AMERICAN METEOROLOGICAL SOCIETY

Journal of Hydrometeorology

EARLY ONLINE RELEASE

The DOI for this manuscript is doi: 10.1175/JHM-D-18-0080.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Seo, B., W. Krajewski, F. Quintero, M. Elsaadani, R. Goska, L. Cunha, B. Dolan, D. Wolff, J. Smith, S. Rutledge, and W. Petersen, 2018: Comprehensive Evaluation of the IFloodS Radar-Rainfall Products for Hydrologic Applications. J. Hydrometeor. doi:10.1175/JHM-D-18-0080.1, in press.

Comprehensive Evaluation of the IFloodS Radar-Rainfall Products for Hydrologic Applications

BONG-CHUL SEO[1], WITOLD F. KRAJEWSKI[1], FELIPE QUINTERO[1], MOHAMED ELSAADANI[1], RADOSLAW GOSKA[1], LUCIANA K. CUNHA[2], BRENDA DOLAN[3], DAVID B. WOLFF[4], JAMES A. SMITH[2], STEVEN A. RUTLEDGE[3], AND WALTER A. PETERSEN[5]

[1]IIHR—Hydroscience & Engineering, University of Iowa, Iowa City, Iowa
[2]Department of Civil & Environmental Engineering, Princeton University, Princeton, New Jersey
[3]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado
[4]NASA Wallops Flight Facility, Wallops Island, Virginia
[5]NASA Marshall Space Flight Center, Huntsville, Alabama

Revised Manuscript Submitted to

Journal of Hydrometeorology

September 2018

Corresponding author address: Bong-Chul Seo, IIHR—Hydroscience & Engineering, University of Iowa, Iowa City, IA 52242.
E-mail: bongchul-seo@uiowa.edu

Abstract

This study describes the generation and testing of a reference rainfall product created from field campaign datasets collected during the NASA Global Precipitation Measurement mission Ground Validation (GPM/GV) Iowa Flood Studies (IFloodS) experiment. The study evaluates ground-based radar-rainfall (R-R) products acquired during IFloodS in the context of building the reference rainfall product. The purpose of IFloodS was not only to attain a high quality ground-based reference for the validation of satellite rainfall estimates but also to enhance understanding of flood-related rainfall processes and the predictability of flood forecasting. We assessed the six R-R estimates (IFC, Q2, CSU-DP, NWS-DP, Stage IV, and Q2-Corrected) using data from rain gauge and disdrometer networks that were located in the broader field campaign area of central and northeastern Iowa. We performed the analyses with respect to time scales ranging from 1-h to the entire campaign period in order to compare the capabilities of each R-R product and to characterize the error structure at scales that are frequently used in hydrologic applications. The evaluation results show that the Stage IV estimates perform superior to other estimates, demonstrating the need for gauge based bias corrections of radar-only products. This correction should account for each product's algorithm-dependent error structure that can be used to build unbiased rainfall products for the campaign reference. We characterized the statistical error structures (e.g., systematic and random components) of each R-R estimates and used them for the generation of campaign reference rainfall product. To assess the hydrologic utility of the reference product, we performed hydrologic simulations driven by the reference product over the Turkey River basin. The comparison of hydrologic simulation results demonstrates that the campaign reference product performs better than Stage IV in streamflow generation.

1. Introduction

Rainfall estimates from ground-based radars are often used as a reference to assess the capabilities and limitations inherent in using space-based rainfall estimates in hydrologic modeling and prediction (e.g., Schumacher and Houze 2000; Chandrasekar et al. 2008; Villarini et al. 2009). During the period of late spring to early summer in 2013, the National Aeronautics and Space Administration (NASA) conducted a hydrology-oriented field campaign called Iowa Flood Studies (IFloodS) in collaboration with the Iowa Flood Center (IFC) at The University of Iowa. This field campaign sought to enhance the understanding of flood-related rainfall processes and the prediction capability in flood forecasting as well as to support activities of Global Precipitation Measurement (GPM) Ground Validation (see e.g., Hou et al. 2014; Skofronick-Jackson et al. 2017). A number of scientific instruments were deployed in central and northeastern Iowa to collect high quality precipitation data and thus improve flood forecasting capabilities (Petersen and Krajewski 2013). Therefore, this unique campaign can be understood in the context of many other NASA's field experiments briefly summarized in Dolan et al. (2018).

While multiple types of rainfall data sets (e.g., satellite, radar, rain gauge and disdrometer, and many others) are available through IFloodS, we focus on evaluating the ground-based radar-rainfall (R-R) composite products. The utility of evaluating the R-R products and characterizing their uncertainties is toward the goal of building a campaign reference product for the satellite data validation and distributed hydrologic modeling (e.g., Reed et al. 2004; Smith et al. 2004). The radar-only products used in the evaluation are the U.S. Next-Generation Radar (NEXRAD) single-polarization (SP) estimates (i.e., National Mosaic and QPE Q2 and IFC) and products generated using dual-polarization

(DP) procedures (i.e., the U.S. National Weather Service operational and Colorado State University experimental blended precipitation processing algorithms). We also compare these radar-only products with rain gauge corrected R-R estimates (Stage IV and Q2-Corrected products). We explore the algorithm-dependent features (e.g., SP vs. DP) among the R-R estimates based on the comprehensive analyses of product inter-comparison. The uncertainty for different temporal and spatial resolution products is also characterized using ground reference of dense rain gauge and disdrometer networks. This multi-scale characterization is required for hydrologic modeling frameworks that assess model predictive abilities as a function of space and time scales. Based on the evaluation and error characterization results, we create the campaign reference product by combining selected R-R estimates with the data from the NASA polarimetric radar (NPOL) that was placed at the center of the campaign domain. We do not include the detailed evaluation of NPOL R-R estimates in this study because the comparison between individual (e.g., NPOL) and composite products would not be fair, and individual radar products are often affected by significant range effects (e.g., Fabry et al. 1992) that are less impactful for composite products. A detailed evaluation and the performance of NPOL estimates are documented in Chen et al. (2017). We also drive the IFC hillslope-link model (HLM) using the reference product over the Turkey River basin in Iowa and assess its predictive capability in flood prediction.

The paper is structured as follows. In Section 2, we introduce the study area in which the IFloodS campaign was conducted and discuss the data sets of the R-R products, rain gauge, and disdrometer. Section 3 describes the methodology we used for the R-R product evaluation and error characterization in this study. In Section 4, we present the

comparison and evaluation results and discuss the observed similarity and discrepancy among the R-R products. In Section 5, we provide a procedure to create the campaign reference product and evaluate its hydrologic utility. Section 6 summarizes and discusses the main findings and required future works.

2. Data

In this section, we briefly introduce the IFloodS study area and describe the R-R products collected during the campaign and the ground reference data sets (i.e., rain gauge and disdrometer data) used for the evaluation of the collected products. The IFloodS domain consists of central and northeast Iowa, and the major basins in the area are the Cedar and Iowa River basins in the middle of the domain and the Turkey River basin near the northeast Iowa border (Fig. 1). As a result of the record flood that occurred in these basins in 2008, they have been used in a number of hydrologic studies to investigate various hydrologic factors (see e.g., Gupta et al. 2010; Cunha et al. 2012; Seo et al. 2013; Smith et al. 2013; Ayalew et al. 2014). The basin areas are visible from the existing network of NEXRAD radars (KARX in La Crosse, Wisconsin; KDMX in Des Moines, Iowa; KDVN in Davenport, Iowa; and KMPX in Minneapolis, Minnesota). Despite the fact that the field deployment of rainfall measuring devices started as early as April, we define the analysis time window as the period of 1 May to 15 June 2013 in order to synchronize different periods of collected data and products. We refer to this time window as the "official" IFloodS period. Further details of the precipitation events that occurred during the period is described in Cunha et al. (2015) and Seo et al. (2015a).

a. Radar-rainfall products

We acquired six NEXRAD based rainfall composite products through the campaign: the Iowa Flood Center (IFC) real-time product; the National Mosaic and QPE next generation Q2; the Colorado State University dual-polarization product (CSU-DP); the National Weather Service real-time DP product (NWS-DP); the National Center for Environmental Prediction (NCEP) Stage IV analysis; and the Q2 product with rain gauge correction (Q2-Corrected). As featured in Table 1, these composite products can be categorized into three types: radar-only SP (IFC and Q2), radar-only DP (CSU-DP and NWS-DP), and rain gauge corrected (Stage IV and Q2-Corrected) products. We provide a brief comparison of space and time resolutions and estimation algorithms of each product in Table 1.

Using processing algorithms documented in Krajewski et al. (2013), the IFC rain rate map is generated every 5-min with a grid spacing of a quarter decimal minute (approximately 0.5 km). Seven NEXRAD radars (KEAX in Kansas City, Missouri; KFSD in Sioux Falls, South Dakota; KOAX in Omaha, Nebraska; and four more radars discussed earlier) are used to cover the entire State of Iowa, whereas the Q2 product is created with a 5-min and one-hundredth decimal degree (approximately 1 km) resolution over the entire United States. While the IFC uses a single NEXRAD Z-R equation ($Z=300R^{1.4}$), the Q2 algorithm uses four different Z-Rs (see Table 1) that depend on the precipitation type classification based on the three dimensional structure of reflectivity and environmental (atmospheric) variables with physically based heuristic rules (Zhang et al. 2011). We note that there have been many changes and improvements in Q2, and it is now called MRMS

(Multi-Radar Multi-Sensor; Zhang et al. 2016). The evaluation of Q2/MRMS and their comparison with Stage IV are reported in Chen et al. (2013) and Zhang et al. (2016).

For the CSU-DP product, the radar Level II volume data (e.g., Kelleher et al. 2007) from four radars (KARX, KDMX, KDVN, and KMPX) were post-processed, not in real-time, using a DP algorithm called CSU-HIDRO (Cifelli et al. 2011) and a hybrid scan algorithm documented in Seo et al. (2011) for combining multiple elevation angle data (for more details on the product generating procedures, refer to Seo et al. 2015a in the same Special Collection). The time and space resolution is identical to that of the IFC product. The CSU-DP product covers only the IFloodS domain and does not provide full coverage of the entire State of Iowa. Regarding the creation of the NWS-DP product, the instantaneous precipitation rate (Level III) products, generated using the algorithm reported in Istok et al. (2009), were collected for involved radars. We applied the procedure described by Cunha et al. (2013) to generate hourly rainfall accumulations based on the instantaneous precipitation rate. We then combined the data from the individual radars into a composite map using the exponential decaying scheme (e.g., Zhang et al. 2005) that assigns weights calculated by the distance from individual radars for a given location. Most DP algorithms are based on the procedures of identifying hydrometeor types and selecting relevant rain rate estimators. Both the CSU (e.g., Lim et al. 2005) and NWS (e.g., Park et al. 2009) identification algorithms use a similar fuzzy logic, but the architecture of the classification procedure is different in terms of the input variables and membership functions employed. These DP identification algorithms contain part of data quality control (see Table 1) and yield categories of non-precipitation radar echoes (e.g., ground clutter and biological returns) as well as hydrometeor types. The comparison of NWS-DP

and -SP products, as well as the effect of hydrometeor identification is documented in Cunha et al. (2013).

The Stage IV product (Lin and Mitchell 2005; Wu et al. 2012) consists of hourly-based rain gauge corrected precipitation analyses with some manual quality controls that are performed by forecasters in the River Forecast Center (RFC). The rainfall maps that cover each individual RFC are collected at NCEP and are then combined into a national coverage based on the 4 km Hydrologic Rainfall Analysis Project (HRAP; see e.g., Reed and Maidment 1999). The Q2-Corrected product represents the hourly Q2 bias correction using a national network of rain gauges (e.g., Kim et al. 2009), and the detailed procedures are documented in Zhang et al. (2011).

A comparison of the IFC and Q2 SP estimation algorithms presented in Table 1 demonstrates that different Z-R relations can be used even for identical meteorological targets, depending on the result of precipitation classification in Q2. This can lead to a major discrepancy between the two SP products. The DP algorithms initially implement a sophisticated quality control method (e.g., Ryzhkov and Zrnic 1998; Park et al. 2009) to eliminate non-precipitation echoes that have been identified during the hydrometeor classification step and then apply a designated relationship between rain rate and measured radar variables (i.e., differential reflectivity, specific differential phase, or horizontal reflectivity) according to the classified types. There are two major differences in defining the rain rate conversion between the CSU and NWS algorithms (we do not discuss the difference in the hydrometeor identification procedures): (1) there is no rain rate estimation in the CSU algorithm when the radar beam observes the melting layer or ice region; and (2) the CSU algorithm uses both the specific differential phase (K_{dp}) and differential

reflectivity (Z_{dr}) for the liquid phase, while the NWS algorithm seems to rely more heavily on Z_{dr}. Section 4 discusses the algorithm-derived differences in rain rate estimation among products in more detail.

b. Rain gauge and disdrometer data

We use rain gauge and disdrometer data as a ground reference to evaluate the collected R-R products. We acquired data from local networks that were operated by NASA, IFC, USDA Agriculture Research Service (ARS), and the University of Wyoming as well as the national networks of Automated Surface Observing System (ASOS; Clark 1995), Automated Weather Observing System (AWOS), and National Weather Service Cooperative Observer Program (NWS COOP; Mosbacher et al. 1989). As illustrated in Fig. 1, we selected rain gauge sites that effectively cover the IFloodS study area and discussed basins.

