

---

01 Aug 2018

## High-resolution QPF Uncertainty And Its Implications For Flood Prediction: A Case Study For The Eastern Iowa Flood Of 2016

Bong Chul Seo

Missouri University of Science and Technology, bongchul.seo@mst.edu

Felipe Quintero

Witold F. Krajewski

Follow this and additional works at: [https://scholarsmine.mst.edu/civarc\\_enveng\\_facwork](https://scholarsmine.mst.edu/civarc_enveng_facwork)



Part of the [Civil and Environmental Engineering Commons](#)

---

### Recommended Citation

B. C. Seo et al., "High-resolution QPF Uncertainty And Its Implications For Flood Prediction: A Case Study For The Eastern Iowa Flood Of 2016," *Journal of Hydrometeorology*, vol. 19, no. 8, pp. 1289 - 1304, American Meteorological Society, Aug 2018.

The definitive version is available at <https://doi.org/10.1175/JHM-D-18-0046.1>

This Article - Journal is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Civil, Architectural and Environmental Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

# High-Resolution QPF Uncertainty and Its Implications for Flood Prediction: A Case Study for the Eastern Iowa Flood of 2016

BONG-CHUL SEO, FELIPE QUINTERO, AND WITOLD F. KRAJEWSKI

*IHR—Hydroscience and Engineering, University of Iowa, Iowa City, Iowa*

(Manuscript received 6 March 2018, in final form 25 June 2018)

## ABSTRACT

This study addresses the uncertainty of High-Resolution Rapid Refresh (HRRR) quantitative precipitation forecasts (QPFs), which were recently appended to the operational hydrologic forecasting framework. In this study, we examine the uncertainty features of HRRR QPFs for an Iowa flooding event that occurred in September 2016. Our evaluation of HRRR QPFs is based on the conventional approach of QPF verification and the analysis of mean areal precipitation (MAP) with respect to forecast lead time. The QPF verification results show that the precipitation forecast skill of HRRR significantly drops during short lead times and then gradually decreases for further lead times. The MAP analysis also demonstrates that the QPF error sharply increases during short lead times and starts decreasing slightly beyond 4-h lead time. We found that the variability of QPF error measured in terms of MAP decreases as basin scale and lead time become larger and longer, respectively. The effects of QPF uncertainty on hydrologic prediction are quantified through the hillslope-link model (HLM) simulations using hydrologic performance metrics (e.g., Kling–Gupta efficiency). The simulation results agree to some degree with those from the MAP analysis, finding that the performance achieved from the QPF forcing decreases during 1–3-h lead times and starts increasing with 4–6-h lead times. The best performance acquired at the 1-h lead time does not seem acceptable because of the large overestimation of the flood peak, along with an erroneous early peak that is not observed in streamflow observations. This study provides further evidence that HRRR contains a well-known weakness at short lead times, and the QPF uncertainty (e.g., bias) described as a function of forecast lead times should be corrected before its use in hydrologic prediction.

## 1. Introduction

Weather-related natural disasters have become more frequent and extreme, likely because of climate change and its accompanying effects (e.g., Meehl et al. 2000; Milly et al. 2002; Van Aalst 2006; Lehmann et al. 2015). These disasters have substantially endangered public safety and weakened community resilience in recent years. To prevent and manage potential threats from the increasing risks and vulnerability to extreme weather and water events, accurate hydrologic prediction is a critical factor in mitigating impacts (i.e., human life and economic losses) of the extreme events. However, operational hydrologic forecasting is challenging because prediction skill is limited by the difficulties in describing interactions of complex human-nature systems due to imperfect models and numerous uncertainties in model parameters and input data (e.g., Li et al. 2009; Pagano et al. 2014).

During the last several years, the Iowa Flood Center (IFC) has been developing and operating a fully automated flood forecasting system for the entire state of Iowa (Krajewski et al. 2017). As part of the system's development, the IFC has deployed a distributed hydrologic model capable of representing the physical processes of transforming rainfall to runoff. The model is called the hillslope-link model (HLM); it uses a decomposition of landscape into channels and hillslopes with an average element size of 0.1 km<sup>2</sup>. In the HLM, the soil properties and land cover are used to determine the split between runoff and infiltration. For evapotranspiration, the HLM uses climatologic estimates using 12 years of North American Land Data Assimilation System data (Mitchell et al. 2004). The HLM is mainly driven by the IFC's radar-derived quantitative precipitation estimates (QPEs), generated in real time with space and time resolutions of 0.5 km and 5 min (see, e.g., Seo and Krajewski 2015; Seo et al. 2015). For comparison, we also use the Multi-Radar Multi-Sensor (MRMS) QPE (Zhang et al. 2016) with 1-km and 1-h resolutions

---

Corresponding author: Bong-Chul Seo, bongchul-seo@uiowa.edu

DOI: 10.1175/JHM-D-18-0046.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

to drive the HLM. Likewise, in 2016, the U.S. National Water Center implemented and continues to test operationally the National Water Model (NWM). The NWM is also a distributed hydrologic model that simulates and forecasts streamflow over the entire United States (e.g., Maidment 2017; Lin et al. 2018). It contains land surface model components that require meteorological forcing data (e.g., incoming shortwave and longwave radiation, humidity, temperature, pressure, wind speed, and precipitation) to simulate terrestrial hydrologic processes. The precipitation data used for the NWM varies with different forecast model cycles (e.g., MRMS for “analysis and assimilation”). The HLM and NWM share some similarities, mainly that both are based on terrain data (see Quintero and Krajewski 2018). The operation of the HLM and NWM reveals the increasing demand for high-resolution hydrologic modeling and forecasting. These systems enable researchers to describe more detailed aspects of the interactions between atmosphere and land surface that have not been explored by conventional approaches (e.g., lumped and mesoscale models). This distributed modeling effort can complement current hydrologic guidance at National Weather Service (NWS) forecast locations and expand forecast capabilities and guidance coverage in underserved locations (Cosgrove et al. 2015, 2016).

The IFC recently added quantitative precipitation forecasts (QPFs) generated by the High-Resolution Rapid Refresh (HRRR) model into the HLM forcing stream to extend streamflow forecast lead times (e.g., Arduino et al. 2005; Li et al. 2017). The HRRR model provides precipitation forecasts up to 18-h lead times with space and time resolutions of 3 km and 1 h, which meet the requirements of the IFC’s high-resolution distributed hydrologic modeling. The NWM operational configuration of “analysis and assimilation” and “short range” model cycles (<http://water.noaa.gov/about/nwm>) also uses the HRRR QPF product. While the QPE products in the HLM and NWM forcing stream have often been evaluated in terms of precipitation accuracy as well as its impact on flood prediction (e.g., Chen et al. 2013; Seo et al. 2013; Gochis 2016; Krajewski et al. 2017), the utility of HRRR for flood prediction has not been widely examined. The uncertainty structure of HRRR QPF is still unknown (even if the uncertainty in the QPF is generally larger than that of QPE), and its potential effect on hydrologic prediction is not yet well understood. In this study, therefore, we aim to improve our understanding by evaluating the precipitation forecast skills of HRRR and the potential effects of its uncertainty on flood prediction. We used a significant Iowa flooding event that occurred in September 2016 to analyze and evaluate

the HRRR model predictions. This case study allowed us to develop a framework to validate numerical weather predictions that are closely related to hydrologic processes (e.g., precipitation).

This paper is structured as follows. In section 2, we provide information on the flooding event, as well as precipitation (QPE and QPF) and streamflow data used in this study. We also provide a brief description of the HLM in the section. Section 3 describes the analysis framework and evaluation metrics we used for the HRRR QPF assessment and error characterization in this study. In section 4, we present the evaluation results on the HRRR QPF error and its propagation through rainfall–runoff processes. Section 5 summarizes and discusses our main findings and limitations of this study, as well as required future work.

## 2. Flooding event, data, and model

In September 2016, extremely heavy rainfall caused significant river and flash flooding in northeastern and central Iowa (see Fig. 1 for daily precipitation analysis). A tropical air mass interacting with a stationary front triggered several rounds of heavy storms in these areas. During the period of 20–23 September, rainfall totals reported for the region were 80–200 mm on average, and for some localities more than 250 mm. A number of towns near major rivers (e.g., Cedar and Wapsipinicon Rivers) and their tributaries in northeastern and central Iowa experienced significant flooding, causing the evacuation of thousands of residents. The river stages on the major rivers were comparable to those of the devastating Iowa flood of 2008 (e.g., Mutel 2010; Smith et al. 2013). Below, we provide more details on the meteorological and hydrologic aspects of this event, as well as collected precipitation and streamflow datasets used to analyze the uncertainty of HRRR QPF for the selected flooding event.

### a. Flooding event summary

On 21 September, a stationary frontal system produced multiple rounds of showers and thunderstorms in northeastern Iowa near Mason City, as shown in Fig. 1a. These thunderstorms developed a supercell that brought significant amounts of rain over the Shell Rock River. The supercell stayed over the river for more than 3 h and produced 200–250 mm of rain, resulting in flash flooding in Floyd County. On 22 September, the frontal system moved southwesterly over central Iowa and delivered 80–120 mm of rain to the Ames area, as shown in Fig. 1b. The rain then spread and moved to northeast Iowa again, delivering 80–150 mm of heavy rain over the same area that the supercell had hit the night before.

