Journal of Hydrology 373 (2009) 463–478

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Validation of NEXRAD multisensor precipitation estimates using an experimental dense rain gauge network in south Louisiana

Emad Habib [a,*], Boone F. Larson [a], Jeffrey Graschel [b]

[a] Department of Civil Engineering, University of Louisiana at Lafayette, P.O. Box 42991, Lafayette, LA 70504, USA
[b] NWS Lower Mississippi River Forecast Center 62300 Airport Rd. Slidell, LA 70460-5243, USA

ARTICLE INFO

Article history:
Received 4 February 2009
Received in revised form 19 April 2009
Accepted 10 May 2009

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Efrat Morin, Associate Editor

Keywords:
NEXRAD
MPE
Validation
Rain gauge network
Bias
Sub-pixel variability

SUMMARY

This study presents a validation analysis of a radar-based multisensor precipitation estimation product (MPE) focusing on small temporal and spatial scales that are of hydrological importance. The 4×4 km^2 hourly MPE estimates are produced at the National Weather Service (NWS) regional River Forecast Centers and mosaicked as a national product known as Stage IV. The validation analysis was conducted during a 3-year period (2004–2006) using a high quality experimental rain gauge network in south Louisiana, United States, that was not included in the development of the MPE product. The dense arrangement of rain gauges within two MPE 4×4 km^2 pixels provided a reasonably accurate approximation of area-average surface rainfall and avoided limitations of near-point gauge observations that are typically encountered in validation of radar-rainfall products.

The overall bias between MPE and surface rainfall is rather small when evaluated over an annual basis; however, on an event-scale basis, the bias reaches up to ±25% of the event total rainfall depth during for half of the events and exceeds 50% for 10% of the events. Negative bias (underestimation) is more dominant (65% of events), which is likely caused by range-related effects such as beam overshooting and spreading over the study site (~120 km from the closest radar site). A clear conditional bias was observed as the MPE estimates tend to overestimate small rain rates (conditional bias of 60–90% for rates lower than 0.5 mm/h) and underestimate large rain rates (up to −20% for rates higher than 10 mm/h). False detections and lack of detection problems contributed to the MPE bias, but were negligible enough to not result in significant false detection or underestimation of rainfall volumes. A significant scatter was observed between MPE and surface rainfall, especially at small intensities where the standard deviation of differences was in the order of 200–400% and the correlation coefficient was rather poor. However, the same statistics showed a much better agreement at medium to high rainfall rates. The MPE product was also successful in reproducing the underlying spatial and temporal organization of surface rainfall as reflected in the assessment of rainfall self-correlations and the extreme tail of the hourly rainfall marginal distribution. The quantitative results of this analysis emphasized the need for multiple gauges within MPE pixels as a prerequisite for validation studies. Using a single gauge within an MPE pixel as a reference representation of surface rainfall resulted in an unrealistic inflation of the actual MPE estimation error by 120–180%, especially during high-variability rainfall cases. This and the enhanced quality of the reference gauge dataset, explain the improved performance by MPE as compared to other previous studies. Compared to previous studies, the current analysis shows a significant improvement in the MPE performance. This is attributed to two main factors: continuous MPE algorithmic improvements and increased experience by its users at the NWS forecast centers, and the use of high-quality and dense rain gauge observations as a validation reference dataset. The later factor ensured that gauge-related errors are not wrongly assigned to radar estimation uncertainties.

Introduction

Accurately capturing the spatial and temporal variability of rainfall has largely been limited by the sparsity of rain gauge stations and the rather low quality of their measurements. Moreover, rain gauges are generally not capable of detecting rainfall at the resolution of most hydrologic applications. Since rainfall is the driving force behind numerous hydrologic forecasting activities such as river stage and flash flood warnings, errors associated with rainfall measurements will be inevitably propagated through such applications. With the average rain gauge density being only

* Correspondence author. Tel.: +1 337 482 6513.
E-mail address: habib@louisiana.edu (E. Habib).

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.05.010

1.3 gauges/1000 km^2 in the United States (Linsley et al., 1992), weather radar systems are becoming increasingly more useful. In the early 1990s, the Next Generation Weather Radar (NEXRAD) system was installed across the United States. This technology transformed the National Weather Service (NWS) implementation of forecast and warning programs (Fulton, 2002). The ability of the radar to capture the spatial details and severity of thunderstorms in near real time makes it very advantageous for a variety of hydrologic and meteorological applications. The NWS River Forecast Center's (RFC) routinely produce regional radar-based multisensor precipitation estimates (MPE) to be used as input to the NWS operational hydrologic forecasting models. The MPE products are mapped onto a polar stereographic projection known as the Hydrologic Rainfall Analysis Project (HRAP) grid, which has a spatial resolution of 4 km $\times$ 4 km. Regional MPE products are eventually mosaicked over the entire continental US (CONUS). However, due to the indirect nature of radar measurement of rainfall, radar-based rainfall products are subject to several sources of uncertainty such as: (1) range effects caused by increase in beam elevation and degradation of resolution due to beam spreading; (2) Anomalous Propagation (AP) (3) bright band contamination; and (3) uncertainty in the choice of a particular Z–R relationship (Wilson and Brandes, 1979; Austin, 1987; Joss and Waldvogel, 1990). Several procedures have been developed and implemented by the NWS and its RFCs to improve the quality of the MPE products such as mean-field bias adjustment, merging of radar and gage estimates, local bias adjustment, and use of satellite estimates. However, the inherent inability of radar systems to accurately measure surface rainfall makes it necessary to evaluate the MPE estimates and assess their uncertainties.