NASA deployed 20 and 5 rain gauge platforms each with double tipping-bucket gauges in the Turkey River basin and South Fork Iowa River basin, respectively. Likewise, 20 NASA-owned disdrometers (14 Autonomous Parsivel Unit (APU) and six two-dimensional video (2DVD) types) were deployed and distributed along the southeast direction from the domain center (some of them were clustered). Additionally, 30 IFC gauges, which were similar to the platforms in the NASA network, were clustered around Iowa City, and four more IFC gauge platforms were deployed in central Iowa. The NASA and IFC rain gauge networks transferred the recorded time-of-tip data, and the accumulated number of tips data with a 5-min resolution were used. The ARS deployed 15 rain gauges within the South Fork Iowa River basin, and all gauges were equipped with double tipping-buckets (Coopersmith et al. 2015). The University of Wyoming group placed four triple

tipping-bucket gauges with soil moisture probes in the vicinity of the IFC network. The ASOS and AWOS data were collected with their original resolutions (i.e., 1- and 5-min, respectively) and accumulated over the designated time intervals (e.g., hourly). The use of the NWS COOP data was limited to the rain total and daily analyses because the network only reports data daily.

3. Methodology

We provide the analysis procedures that are associated with product evaluation and error characterization with respect to multi time scale ranging from 1-h to the entire campaign period. As shown in Table 1, with respect to radar-rain gauge (R-G) comparison, the majority of the R-R products has comparable spatial resolutions (0.5 and 1 km), so it is assumed that point rainfall measurements from rain gauges well represent the areal rainfall over such spatial scales. This assumption can be justified for given time scales (e.g., hourly) of the analyses because rainfall spatial variability is relatively small at such spatial (even for the 4 km resolution of Stage IV) and temporal scales (e.g., Villarini et al. 2008). This enables direct R-G comparison without considering a spatial sampling disagreement (e.g., Seo and Krajewski 2011) between different measuring devices (e.g., radar vs. rain gauge). Sub-hourly scale (e.g., 15- and 30-min) evaluation may require much denser rain gauge networks because gauge representativeness decreases (rainfall spatial variability increases) at finer temporal scales.

a. Product evaluation

The evaluation began by examining and comparing accumulated rain totals for the entire campaign period. We present and discuss the observed discrepancies that arise from

the different estimation algorithms among all acquired R-R products. In addition, we perform an R-G comparison analysis to assess campaign totals at the ground reference locations. We also use the PRISM (Parameter-elevation Regressions on Independent Slopes Model) rain gauge interpolated analysis (Daly et al. 2008) as a gridded reference for the campaign total (we assume that the rainfall spatial variability at the time scale of the entire period and spatial scale of 4 km is relatively small), which enables us to explore the spatial error structure of each R-R product at the six-week period scale.

While the campaign total analysis solely reveals overall agreement with ground reference data, the temporal variation of the error (over- or under-estimation depending on individual events) might be somewhat compensated for and concealed in this analysis. In fact, different rainfall estimators (see Table 1) determined by the classification procedures and their outcomes tend to be sensitive to individual rain events. The classification and resulting estimators among different R-R products that are associated with meteorological regimes and storm types could become major factors of estimation errors (e.g., Rosenfeld et al. 1995; Anagnostou 2004). Consequently, we selected two precipitation cases that were identified as the snow/mix (with stratiform rain) and mesoscale convective system events, respectively. We present the results of the R-G comparison analysis and discuss possible reasons that algorithm-derived discrepancies were detected. This event-based analysis allows for the persuasive assessment of the potential benefits of using DP vs. SP algorithms as well as exposes the basic performance of each algorithm.

In the multi-scale R-G comparison, we use time resolutions of 1-, 3-, 6-, 12-, and 24-h. For those accumulation times, we integrated the R-R products and ground reference data over the corresponding time intervals from the original data resolutions. If missing

minutes or hours in a specific accumulation window exceed 10% of the designated time interval, the corresponding hour data are regarded as missing and are excluded from the analysis. We define the systematic tendency of the R-R products using the overall and conditional bias terms. We also employ two more statistical metrics, the correlation coefficient and root mean square error normalized by the mean of rain gauge data (normalized RMSE), to measure the accuracy across the presented time scales and to compare performance among the R-R products.

b. Error characterization

In general, the error is identified as the discrepancy between the true and estimated rainfall, and we use rain gauge measurements as a reference against R-R estimates at radar pixels that are collocated with the gauge location. As we discussed earlier in this section, sampling scale disagreements between rain gauges and radars are less impactful with respect to the temporal (1- to 24-h) and spatial (0.5 - 4 km) scales used in this analysis. The R-R estimation error is typically defined using two mathematical notions of multiplicative and additive terms (represented as the proportions/ratios and differences, respectively), and both terms have been employed numerous times in the literature. In this study, we adopt the multiplicative term of the error/bias to characterize the error structure of the acquired R-R products. The full procedures of error characterization conforms to the one in Ciach et al. (2007).

As an initial step in the error characterization, we estimate and eliminate a systematic or climatological tendency, which is described as the overall bias factor (B):

$$B = \frac{\sum_t R_g(t)}{\sum_t R_r(t)} \quad (1)$$

where $R_g(t)$ and $R_r(t)$ denote rain gauge and radar rainfall at a time step t, and R-G data pairs are aggregated for all of the time steps in the period in order to calculate the overall bias factor. This value should be unique for the same R-R product regardless of the data accumulation time scales if the R-G data pairs at any time scale are not significantly affected by missing data or gaps. After removing this overall bias, we need to account for the over- or under-estimation that occurs depending on the R-R magnitude (e.g., Katz and Murphy 1997; Ciach et al. 2000). This behavior can be determined by the concept of conditional expectation function, $h(\cdot)$:

$$h(r_r) = E\{R_{true}|R_r = r_r\} \tag{2}$$

where $E\{\cdot\}$ denotes an expectation function, R_r is a random variable, and r_r is a specific R-R value. The function, $h(r_r)$, implies a systematic distortion describing the conditionality of error on the R-R magnitude. This tendency can be estimated using the non-parametric kernel smoothing regression (e.g., Nadaraya 1964) or the two-parameter (a_h and b_h) power law function:

$$h(r_r) = a_h r_r^{b_h} \tag{3}$$

Although Eqs. (1) and (3) account for the systematic behaviors of R-R estimates, there is a remaining component describing a stochastic process of random uncertainties. We address this random component by estimating conditional variance of the error in Eq. (4) and use three-parameter (σ_{0e}, a_e, and b_e) function in Eq. (5) to take into account the random factor:

$$\sigma_e^2(r_r) = E\left\{\left(\frac{R_{true}}{h(r_r)} - 1\right)^2 |R_r = r_r\right\} \tag{4}$$

$$\sigma_e(r_r) = \sigma_{0e} + a_e r_r^{b_e} \tag{5}$$

where σ_e denotes the standard deviation of the multiplicative error, and the estimated random feature is used to combine selected R-R estimates for the campaign reference products. In our error characterization, we assumed that the conditional mean and standard deviation of the R-R error are stationary for convenience in modeling because accounting for non-stationarity in the modeling procedure is challenging. We also note that other factors (e.g., radar beam altitude) can be considered in modeling errors depending on product types (e.g., individual vs. composite) while we used rain rate as a main factor in this study.

4. R-R product evaluation

In this section, we present the analysis results of product rain totals (for the entire period and two selected events) and the statistical evaluation of the products with respect to diverse time scales (1-, 3-, 6-, 12-, and 24-h). The former analysis compares the total amounts of rainfall for the specified periods among the R-R products and the ground reference data and assesses the algorithm-dependent strengths and weaknesses of each product. The latter approach discloses the statistical structure of the product error and provides useful information for the reference product generation and hydrologic modeling.

a. Difference among products

1) CAMPAIGN TOTALS

The rain maps of the campaign totals for the "official period" (1 May through 15 June) are illustrated in Fig. 2, in which the rain gauge-corrected (Stage IV and Q2-Corrected), radar-only SP based (IFC and Q2), and radar-only DP based (CSU-DP and NWS-DP) products are aligned from the left to the right panels. While we evaluate the R-

R products using the ground reference data for the spatial domain, as shown in Fig. 1, we present the campaign totals for the entire State of Iowa in Fig. 2. For that reason, the CSU-DP map in Fig. 2 that we created using the data from the four NEXRAD radar only (see Section 2) shows some of the missed (gray) rain area that is not covered by the four radars. The main features of the rainfall spatial structure is captured in most composite products, with some differences. The CSU-DP product shows certain range rings at far ranges from the individual radar locations because the CSU-DP algorithm does not estimate rain rate when the radar beam interacts with the melting layer or ice regions, and the chance of detecting the cold regions increases at far range with higher sampling altitudes. On the other hand, the NWS-DP exposes the inconsistency among individual radar observations (maybe due to radar calibration errors, and we will discuss this issue in the conclusion section) as well as quality control issues such as the wind farm effects discussed in Seo at al. (2015b). The wind farm locations are also clearly visible in the Q2 and Q2-Corrected products in Fig. 2.

To ensure that the PRISM rain gauge interpolation analysis can be used as gridded reference at the campaign total scale, we first evaluate the PRISM rain totals with rain gauge observations (Fig. 3). Despite the fact that only the ASOS and NWS COOP rain gauge network data are incorporated in the PRISM analysis shown in Fig. 3 (left panel), the PRISM analysis agrees well with the IFC, NASA, and ARS network data near the one-to-one line as shown in Fig. 3 (right panel). This agreement with the independent network data confirms that the PRISM estimation is reliable as a reference (only at the scale of rain totals) and allows a further analysis to show the spatial pattern of product error using the normalized error/difference term:

$$\frac{RR_{total} - PRISM_{total}}{PRISM_{total}} \quad (6)$$

where RR_{total} denotes the campaign totals from the six R-R products presented in Fig. 2. The R-R products are resampled (averaged) with the same spatial resolution of the PRISM (4 km), and the normalized error calculated by Eq. (6) is mapped in Fig. 4. The blue and red colors used in Fig. 4 distinguish under- and over-estimation patterns, respectively. Since the campaign totals of Stage IV (upper-left panel in Fig. 2) and PRISM (left panel in Fig. 3) look quite similar, the Stage IV error shown in Fig. 4 is even less than in the others, which implies that the rain gauge correction in Stage IV was successful. We note that a small number of ASOS rain gauges (e.g., 15 in Iowa) are commonly used for both PRISM and Stage IV. The Q2-corrected estimates tend to reduce the error that was originally observed in the Q2, but some errors remain. The IFC in Fig. 4 shows underestimation mostly in the northeast and some overestimation within the domain of the KOAX radar. The area covered by the KFSD represents some radar beam blockage effects and significant differences with surrounding radars (e.g., KOAX and KDMX). It is likely that the KFSD difference from other radars detected in the IFC product was suitably handled in the Q2 and NWS-DP algorithms. However, the Q2 and NWS-DP, shown in Fig. 4, introduce other questions regarding substantial overestimation in the KOAX and KDMX regions. The CSU-DP shows mostly underestimation due to range effects.

In Fig. 5, we present rain gauge comparison results. The rain gauge corrected products show relatively good agreement but indicate slight overestimation. While the Q2 and NWS-DP demonstrate significant overestimation, as seen in Fig. 4, the IFC and CSU-DP appear as underestimation. Although the dots denoting the R-G pairs in the IFC are clustered along the one-to-one line in Fig. 5, the scatter is relatively larger than that in the

rain gauge corrected one (e.g., Stage IV), and many more dots (rain gauges in the blue area in the upper-middle panel of Fig. 4) are densely placed in the underestimation area.

2) EVENT TOTALS

We selected two example precipitation cases to demonstrate the algorithm-derived capabilities of R-R estimation. The first case is defined as a snow/mix case with stratiform rain during the period of 2–4 May. The second one, which took place during 27–30 May, was relatively wetter and is characterized by some convective systems. Some of the convective storms were followed by widespread stratiform storms. For the detailed meteorological characteristics of these two events, refer to Seo et al. (2015a) in the same Special Collection. Figure 6 shows three- and four-day event rain totals of the R-R products with the same configuration that is seen in Fig. 2. Figure 7 also presents the event based R-G comparison.

Regarding the snow/mix case, it is hard to conclude that the DP algorithm shows superior performance (see Fig. 7a) and, in particular, the CSU-DP shows significant low estimation, which is due to the range limitation arising from the detection of a low level melting layer. However, we note that the rain gauge measurement in such a cold case might contain errors as well because the rain gauges used in this study are mostly non-heated tipping bucket types. We think that these probable errors, if any, were not considerable because the frozen and mixed snow transitioned to stratiform rain after a short duration of snow. For the SP algorithm comparison, the IFC does not capture the rainfall feature in the northeast (Fig. 6a), which appears in the Q2 and Stage IV. It is likely that the rain type classification and the application of different *Z*-*R*s in Q2 (see Fig. 8a) lead to this observed difference. The *Z*-*R* curves illustrated in Fig. 8a demonstrate that the snow

and stratiform (represented as "M-P" in Fig. 8a) types result in a larger rain rate at the lower reflectivity range (e.g., 0-30 dBZ) than the unique *Z*-*R* ("NEXRAD") used in the IFC algorithm does.