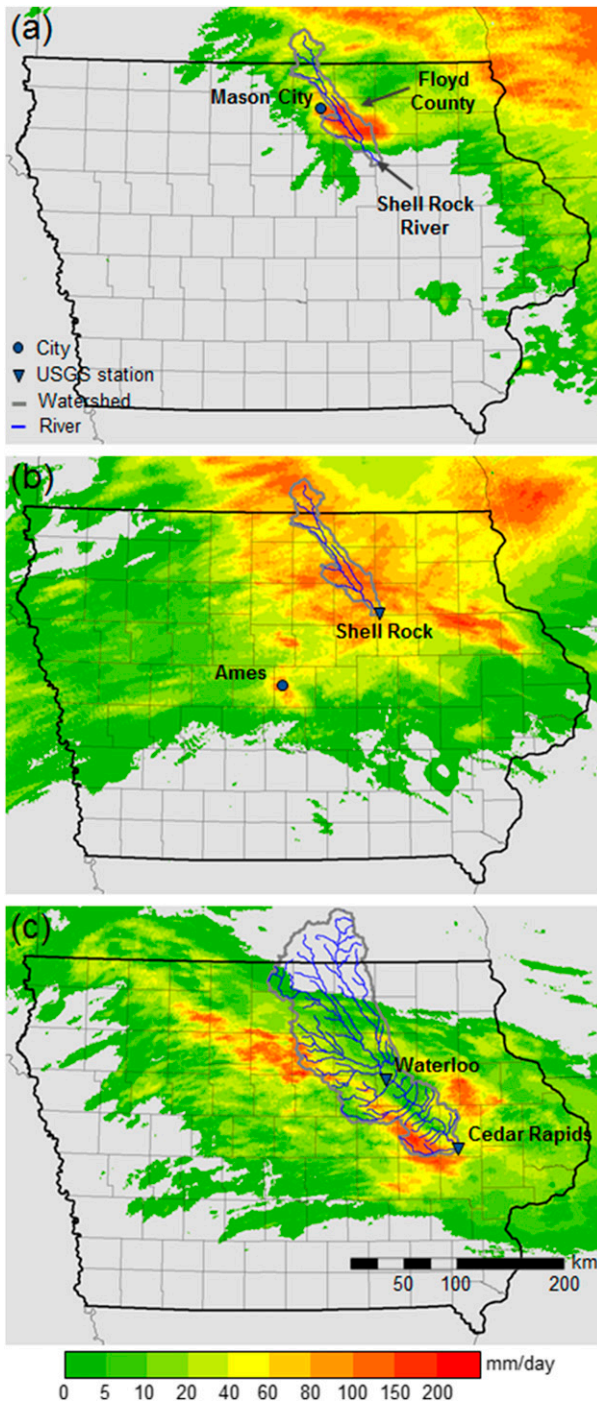


FIG. 1. Daily rainfall accumulation for (a) 21, (b) 22, and (c) 23 Sep 2016. The daily analysis maps were created using MRMS QPE with rain gauge correction.

This resulted in new flash flooding and record water levels on the Shell Rock River at Shell Rock. On 23 September, another round of widespread showers and heavy storms generated by the frontal system delivered more than

130 mm of rain near Cedar Rapids, as shown in Fig. 1c. This heavy rain resulted in the second-highest crest ever recorded on the Cedar River at Cedar Rapids on 30 September. The flooding during this period was the result of intense and slow-moving storms that extensively covered the areas shown in Fig. 1, along with abnormally wet soil conditions during August 2016.

#### b. Precipitation and streamflow data

HRRR is a real-time convection-allowing atmospheric model updated hourly and initialized by 3-km grids with radar assimilation (Benjamin et al. 2009; Alexander et al. 2010). HRRR is fully dependent on its parent models, the radar-assimilating Rapid Refresh and radar-enhanced Rapid Update Cycle. The HRRR model provides hourly 3-km QPF for up to 18 h (while it also creates an experimental subhourly product, we used the hourly product in this study). We collected the HRRR QPF product for the 6-month period of May–October 2016 that contains the significant flooding events. We used the entire 6-month product for the conventional verification of HRRR precipitation forecast skills and examined the product for the flooding period to evaluate its impact on streamflow prediction.

MRMS integrates base radar data across the conterminous United States with satellite, lightning, and rain gauge observations, as well as atmospheric environmental data (Zhang et al. 2016). MRMS generates a suite of weather and QPE products (e.g., rainfall rate, accumulation, and precipitation type) with enhanced time resolutions ranging from 2 min to 1 day, with spatial resolution of approximately 1 km. Because of its national coverage and high resolution, MRMS products are extensively used to strengthen severe weather warnings and forecasting. In particular, MRMS QPE is not only fed into the HLM and NWM, but is also used for other operational hydrologic and weather models for improved flash flood and weather forecasting (e.g., Gourley et al. 2017). In this study, we used the rain gauge-corrected MRMS QPE (hereafter, MRMS-GC) as a reference to assess the forecast skills and performance of HRRR QPF. MRMS-GC is “1-h local gauge bias-corrected radar precipitation accumulations” referred to as “Q3GC\_SHSR\_1H” in Zhang et al. (2016). We collected the hourly MRMS-GC product for the same 6-month period.

The IFC product is a radar-only (e.g., without rain gauge correction) QPE, generated by combining data from seven Weather Surveillance Radar-1988 Doppler (WSR-88D) radars that cover the entire Iowa domain. The IFC QPE algorithms (e.g., Seo et al. 2011) construct a composite rain rate map every 5 min after applying polarimetric data quality control (Seo et al. 2015).

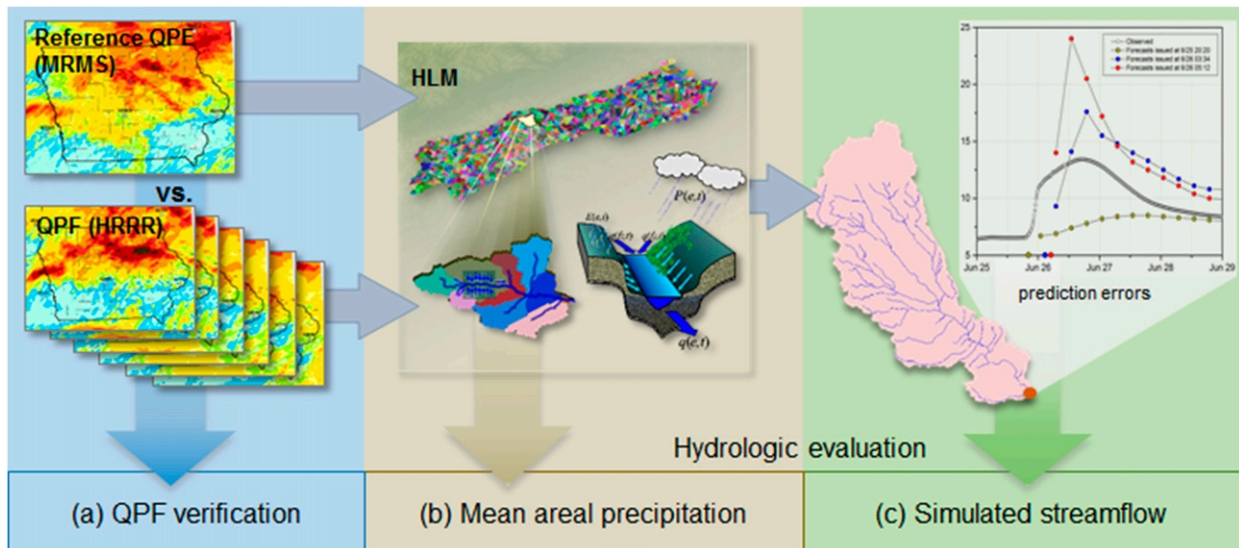


FIG. 2. Schematic overview of HRRR QPF evaluation procedures. The QPF evaluation involves three steps: (a) conventional QPF verification by comparing HRRR QPF with reference QPE (MRMS), (b) analysis of mean areal precipitation estimated for basin units used in the HLM, and (c) performance evaluation of simulated streamflow driven by the precipitation forcing products.

The algorithms then generate hourly accumulations using an advection procedure (Seo and Krajewski 2015) to correct radar temporal sampling errors. We used the IFC product to simulate streamflow using the HLM and to compare the simulation results with those driven by MRMS-GC.

We also collected USGS streamflow observations for the evaluation of hydrologic model simulations driven by the multiple precipitation forcing products (e.g., QPE and QPF). As shown in Fig. 1, the heavy storms hit the domain of the Cedar River basin and caused significant flooding in the area. Therefore, we selected two major USGS stations within the Cedar River basin and acquired stream discharge data for the flooding period. In Fig. 1c, we present the two selected USGS stations, Waterloo and Cedar Rapids, with the basin boundary. The Cedar River basin is one of the largest basins in Iowa, with an upstream catchment area of approximately 17 000 km<sup>2</sup>.

### c. Hydrologic model

The HLM builds on the concept of landscape decomposition into hillslopes and channels (Mantilla and Gupta 2005). The HLM allows for flexible structure and the representation of the physical processes of runoff generation and water transport; these processes include initial abstraction, infiltration, overland flow, percolation, base flow, and channel routing. The HLM is calibration-free, that is, a common configuration of parameters determined a priori applies to all the hillslopes. Each hillslope contains four water storage components:

channel storage, water ponded on hillslope surface, effective water depth in the topsoil layer, and effective water depth in hillslope subsurface. The mass conservation equations of the water storage are defined in terms of ordinary differential equations. Channel streamflow comprises several flow components: 1) overland flow from the water ponded on hillslope surface, 2) interflow from the water depth in the top soil layer, and 3) baseflow from the hillslope subsurface. The mass transport for each channel link in the network is defined as a power-law relation that describes flow velocity as a function of discharge and drainage area (Ayalew et al. 2014). More details about the HLM equations, configuration, and numerical solver are provided in Small et al. (2013) and Krajewski et al. (2017).

## 3. Methodology

In this section, we describe evaluation approaches to quantify the precipitation forecast skills of HRRR QPF and the effects of its error or uncertainty on the streamflow prediction. Figure 2 illustrates a schematic view of the evaluation procedures in this study. First of all, we assess the HRRR precipitation forecasts using common verification skill scores (see, e.g., Schaefer 1990) on forecast lead time. Before we apply the hydrologic model for streamflow simulations, we also characterize basin-scale uncertainties of the QPF forcing data. We use statistical metrics based on a meaningful hydrologic factor, that is, mean areal precipitation (MAP; Johnson et al. 1999), that is closely connected

with the streamflow generation. We then drive the HLM with the forcing products (e.g., MRMS QPE and HRRR QPF) and evaluate simulated streamflow with input from HRRR QPF. We assess the streamflow simulation results with respect to HRRR lead time and basin scale. We do not use any calibration procedure in the HLM simulation. In real-time operations, it is common to drive a hydrologic model with a combination of QPE at present time and QPF for the next couple of hours or days and generate hydrologic forecasts. In this study, however, we split the QPE and QPF forcing data and drive the model individually. This forcing data separation prevents blending errors in hydrologic forecasts contributed by both the QPE and QPF forcing products and allows us to examine the errors driven solely by the QPF product.