Numerous studies have looked into the evaluation of various radar-based rainfall estimation products that have been developed over the last two decades. In the following, we provide a brief discussion of these studies focusing on those concerned with the NWS MPE algorithm. Young et al. (2000) examined two earlier versions of the MPE (Stage III and P1 products) over the Arkansas-Red Basin RFC (ABRFC) using a total 111 gauges. The evaluation was done at the native resolution of both products (hourly and 4-km). Results indicated that Stage III had 41% probability in detecting gauge rainfall greater than zero, while P1 had a probability of 86%. Probability of detection in either estimation method did not reach 100% until gauge threshold $\geqslant$25 mm. Stage III was also found to significantly underestimate gauge rainfall ($-$21.5% bias). P1 estimates tended to slightly overestimate gauge rainfall (5.2% bias). Similar results on these two products were also reported by Wang et al. (2001). Grassotti et al. (2003) performed an inter-comparison study over the Arkansas-Red River and the Illinois River basins using three rainfall datasets: hourly P1 estimates, NEXRAD-based NOWrad estimates produced by the Weather Services International Corporation (WSI) with a 15 min 2-km resolution, and conventional hourly rain gauge measurements. Overall, the P1 product was found to have better agreement with rain gauges than WSI product at daily and monthly timescales. Jayakrishnan et al. (2004) assessed daily Stage III precipitation data over the Texas-Gulf basin for the period of 1995–1999 using data from 545 daily rain gauges. Overall, the Stage III product underestimated precipitation at most gauge locations in the Texas-Gulf basin. Differences in total rainfall depth were within ±500 mm at 42% of the rain gauges and the root mean square difference was greater than 10 mm at 78% of the rain gauges. Dyer and Garza (2004) reported significant underestimation at a basin-average scale by the Stage III product over Florida with seasonal and annual non-stationarity of the product bias. Xie et al. (2006) performed an evaluation study using 4 years of Stage III 4-km resolution data (1998–2001) over central New Mexico. Their results indicated that Stage III overestimated seasonal precipitation accumulation by 11–88% during monsoon season and underestimated by 18–89% during non-monsoon season. Stage III performance was found to improve since the 1998 monsoon season. Westcott et al. (2008) compared MPE estimates and gauge data at monthly and daily temporal scales and at county and MPE grid spatial scales for the Midwestern United States over a 41-month period (2002–2005). Compared to gauge data, county-averaged monthly MPE estimates had an underestimation bias that varied across the nine states covering the study area. At the daily scale, MPE estimates tended to overestimate gauge data at low precipitation values and be the same or less than the gauge amounts at higher precipitation values. Issues related to range degradation, terrain blockage, and quality of gauge data used within the MPE algorithm have also been proven to be of importance when assessing the performance of the products (Stellman et al., 2001; Marzen and Fuelberg, 2005; Boyles et al., 2006).

Recent studies have also made comparisons between Stage III and MPE products during the Stage III-to-MPE transition period (2000–2003). Yilmaz et al. (2005) reported an overall improvement in basin-averaged MPE estimates compared to Stage III especially in winter months which are characterized by lesser degrees of spatial variability. However, both products showed underestimation in the cold season and overestimation in the warm season. Over et al. (2007) compared Stage III and MPE product against 27 radio-telemetered rain gauges in northeastern Illinois from July 1997 to September 2005. The analysis was done at 4-km and daily scales. Their results indicated varying degrees of overestimation and underestimation of both products, with underestimation being smallest ($\sim$3%) during the later period of the analysis when transition from Stage III to MPE began. Wang et al. (2008) compared the performance of Stage III and MPE products over a large river basin in central Texas. Due to gauge near-point sampling limitations, their study focused on cases of uniform rainfall only and found that compared to Stage III, MPE has better detection, higher correlation and a lower bias (7%). Young and Brunsell (2008) evaluated Stage III and MPE precipitation estimates for the Missouri River Basin using daily gauge data and found continuous improvement in the estimates over the period 1998–2004 especially during warm season.

With a few exceptions, most previous studies that focused on evaluation of MPE estimates at small spatial and temporal scales, have relied on observations from a single gauge located within the domain of the MPE pixel. However, in view of experimental findings about small-scale rainfall variability (e.g., Krajewski et al., 2003; Miriovsky et al., 2004; Habib and Krajewski, 2002), relying on a single gauge to evaluate the accuracy of radar estimates can be inconclusive and misleading. This is due to the fact that rain gauges are limited by their near-point observational nature and their poor representation of surface areal rainfall (Kitchen and Blackall, 1992). This problem has been highlighted by Young et al. (2000) who emphasized the importance of isolating the effect of sub-grid variability caused by differences between point gauge measurements and areal averages. The same issue has lead Wang et al. (2008) to limit their conclusions on MPE evaluation to spatially uniform rainfall cases only. Therefore, a more conclusive approach for MPE validation requires access to reference data that provides accurate representation of the true surface areal rainfall. Such reference data can be available through a dense network of gauges that are arranged with separation distances smaller than the resolution of the radar products. In addition, the surface reference data needs to be independent from the MPE products under evaluation (i.e., not used by NWS in developing or adjusting the bias of MPE estimates).

To overcome the limitations that result from sparsity and lack of independence of most operational rain gauge networks, the current study will validate the MPE products against surface observations from a small-scale, independent, dense rain gauge experimental

network located in south Louisiana, United States. The dense arrangement of the gauges (two MPE pixels with four gauges in each) will provide reference rainfall information that is more representative of the true areal-rainfall over an MPE pixel than what is typically available from a single gauge. Comparison of the MPE products with the pixel-average gauge rainfall is expected to reduce the effect of single-gauge uncertainty, leading to an accurate assessment of the MPE error levels (Ciach and Krajewski, 1999a,b; Anagnostou et al., 1999; Young et al., 2000; Habib and Krajewski, 2002). However, it is noted that the results of this study will be limited by the fact that only two MPE pixels are considered in the analysis, which does not allow to fully investigate some important MPE estimation problems (e.g., range-dependent biases). Another desired feature of this surface reference network is the high-quality of its gauge measurements, which is a result of a special dual-gauge setup and frequent equipment service and maintenance. The study presents both visual and statistical analysis of the differences between MPE estimates and the corresponding surface rainfall quantities over a 3-year period (2004–2006). The performance of the MPE product is evaluated in terms of systematic bias (on an annual as well as event basis), random differences, linear association and correspondence, and probabilities of successful and false detection. The ability of the MPE product to reproduce the distribution and spatio-temporal organization of surface rainfall is also examined. To gain insight into the experimental and data requirements needed for viable MPE validation analysis, we also present a detailed analysis of the significance of using accurate estimates of areal surface rainfall and quantify the contribution of sub-grid variability in the assessment of MPE uncertainty. The results of this study will help provide the MPE user community and algorithm developers with desired information about the limitations and accuracy levels of radar-based rainfall estimates, as well as the implied effects on their usage in various hydrologic and water resources management applications.

Study area and data resources

Fig. 1 shows the area of this study, which is the mid-size Isaac-Verot watershed (~35 km^2) located in the city of Lafayette, in southern Louisiana. The watershed is a sub-drainage area of the Vermilion river basin which drains into the Gulf of Mexico (Habib and Meselhe, 2006). The watershed is frequently subject to frontal systems, air-mass thunderstorms, and tropical cyclones with annual rainfall of about 140–155 cm and monthly accumulations as high as 17 cm. The area of the watershed is within the boundaries of the NWS Lower Mississippi River Forecast Center (LMRFC).