In Fig. 7b, the convective example shows better agreement in both the SP and DP algorithms than the cold one. In this case, the DP tends to work better than the SP in terms of the scatter and the R-G pair alignment on the one-to-one line. Particularly, the superior performance of the CSU-DP is noticeable (upper-right panel in Fig. 7b). Most dots are aligned and concentrated on the line (indicates very good agreement) except for some NWS COOP gauge locations (yellow dots). As the COOP gauges are well distributed over the analysis domain, the disagreement can be interpreted by the observed range issue (in the upper-right panel in Fig. 2) due to the fact that some of the gauges are located far from the radars. Regarding the observed difference between the DP estimates in Fig. 7b, one probable deriving factor could be DP variables used in rain rate estimation. Since the CSU-DP shows better agreement at its observable range, the rain rate estimation in the CSU-DP based on both K_{dp} and Z_{dr} for the liquid phase (see Fig. 8b) seems more reliable than that in the NWS-DP, mostly based on Z_{dr} and Z_h. The detailed equations related to each phase in Fig. 8b are listed in Table 1. Concerning the SP estimate comparisons in Fig. 7b, it seems likely that the use of the "tropical" *Z*-*R* (Fig. 8a) in Q2 generates some overestimation and differences between the IFC and Q2.

b. Multi-scale comparison

We evaluated the R-G agreement with respect to the diverse accumulation time scales (1-, 3-, 6-, 12-, and 24-h) that are frequently used for various hydrologic models. Figure 9 shows two-dimensional histograms of the hourly R-G comparison. The different

colors in Fig. 9 indicate data occurrences for the given R-G magnitude with a 1mm resolution. The overall bias values are placed in the upper-right corner of each panel and imply under- (>1) or over-estimation (<1) of the R-R estimates. The overall tendency of under- or over-estimation presented in Fig. 9 is similar to that observed in Figs. 4 and 5. We performed the same analysis for other accumulation time scales and confirmed that the bias values were in the same range and exhibited smaller scatter as time scale increases. The Stage IV in Fig. 9 reveals relatively frequent false detection on the x (radar) axis at ranges smaller than 30 mm. We speculated that the false detection in Stage IV might arise from a mismatch of spatial scales (point vs. 4×4 km^2) and the small scale variability of rainfall.

In Table 2, we present three statistical metrics (overall bias, correlation coefficient, and normalized RMSE) from the six R-G data sets that pertain to five time scales. The overall bias should not change with time scale if there is no significant effect from missing data or gaps in data. Therefore, we calculated the overall bias values presented in Table 2 from the hourly R-G data. Figure 10 illustrates the change in correlation and RMSE with respect to time scale and demonstrates that temporal aggregation results in better R-G agreement with increasing correlation and decreasing RMSE. However, the correlation at a longer time span (e.g., 24-h) slightly decreases for most products in Fig. 10 (left panel), probably because of adding the NWS COOP data set to the analysis. Particularly, the most significant correlation drop observed in the CSU-DP at the 24-h scale is caused by the fact that some COOP gauges located far from the radars are not within the observable range of the CSU-DP, as seen in Fig. 2. Overall, the rain gauge corrected (Stage IV) product shows statistically superior performance in all metrics, assuming that the difference in spatial

resolution (4 vs. 0.5 and 1 km) is negligible. Despite the given range limitation, the CSU-DP agrees well with rain gauge observations at all scales, and its agreement is comparable to that of the Stage IV in both correlation and RMSE. Based on the presented metrics, the NWS-DP does not seem much better than the SP products.

c. Error characterization

The error structure of the R-R products is characterized for the aforementioned time scales. As discussed in Section 3, the overall systematic tendency in the R-R field is first eliminated by a simple multiplication of the bias value (see Table 2) by the R-R estimates. For the next step, we use both non-parametric and parametric regression methods to model the remaining conditional bias. The advantage of employing the non-parametric approach is that the bias structure/behavior is not restricted by the predefined function as used in the parametric approach. However, the curve (conditional bias) pattern estimated by the non-parametric Gaussian smoothing was inconsistent and showed abrupt changes at large R-R range with shorter time scales (e.g., 1-h). This behavior can be attributed to the limited sample size, which implies that there are few large R-R values at the hourly scale for the given six-week period (but temporal aggregation increases the number of large R-R values). For that reason, we present the results from only parametric application in Fig. 11 for the R-R products and all time scales. The common aspect observed in Fig. 11 is that the R-R data aggregation over longer time spans reduces the conditional bias. Table 3 presents the parameters of the power function defined in (3). The estimated curve feature shown in Fig. 11 and parameter values in Table 3 are comparable to those in Ciach et al. (2007). The presented conditional structure is useful for hydrologic applications that are forced by the R-R estimates because the systematic difference in rainfall volume tends to significantly

affect errors in streamflow simulations/predictions (see e.g., Seo et al. 2013). We describe the results of random error structure in Section 5 because the random structure is used to combine different R-R estimates in generating the reference product. We indicate that the error models provided in this study represent the uncertainty features averaged over the six-week campaign period, and the uncertainty features may vary with different events or seasons.

5. Reference product

a. Reference product generation

This section describes the procedures for creating the campaign reference rainfall product. These procedures involve the systematic error (overall and conditional biases) correction of the R-R estimates and their weighted combinations, calculated using the relative magnitude of random errors. The random errors are characterized by the standard deviation of remaining errors after correcting the R-R estimates for the overall and conditional biases. We assumed the random error as a normal distribution because we removed the effects of bias and skewness in the R-R estimates. We tested Stage IV, Q2-Corrected, IFC, CSU-DP, and two NPOL DP products as the ingredients of reference product. The NPOL estimates are each called NPOL-RR and -RC, identified by the data processing and QPE algorithms known as DROPS2 (Pippitt et al. 2015; Chen et al. 2017) and CSU-HIDRO (Cifelli et al. 2011). We excluded NWS-DP and Q2 because of the relatively low performance shown in the evaluation and the presence of Q2-Corrected, respectively. As shown in Fig. 12, we estimated the random error function in Eq. (5) for the four selected R-R composite and two NPOL products. The parameters of conditional

bias function in Eq. (3) were also estimated for the NPOL products. We conducted the parameter estimation for both conditional mean and random components for the hourly scale at which the reference product is generated.

We examined a variety of combinations using all the ingredient products, and the resultant candidates for the reference product looked more or less similar (at the scale of a six-week period), mainly because of the bias correction used in the combining procedure. We then selected the statistically better ones through an independent evaluation at the scale of campaign totals. For the selected reference, we used and combined Stage IV, Q2-Corrected, IFC, CSU-DP, and NPOL-RC. Figure 13 illustrates the map of campaign totals of the selected reference product and its independent evaluation using NWS COOP and CoCoRaHS (Cifelli et al. 2005) observations. The term "independent" is justified here because the both network stations collect daily reports only, and their observations were not used to quantify uncertainties at the hourly scale shown in Figs. 11 and 12. As shown in Fig. 13, part of the CoCoRaHS observations contains a quality control issue (e.g., missing), and we did not include data from this network in a simple quantitative/statistical evaluation. The calculated bias (G/R) and mean absolute error of the reference product with the COOP observations are 0.97 and 28.3 mm (9.4% of the mean of COOP totals), respectively. Based on the observed agreement with the COOP data (Fig. 13), the reference product appears almost unbiased, which is the most significant element required for hydrologic prediction (e.g., Seo et al. 2013).

b. Hydrologic evaluation

We created the reference product by correcting major uncertainty features (e.g., overall and conditional biases) of the selected R-R products. A direct evaluation or

verification of the reference product at finer scale (e.g., hourly) was not feasible because of the lack of independent ground reference data at the required scale. Rain gauge and disdrometer data collected during IFloodS were all included in the R-R uncertainty characterization and used in the reference product generation procedures. Therefore, in this section, we force a hydrologic model using the reference product and assess its predictive capability in flood forecasting.

We used the IFC hillslope-link model (HLM) to simulate streamflow during the campaign period. This distributed hydrologic model is based on landscape decomposition into hillslopes and channels, and its configuration and governing equations are documented in Krajewski et al. (2017). Here, suffice to say that the model is terrain-based; i.e., it respects water transport in the stream and river network. The key components are: (1) rainfall to runoff transformation at the hillslopes; and (2) water routing in the river channels. The main feature of the HLM is that it is calibration free: the model parameters are determined *a priori*; therefore, the model does not "favor" any particular input product. Model calibration may conceal different aspects in streamflow generation driven by different precipitation forcing products. The use of HLM can be understood in the context of the Prediction in Ungauged Basin (PUB; Sivapalan 2003) initiative because the HLM predictions are not limited to the locations/stations where streamflow observations exist. Although such a physics-based model does not always guarantee accurate predictions, our earlier and ongoing evaluations of HLM (e.g., Cunha et al. 2012; Seo et al. 2013, 2018; Ayalew et al. 2014; Quintero et al. 2016; Krajewski et al. 2017) have indicated its acceptable performance. We selected the Turkey River basin for this hydrologic evaluation because there are five USGS stream gauges providing discharge at a range of spatial scales

from about 450 to about 4,000 km^2. In addition, 20 NASA rain gauges densely deployed within the basin (see Fig. 1) allowed us to test and compare the simulation results driven by gauge-based gridded estimates with those driven by the reference product.

We created a gauge interpolated rainfall product at the hourly scale using a geostatistical procedure known as the optimal interpolation technique, ordinary kriging (e.g., Tabois and Salas 1985). Figure 14 shows the campaign rainfall totals of the reference and gauge interpolation products over the Turkey River basin and also indicates the locations of the USGS stations and NASA rain gauges. We then ran the HLM with the rainfall forcing of the reference, gauge interpolation and Stage IV products. We compare each simulated hydrograph at the five USGS stations with streamflow observations in Fig. 15 and present performance metrics in Table 4 to quantitatively assess the hydrologic prediction capability associated with each rainfall product. We indicate that the rating curve uncertainty was not accounted for in the analysis. The performance metrics used here are Kling-Gupta efficiency (KGE; Gupta et al. 2009), correlation, and normalized RMSE. All simulations started with the same initial conditions; i.e., the amount of water in the soil and in the channels. We can observe from Fig. 15 that the simulation results driven by the gauge interpolation product better agree with the USGS observations than those driven by the R-R products. The gauge interpolation product simulation tends to capture small peaks in May well, while both the reference and Stage IV simulations overestimate these somewhat (the overestimation is more significant in Stage IV). We think that these streamflow overestimations were not caused by the systematic rainfall overestimation in the reference products, but rather by complicated hydrologic processes and interactions between initial soil water content and dynamic changes of rainfall space-

time distribution. We confirmed that there was little difference between the reference and gauge interpolation products in the total amounts of mean areal precipitation (particularly for the event in early May) at all five scales. Regarding the significant event in late May and early June, the simulations driven by the reference product captured the flood peak and timing well for the relatively smaller-scale basin (e.g., at Spillville in Fig. 15). The observed delay of the streamflow peak at Elkader and the noticeable underestimation at Garber do not look like a rainfall issue because all forcing products led to the similar results. Given the evaluation metrics provided in Table 4, we concluded that the overall performance of the reference product in generating streamflow is superior to that of Stage IV.

While the best performance of gauge-only rainfall product may come as a surprise, the set-up in terms of gauge density (one gauge per 200 km^2) and quality (double gauges at each location) would be difficult to repeat in an operational environment. For example, in Iowa, this would require some 800 rain gauge sites. It seems that the best strategy is what has been implemented operationally, i.e., rain gauge corrected radar rainfall. Good performance of the Stage IV and campaign reference products offers solid evidence to support this approach.

6. Summary and conclusions

We evaluated the R-R composite products collected during the NASA IFloodS campaign, which was designed to serve as a high-quality ground-based reference for the validation of satellite rainfall estimates. We characterized the acquired R-R products as the SP (IFC and Q2), DP (CSU-DP and NWS-DP), and rain gauge corrected (Stage IV and Q2-Corrected) estimates. We used data from a number of rain gauge and disdrometer

networks (NASA, IFC, USDA ARS, University of Wyoming, ASOS, AWOS, and NWS COOP) as ground reference to assess the algorithm-derived capability of the R-R products and their potential benefit for hydrologic prediction. Some of these networks were newly deployed, while others were pre-existing within the campaign area. We implemented the performance evaluation and error characterization of the R-R products with respect to multi-scale ranging from 1-h to the entire campaign period.