In the analysis procedures shown in Fig. 2, we use MRMS-GC as reference to access the prediction capability of HRRR QPF. The MRMS-GC product incorporates hourly rain gauge data from the hydrometeorological automated data system (HADS; Kim et al. 2009). The performance evaluation of this reference requires an independent hourly gauge network that thoroughly covers the study domain. Unfortunately, there is no such network in the area, and thus we were not able to perform the reference product evaluation at the hourly scale. Instead, we performed a simple comparison with the results presented in Zhang et al. (2016), which reported that MRMS-GC and Stage IV products showed similar performance at the daily scale for the cold months of 2014, while MRMS-GC showed larger errors for warm months. We evaluated both Stage IV and MRMS-GC products for our study period, using the NWS Cooperative Observer Program (COOP; Mosbacher et al. 1989) network that reports daily precipitation values. We found that the daily averages of mean absolute error (MAE) of MRMS-GC (about 1.24 mm) and Stage IV were almost equivalent. This MAE value was

comparable to the one for cold months (approximately 2.54 mm) in Zhang et al. (2016), which showed better performance than it did in the warm months. We also looked at an hourly gauge interpolation analysis based on HADS and observed that gauge interpolation was not effective to capture rainfall spatial variability at the hourly scale. These facts may justify the use of MRMS-GC as reference in this study. We note that radar–gauge merging schemes (e.g., Velasco-Forero et al. 2009; Nanding et al. 2015) that preserve the spatial variability of the radar field with maintaining the accuracy of rain gauge measurements might be potential options to obtain a more reliable reference product. However, we think that those merging schemes may not yield better results than MRMS-GC (local gauge bias correction) for the study area because of the low density of HADS rain gauges in the Cedar River basin (e.g., 18 gauges in a 17 000 km<sup>2</sup> area).

#### a. Skill scores for QPF verification

We use four skill scores frequently employed in the forecast verification analysis to objectively assess the performance of precipitation forecasts. The scores include hit rate (HR), false alarm rate (FAR), frequency bias (FB), and Gilbert skill score (GSS). We calculate these scores based on a contingency table that describes the number of forecasting successes and failures depending on observations (event or nonevent), as presented in Table 1. The definition of the skill scores using Table 1 are

$$\text{HR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}}, \quad (2)$$

$$\text{FB} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FN}}, \quad \text{and} \quad (3)$$

$$\text{GSS} = \frac{\text{TP} \times \text{TN} - \text{FN} \times \text{FP}}{(\text{FN} + \text{FP})(\text{TP} + \text{FN} + \text{FP} + \text{TN}) + (\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}. \quad (4)$$

The values of TP, FN, FP, and TN in Table 1 are determined by counting the number of grids (in MRMS-GC and HRRR QPF) identified as rain or no-rain. For instance, a rain grid is defined when its rainfall value exceeds a specific rainfall threshold. The HR is equivalent to the probability of detection (POD) and indicates the ratio of the correctly forecasted number of grids to the total number of observed rain grids. The FAR is a

measure of a forecast failure defined as the proportion of the number of failures to the total number of forecasted rain grids. The FB is the ratio of the total number of forecasted rain grids to the total number of observed rain grids, and an FB value greater than unity indicates the overprediction of a rainfall coverage. To evaluate overall forecast skill, we use the GSS instead of, but similar to, the critical success index

TABLE 1. Contingency table to define forecast skill scores. TP, FN, FP, and TN denote true positive, false negative, false positive, and true negative, respectively.

| Observations | Forecasts |    |  |
|--------------|-----------|----|--|
|              | Yes       | No |  |
| Yes          | TP        | FN |  |
| No           | FP        | TN |  |

(CSI; Donaldson et al. 1975), because the CSI is sometimes biased and dependent on the occurrence rate of rain events being forecasted. Schaefer (1990) provides a detailed discussion about the GSS and CSI.

### b. Statistical metrics for hydrologic evaluation

As shown in Fig. 2, we divide the hydrologic evaluation of HRRR QPF into two parts based on 1) the analysis of mean areal precipitation estimated for basin units used in the HLM and 2) the HLM streamflow simulation driven by the QPF forcing. As references for both evaluations, we compute MAP using MRMS-GC and simulate streamflow discharge with MRMS-GC forcing.

We based our calculation of the verification scores of Eqs. (1)–(4) on binary classification (e.g., rain or no rain), and this does not imply the magnitude of errors in the precipitation forecasts that can directly affect the rainfall–runoff translation processes. Because precipitation forecasts contain significant uncertainties and are generally biased (e.g., Buizza et al. 2005), we introduce factors to describe errors/uncertainties of HRRR QPF that might affect errors in streamflow generation. The statistical metrics for the MAP analysis include bias  $B$ , correlation coefficient  $R$ , and root-mean-square error (RMSE). In this study, we define the bias as a multiplicative term of errors as seen in Eq. (5). The bias describes a systematic tendency of the precipitation forecasts that can erroneously generate water volume in hydrologic prediction. The correlation coefficient describes agreement or linear dependence between reference and forecasts, and the RMSE is used as a measure of the differences between them:

$$B = \frac{\sum_{s,t} R_{\text{HRRR}}(s,t)}{\sum_{s,t} R_{\text{MRMS-GC}}(s,t)}, \quad (5)$$

where  $R_{\text{HRRR}}(s,t)$  and  $R_{\text{MRMS-GC}}(s,t)$  denote rainfall values at a time step  $t$  and a grid location  $s$  within a specific basin.

To assess the impact of the QPF forcing on streamflow prediction, we define three evaluation metrics: Nash–Sutcliffe efficiency (NSE) defined by Nash and Sutcliffe (1970),  $R$ , and Kling–Gupta efficiency (KGE) proposed

by Gupta et al. (2009). NSE typically describes the predictive power of hydrologic models and has been used steadily in numerous studies (e.g., Krause et al. 2005; Seo et al. 2013) for the evaluation of model performance. NSE is defined as

$$\text{NSE} = 1.0 - \frac{\sum_t (S_t - O_t)^2}{\sum_t (O_t - \bar{O})^2}, \quad (6)$$

where  $S$  and  $O$  denote simulated and reference (observed) streamflow. NSE ranges from negative infinity to 1.0. The negative NSE values imply that the mean of the reference streamflow is a better predictor than model simulation, whereas values close to 1.0 indicate more accurate model performance. KGE is an alternative metric that was proposed to improve a deficiency in NSE, namely, that the peak is likely to be underestimated when NSE is used in optimization (e.g., calibration). A more detailed comparison of NSE and KGE can be found in Gupta et al. (2009). In Eq. (7), we calculate KGE using correlation and the ratios of mean and standard deviation of simulated and observed streamflow. The index can be decomposed into easy-to-understand terms:

$$\text{KGE} = 1.0 - \sqrt{(\rho - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}, \quad (7)$$

where  $\rho$ ,  $\alpha$ , and  $\beta$  denote correlation, the ratio of standard deviation ( $\sigma_s/\sigma_o$ ), and the ratio of mean ( $\mu_s/\mu_o$ ) between simulated and observed streamflow, respectively. Similar to NSE, a KGE value close to 1.0 implies an optimal estimation.

## 4. Results

In Fig. 3, we present the accumulated rainfall maps of MRMS-GC and HRRR QPF for the flooding period of 14–23 September 2016. We use the MRMS-GC product as reference because it was corrected for errors using hourly rain gauge observations (Zhang et al. 2014). Regarding the HRRR maps shown in Fig. 3, we accumulated a set of the same lead time products issued at different times and show the generated maps for 1-, 2-, 5-, and 6-h lead times. A visual inspection reveals similarities and significant discrepancies between MRMS and HRRR in Fig. 3. While the observed overall patterns that show relatively low rain in the south and high rain in the north of the domain tend to be similar, the exact locations of heavy rain and rainfall spatial distribution look quite different. HRRR QPF seems to be somewhat biased (notably, overestimated for short

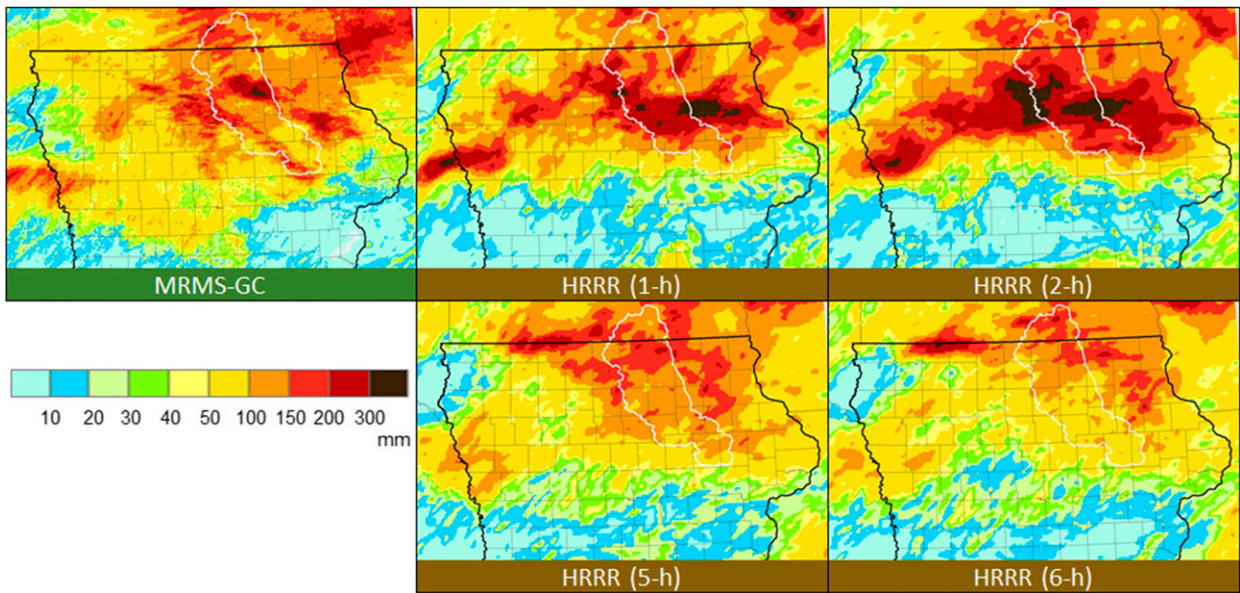


FIG. 3. Comparison of 10-day rainfall accumulation maps for MRMS-GC and HRRR for a significant flooding event in September 2016. The HRRR data were accumulated using a set of the same lead time products issued at different times. The numeric values with HRRR maps indicate QPF lead times.

lead times) depending on lead times and locations. Heavy rainfall events might have larger conditional biases (on magnitude, season, storm type, and other attributes) that may have significant impact on flood forecasting (e.g., Brown et al. 2012).