NEXRAD multisensor precipitation estimates

The source of multisensor precipitation estimates used in this study is the Stage IV dataset available from the National Center for Environmental Prediction (NCEP). The Stage IV product is a national mosaic of regional multisensor precipitation estimates that are routinely produced at the NWS RFCs for operational hydrologic forecasting. The area of the current study is fully within the boundaries of the Lower Mississippi River Forecast Center (LMRFC). The multisensor estimates are produced by combining data from several Weather Surveillance Radar-1988 Doppler (WSR-88D) radars with real-time surface rain gauge observations. The closest two Next Generation Radar (NEXRAD) WSR-88D radars to the

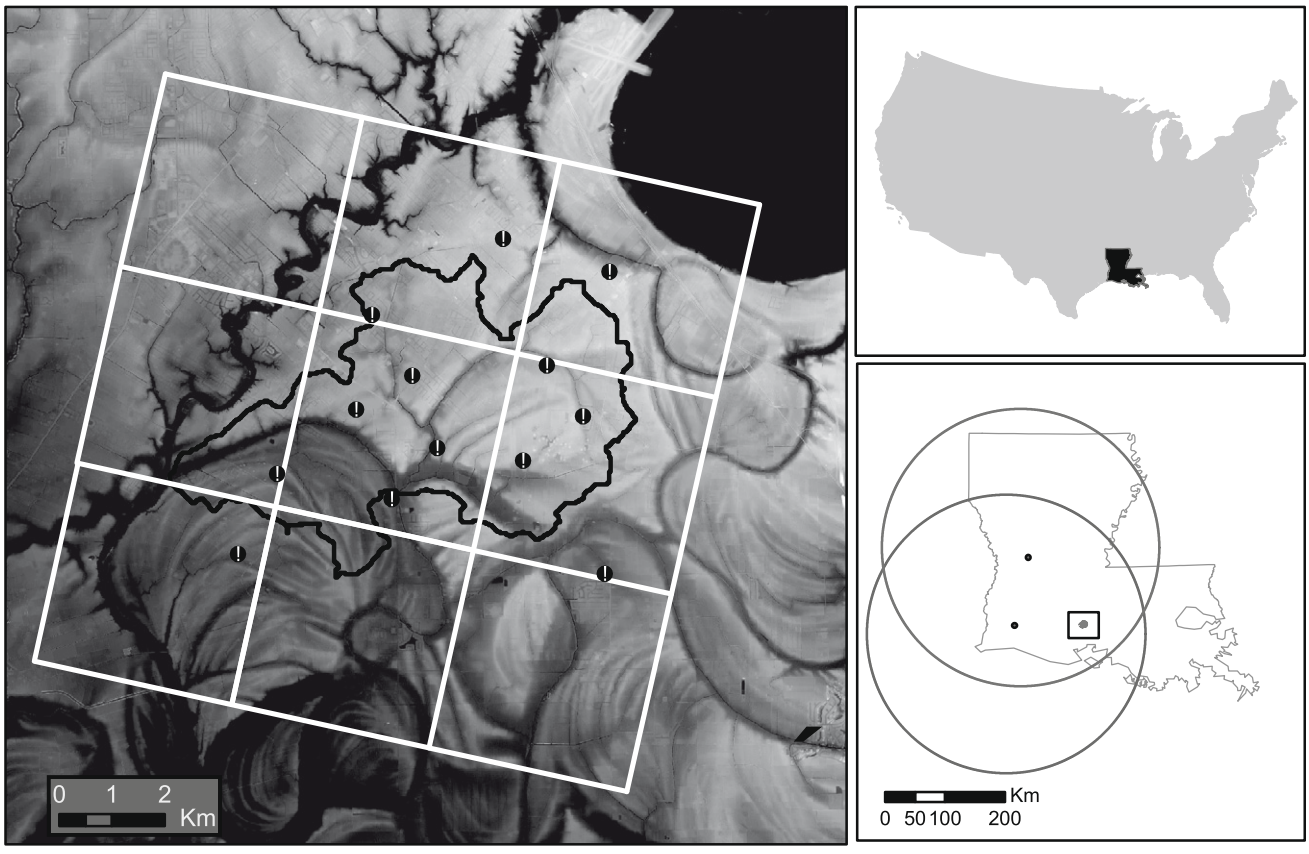

Fig. 1. Rain gauge sites (circles) within the Isaac-Verot watershed in Lafayette, LA. Each gauge site is equipped with two gauges located side-by-side. The 4 × 4 km^2 HRAP grid is superimposed over the watershed. The two circles on the bottom right panel indicate the 250-km umbrellas of the two closest WSR-88D radars in Lake Charles (KLCH) and Fort Polk (KPOE).

Isaac-Verot experimental network are KLCH in Lake Charles (~120 km), and KPOE in Fort Polk (~150 km). At the KLCH distance of 120 km, the height of the lowest radar beam is about 1.82 km above the ground surface. According to the radar coverage map used by the LMRFC, MPE products over the rain gauge network domain are primarily derived from the KLCH radar.

Prior to August 2003, the multisensor estimates were available as a Stage III product with hourly rainfall accumulations over a grid of approximately $4 \times 4\,km^2$ (HRAP grid). Description of the WSR-88D estimation and processing algorithms is available in several previous studies (e.g., Fulton et al., 1998; Seo et al., 1999; Breidenbach and Bradberry, 2001; Lawrence et al., 2003) and only a brief overview is given here. Development of the Stage III data is a three-step process that begins with the raw radar reflectivity field (Z). In Stage I, a power law *Z–R* relationship is applied to the raw reflectivity data to calculate precipitation estimates, which are integrated over time to produce hourly accumulation. The result of the Stage I process is a radar-only product known as Digital Precipitation Product (DPA). Stage II consists of the merging of radar data with rain gauge data to generate bias-adjusted radar precipitation estimates over the grid cells. This includes a mean-field bias adjustment (Seo et al., 1999) where a time-dependent, radar-dependent bias factor is applied as a multiplier to each pixel in the DPA radar product. The final stage, Stage III, involves the regional mosaicking of these radar/gauge estimates as well as the application of quality control measures.

As of 2003, the LMRFC began using a new algorithm called the Multisensor Precipitation Estimator (MPE) as a replacement to the original Stage II and Stage III processes, which offered some improvements over the previous products (Fulton, 2002). Since the operation of the experimental rain gauge network started in 2004, the multisensor estimates used in this study are entirely based on the MPE algorithm. The MPE algorithm allows for an optimal mosaicking technique in areas of overlap, in which the reflectivity value is extracted from the radar with the lowest unblocked beam instead of being taken as the maximum or the average reflectivity value as previously used in Stage III. MPE also makes use of radar coverage delineation to distinguish areas of blockage and deteriorating detection ability so that bias-adjusted estimates are not tainted. In addition to mean-field bias adjustment, the MPE algorithm has includes a new local bias correction capability (Seo and Breidenbach, 2002) that accounts for spatially non-uniform biases that may exist in the individual DPA radar products. As with Stage II and Stage III, MPE also provides interactive quality control for editing the data and rerunning the algorithms. In the final stage of processing, Stage IV, the MPE estimates (Stage III prior to 2003) undergo an additional process which involves the mosaicking of the estimates from all the RFCs on a national scale. The final product is known as the Stage IV hourly precipitation dataset, which can be obtained from the NCEP archives.