The analysis of rain totals for the entire period showed significantly different spatial patterns (Fig. 2 and 4) among the R-R products. The R-G comparison analysis verified this discrepancy (Fig. 5), and the rain gauge corrected products (Stage IV and Q2-Corrected) seemed fairly close to the rain gauge observations. All other products exposed either over- or under-estimation properties. In particular, the CSU-DP showed a range limitation because of an algorithm component in which rain rate was not estimated when the radar beam interacted with regions of ice or melting ice. In the event-based analysis, the heavy rain case performance looked better in the DP-based algorithms (based on the R-G comparison in Fig. 7b), but the DP results were not superior to the SP for a presented snow/mix with the stratiform case (Fig. 7a). This implies that the DP algorithms still need improvement (for more detailed evaluation of the DP products and algorithms, refer to Cunha et al. 2015; Seo et al. 2015a in the same Special Collection). In the comparison of the DP algorithms (see Fig. 7b with the exclusion of daily COOP gauges for a fair comparison), it is likely that the algorithm using both K_{dp} and Z_{dr} (CSU-DP) better represents heavy rain than that based on Z_{dr} and Z_h (NWS-DP). The significant relative bias observed around the KOAX radar (NWS-DP in Fig. 6b) seemed to be affected by the calibration errors in either Z_{dr} or Z_h. We confirmed with the Radar Operations Center (ROC)

that the Z_h values of the KOAX radar were somewhat hotter (1.0 – 1.5 dBZ) than those of adjacent radars (e.g., KDMX) for May 2013. The observed underestimation around the KFSD radar for the IFC product is also explained by the relative Z_h bias of -1.5 to -1.0 dBZ. We note that this calibration error is a challenging issue, particularly in the real-time application, and hope that the new Dual-frequency Precipitation Radar (DPR) that was recently launched by the GPM program will help address this problem in radar QPE (e.g., Schwaller and Morris 2011; Warren et al. 2018).

We performed the multi-scale R-G comparison using three statistical metrics: multiplicative bias, the correlation coefficient, and normalized RMSE. As seen in the precedent analyses, the rain gauge corrected product (Stage IV) showed statistically superior results when compared to the radar-only products. This implies that radar-only products should be corrected in a way (e.g., Steiner et al. 1999; Seo and Breidenbach 2002) that addresses their intrinsic error structure before they are used in hydrologic applications. However, the comparison result from one of the radar-only products, i.e., the CSU-DP, demonstrates its noticeable capability and potential in spite of the presented radar range restriction. We expect that even a simple application of the relation (in the literature) between rain rate and observed radar variables for some cold precipitation types in the CSU-DP algorithm may improve upon the current state for an operational purpose. The VPR (vertical profile of reflectivity) approach (e.g., Krajewski et al. 2011) has the potential to remedy the previously discussed melting layer issue.

We quantitatively characterized the error structure of the R-R products using a framework documented in Ciach et al. (2007). Using the characterized error structure, we removed systematic errors (overall and conditional biases) of the selected R-R products

(Stage IV, Q2-Corrected, IFC, CSU-DP, and NPOL-RC) and combined them using their random error features to create the campaign reference product. We evaluated the created reference product through the HLM streamflow simulations. The streamflow simulation results and evaluation metrics presented in Fig. 15 and Table 4 demonstrate that the reference product created in this study performs better than Stage IV, which was selected as the best R-R composite product in our evaluation. We hope that our findings and understanding, as well as our developments (e.g., the campaign reference product) that have been gained from this unique field campaign will be useful for satellite product validation and various hydrologic modeling efforts.

Acknowledgments

This study was supported by NASA's Iowa Flood Studies in collaboration with the Iowa Flood Center. The National Science Foundation provided partial support under award EAR-1327830. The authors would like to thank Carrie Langston and Pierre Kirstetter at the U.S. National Severe Storms Laboratory for providing the real-time Q2 and rain gauge corrected Q2 products. The Stage IV products were provided by NCAR/EOL under the sponsorship of the National Science Foundation. We also would like to thank Ali Tokay at the NASA Goddard Space Flight Center for providing the disdrometer observations and Munsung Keem at The University of Iowa for the meticulous quality controls of the disdrometer data. The authors are also grateful to the PMM/GPM Program Management for their support of IFloodS and this research.

References

Anagnostou, E., 2004: A convective/stratiform precipitation classification algorithm for volume scanning weather radar observations. *Meteorological Applications*, **11**, 291–300.

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. *Advances in Water Resources*, **71**, 32–43.

Chandrasekar, V., V. N. Bringi, S. A. Rutledge, A. Hou, E. Smith, G. S. Jackson, E. Gorgucci, and W. A. Petersen, 2008: Potential role of dual-polarization radar in the validation of satellite precipitation measurements: Rationale and opportunities. *Bulletin of American Meteorological Society*, **89**, 1127–1145.

Chen, H, V. Chandrasekar, and R. Bechini, 2017: An improved dual-polarization radar rainfall algorithm (DROPS2.0): Application in NASA IFloodS field campaign. *Journal of Hydrometeorology*, **18**, 917–937.

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. *Journal of Hydrometeorology*, **14**, 1308–1322.

Ciach, G. J., M. L. Morrissey, and W. F. Krajewski, 2000: Conditional bias in radar rainfall estimation. *Journal of Applied Meteorology*, **39**, 1941–1946.

Ciach, G. J., W. F. Krajewski, and G. Villarini, 2007: Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. *Journal of Hydrometeorology*, **8**, 1325–1347.

Cifelli, R., N. Doesken, P. Kennedy, L. D. Carey, S. A. Rutledge, C. Gimmestad, and T. Depue, 2005: The Community Collaborative Rain, Hail, and Snow Network: Informal education for scientists and citizens. *Bulletin of American Meteorological Society*, **86**, 1069–1077.

Cifelli, R., V. Chandrasekar, S. Lim, P. C. Kennedy, Y. Wang, and S. A. Rutledge, 2011: A new dual-polarization radar rainfall algorithm: Application in Colorado precipitation events. *Journal of Atmospheric and Oceanic Technology*, **28**, 352–364.

Clark, P., 1995: Automated surface observations, new challenges - new tools. Preprints, *6th Conf. on Aviation Weather Systems*, Dallas, TX, Amer. Meteor. Soc.

Coopersmith, E. J., M. H. Cosh, W. A. Petersen, J. Prueger, and J. J. Niemeier, 2015: Soil moisture model calibration and validation: An ARS watershed on the South Fork Iowa River. *Journal of Hydrometeorology*, **16**, 1087–1101.

Cunha, L. K., P. V. Mandapaka, W. F. Krajewski, R. Mantilla, and A. A. Bradley, 2012: Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model. *Water Resources Research*, **48**, W10515, doi:10.1029/2012WR012138.

Cunha, L. K., J. A. Smith, W. F. Krajewski, M. L. Baeck, and B.-C. Seo, 2015: NEXRAD NWS polarimetric product evaluation for IFloodS. *Journal of Hydrometeorology*, **16**, 1676–1699.

Cunha, L. K., J. A. Smith, M. L. Baeck, W. F. Krajewski, 2013: An early performance evaluation of the NEXRAD dual-polarization radar rainfall estimates for urban flood applications. *Weather and Forecasting*, **28**, 1478–1497.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Dogget, G. H. Tayor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031–2064.

Dolan, B., B. Fuchs, S. A. Rutledge, E. A. Barnes, and E. J. Thompson, 2018: Primary modes of global drop size distribution. *Journal of the Atmospheric Sciences*, **75**, 1453–1476.

Fabry, F., G. L. Austin, and D. Tees, 1992: The accuracy of rainfall estimates by radar as a function of range. *Quarterly Journal of the Royal Meteorological Society*, **118**, 435–453.

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. *Journal of Hydrology*, **377**, 80–91.

Gupta, V. K., R. Mantilla, B. M. Troutman, D. Dawdy, and W. F. Krajewski, 2010: Generalizing a nonlinear geophysical flood theory to medium-size river networks. *Geophysical Research Letters*, **37**, L11402, doi:10.1029/2009GL041540.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi, 2014: Global Precipitation Measurement (GPM) Mission, *Bulletin of the American Meteorological Society*, **95**, 701–722.

Istok, M., and Coauthors, 2009: WSR-88D dual-polarization initial operational capabilities. Preprints, *25th Conf. on Int. Interactive Information and Processing Systems (IIPS) in Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, American Meteorological Society, 15.5. [Available online at http://ams.confex.com/ams/pdfpapers/148927.pdf.]

Katz, R. W., and A. H. Murphy, Eds., 1997: Economic Value of Weather and Climate Forecasts. *Cambridge University Press*, 222 pp.

Kelleher, K. E., and Coauthors, 2007: A real-time delivery system for NEXRAD Level II data via the internet. *Bulletin of the American Meteorological Society*, **88**, 1045–1057.

Kessinger, C., S. Ellis, and J. Van Andel, 2003: The radar echo classifier: A fuzzy logic algorithm for the WSR-88D. Preprints, *3rd Conf. on Artificial Intelligence Applications to the Environmental Sciences*, Long Beach, CA, American Meteorological Society, P1.6 [Available online at http://ams.confex.com/ams/pdfpapers/54946.pdf

Kim, D., B. Nelson, and D.-J. Seo, 2009: Characteristics of reprocessed hydrometeorological automated data system (HADS) hourly precipitation data. *Weather and Forecasting*, **24**, 1287–1296.

Krajewski, W. F., A. Kruger, S. Singh, B.-C. Seo, and J. A. Smith, 2013: Hydro-NEXRAD-2: Real time access to customized radar-rainfall for hydrologic applications. *Journal of Hydroinformatics*, **15**, 580–590.

Krajewski, W. F., B. Vignal, B.-C. Seo, and G. Villarini, 2011: Statistical model of the range-dependent error in radar-rainfall estimates due to the vertical profile of reflectivity. *Journal of Hydrology*, **402**, 306–316.

Krajewski, W.F., D. Ceynar, I. Demir, R. Goska, A. Kruger, C. Langel, R. Mantilla, J. Niemeier, F. Quintero, B.-C. Seo, S. Small, L. Weber, and N. Young, 2017: Real-time flood forecasting and information system for the State of Iowa. *Bulletin of the American Meteorological Society*, **98**, 539–554.

Lim, S., V. Chandrasekar, and V. N. Bringi, 2005: Hydrometeor classification system using dual-polarization radar measurements: model improvements and in situ verification. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 792–801.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications, *19th Conference on Hydrology*, American Meteorological Society, San Diego, USA.

Mosbacher, R., W. E. Evans, and E. W. Friday, Jr., 1989: Cooperative station observations, *National Weather Service Observing Handbook No. 2*, Silver Spring, Maryland.

Nadaraya, E. A., 1964: On non-parametric estimates of density functions and regression curves. *Theory of Probability and its Applications*, **10**, 186–190.

Park, H., A. V. Ryzhkov, D. S. Zrnic, and K.-E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS. *Weather and Forecasting*, **24**, 730–748.

Petersen, W. A, and W. F. Krajewski, 2013: Status update on the GPM Ground Validation Iowa Flood Studies (IFloodS) field experiment. *Geophysical Research Abstract,* Vol. 15, Abstract EGU2013-13345. [Available online at http://meetingorganizer.copernicus.org/EGU2013/EGU2013-13345.pdf.]

Pippitt, J. L., D. B. Wolff, W. Petersen, and D. A. Marks, 2015: Data and operational processing for NASA's GPM Ground Validation program. Preprints, *37th Conf. on Radar Meteorology*, Norman, OK, American Meteorological Society. [Available online at https://ams.confex.com/ams/37RADAR/webprogram/Paper275627.html]

Quintero, F., W. F. Krajewski, R. Mantilla, S. Small, and B.-C. Seo, 2016: A spatial–dynamical framework for evaluation of satellite rainfall products for flood prediction. *Journal of Hydrometeorology*, **17**, 2137–2154.

Reed, S. M., and D. Maidment, 1999: Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling. *Journal of Hydrologic Engineering*, **4**, 174–182.

Reed, S., V. Koren, M. S. Smith, Z. Zhang, F. Moreda, and D.-J. Seo, 2004: Overall distributed model intercomparison project results. *Journal of Hydrology*, **298**, 27–60.

Rosenfeld, D., E. Amitai, and D. B. Wolff, 1995: Classification of rain regimes by the three-dimensional properties of reflectivity fields. *Journal of Applied Meteorology*, **34**, 198–211.

Ryzhkov, A. V., and D. S. Zrnic, 1998: Polarimetric rainfall estimation in the presence of anomalous propagation. *Journal of Atmospheric and Oceanic Technology*, **15**, 1320–1330.

Schumacher, C. and R. A. Houze, Jr., 2000: Comparison of radar data from the TRMM satellite and Kwajalein oceanic validation site. *Journal of Applied Meteorology*, **39**, 2151–2164.

Schwaller M. R., K. R. Morriss, 2011: A ground validation network for the Global Precipitation Measurement mission. *Journal of Atmospheric and Oceanic Technology*, **28**, 301-319.

Sivapalan, M., 2003: Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*, **17**, 3163–3170.

Skofronick-Jackson, G., and Coauthors, 2017: The Global Precipitation Measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*, **98**, 1679–1695.