#### a. Conventional QPF verification

We calculated the skill scores of HRRR QPF using Eqs. (1)–(4). We used the MRMS QPE product, MRMS-GC, as reference (“observation” in Table 1) because the number of ground observations (e.g., rain gauge measurements) is limited, and it is hard to extensively evaluate the skill with limited locations over a large domain. Since the reference product has higher spatial resolution (e.g., 1 vs 3 km), we assigned the same HRRR values to the multiple MRMS grids that are collocated within an identical HRRR grid for a one-to-one match between the two products. We used multiple rainfall threshold values (1, 2, 5, and 10 mm) and individually defined all matched grid cells in both products as rain or no-rain. We then counted all the categories shown in Table 1 for the 10-day flooding event and estimated the scores based on each MRMS grid. Because of the short sample period, the geographic maps of the calculated scores exposed spatially discontinuous patterns (e.g., some spikes and sinks in southeastern Iowa), particularly with HR and FB (we do not present the maps here). To resolve this issue, we extended the sampling period to six months (May–October), which

included the main flooding event, and we were able to eliminate all spikes and sinks in the maps. Figure 4 shows the spatially averaged skill scores over the entire domain shown in Fig. 3, and we present the scores for both 10-day and 6-month periods with respect to forecast lead time. In Fig. 4, we can observe that the forecast skill decreases with increasing rainfall threshold as the forecast time span becomes longer. The decreasing pattern and magnitude of observed FB and overall skill (GSS) are somewhat comparable to those reported in Moser et al. (2015). We can also recognize that the accuracy or performance of HRRR QPF does not seem sufficient to justify its use in hydrologic prediction because the GSS in Fig. 4b is only about 30% even with just 1-h lead time, at which the score reaches its best performance. The sudden jumps and drops detected in Fig. 4b after the 15-h lead time arise from the fact that the HRRR forecast lead time was extended up to 18 h in late August (the previous lead time was up to 15 h). The proportion of the QPF data with full 18-h lead time is about 40% of the entire analysis period.

#### b. Hydrologic evaluation

The QPF verification scores shown in Fig. 4 were obtained from a binary rainfall identification and do not provide straightforward evidence to describe the QPF’s effects on streamflow predictions. In this section, we focus on one of the most significant hydrologic



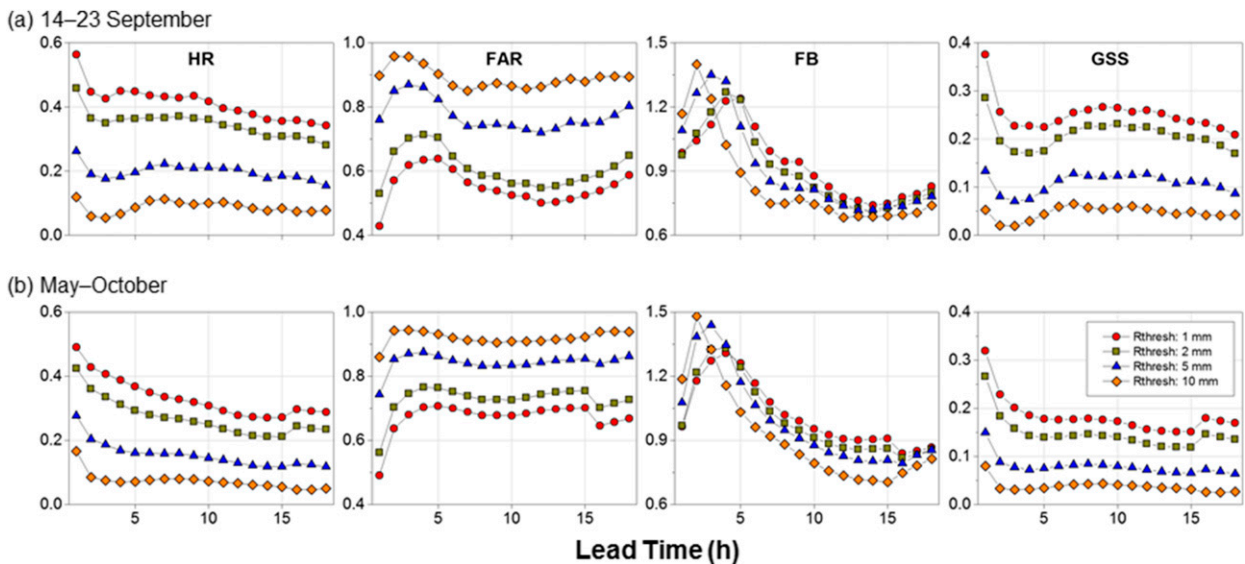


FIG. 4. Forecast skill scores of HRRR QPF with respect to forecast lead time for the periods of (a) 14–23 Sep and (b) May–October. A rainfall threshold is required to calculate the skill scores. “Rthresh” in the legend denotes the rainfall threshold values used in the analysis.

parameters, namely, MAP, in generating streamflow (see, e.g., Quintero et al. 2016) and characterize the statistical error structure of HRRR QPF. We also drive the HLM with HRRR QPF and investigate the QPF error’s contribution on the streamflow generation.

### 1) MEAN AREAL PRECIPITATION

Figure 5 shows statistical properties of the HRRR QPF error (e.g.,  $B$ ,  $R$ , and RMSE) with respect to lead time, and here we present the analysis results up to 6-h lead time. For this analysis, we accumulated rainfall amounts over the Cedar River basin as shown in Fig. 1c for the same period as shown in Fig. 3. The bias is defined as a multiplicative term of QPF error as described in Eq. (5), and the value greater than 1.0 implies an overestimation of HRRR QPF versus MRMS. In Fig. 5, all nested drainage areas ranging from 0.1 to 17 000 km<sup>2</sup> within the Cedar River basin are included in the box plots, and the statistical features of QPF error are compared across the basin scale. As one can see in Fig. 5, HRRR QPF tends to overestimate at short lead times, and the tendency of overestimation and high variability at small scales gradually decreases with longer lead times. This behavior agrees to some extent with the widely acknowledged fact that numerical weather prediction (NWP) models have difficulty in making predictions at short lead times because they do not capture the initial precipitation distribution and amounts well; the models perform better as they dynamically resolve the large-scale flow (e.g., Lin et al. 2005). The RMSE shows similar behavior observed in  $B$ , whereas

$R$  shows the opposite properties against  $B$  and RMSE: at the smaller-scale basin, variability increases with longer lead times, and  $R$  values tend to be somewhat higher as forecast lead time increases.

In Fig. 6, we aggregate the results presented in Fig. 5. Figure 6 shows the statistical error structure of forecasted MAP characterized by forecast lead time and drainage basin scale. For the metrics shown in Fig. 6, the prediction skill generally improves as basin scale grows larger, although there is an exception with the bias. We can also observe that the skill is not sufficiently good at the initial lead times; it approaches its best performance at 5- or 6-h lead time, and then decreases. Since we based this analysis on a specific event that induced significant flooding in Cedar Rapids, Iowa, the error characteristics may vary depending on the season, geographic location, and precipitation regime.

### 2) STREAMFLOW

We drove the IFC HLM with the individual forcings of MRMS-GC, IFC, and HRRR QPF and present observed and simulated streamflow hydrographs. We determined the initial model states from a spinup run that uses a continuous forcing of the previous 6-month MRMS-GC product. Figure 7 shows the simulation results driven by QPE products (e.g., MRMS and IFC) to demonstrate the HLM’s capability compared to the streamflow observations at the two USGS stations, Waterloo and Cedar Rapids. We also present the NWM cycle of “analysis and assimilation” in Fig. 7. In this figure, the simulated hydrographs driven by the MRMS