Dense rain gauge network

Evaluation of the MPE products will be achieved through comparison against observations from an experimental surface rain gauge network. The network is owned and operated by the Civil Engineering Department of the University of Louisiana at Lafayette and is located within the Isaac-Verot watershed (Fig. 1). The network is composed of a total of 13 tipping-bucket rain gauge sites, with each site having a dual-gauge setup. The gauges are of the same type used in the Oklahoma Mesonet (Brock et al., 1995) and have been specially modified by the manufacturer to improve their accuracy and reliability. Each gauge has an orifice size of 30.5 cm (12 in.) and is equipped with a digital data logger that records the time of occurrence of successive 0.254 mm (0.01 in.) tips from which rainfall intensities can be calculated using an interpolation procedure (Habib et al., 2001a,b; Ciach, 2003). The gauge intensities were accumulated to an hourly scale to match the resolution of the multisensor precipitation products under evaluation in this study. Before being deployed in the field, gauges undergo a careful procedure of both static and dynamic calibration (Humphrey et al., 1997). Rain gauges are also known for various operational problems (e.g., mechanical and electronic failure, clogging, etc.). Therefore, the network has been designed using a dual-gauge setup where each site has two gauges located side-by-side. The dual-gauge setup has been recommended by previous validation studies (e.g., Krajewski et al., 2003; Steiner et al., 1999) to significantly improve the quality of the observations through data redundancy and double-checking. In addition, the quality of the data is highly improved through frequent maintenance and downloads to ensure early fault detection and continuous data records. The gauge locations within the watershed were selected in such a way that two HRAP pixels are populated with four dual-gauge stations each, while the other pixels have one or two gauge stations. It is noted that the gauge arrangement was restricted by site availability and accessibility limitations. As explained later, the four-gauge pixels will provide an improved approximation of the

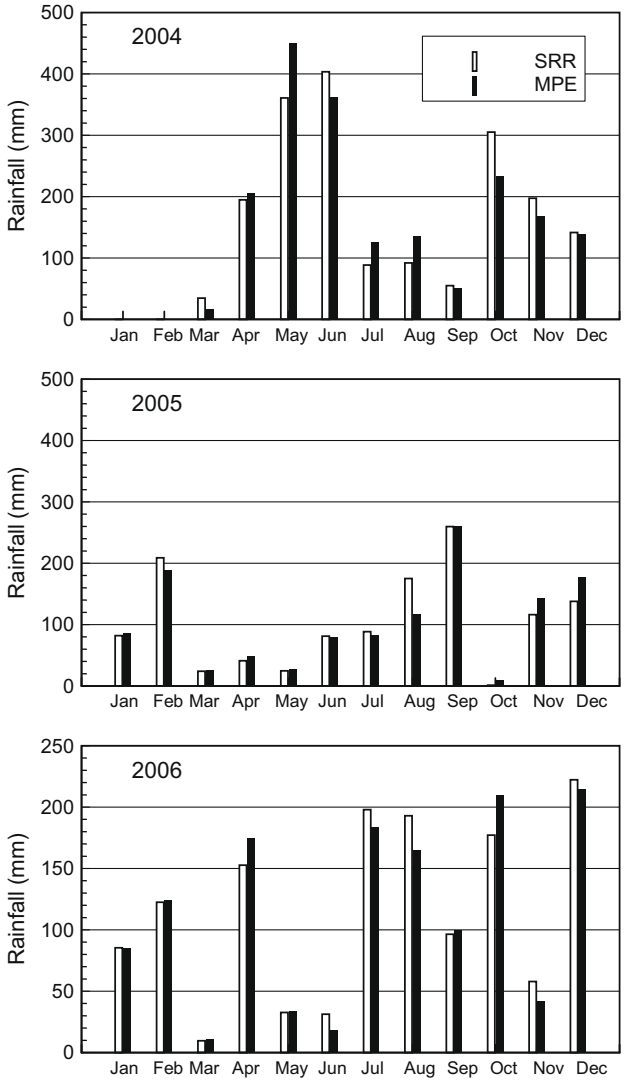

Fig. 2. Monthly rainfall accumulations over the study site for the surface reference rain (SRR) and MPE product in 2004, 2005, and 2006. Gauge data used to establish SRR was not available for January and February of 2004.

unknown true area-average surface rainfall over the MPE pixel scale.

The current study is based on a 3-year period (2004–2006) of gauge and MPE data. Since NCEP Stage IV data have been archived since 2002, there are no missing periods in the MPE dataset. The rain gauge dataset has gaps in several months of 2004 due to the late installation of some gauges. Gauges that are located in the multiple-gauge HRAP pixels were not fully functional until September of 2004. However, other gauges were active as early as March of 2004. Except for short breakdown periods, rain gauge observations were fully available during 2005 and 2006. The year 2004 was a relatively wet year with an annual cumulative rainfall of about 2000 mm (~78 in.) while about 1300 mm (~50 in.) of rain was recorded in each of 2005 and 2006. Calculated monthly accumulations based on the gauge and MPE data (Fig. 2) show significant variations within each year and from 1 year to another. Fig. 3 shows some summary statistics on the distribution of events and rainfall intensities observed during the study period. A large number of events were recorded (Fig. 3), where an event is defined as continuous raining period interrupted by no longer than 6 h of no rain and with a rainfall depth of at least 5 mm. Rainfall events included two tropical-related events (Tropical Storm Matthew in October 2004, and Hurricane Rita in September 2005) each of which generated more than 10 in. of rainfall over the network site, as well as several other storms of varying types, intensities and durations. A characteristic of the plotted histogram of hourly rainfall intensities recorded by the network gauges is the occurrence of intense rain in excess of 25 mm/h (~1 in./h).

