang (2013) indicated that 3DVAR schemes can have problems analyzing the low-level cold pool, its leading edge convergence, and midlevel latent heating. Their study also found that four-dimensional variational data assimilation (4DVAR), which uses data at more than one time step and employs strong dynamic constraints, results in improved analysis of cold pools, a better radial velocity assimilation, and reduction of issues with dynamic adjustment, and produces a better forecast than 3DVAR.

A flood severity index was devised to determine where a hydrology model would predict flash flooding when forced with rainfall from stage IV, QPF-Hot, and QPF-Cold. The QPF-Hot and the stage IV rainfall yielded flash flooding in each of the 12 cases, whereas QPF-Cold failed to result in any substantial flash flooding for several of the cases and underpredicted flash flooding in the others. Even though the hydrologic
model predicted flash flooding in each of the cases when forced with QPF-Hot, discrepancies in placement, coverage, and intensity of the flooding existed when compared with runs using stage IV rainfall. In a few cases, such as 26 August 2009, the general location and intensity of flash flooding were predicted fairly accurately, but in others like 5 June 2008 large discrepancies existed in placement and intensity of the flash flooding.

Overall, the use of radar data yielded a noticeable, albeit varying degree of, improvement in the high-resolution model rainfall simulations for the 12 cases. This improvement also carried over into the hydrology model’s flash flood prediction, where the improvement in the predictions once again varied case by case for QPF-Hot. One consistency between the hydrology model runs using QPF-Hot as input is that flash flooding was predicted during the 24-h simulations in all 12 cases, which is also true when the stage IV rainfall was used. As mentioned earlier, this is an encouraging result, although it must be noted that testing should be conducted on rainfall events that had no flooding associated with them to see if the radar assimilation correctly produced no flooding. Even with this encouraging result, some important complications with the hot-start runs need to be studied further. For

Fig. 21. As in Fig. 18, but for case-averaged areal rainfall coverage (grid points in thousands) and including stage IV data. Note different scale for 0.254-mm threshold.
instance, slight changes may need to be made to some tunable parameters in the ARPS 3DVAR radar data assimilation program in order to alleviate problems with rainfall intensity early on. Another option would be to use 4DVAR instead of 3DVAR to assimilate the data, although this would require significantly more computer resources. Finally, the assimilation of radar data at more than one time, along with the assimilation of meteorological data from other sources, may be necessary to facilitate further improvement in these high-resolution WRF Model runs used for rainfall prediction.

Acknowledgments. This work was supported by the Iowa Flood Center. We thank Dave Flory at Iowa State University for his assistance in running the WRF system and for sharing his scripts to extract the rainfall data from the WRF output. We also thank Kevin Thomas from CAPS at the University of Oklahoma for his help in implementing the necessary ARPS programs for this project. Finally, we thank Scott Small and Radek Goska at the Iowa Flood Center for their help in setting up the hydrological model runs, and Prof. Witold Krajewski for his support and constructive comments that helped improve this manuscript.
REFERENCES