Seo, B.-C., B. Dolan, W. F. Krajewski, S. Rutledge, and W. Petersen, 2015a: Comparison of single and dual polarization based rainfall estimates using NEXRAD data for the NASA Iowa Flood Studies Project. *Journal of Hydrometeorology*, **16**, 1658–1675.

Seo, B.-C., 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 Resources Research*, **49**, 2747–2764.

Seo, B.-C., and W. F. Krajewski, 2011: Investigation of the scale-dependent variability of radar-rainfall and rain gauge error covariance. *Advances in Water Resources*, **34**, 152–163.

Seo, B.-C., W. F. Krajewski, and K. V. Mishra, 2015b: Using the new dual-polarimetric capability of WSR-88D to eliminate anomalous propagation and wind turbine effects in radar-rainfall. *Atmospheric Research*, **153**, 296–309.

Seo, B.-C., F. Quintero, and W. F. Krajewski, 2018: High-resolution QPF uncertainty and its implications for flood prediction: A case study for the Eastern Iowa flood of 2016. *Journal of Hydrometeorology*, in press, doi: 10.1175/JHM-D-18-0046.1.

Seo, D.-J., and J. P. Breidenbach, 2002: Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology*, **3**, 93–111.

Smith, J. A., D.-J. Seo, M. L. Baeck, and M. D. Hudlow, 1996: An intercomparison study of NEXRAD precipitation estimates. *Water Resources Research*, **32**, 2035–2045.

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. *Journal of Hydrometeorology*, **14**, 1810–1825.

Smith, M. B., D.-J. Seo, V. I. Koren, S. M. Reed, Z. Zhang, A. Duan, F. Moreda, and S. Cong, 2004: The distributed model intercomparison project (DMIP): Motivation and experiment design. *Journal of Hydrology*, **298**, 4–26.

Steiner, M., J. A. Smith, S. J. Burges, C. V. Alonso, and R. W. Darden, 1999: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resources Research*, **35**, 2487–2503.

Tabois, G. Q., and J. D. Salas, 1985: A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resources Bulletin*, 21, 365–380.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research*, **113**, D11102, doi:10.1029/2007JD009214.

Villarini, G., W. F. Krajewski, and J. A. Smith, 2009: New paradigm for statistical validation of satellite precipitation estimates: Application to a large sample of the TMPA 0.25° 3-hourly estimates over Oklahoma. *Journal of Geophysical Research*, **114**, D12106, doi:10.1029/2008JD011475.

Warren, R. A., A. Protat, S. T. Siems, H. A. Ramsay, V. Louf, M. J. Manton, and T. A. Kane, 2018: Calibrating ground-based radars against TRMM and GPM. *Journal of Atmospheric and Oceanic Technology*, **35**, 323–346

Wu, W., D. Kitamiller, and S. Wu, 2012: Evaluation of radar precipitation estimates from the National Mosaic and Multisensor Quantitative Precipitation Estimation system and the WSR-88D Precipitation Processing System over the conterminous United States. *Journal of Hydrometeorology*, **13**, 1080–1093.

Zhang, J., K. Howard, C. Langston, B. Kaney, Y. Qi, L. Tang, H. Grams, Y. Wang, S. Cocks, S. Martinaitis, A. Arthur, K. Cooper, J. Brogden, and D. Kitzmiller, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97**, 621–638.

Zhang, J., K. Howard, C. Langston, S. Vasiloff, B. Kaney, A. Arthur, S. V. Cooten, K. Kelleher, D. Kitzmiller, F. Ding, D.-J. Seo, E. Wells, and C. Dempsey, 2011: National Mosaic and multi-sensor QPE (NMQ) system: Description, results, and future plans. *Bulletin of the American Meteorological Society*, **92**, 1321–1338.

Zhang, J., K. Howard, and J. J. Gourley, 2005: Constructing three-dimensional multiple-radar reflectivity mosaics: Examples of convective storms and stratiform rain echoes. *Journal of Atmospheric and Oceanic Technology*, **22**, 30–42.

List of Tables

List of Figures

FIG. 7. R-G comparison of event rain totals shown in Fig. 6 for the two selected events characterized by (a) snow/mix (2 May to 4 May) with stratiform rain and (b) mesoscale convective system (27 May to 30 May).

FIG. 8. Comparison of rain rate estimation functions in the SP and DP algorithms: (a) *Z*-*R* relation curves show the difference in rain rate estimation between the IFC and Q2 algorithms (the inset shows a zoom-in view for the reflectivity range of 0-30 dBZ) and (b) rain rate estimation functions according to identified hydrometeor types in the CSU-DP and NWS-DP algorithms. The coefficient "A" for the ice and snow types in the NWS-DP rain rate estimation changes according to hydrometeor classes.

FIG. 9. Two-dimensional histograms of the hourly R-G comparison. Different colors indicate data occurrences for given R-G pairs with a 1mm resolution. Overall bias values are provided in the upper-right corner of each panel. The solid black lines represent the averaged tendency described by the presented overall bias values.

FIG. 10. Two statistical metrics of multi-scale R-G comparison: correlation coefficient (left panel) and normalized RMSE (right panel). The CSU-DP correlation drop at 24-h is caused by the NWS COOP rain gauges that are located outside of the observable range shown in Fig. 2.

FIG. 11. Conditional bias of the R-R products represented by a power-law function with respect to time scale. Table 3 presents the power-law function parameters.

FIG.12. Random error structures at the hourly scale for the four composite and two NPOL DP products. Model parameters are provided in the figure.

FIG. 13. Rain total maps of the campaign reference product and its independent evaluation using the NWS COOP and CoCoRaHS rain gauge data. The scatter plots show R-G comparison of the campaign totals between the reference product and rain gauge observations. The CoCoRaHS observations show a quality control issue (e.g., missing).

FIG. 14. The maps of campaign rain totals of the (a) reference and (b) gauge interpolation products over the Turkey River basin.

FIG. 15. Hydrologic simulation results driven by the gauge interpolation, campaign reference, and Stage IV products at the five USGS stream gauge stations in the Turkey River basin.

TABLE 1. The R-R composite products evaluated and their resolution and algorithm comparison.

Product type	QPE Product	Resolution		Algorithms		
		Time	Space (km)	Quality Control (QC)	Classification	Rain Rate Estimator
Radar-Only SP	IFC	5-min	0.5	Vertical and horizontal continuity of Z and thresholds on ρ_{hv} and dispersion of Φ_{dp} (Seo et al. 2015b)	N/A	$R(Z)=0.017Z^{0.714}$; $Z=300R^{1.4}$
	Q2	5-min	1.0	Neural network and heuristic rules (Zhang et al. 2011)	Precipitation type • Convective • Stratiform • Warm rain • Snow	$R(Z)=0.017Z^{0.714}$; $Z=300R^{1.4}$ $R(Z)=0.036Z^{0.625}$; $Z=200R^{1.6}$ $R(Z)=0.013Z^{0.8}$; $Z=230R^{1.25}$ $R(Z)=0.115Z^{0.5}$; $Z=75R^{2.0}$
Radar-Only DP	CSU-DP	5-min	0.5	Thresholds on dispersion of Φ_{dp}, ρ_{hv}, and SNR (Cifelli et al. 2011)	Hydrometeor • Heavy rain • Hail • Moderate rain • Drizzle • Mixture • Ice	$R(K_{dp},Z_{dr})=90.8K_{dp}^{0.93}Z_{dr}^{-0.169}$ $R(K_{dp})=40.5K_{dp}^{0.85}$ $R(Z,Z_{dr})= 0.0067Z^{0.927}Z_{dr}^{-3.43}$ $R(Z)=0.017Z^{0.714}$ $R(K_{dp})=40.5K_{dp}^{0.85}$ N/A
	NWS-DP	1-h	1.0	Fuzzy logic using polarimetric observables (HCA; Park et al. 2009)	Hydrometeor • Heavy rain • Light/Moderate rain • Hail, mixed with rain • Wet snow • Graupel • Dry snow and ice	$R(Z,Z_{dr})=0.0067Z^{0.927}Z_{dr}^{-3.43}$ $R(Z,Z_{dr})=0.0067Z^{0.927}Z_{dr}^{-3.43}$ $R(K_{dp})=44.0\lvert K_{dp}\rvert^{0.822}sign(K_{dp})$ 0.6*R(Z) 0.8*R(Z) 2.8*R(Z)
Gauge-Corrected	Stage IV	1-h	4.0	Fuzzy logic using Z_h and Doppler information (REC; Kessinger et al. 2003) with manual QC (Lin et al. 2005)	N/A	$R(Z)=0.017Z^{0.714}$; $Z=300R^{1.4}$
	Q2-Corrected	1-h	1.0	same as Q2	same as Q2	same as Q2

TABLE 2. R-G comparison results with respect to time scale: three statistical metrics (overall bias, correlation coefficient, and normalized RMSE).

Statistical metrics	R-R Product	Time scale (h)				
		1	3	6	12	24
Overall bias (G/R)	Stage IV	0.96				
	Q2-Corrected	0.93				
	IFC	1.18				
	Q2	0.73				
	CSU-DP	1.40				
	NWS-DP	0.87				
Correlation	Stage IV	0.83	0.89	0.91	0.92	0.91
	Q2-Corrected	0.84	0.87	0.90	0.90	0.90
	IFC	0.81	0.85	0.87	0.87	0.86
	Q2	0.83	0.87	0.89	0.89	0.88
	CSU-DP	0.84	0.87	0.89	0.89	0.86
	NWS-DP	0.82	0.85	0.87	0.86	0.85
Normalized RMSE	Stage IV	3.23	1.98	1.46	1.04	0.82
	Q2-Corrected	3.71	2.46	1.73	1.26	0.97
	IFC	3.44	2.26	1.73	1.29	1.05
	Q2	4.20	2.90	2.27	1.67	1.30
	CSU-DP	3.08	2.08	1.60	1.22	1.06
	NWS-DP	3.83	2.65	2.06	1.53	1.22

TABLE 3. Estimated power law function parameters describing the R-R conditional bias with respect to time scale.

Parameter	R-R Product	Time scale (h)				
		1	3	6	12	24
a_h	Stage IV	1.08	1.08	1.06	1.21	1.03
	Q2-Corrected	1.31	1.54	1.60	1.71	1.72
	IFC	1.15	1.42	1.51	1.67	1.62
	Q2	1.17	1.31	1.35	1.42	1.39
	CSU-DP	1.30	1.57	1.66	1.83	1.89
	NWS-DP	1.17	1.39	1.48	1.67	1.70
b_h	Stage IV	0.93	0.96	0.98	0.93	0.99
	Q2-Corrected	0.80	0.80	0.82	0.81	0.83
	IFC	0.82	0.80	0.81	0.79	0.82
	Q2	0.88	0.87	0.88	0.87	0.89
	CSU-DP	0.79	0.78	0.79	0.77	0.79
	NWS-DP	0.85	0.82	0.83	0.80	0.82

TABLE 4. Performance metrics for hydrologic simulations at the five USGS stream gauge locations in the Turkey River basin.

Statistical Metrics	Forcing Product	USGS stations				
		Spillville	Littleport	Eldorado	Elkader	Garber
KGE	Gauge Interpolation	0.70	0.74	0.72	0.55	0.70
	Campaign Reference	0.56	0.74	0.65	0.43	0.66
	Stage IV	0.41	0.70	0.64	0.30	0.63
Correlation	Gauge Interpolation	0.78	0.92	0.78	0.73	0.83
	Campaign Reference	0.62	0.78	0.67	0.62	0.73
	Stage IV	0.55	0.78	0.65	0.55	0.69
Normalized RMSE	Gauge Interpolation	0.47	0.34	0.52	0.47	0.43
	Campaign Reference	0.60	0.47	0.62	0.57	0.49
	Stage IV	0.72	0.49	0.67	0.65	0.53

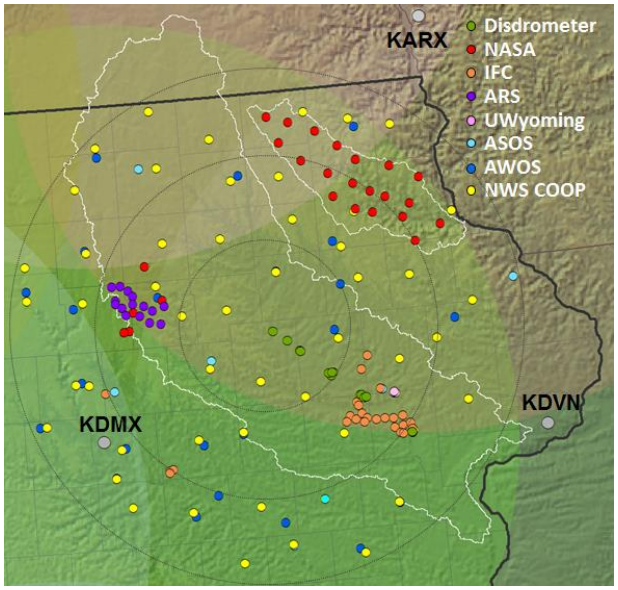

FIG. 1. IFloodS spatial domain and the distribution of the rain gauge and disdrometer networks used in the R-R product evaluation. The shaded circular areas indicate the 230 km range domain from the involved NEXRAD radars. The circular lines in the middle of the domain demarcate every 50 km from the NPOL radar location. The Cedar/Iowa and Turkey River basins are presented in the middle of the domain and in the northeast close to the Iowa border, respectively.