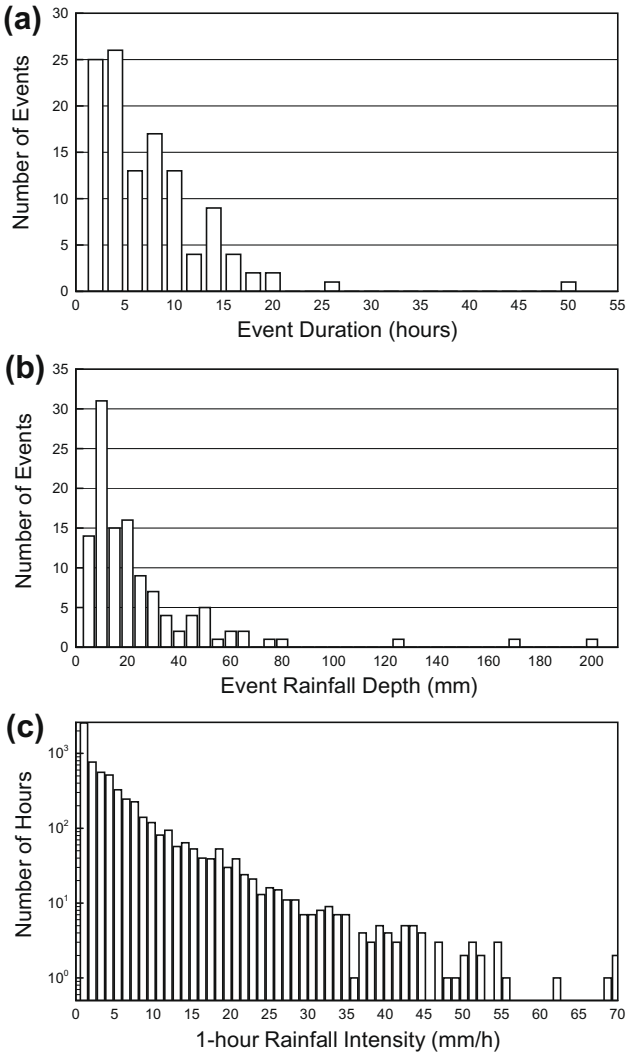

Fig. 3. Summary statistics of rainfall events used in the current study: (a) distribution of event duration (b) distribution of event rainfall cumulative depth, and (c) distribution of hourly surface rainfall intensities recorded by the different rain gauges in the network.

Approach and methods

Surface reference rainfall

Three time scales (hourly, daily, and monthly) are considered in the MPE validation with a special emphasis on the hourly scale due to its hydrological relevance. At hourly or smaller scales, it is recognized that rain gauges may be limited by their near-point sampling and may not provide acceptable approximation of surface rainfall over the MPE 4×4 km^2 scale. Therefore, we limited the hourly-scale analysis to the two pixels that have four gauges each. Within each MPE pixel, hourly observations are averaged from the four gauges to provide an estimate of surface rainfall and then compared to the MPE estimates. The adequacy of these four gauges as a ground reference for MPE evaluation can be quantitatively assessed using the variance reduction factor, VRF (Bras and Rodriguez-Iturbe, 1993). The VRF provides a relative measure of the error variance of areal-rainfall approximations obtained by averaging observations from individual gauges within an area of interest. The VRF depends primarily on the number and configuration of gauges within the area of interest (an individual MPE pixel in this case). However, the VRF, and the adequacy of four gauges to represent areal-rainfall over the MPE pixel scale, will also depend on the degree of rainfall natural spatial variability encountered during the study period. Therefore, the VRF should be assessed for different rainfall spatial variability conditions (i.e., uniform versus variable rainfall). To do this, we calculated the coefficient of variation (CV) of rainfall intensity within the MPE pixel for every individual hour throughout the entire observational period. The hourly CV statistic was calculated as the ratio of the standard deviation of hourly rainfall calculated from the four gauges located within each pixel to the mean rainfall of the same gauges. We then divided the sample according the hourly CV value into three sub-samples that reflect different spatial variability conditions: uniform rain ($CV \leqslant 0.2$), medially variable rain ($0.2 < CV \leqslant 0.5$), and highly variable rain ($CV > 0.5$) and calculated VRF for each condition. Calculation of

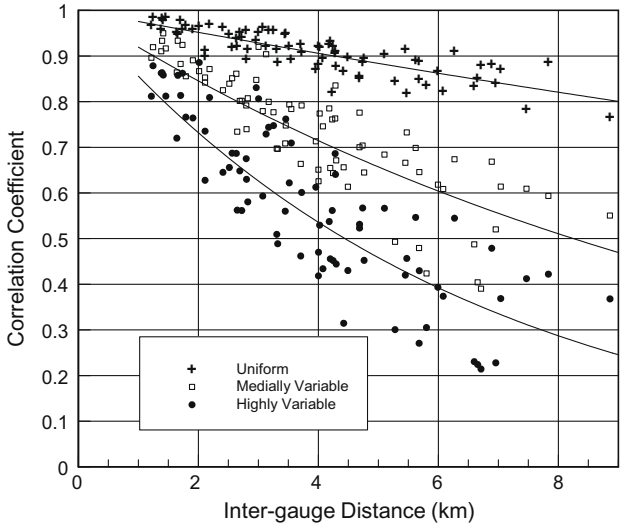

Fig. 4. Spatial correlation functions at hourly scale for three sub-variability conditions: uniform rain ($CV \leqslant 0.2$), medially variable rain ($0.2 < CV \leqslant 0.5$) and highly variable rain ($CV > 0.5$).

the VRF also requires specification of the rainfall spatial correlation function, which was estimated from inter-gauge hourly correlations calculated from the rain gauge network. The correlation functions were developed for each CV condition separately (Fig. 4). We used a common isotropic exponential model of the spatial correlations as a function of distance and fitted it to the calculated correlation coefficients. It should be noted that the selection of the specific CV thresholds was not pre-determined but was rather driven by which thresholds would result in distinctly different rainfall conditions as reflected by the fitted correlation functions. The VRF was computed for the three rainfall conditions and for each of the two multiple-gauge pixels in the network. For comparison, we also computed the VRF for the individual gauges located in the same two pixels. For the uniform rain case ($CV \leqslant 0.2$), the VRF decreased from 3% to 5% for individual gauges to less than 1% when four gauges were used. For the medially variable rain ($0.2 < CV \leqslant 0.5$), the VRF decreased from 10% to 22% with individual gauges to 2–4% when four gauges were used. The improvement was noticeably significant for the highly variable sub-sample ($CV > 0.5$) with the VRF decreasing from 22% to 42% for individual gauges to less than 5% for the four gauge average case.

In the remainder of this paper, reference area-average surface rainfall used for MPE validation will be referred to as SRR. Based on the VRF analysis, it is reasonable to assume that for hourly scale, averaging gauge observations within the two multiple-gauge pixels can provide relatively accurate approximation of SRR. However, at daily or monthly scales, observations from other individual-gauge pixels will also be included since the problem of sub-pixel rainfall spatial variability is significantly smoothed out at the small spatial scales considered herein (Huff and Shipp, 1969). Accordingly, daily and monthly samples will be composed of six MPE/SRR pixels while hourly samples will be limited to two pixels only. The number of MPE-SRR pairs (with at least one of them reporting non-zero rain) is shown in Table 1 for each of the three time scales considered in the study.

Validation metrics and methods

In the upcoming analyses, errors of the MPE product are defined as the differences between MPE estimates and the corresponding reference surface rainfall. The MPE-reference samples are based upon paired datasets where either the reference value or the MPE value is greater than zero. To assess the accuracy of the MPE product with respect to the SRR, we used a suite of both graphical and statistical techniques. Graphical comparisons include scatter plots of rainfall rates/depths to visually inspect how the MPE product compares against the surface reference rainfall. Double-mass plots are generated to examine the progression of cumulative agreement over the three analyzed years. The probability distributions of MPE hourly rainfall rates are compared to those of the reference rainfall by analyzing the probability of exceedance of each dataset. Systematic and random differences between MPE and SRR are evaluated using several statistical metrics that include continuous and categorical measures, unconditional and conditional metrics, and rainfall self-correlation in time and space.