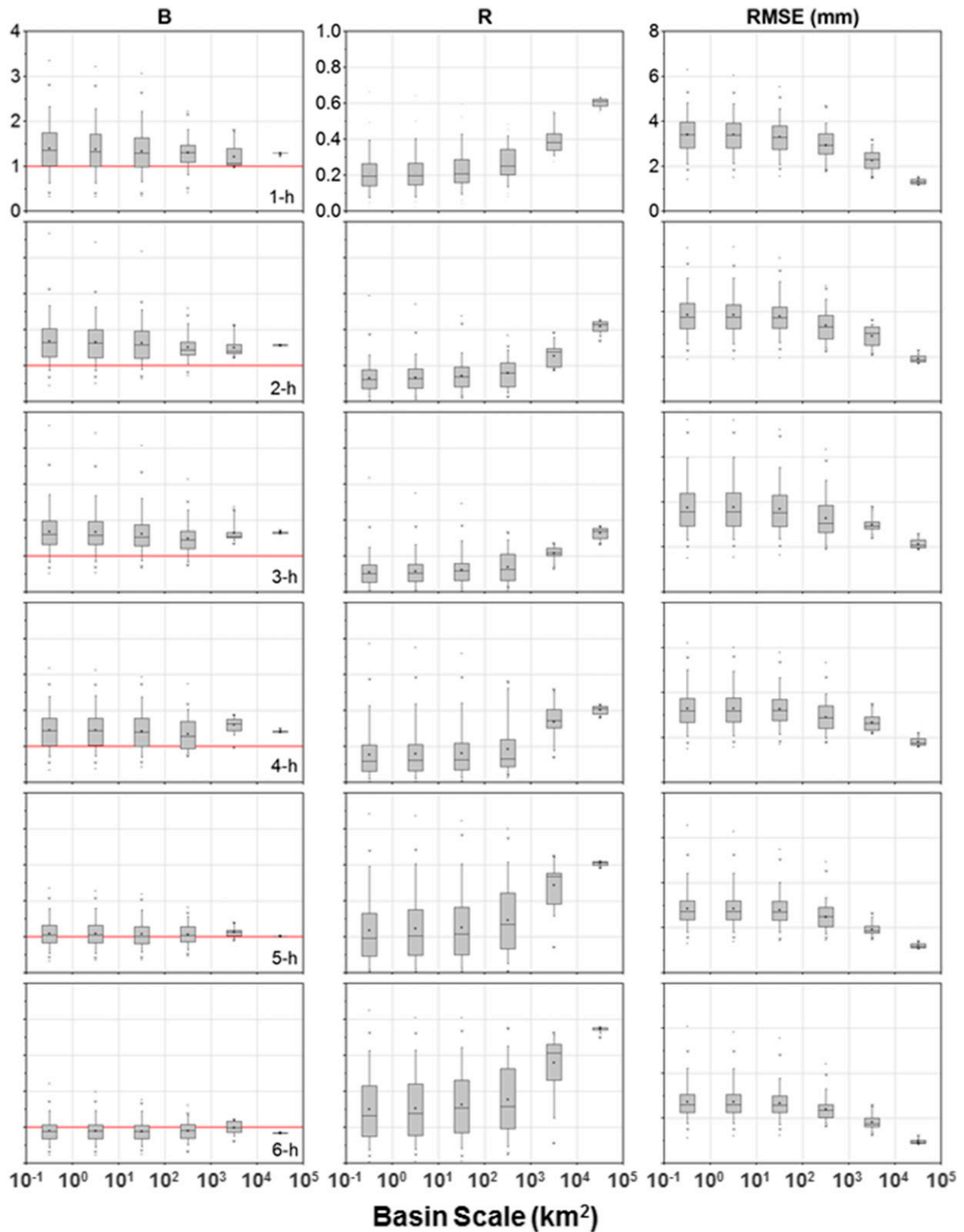


FIG. 5. Statistical properties of HRRR QPF error characterized using MAP regarding basin scale and forecast lead time. All nested catchment areas within the Cedar River basin are included in the analysis. Column heads *B*, *R*, and RMSE denote the multiplicative bias, correlation, and RMSE, respectively.

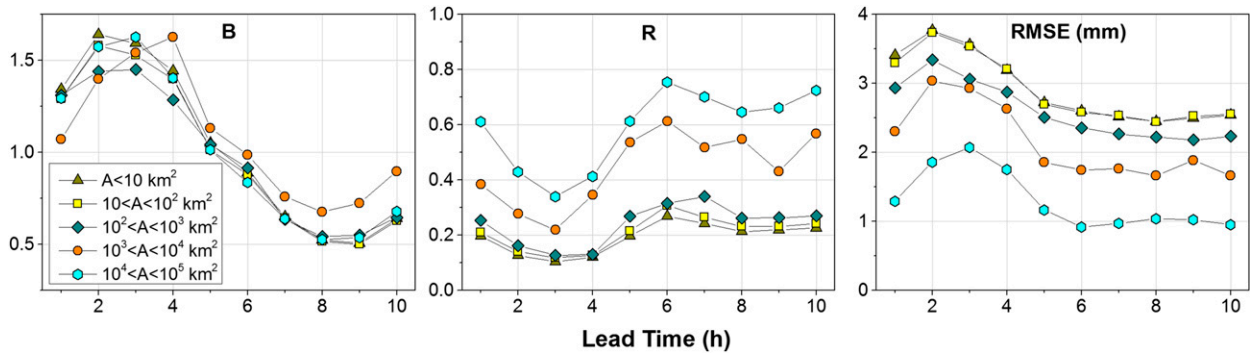


FIG. 6. Statistical error structure of forecasted MAP characterized by forecast lead time and drainage basin scale. The symbols indicate the median values at each basin scale and forecast lead time.

and IFC products look very similar and close, but show some early recession and two consecutive peaks when compared to the USGS streamflow observation. We closely investigated the double peaks and concluded that they were caused by rainfall event separation to which the actual streamflow did not sensitively respond. This is also associated with the HLM's early recession issue discussed above. We will discuss the event separation and resulting peaks later in this section. The NWM results presented in Fig. 7 show that the falling limbs at both locations agree with the USGS observations well because the observation data were assimilated to initialize the forecast model cycles (e.g., short and medium range). We noted that the NWM also yields two peaks at Cedar Rapids that are similar to those from the HLM simulations, but the first peak of the NWM seems relatively sharp. We speculate that the sharp peak arises from the effect of the data assimilation (e.g., update of the model state using streamflow observations that are quite different from model forecasts). For Waterloo, the NWM significantly overestimated the peak without the doubled pattern observed from the HLM simulations. We present statistical metric values (e.g., NSE,  $R$ , and KGE) in Table 2 to illustrate the performance of streamflow simulations shown in Fig. 7. Table 2 shows that the NSE and KGE values of HLM simulations driven by MRMS and IFC are comparable to those of NWM, for which streamflow observation data were used for calibration. This enabled us to use the MRMS simulation results as reference for ungauged locations where USGS stations do not exist.

Figure 8 illustrates simulation results driven by HRRR QPF with lead times of 1-, 2-, 5-, and 6-h QPF. In Fig. 8, we present the USGS observation and MRMS result as references. We used the MRMS result as an alternative reference because USGS streamflow observations are not available at all catchment outlet

locations employed for the MAP analysis in Fig. 5. We can only acquire USGS observations at some designated locations, which does not allow us to explore the impact of QPF error on the flood prediction at various catchment scales. For the locations where USGS observations do not exist, we captured the simulated hydrographs driven by MRMS for all hillslope links (almost 40 000 locations) within the Cedar River basin. These are used

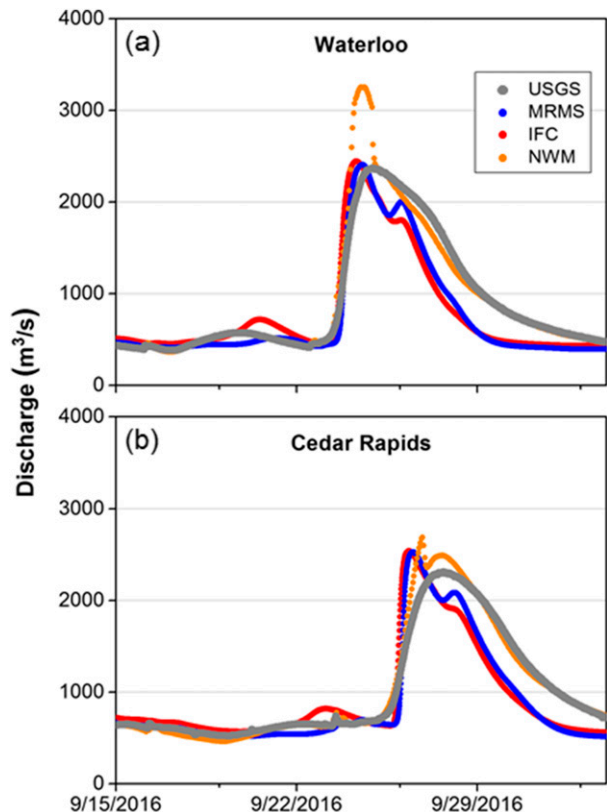


FIG. 7. Simulated streamflow driven by the QPE (MRMS and IFC) products at (a) Waterloo and (b) Cedar Rapids. The NWM cycle of analysis and assimilation is also presented for reference.

TABLE 2. Performance metrics of streamflow simulations driven by the QPE products (MRMS and IFC). The NWM cycle of analysis and assimilation is also presented for reference.

| Metrics  | Waterloo |      |      | Cedar Rapids |      |      |
|----------|----------|------|------|--------------|------|------|
|          | MRMS     | IFC  | NWM  | MRMS         | IFC  | NWM  |
| NSE      | 0.85     | 0.78 | 0.89 | 0.79         | 0.72 | 0.97 |
| R        | 0.95     | 0.90 | 0.96 | 0.92         | 0.86 | 0.99 |
| KGE      | 0.80     | 0.79 | 0.85 | 0.85         | 0.82 | 0.90 |
| $\alpha$ | 0.90     | 0.87 | 1.14 | 0.97         | 0.91 | 1.10 |
| $\beta$  | 0.83     | 0.88 | 1.03 | 0.88         | 0.92 | 1.01 |

to quantify the performance of flood prediction driven by HRRR QPF. In Figs. 8a and 8c, 1- and 2-h lead time results show erroneous early peaks before 23 September as well as significant overestimation at the actual peak time shown in the MRMS simulation. On the other hand, 5- and 6-h lead time results do not generate the early peak detected with 1- and 2-h lead time QPF, but the peaks at both stations seem to be underestimated and delayed. In Tables 3 and 4, we present the performance metrics of simulations shown in Figs. 8a and 8c. For the calculation of these metrics, we used the USGS streamflow observations as reference while employing the MRMS forcing results as reference for ungauged locations in the further analysis. In Tables 3 and 4, we recognize that the performance decreases at short lead times and starts increasing from 4-h lead time, which agrees to some extent with the MAP results presented in Fig. 6. We speculate that the exceptional performance at 1-h lead time is quite different from that at other initial lead times (e.g., 2 and 3 h) because of the radar assimilation in the initialization of HRRR (e.g., Benjamin et al. 2009).