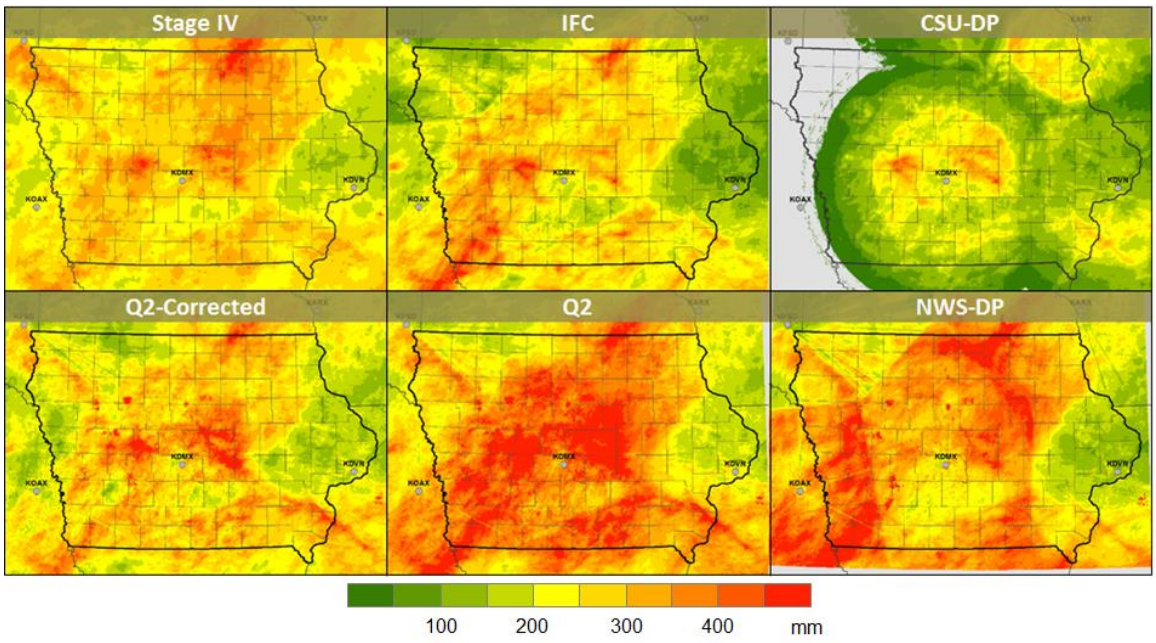

FIG. 2. Rain total maps of the R-R products accumulated over the entire campaign period (May 1st through June 15th). The rain gauge-corrected (Stage IV and Q2-Corrected), radar-only SP (IFC and Q2), and radar-only DP (CSU-DP and NWS-DP) products are aligned from the left to the right panels. Since the CSU-DP product was created using only four radars (KARX, KDMX, KDVN, and KMPX), the coverage of this product is limited to the central and eastern regions of Iowa.

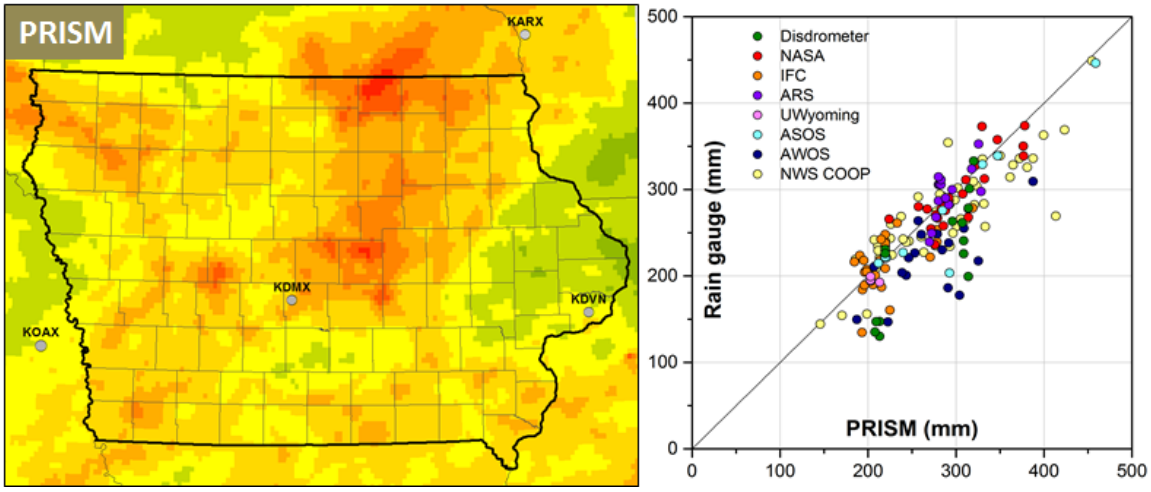

FIG. 3. Rain total map of the PRISM rain gauge interpolation analysis (left panel) and rain gauge comparison of PRISM rain totals (right panel). Independent gauges (e.g., NASA, IFC, and ARS) show good agreement with the PRISM data.

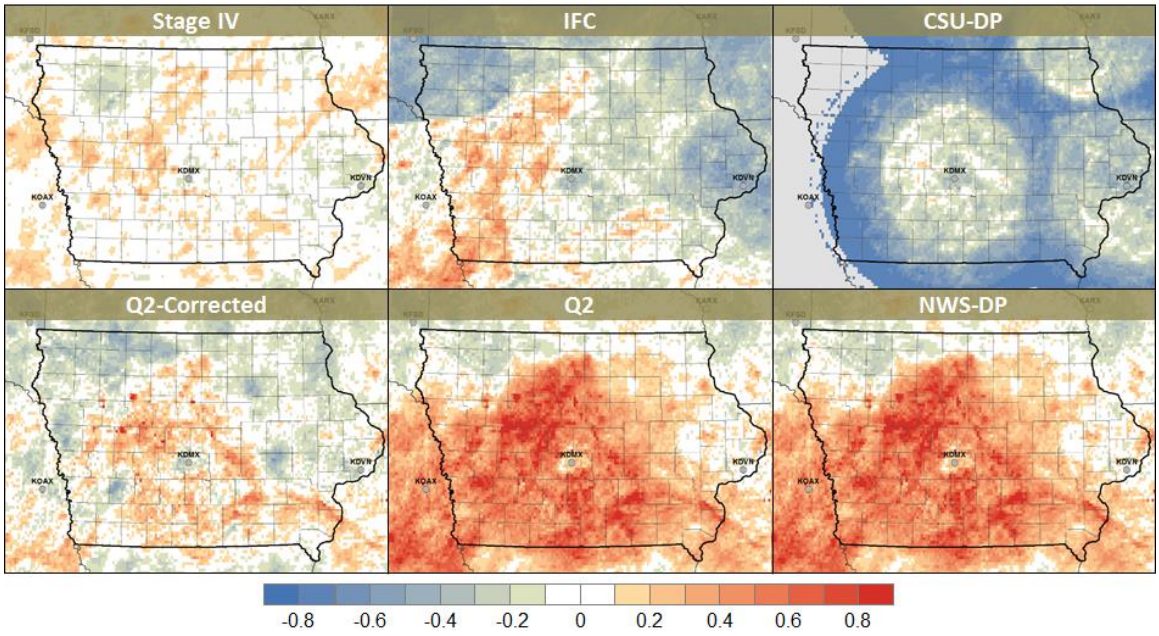

FIG. 4. Normalized error/difference maps estimated by (6) for the campaign totals. Red and blue colors indicate over- and under-estimation, respectively. The map alignment is the same as in Fig. 2.

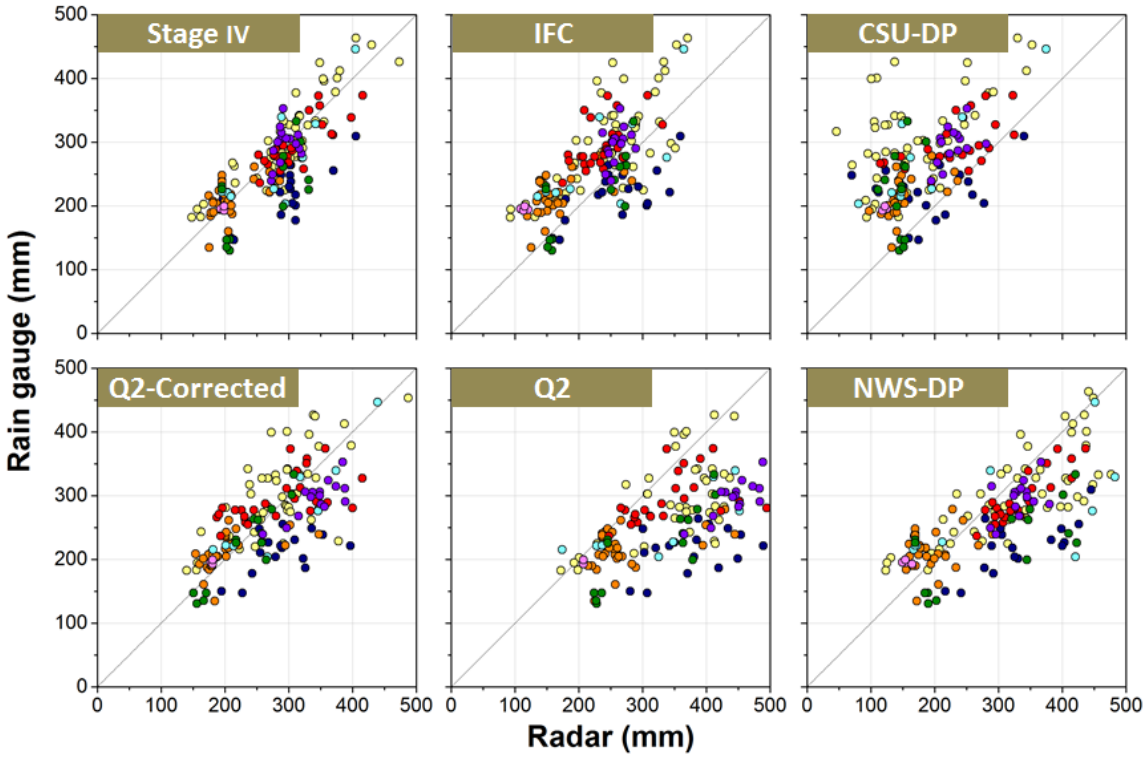

FIG. 5. R-G comparison of the campaign totals. The rain gauge color code is the same as in Fig. 3 (right panel). The rain gauge-corrected (Stage IV and Q2-Corrected), radar-only SP (IFC and Q2), and radar-only DP (CSU-DP and NWS-DP) products are aligned from the left to the right panels.

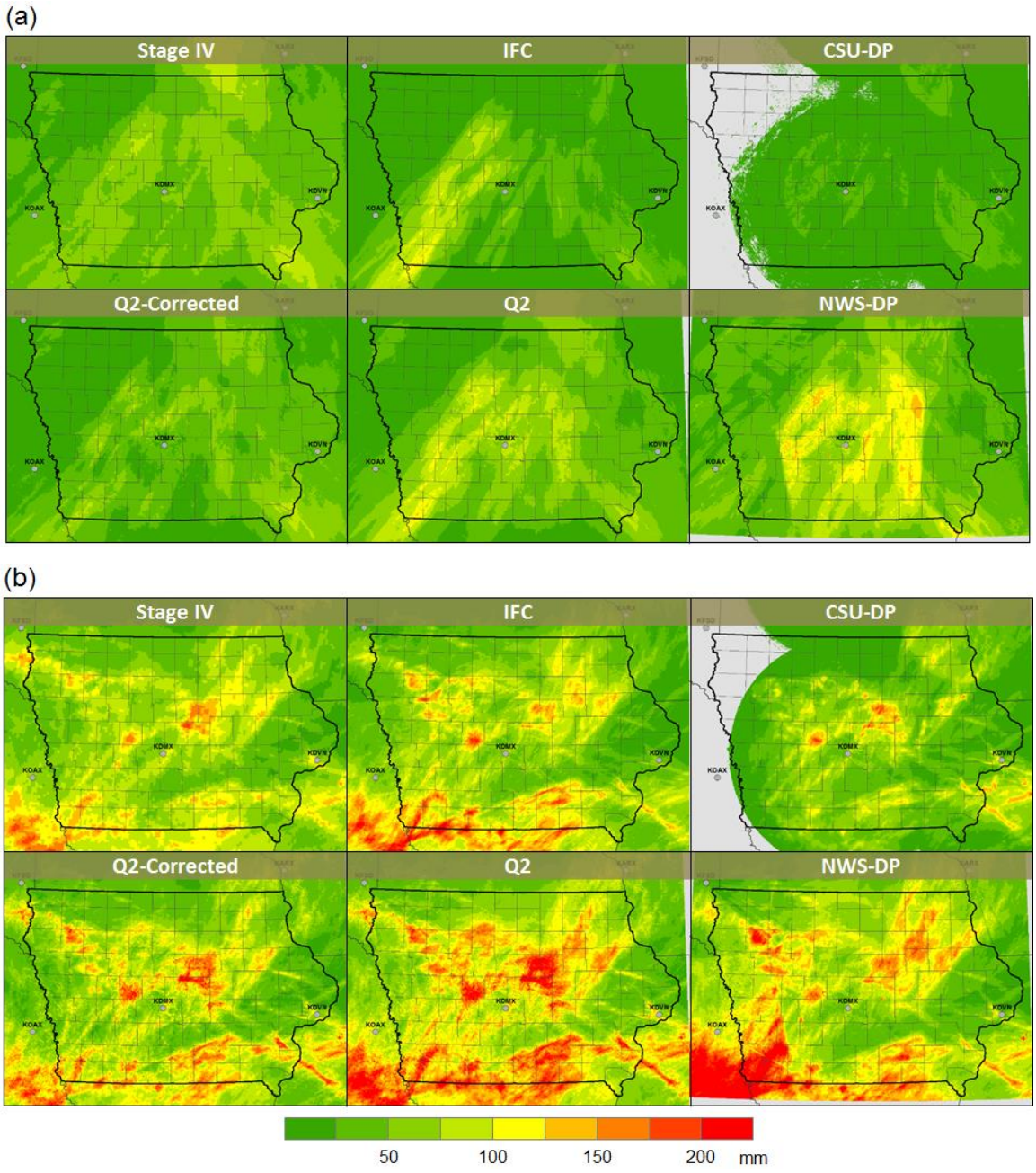

FIG. 6. Rain total maps for the two selected events characterized by (a) snow/mix with stratiform rain (2 May to 4 May) and (b) mesoscale convective system (27 May to 30 May). The map alignment is the same as in Fig. 2.