Categorical statistics

The categorical statistics used in the comparisons of the MPE estimates with the reference dataset are the probability of detection (POD) and the probability of false detection (POFD). The POD represents the ratio of the number of correct detections of rainfall by MPE to the total number of SRR rainfall occurrences while the POFD represents the ratio of the number of false identifications of rainfall to the total number of non-rainfall occurrences. Both POD and POFD range from 0 to 1, with 1 being a perfect POD and 0 being a perfect POFD. To investigate dependence on the magnitude of rain rate, the POD and POFD are broken down through conditioning on various SRR thresholds. The volumes of rainfall correctly and incorrectly identified by MPE are also observed as an indication of the significance of either probability in the overall performance of the MPE product.

Continuous statistics

The continuous statistics used to quantify the differences between the MPE estimates and the reference dataset are the mean difference (a measure of the bias), the standard deviation of differences (a measure of the random error), and the Pearson's correlation coefficient (a measure of linear association) defined as follows:

$$\text{Bias}: B = \overline{R}_{\text{MPE}} - \overline{R}_{\text{SRR}} \tag{1a}$$

$$\text{Relative bias}: RB = \frac{\overline{R}_{\text{MPE}} - \overline{R}_{\text{SRR}}}{\overline{R}_{\text{SRR}}} \tag{1b}$$

$$\text{Standard deviation of difference } \sigma_{(R_{\text{MPE}} - R_{\text{SRR}})} \tag{2a}$$

$$\text{Relative standard deviation of difference}: \frac{\sigma_{(R_{\text{MPE}} - R_{\text{RR}})}}{\overline{R}_{\text{SRR}}} \tag{2b}$$

$$\text{Pearson's correlation coefficient}: R^2 = \frac{\overline{(R_{\text{MPE}} - \overline{R}_{\text{MPE}})(R_{\text{SRR}} - \overline{R}_{\text{SRR}})}}{\sigma_{\text{MPE}}\sigma_{\text{SRR}}} \tag{3}$$

where R_{MPE} is the MPE rainfall rate (or accumulation) and R_{SRR} is the corresponding reference value. The overbar and σ symbols denote the sample mean and standard deviation, respectively. These statistics will be computed by combining the MPE/SRR samples for each of the three analyzed years. The bias will also be calculated at an event scale. To investigate whether the MPE error depends on rain-

Table 1
Statistical measures based on the comparison of reference rainfall (SRR) and the MPE dataset for 2004–2006 at hourly, daily, and monthly time scales.

	2004			2005			2006		
	Hourly	Daily	Monthly	Hourly	Daily	Monthly	Hourly	Daily	Monthly
Sample size	611	589	43	1267	938	69	1767	1170	68
Mean (MPE) (mm)	2.12	12.45	167.96	2.02	7.76	105.49	1.53	6.50	110.33
Mean (SRR) (mm)	2.29	12.30	164.97	1.96	7.71	104.75	1.59	6.51	110.25
σ_{MPE} (mm)	4.78	21.72	103.40	4.21	16.32	71.36	4.03	12.68	68.89
σ_{RR} (mm)	5.61	23.15	115.30	4.37	17.19	75.64	4.76	13.38	67.67
Bias (mm)	−0.17	0.15	2.99	0.06	0.05	0.74	−0.06	−0.01	0.08
Relative bias	−0.07	0.01	0.02	0.03	0.01	0.01	−0.04	0.00	0.00
$\sigma_{(\text{MPE}-\text{SRR})}$	3.25	9.94	46.67	2.36	6.45	28.71	1.84	4.62	20.70
Relative $\sigma_{(\text{MPE}-\text{SRR})}$	1.42	0.81	0.28	1.20	0.84	0.27	1.16	0.71	0.19
Correlation	0.82	0.90	0.91	0.85	0.93	0.93	0.93	0.94	0.95

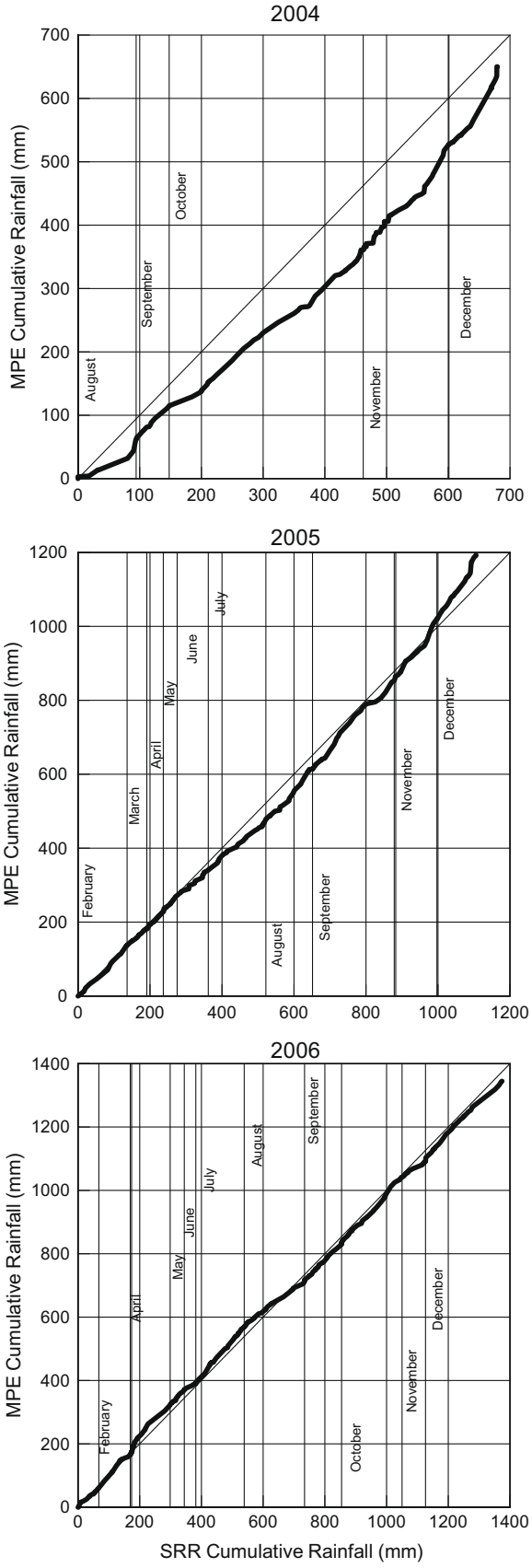

Fig. 5. Double mass curves showing the cumulative rainfall of the MPE and SRR as a function of time for 2004, 2005, and 2006 using data from one of the 4-gauge MPE pixels. Vertical lines indicate the beginning of each month.

fall magnitudes, the statistical metrics will also be analyzed by conditioning on various ranges of the reference rainfall.