To inspect the discovered features from the simulated hydrographs, we present time series of MAP for the upstream basins of Waterloo and Cedar Rapids in Figs. 8b and 8d. The numeric values in Figs. 8b and 8d indicate accumulated MAP for the two rain events. A MAP difference of about 60 mm between MRMS and 2-h lead time QPF for the first rain event at Cedar Rapids resulted in a huge difference in water volume for the given catchment area of 17 000 km<sup>2</sup> and led to the erroneous peak. The accumulated rainfall amounts (for the first event) over time and space as represented by the river networks are illustrated in Fig. 9a. The color on a specific location of the river networks represents MAP for the upstream catchment area of the location. Figure 9a demonstrates that the erroneous early peaks generated by 1- and 2-h lead time QPF (Fig. 8) were mainly caused by major overestimation in the north-western upstream areas of the Cedar River basin. In Figs. 8b and 8d, the second rain event shows rainfall

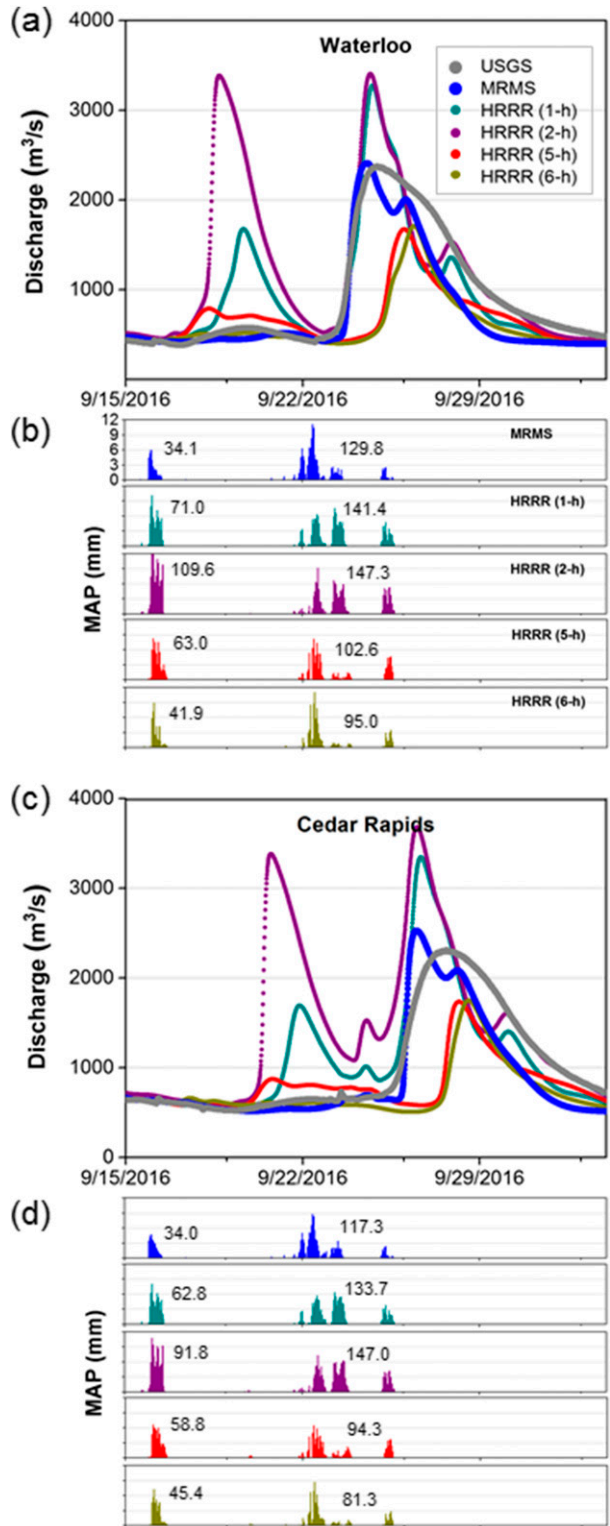


FIG. 8. Simulated streamflow driven by HRRR QPF at (a) Waterloo and (c) Cedar Rapids and time series of MAP for the upstream area at (b) Waterloo and (d) Cedar Rapids. The numeric values in (b) and (d) represent accumulated MAP for each rain event.

TABLE 3. Performance metrics of streamflow simulations at Waterloo driven by HRRR QPF up to 6-h lead time.

| Metrics  | HRRR QPF lead time (h) |       |       |       |      |      |
|----------|------------------------|-------|-------|-------|------|------|
|          | 1                      | 2     | 3     | 4     | 5    | 6    |
| NSE      | 0.67                   | -0.49 | -1.19 | -0.09 | 0.30 | 0.24 |
| <i>R</i> | 0.85                   | 0.59  | 0.37  | 0.39  | 0.69 | 0.72 |
| KGE      | 0.83                   | 0.35  | 0.17  | 0.36  | 0.33 | 0.32 |
| $\alpha$ | 1.08                   | 1.38  | 1.41  | 0.85  | 0.47 | 0.47 |
| $\beta$  | 1.03                   | 1.33  | 1.35  | 1.10  | 0.75 | 0.67 |

TABLE 4. Performance metrics of streamflow simulations at Cedar Rapids driven by HRRR QPF up to 6-h lead time.

| Metrics  | HRRR QPF lead time (h) |       |       |       |      |      |
|----------|------------------------|-------|-------|-------|------|------|
|          | 1                      | 2     | 3     | 4     | 5    | 6    |
| NSE      | 0.47                   | -1.22 | -1.44 | -0.11 | 0.30 | 0.19 |
| <i>R</i> | 0.79                   | 0.51  | 0.35  | 0.39  | 0.68 | 0.67 |
| KGE      | 0.73                   | 0.19  | 0.14  | 0.36  | 0.34 | 0.33 |
| $\alpha$ | 1.16                   | 1.49  | 1.41  | 0.85  | 0.46 | 0.49 |
| $\beta$  | 1.09                   | 1.41  | 1.38  | 1.13  | 0.80 | 0.73 |

separation by about a day, which verifies that this event generated two consecutive peaks associated with rapid recession of HLM. We also present the aggregated rain for the second event over the river networks in Fig. 9b. Figure 9b shows significant differences in accumulated rainfall (e.g., over the Shell Rock River basin and upstream basins near Waterloo) and its spatial distribution among MRMS and HRRR QPF with different lead times. This led to the overestimation (1 and 2 h) and underestimation (5 and 6 h) of the peaks. Because of the large contributing area, the stream locations near the basin outlet (Cedar Rapids) reveal relatively small MAP when compared to some upstream tributaries.

Figure 10 demonstrates the statistical performance of hydrologic simulations driven by HRRR QPF with respect to forecast lead time and basin scale. The number of drainage areas included in Fig. 10 is identical to that in Fig. 5, and we present the results up to 6-h lead time. We note that some negative ranges for NSE and KGE are not fully shown in short lead times because negative values were widely distributed, particularly at the smaller-scale basins. In Fig. 10, we do not include the ingredients of KGE,  $\alpha$ , and  $\beta$ , presented in Tables 3 and 4. Similar to the results shown in Fig. 5 and Tables 3 and 4, the performance gradually decreases at the initial lead times of 2 and 3 h, and then starts increasing at 4–6-h lead times, particularly at smaller basin scales. It is hard to find any scale-dependent property clearly, but large-scale basins demonstrate less variability and better performance, except for the performance of the largest basin scale at 3–6-h lead times.

## 5. Summary and discussion

To enhance hydrologic prediction capabilities, high-resolution QPF (e.g., HRRR) has recently been appended into the procedures of operational hydrologic forecasting based on distributed hydrologic modeling (e.g., Maidment 2017; Krajewski et al. 2017). In this study, we aimed to evaluate HRRR QPF in terms of the precipitation forecast skill and the effects of forecast uncertainty on flood prediction using an Iowa flooding event

that occurred in September 2016. We statistically quantified the uncertainty of HRRR QPF and the performance of hydrologic simulation results driven by the QPF product. We used conventional verification skill scores employed in a number of meteorological studies (e.g., Schaefer 1990; Moser et al. 2015) for the precipitation forecast evaluation. For hydrologic evaluation, we examined MAP estimated from HRRR QPF against that from the gridded reference (MRMS), as well as hydrologic simulation results driven separately by HRRR and MRMS.

Verification of precipitation forecasts reveals that the skill gradually decreases as the forecast lead time increases (Fig. 4). The analysis also shows that the best skill is achieved at 1-h lead time (but about 30% skill based on GSS is not good enough to justify further streamflow forecasting) and significantly drops after that, implying that some improvement for 1-h as well as further lead times is necessary to apply the QPF product to hydrologic prediction. The MAP analysis shown in Figs. 5 and 6 clarifies the quantitative uncertainty features of HRRR QPF regarding catchment scale and forecast lead time. Overall, the bias and variability of HRRR QPF at smaller-scale basins steadily decrease as forecast lead times become longer (e.g., up to 6-h lead time). This indicates that the prediction capability of numerical models is limited at short lead times because of the challenge in model initialization (e.g., Lin et al. 2005). HRRR appeared to improve the initialization issue somewhat using radar data assimilation (Benjamin et al. 2009) because the QPF uncertainty at 1-h lead time was remarkably lower than it was at other initial lead times such as 2, 3, and 4 h (see, e.g., Fig. 6). Three-dimensional radar reflectivity data created as part of the MRMS data suite are assimilated into the HRRR model to initialize storm information at the 3-km scale and thus to improve predictions regarding ongoing convection at the earlier forecast lead times. While we are uncertain of the assimilation performance using the radar data in HRRR, Pinto et al. (2015) reported that the model predicted too many convective systems within 4-h lead time over the Great Plains.