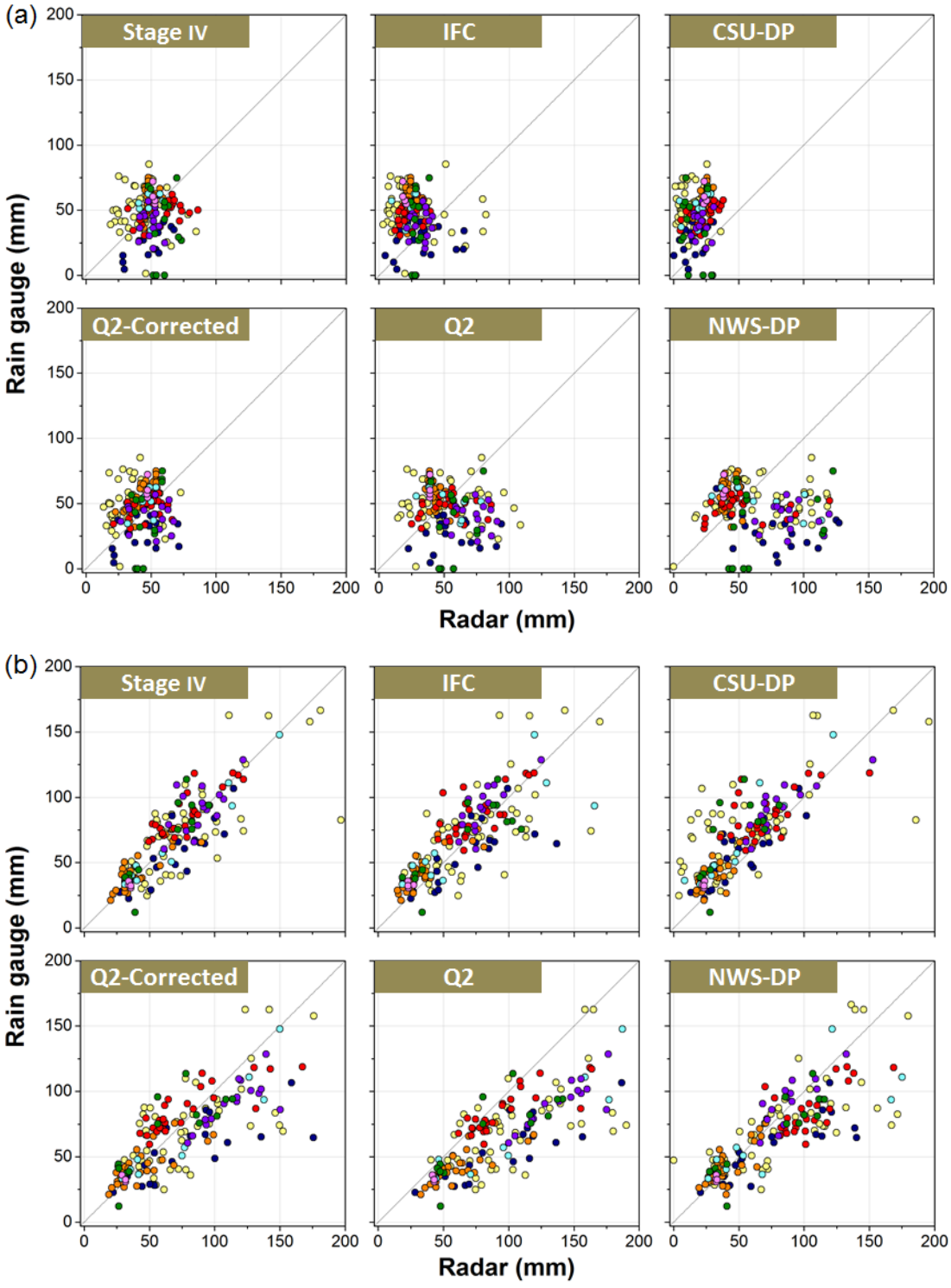

FIG. 7. R-G comparison of event rain totals shown in Fig. 6 for the two selected events characterized by (a) snow/mix (2 May to 4 May) with stratiform rain and (b) mesoscale convective system (27 May to 30 May).

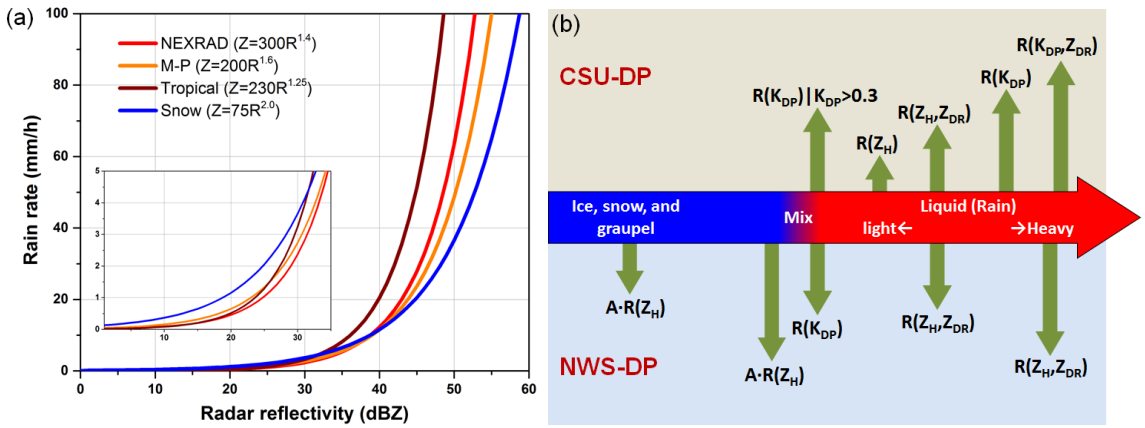

FIG. 8. Comparison of rain rate estimation functions in the SP and DP algorithms: (a) *Z*-*R* relation curves show the difference in rain rate estimation between the IFC and Q2 algorithms (the inset shows a zoom-in view for the reflectivity range of 0-30 dBZ) and (b) rain rate estimation functions according to identified hydrometeor types in the CSU-DP and NWS-DP algorithms. The coefficient "A" for the ice and snow types in the NWS-DP rain rate estimation changes according to hydrometeor classes.

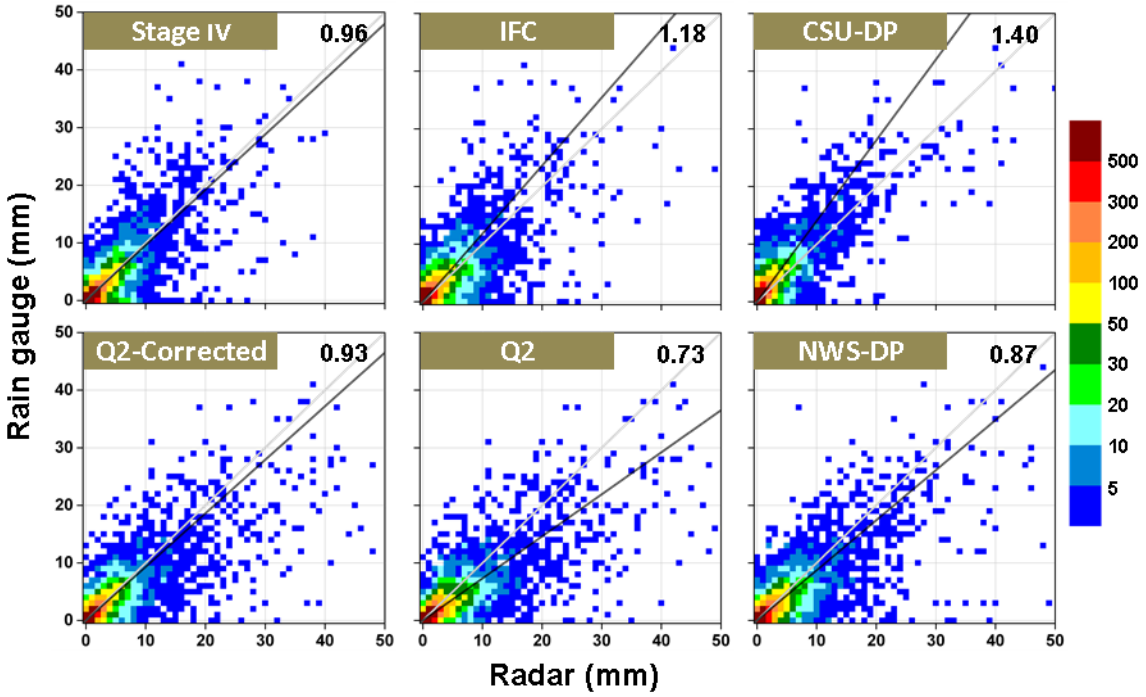

FIG. 9. Two-dimensional histograms of the hourly R-G comparison. Different colors indicate data occurrences for given R-G pairs with a 1mm resolution. Overall bias values are provided in the upper-right corner of each panel. The solid black lines represent the averaged tendency described by the presented overall bias values.

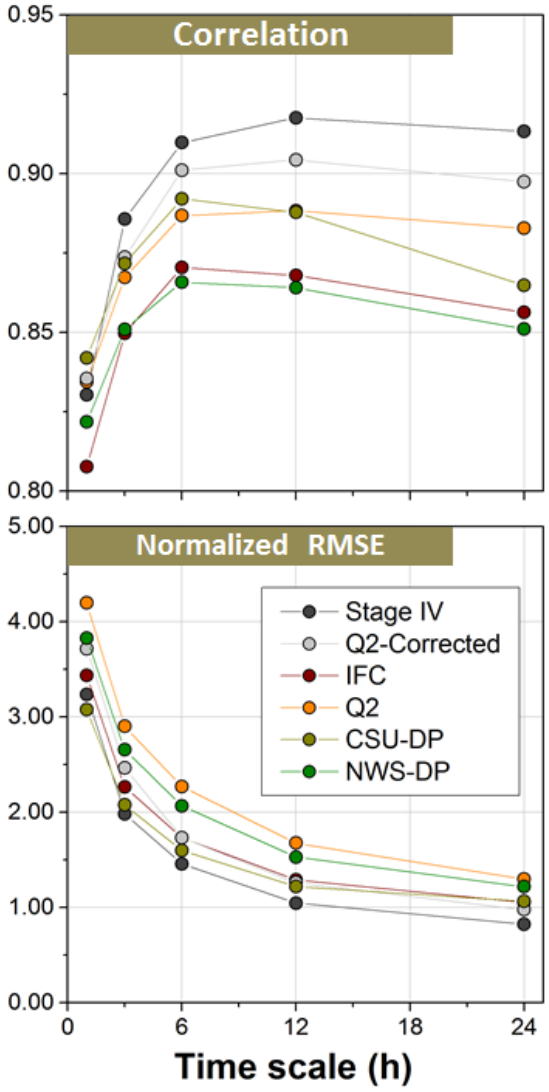

FIG. 10. Two statistical metrics of multi-scale R-G comparison: correlation coefficient and normalized RMSE. The CSU-DP correlation drop at 24-h is caused by the NWS COOP rain gauges that are located outside of the observable range shown in Fig. 2.