Self-correlation statistics

The validation metrics described so far are based on comparisons of individual MPE pixel estimates to the corresponding SRR values over the same pixel. Another approach focuses on the analysis of the underlying structure of the rainfall fields (e.g., Gebremichael and Krajewski, 2004; Germann and Joss, 2001; Harris et al., 2001; Ebert and McBride, 2000; Marzban and Sandgathe, 2009). Such an analysis (e.g., variograms or correlation functions) provides insight into how the MPE product can reproduce the spatial and temporal organization of the surface rainfall. A full examination of this aspect of the validation process across a wide range of scales is not possible in the current study due to the limited spatial extent of the reference network, which covers only few MPE pixels. Therefore, we examine the self-correlation present in the reference and MPE datasets by calculating the spatial auto-correlation at one lag only (4-km). The temporal auto-correlation can be calculated at several lags starting from 1 h.

Error decomposition

To gain more insight on the source of MPE errors, the overall bias between the MPE estimates and the reference rainfall can be further decomposed as follows. The bias calculated using Eq. (1) is based on aggregation of differences in rainfall volume over the entire sample and does not provide information on the source of such differences. Therefore, it is desirable to break down the total bias into three components consisting of the bias associated with successful detections (hits), bias due to rainfall misses, and bias due to false detections:

$$\text{Hit bias (HB)} = \sum (R_{\text{MPE}}(R_{\text{MPE}} > 0 \;\&\; R_{\text{SRR}} > 0) - R_{\text{SRR}}(R_{\text{MPE}} > 0 \;\&\; R_{\text{SRR}} > 0)) \quad (4)$$

$$\text{Missed-rain bias (MB)} = \sum R_{\text{SRR}}(R_{\text{MPE}} = 0 \;\&\; R_{\text{SRR}} > 0) \quad (5)$$

$$\text{False-rain bias (FB)} = \sum R_{\text{MPE}}(R_{\text{MPE}} > 0 \;\&\; R_{\text{SRR}} = 0) \quad (6)$$

Such decomposition can distinguish among the three possible bias sources, whose values can be cancelled by opposite signs if the total bias is only evaluated. The summation of these three components adds up to the total bias (TB $= \sum(R_{\text{MPE}} - R_{\text{SRR}})$). The proportion of total bias attributed to each bias source can be described by the ratio of the particular bias component to the total bias (e.g., HB/TB, MB/TB, and FB/TB), with the three ratios adding up to 1.

Results and discussion

We start by showing some graphical comparisons between the MPE product and the reference rainfall dataset. As shown earlier, monthly comparisons of MPE versus the reference rainfall dataset (Fig. 2), indicate that the MPE estimates were able to capture the overall monthly trends and magnitudes. Cumulative rainfall observed by the MPE and the reference dataset SRR as a function of time are shown in the form of double-mass curves (Fig. 5). These plots are produced by accumulating MPE and surface rainfall over one of the four-gauge pixels. Ideal agreement on double-mass curves is identified when the plotted curve is parallel to the 1:1 line; any deviation from this direction is an indication of cumulative drift (either overestimation or underestimation) by the MPE estimates in comparison to the reference rainfall. In 2004, signs of such drifts are evident especially in August, early October, late November and December. The comparison is relatively better in 2005 and 2006 where approximate equality between MPE and sur-

face rainfall is noticed in several months. Staircase-like features, which indicate detection problems, are also observed (e.g., during November months in 2004 and 2005).

Analysis of rainfall distributions

Distribution-based comparisons can be made by examining scatter plots of the MPE estimates versus the reference values (Fig. 6). These graphs are generated by pooling data pairs from all relevant pixels (2 pixels for the hourly scale and 6 pixels for the daily and monthly scales). Significant scatter exists between the MPE estimates and the corresponding reference values at the hourly scale where differences as large as 15–20 mm/h are not uncommon. Several instances of failed and false detection by MPE are observed. As expected, the scatter is reduced as the MPE estimates are accumulated to longer time scales such as daily and monthly. However, significant differences are still observed at such scales (up to 20–25 mm/day at the daily scale and up to 50 mm/month at the monthly scale). An improvement over time is apparent in the relatively reduced scatter in 2006 compared to 2004 and 2005.

To further examine the distributional agreement, the marginal distributions of the MPE estimates and the surface reference rainfall were analyzed by examining the probability of exceedance of each dataset (Fig. 7). The probability of exceedance is defined and calculated as the probability that an estimate exceeds a certain threshold. This analysis is relevant from a hydrologic prediction point of view since it examines the exceedance of extreme rainfall thresholds, which are usually responsible for triggering flash floods or other extreme natural events. For this reason and sample size limitations, results from the hourly and daily scales only are presented. Consider the extreme tail of the distributions, which can be loosely defined by the hourly rain rate exceeding 8–10 mm/h (~5–10% hourly exceedance probability) and the daily rainfall