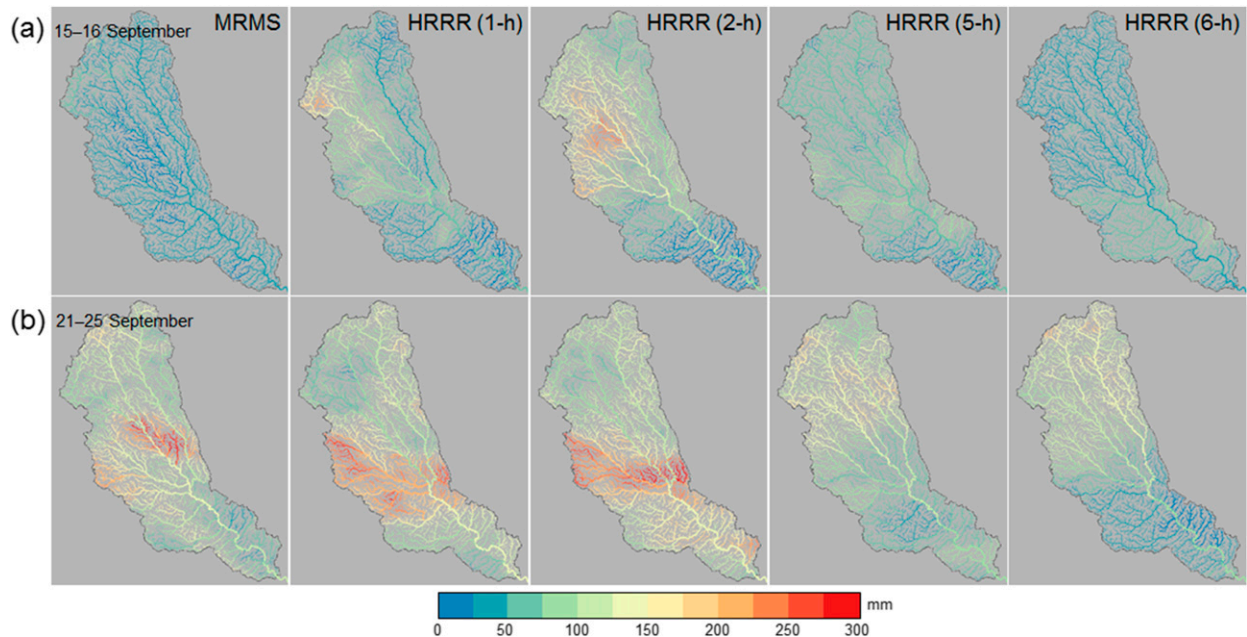


FIG. 9. Aggregated MRMS and HRRR QPF rainfall over time and space represented by the river networks: (a) the first event starting on 15 Sep and (b) the second event starting on 21 Sep, both of which are shown in Figs. 8b and 8d.

In operational hydrologic forecasting, a combination of available QPE at present time and QPF for the next couple of hours or days is used to drive a hydrologic model. This general procedure is repeated at each forecast time and generates streamflow forecasts. In this study, however, we separately drove the HLM with each QPE and QPF forcing product (see Fig. 2) to examine hydrologic prediction errors contributed solely by the QPF uncertainty. The QPF forcing products were organized by the same lead time data issued at different times. We note that this setup may exaggerate the prediction errors, and we may not see the same or similar extent of errors in actual operations as or to those presented in this study. The hydrologic simulation results based on performance metrics (Tables 3, 4) demonstrate that the result driven by 1-h lead time QPF is relatively better than other lead time results, and the performance sharply drops at initial lead times (e.g., 2 and 3 h). The best performance achieved at 1-h lead time does not seem acceptable because it significantly overestimates the flood peak along with the generation of an early erroneous peak that is not detectable in the streamflow observation (Fig. 8). The hydrologic simulation shows a similar tendency discussed in the MAP analysis in which the QPF simulation performance sharply decreases at 1–3-h lead times and starts increasing from 4- through 6-h lead times. The scale dependence shown in the MAP analysis (Fig. 5) is not clearly detected in the hydrologic simulations, while large-scale basins show better

performance and smaller variability, particularly at 5- and 6-h lead times (Fig. 10).

We based this study on a specific event that caused significant flooding in Cedar Rapids, Iowa. The HRRR error characteristics may vary depending on the season, geographic location, and precipitation regime (see, e.g., Pinto et al. 2015). Therefore, for comprehensive understanding, we need further study to collect multiyear products and examine error features of HRRR QPF. We also know that the performance of persistence-based precipitation forecasts tends to be better than that of numerical weather predictions at very short lead times (e.g., Wilson et al. 1998; Lin et al. 2005). The range of these short lead times may vary with different forecast schemes and models.

Our future investigation will study which lead time shows the performance transition between HRRR QPF and persistence-based (e.g., advection) forecasts by comparing their forecast skills. This will provide researchers and forecasters with useful information on whether they should replace or blend model QPF with persistence-based QPF to improve prediction skills at short lead times. The evaluation framework presented in this study demonstrates that the QPF error is a function of forecast lead time. We note that the QPF error (e.g., bias) should be individually corrected regarding forecast lead time before its application to hydrologic prediction, and the required procedure for error characterization (e.g., bias and variability) can be

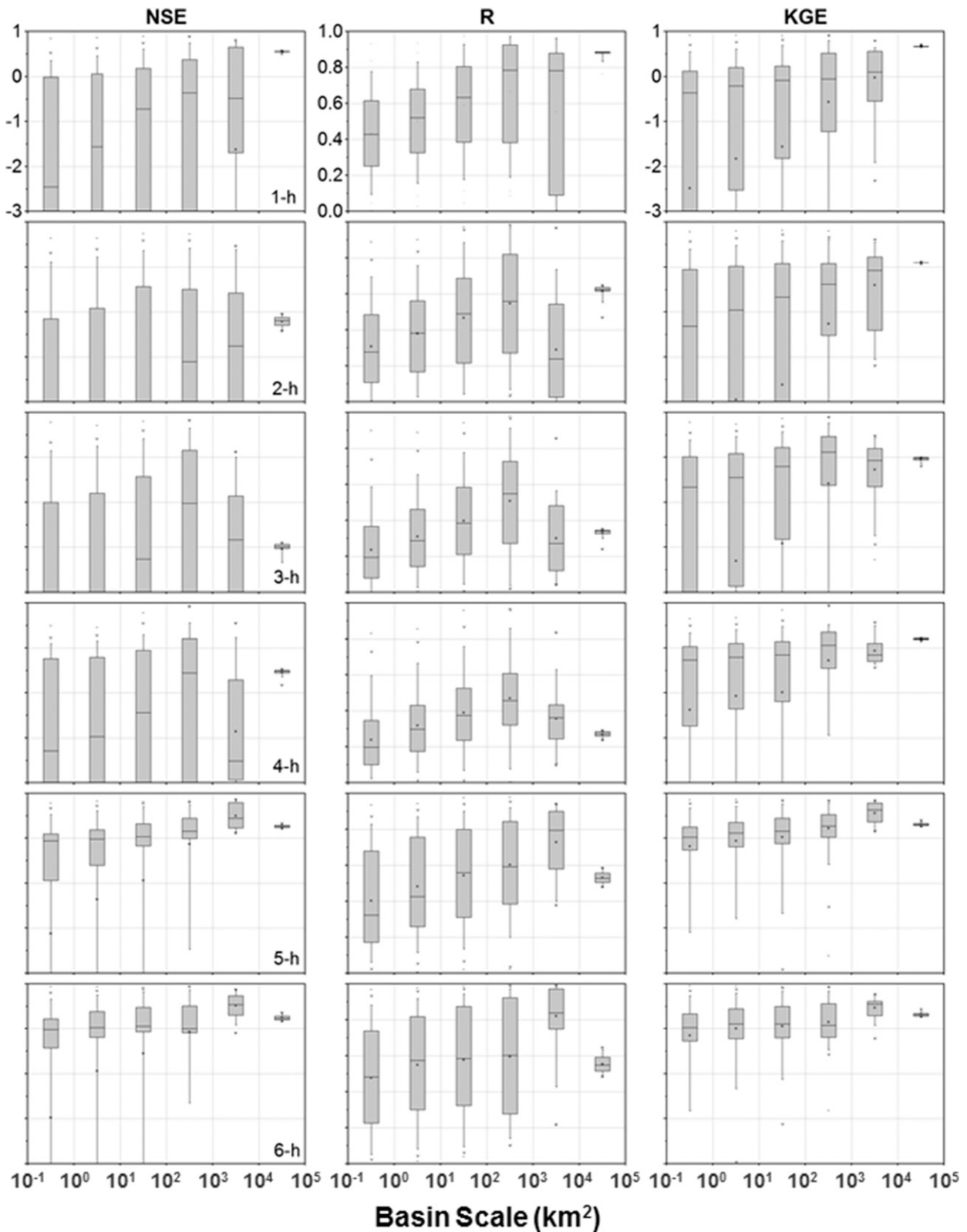


FIG. 10. Statistical performance of hydrologic simulation driven by HRRR QPF. The performance is measured by NSE,  $R$ , and KGE regarding basin scale and forecast lead time. As in Fig. 5, all nested catchment areas in the Cedar River basin are included in the analysis.

the basis of an ensemble nowcasting framework (e.g., Bowler et al. 2006).

**Acknowledgments.** This study was supported by the Iowa Flood Center at the University of Iowa and the Hydrometeorology Testbed (HMT) Program within NOAA/OAR Office of Weather and Air Quality under Grant NA17OAR4590131. The authors are grateful to the University Information Technology Services and IFC staff who facilitated the HLM simulations using the clusters of high-performance computing resources.