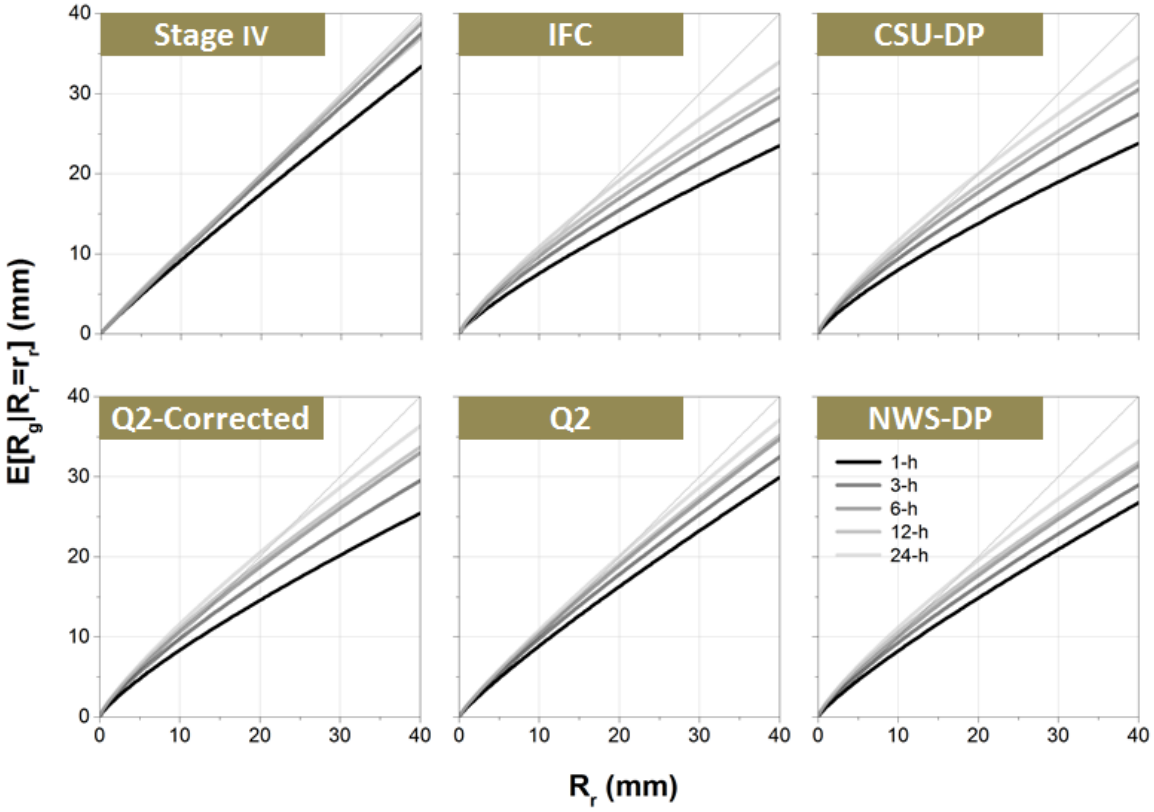

FIG. 11. Conditional bias of the R-R products represented by a power-law function with respect to time scale. Table 3 presents the power-law function parameters.

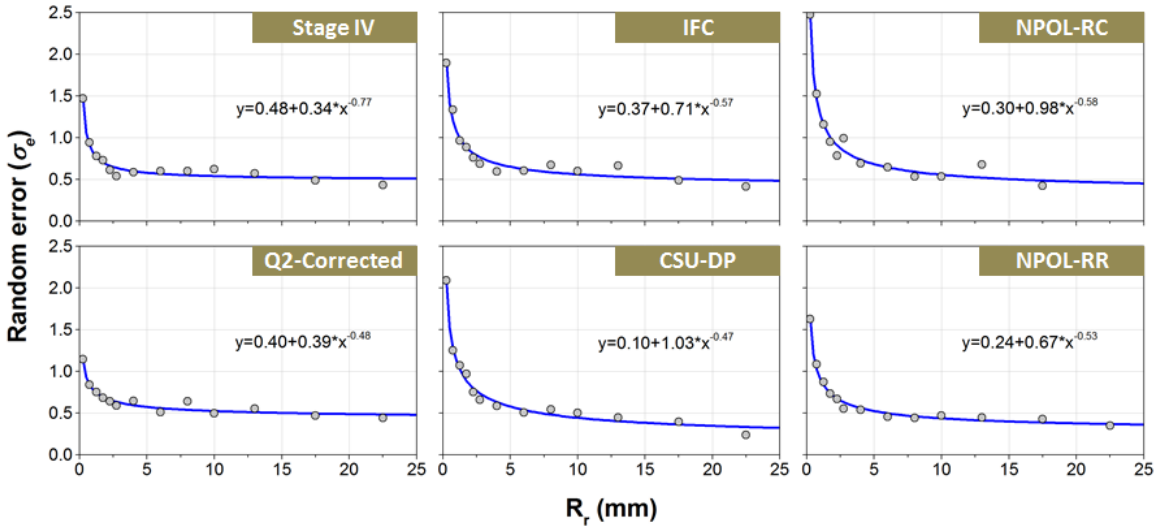

FIG.12. Random error structures at the hourly scale for the four composite and two NPOL DP products. Model parameters are provided in the figure.

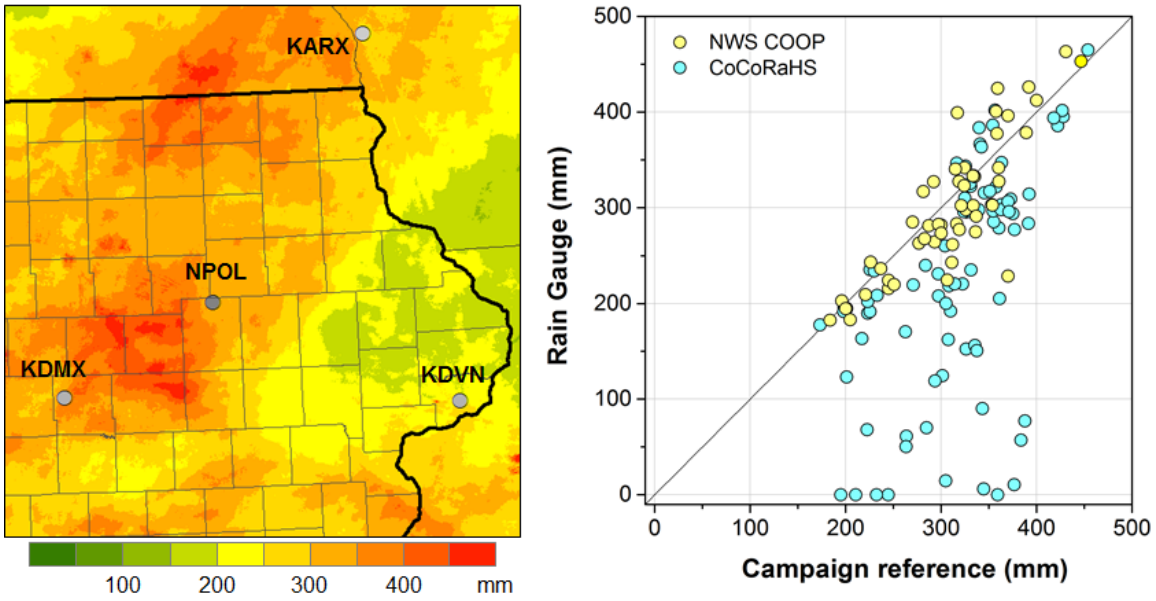

FIG. 13. Rain total maps of the campaign reference product and its independent evaluation using the NWS COOP and CoCoRaHS rain gauge data. The scatter plots show R-G comparison of the campaign totals between the reference product and rain gauge observations. The CoCoRaHS observations show a quality control issue (e.g., missing).

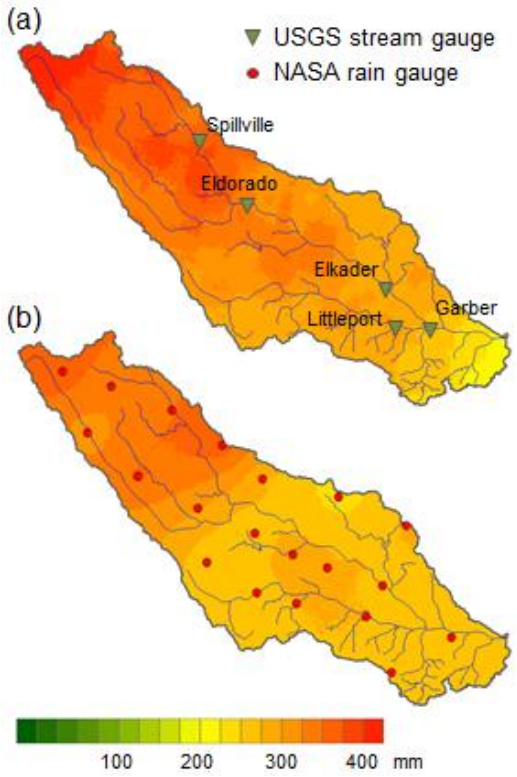

FIG. 14. The maps of campaign rain totals of the (a) reference and (b) gauge interpolation products over the Turkey River basin.

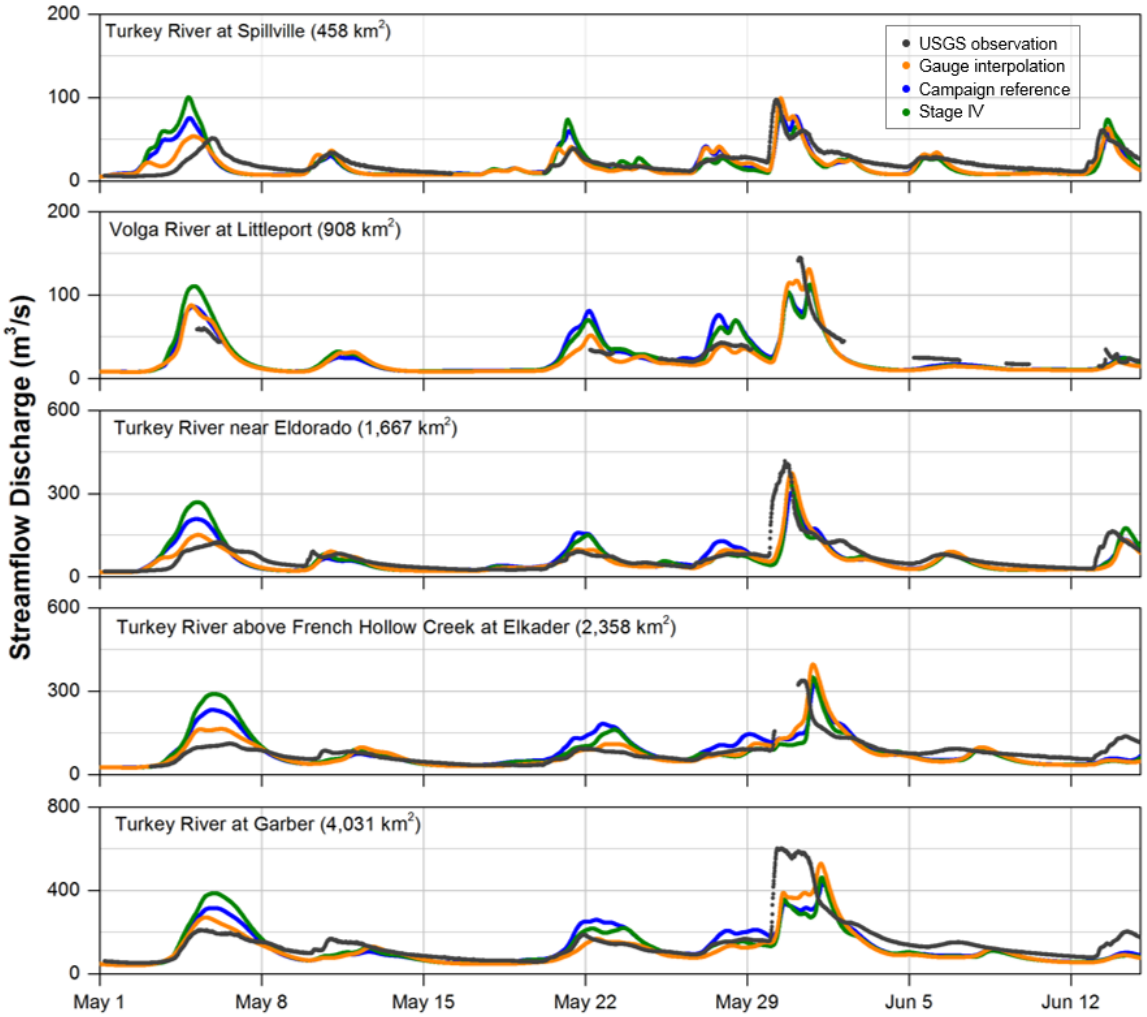

FIG. 15. Hydrologic simulation results driven by the gauge interpolation, campaign reference, and Stage IV products at the five USGS stream gauge stations in the Turkey River basin.