#### REFERENCES

- Alexander, C. R., S. S. Weygandt, T. G. Smirnova, S. Benjamin, P. Hofmann, E. P. James, and D. A. Koch, 2010: High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2010 convective season. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 9.2, [https://ams.confex.com/ams/25SLS/techprogram/paper\\_175722.htm](https://ams.confex.com/ams/25SLS/techprogram/paper_175722.htm).
- Arduino, G., P. Reggiani, and E. Todini, 2005: Recent advances in flood forecasting and flood risk assessment. *Hydrol. Earth Syst. Sci.*, **9**, 280–284, <https://doi.org/10.5194/hess-9-280-2005>.
- Ayalew, T. B., W. F. Krajewski, and R. Mantilla, 2014: Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties. *Adv. Water Resour.*, **71**, 32–43, <https://doi.org/10.1016/j.advwatres.2014.05.009>.
- Benjamin, S. G., T. G. Smirnova, S. S. Weygandt, M. Hu, S. R. Sahm, B. D. Jamison, M. M. Wolfson, and J. O. Pinto, 2009: The HRRR 3-km storm-resolving, radar-initialized, hourly updated forecasts for air traffic management. *Aviation, Range and Aerospace Meteorology Special Symp. on Weather-Air Traffic Management Integration*, Phoenix, AZ, Amer. Meteor. Soc., Pl.2, [https://ams.confex.com/ams/89annual/techprogram/paper\\_150430.htm](https://ams.confex.com/ams/89annual/techprogram/paper_150430.htm).
- Bowler, N. E., C. E. Pierce, and A. W. Seed, 2006: STEPS: A probabilistic precipitation forecasting scheme which merges as extrapolation nowcast with downscaled NWP. *Quart. J. Roy. Meteor. Soc.*, **132**, 2127–2155, <https://doi.org/10.1256/qj.04.100>.
- Brown, J. D., D.-J. Seo, and J. Du, 2012: Verification of precipitation forecasts from NCEP's Short-Range Ensemble Forecast (SREF) system with reference to ensemble streamflow prediction using lumped hydrologic models. *J. Hydrometeorol.*, **13**, 808–836, <https://doi.org/10.1175/JHM-D-11-036.1>.
- Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, <https://doi.org/10.1175/MWR2905.1>.
- Chen, S., and Coauthors, 2013: Evaluation and uncertainty estimation of NOAA/NSSL next-generation National Mosaic Quantitative Precipitation Estimation product (Q2) over the continental United States. *J. Hydrometeorol.*, **14**, 1308–1322, <https://doi.org/10.1175/JHM-D-12-0150.1>.
- Cosgrove, B., and Coauthors, 2015: Hydrologic modeling at the National Water Center: Operational implementation of the WRF-Hydro model to support National Weather Service hydrology. *2015 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H53A-1649.
- , and Coauthors, 2016: An overview of the National Weather Service National Water Model. *2016 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H42B-05.
- Donaldson, R. J., Jr., R. M. Dyer, and M. J. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, *9th Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 321–326.
- Gochis, D., 2016: Research to operations: Implementation and evolution of the NOAA National Water Model using the community WRF-Hydro system. *2016 Fall Cyberseminar Series*, Cambridge, MA, Consortium of Universities for the Advancement of Hydrologic Sciences, Inc., <https://www.cuahsi.org/education/cyberseminars/fall-cyberseminar-implementation-and-evolution-of-the-noaa-national-water-m>.
- Gourley, J. J., and Coauthors, 2017: The FLASH project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372, <https://doi.org/10.1175/BAMS-D-15-00247.1>.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Johnson, D., M. Smith, V. Koren, and B. Finnerty, 1999: Comparing mean areal precipitation estimates from NEXRAD and rain gauge networks. *J. Hydrol. Eng.*, **4**, 117–124, [https://doi.org/10.1061/\(ASCE\)1084-0699\(1999\)4:2\(117\)](https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(117)).
- Kim, D., B. Nelson, and D.-J. Seo, 2009: Characteristics of reprocessed hydrometeorological automated data system (HADS) hourly precipitation data. *Wea. Forecasting*, **24**, 1287–1296, <https://doi.org/10.1175/2009WAF2222227.1>.
- Krajewski, W. F., and Coauthors, 2017: Real-time flood forecasting and information system for the state of Iowa. *Bull. Amer. Meteor. Soc.*, **98**, 539–554, <https://doi.org/10.1175/BAMS-D-15-00243.1>.
- Krause, P., D. P. Boyle, and F. Bäse, 2005: Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.*, **5**, 89–97, <https://doi.org/10.5194/adgeo-5-89-2005>.
- Lehmann, J., D. Coumou, and K. Frieler, 2015: Increased record-breaking precipitation events under global warming. *Climatic Change*, **132**, 501–515, <https://doi.org/10.1007/s10584-015-1434-y>.
- Li, H., L. Luo, E. F. Wood, and J. Schaake, 2009: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *J. Geophys. Res.*, **114**, D04114, <https://doi.org/10.1029/2008JD010969>.
- Li, J., Y. Chen, H. Wang, J. Qin, J. Li, and S. Chiao, 2017: Extending flood forecasting lead time in a large watershed by coupling WRF QPF with a distributed hydrological model. *Hydrol. Earth Syst. Sci.*, **21**, 1279–1294, <https://doi.org/10.5194/hess-21-1279-2017>.
- Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki, 2005: Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.*, **32**, L14801, <https://doi.org/10.1029/2005GL023451>.
- Lin, P., M. A. Rajib, Z.-L. Yang, M. Somos-Valenzuela, V. Merwade, D. R. Maidment, Y. Wang, and L. Chen, 2018: Spatiotemporal evaluation of simulated evapotranspiration and streamflow over Texas using the WRF-Hydro-RAPID modeling framework. *J. Amer. Water Resour. Assoc.*, **54**, 40–54, <https://doi.org/10.1111/1752-1688.12585>.
- Maidment, D. R., 2017: Conceptual framework for the National Flood Interoperability Experiment. *J. Amer. Water Resour. Assoc.*, **53**, 245–257, <https://doi.org/10.1111/1752-1688.12474>.
- Mantilla, R., and V. K. Gupta, 2005: A GIS numerical framework to study the process basis of scaling statistics in river networks. *IEEE Geosci. Remote Sens. Lett.*, **2**, 404–408, <https://doi.org/10.1109/LGRS.2005.853571>.



- Meehl, G. A., F. Zwiers, J. Evans, T. Knutson, L. Mearns, and P. Whetton, 2000: Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. *Bull. Amer. Meteor. Soc.*, **81**, 427–436, [https://doi.org/10.1175/1520-0477\(2000\)081<0427:TIEWAC>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0427:TIEWAC>2.3.CO;2).
- Milly, P. C. D., R. T. Weatherald, K. A. Dunne, and T. L. Delworth, 2002: Increasing risk of great floods in a changing climate. *Nature*, **415**, 514–517, <https://doi.org/10.1038/415514a>.
- Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GICP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, <https://doi.org/10.1029/2003JD003823>.
- Mosbacher, R., W. E. Evans, and E. W. Friday Jr., 1989: Cooperative station observations. National Weather Service Observing Handbook No. 2, 6–25.
- Moser, B. A., W. A. Gallus Jr., and R. Mantilla, 2015: An initial assessment of radar data assimilation on warm season rainfall forecasts for use in hydrologic models. *Wea. Forecasting*, **30**, 1491–1520, <https://doi.org/10.1175/WAF-D-14-00125.1>.
- Mutel, C., 2010: *A Watershed Year: Anatomy of the Iowa Floods of 2008*. University of Iowa Press, 284 pp.
- Nanding, N., M. A. Rico-Ramirez, and D. Han, 2015: Comparison of different radar-rain gauge rainfall merging techniques. *J. Hydroinf.*, **17**, 422–445, <https://doi.org/10.2166/hydro.2015.001>.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I—a discussion of principles. *J. Hydrol.*, **10**, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Pagano, T. C., and Coauthors, 2014: Challenges of operational river forecasting. *J. Hydrometeorol.*, **15**, 1692–1707, <https://doi.org/10.1175/JHM-D-13-0188.1>.
- Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the high-resolution rapid refresh model's ability to predict mesoscale convective systems using object-based evaluation. *Wea. Forecasting*, **30**, 892–913, <https://doi.org/10.1175/WAF-D-14-00118.1>.
- Quintero, F., and W. F. Krajewski, 2018: Conflating the Iowa Flood Center high-resolution river network with the National Water Center NHDPlus network. *J. Amer. Water Resour. Assoc.*, **54**, 28–39, <https://doi.org/10.1111/1752-1688.12554>.
- , —, R. Mantilla, S. Small, and B.-C. Seo, 2016: A spatial-dynamical framework for evaluation of satellite rainfall products for flood prediction. *J. Hydrometeorol.*, **17**, 2137–2154, <https://doi.org/10.1175/JHM-D-15-0195.1>.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, [https://doi.org/10.1175/1520-0434\(1990\)005<0570:TCSIAA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2).
- Seo, B.-C., and W. F. Krajewski, 2015: Correcting temporal sampling error in radar-rainfall: Effect of advection parameters and rain storm characteristics on the correction accuracy. *J. Hydrol.*, **531**, 272–283, <https://doi.org/10.1016/j.jhydrol.2015.04.018>.
- , —, A. Kruger, P. Domaszczynski, J. A. Smith, and M. Steiner, 2011: Radar-rainfall estimation algorithms of Hydro-NEXRAD. *J. Hydroinf.*, **13**, 277–291, <https://doi.org/10.2166/hydro.2010.003>.
- , L. K. Cunha, and W. F. Krajewski, 2013: Uncertainty in radar-rainfall composite and its impact on hydrologic prediction for the eastern Iowa flood of 2008. *Water Resour. Res.*, **49**, 2747–2764, <https://doi.org/10.1002/wrcr.20244>.
- , W. F. Krajewski, and K. V. Mishra, 2015: Using the new dual-polarimetric capability of WSR-88D to eliminate anomalous propagation and wind turbine effects in radar-rainfall. *Atmos. Res.*, **153**, 296–309, <https://doi.org/10.1016/j.atmosres.2014.09.004>.
- Small, S. J., J. O. Laurent, R. Mantilla, R. Curtu, L. K. Cunha, M. Fonley, and W. F. Krajewski, 2013: An asynchronous solver for systems of ODEs linked by a directed tree structure. *Adv. Water Resour.*, **53**, 23–32, <https://doi.org/10.1016/j.advwatres.2012.10.011>.
- Smith, J. A., M. L. Baeck, G. Villarini, D. B. Wright, and W. F. Krajewski, 2013: Extreme flood response: The June 2008 flooding in Iowa. *J. Hydrometeorol.*, **14**, 1810–1825, <https://doi.org/10.1175/JHM-D-12-0191.1>.
- Van Aalst, M. K., 2006: The impact of climate change on the risk of natural disasters. *Disasters*, **30**, 5–18, <https://doi.org/10.1111/j.1467-9523.2006.00303.x>.
- Velasco-Forero, C. A., D. Sempere-Torres, E. F. Cassiraga, and J. J. Goimez-Hernandez, 2009: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Adv. Water Resour.*, **32**, 986–1002, <https://doi.org/10.1016/j.advwatres.2008.10.004>.
- Wilson, J. W., N. A. Crook, C. K. Mueller, J. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: A status report. *Bull. Amer. Meteor. Soc.*, **79**, 2079–2099, [https://doi.org/10.1175/1520-0477\(1998\)079<2079:NTASR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<2079:NTASR>2.0.CO;2).
- Zhang, J., Y. Qi, C. Langston, B. Kaney, and K. Howard, 2014: A real-time algorithm for merging radar QPEs with rain gauge observations and orographic precipitation climatology. *J. Hydrometeorol.*, **15**, 1794–1809, <https://doi.org/10.1175/JHM-D-13-0163.1>.
- , and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, <https://doi.org/10.1175/BAMS-D-14-00174.1>.