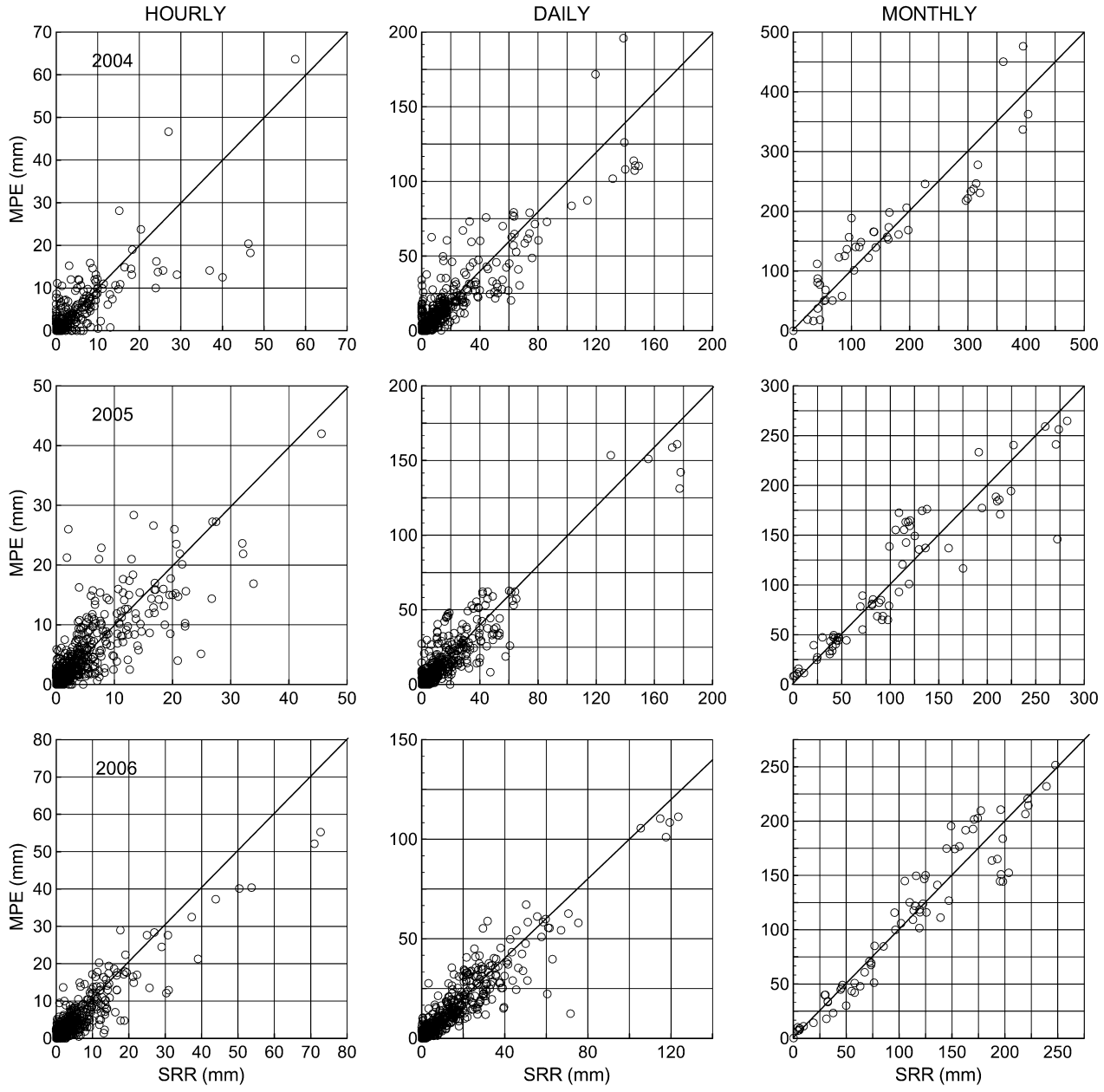

Fig. 6. Scatter plots of MPE versus SRR for 2004–2006 at hourly, daily, and monthly time scales.

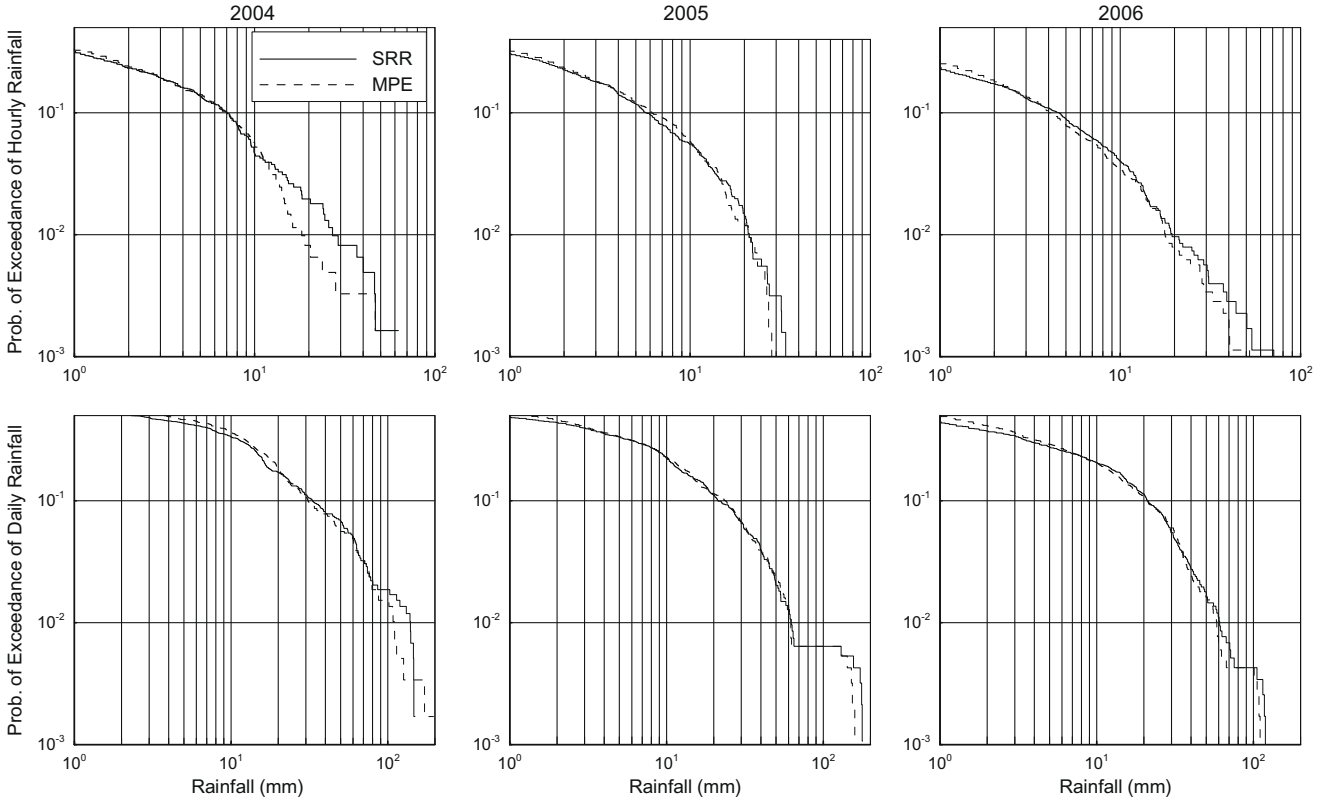

Fig. 7. Probability of exceedance plots of SRR and MPE at hourly (upper panels) and daily (lower panels) time scales.

depth exceeding 20–30 mm (~5–10% daily exceedance probability). A close agreement between the tails of the two distributions is clear (especially in 2005 and 2006), which indicates the ability of MPE estimates to represent the occurrences of extreme rainfall values. The two distributions deviate at the extreme tail (1% exceedance or lower) where the MPE distribution shows less number of occurrences of such extreme intensities. However, it should be noted that the estimation of the most extreme tail of the distributions may be subject to sampling effects.

Analysis of MPE Bias

The overall total bias between the MPE estimates and the reference rain is quantified by calculating the difference between their arithmetic means over every year and is expressed in absolute and relative units (Eqs. (1a) and (1b)); Table 1. When aggregated over the full sample of each year, the MPE estimates have relatively small bias values. These results agree with earlier MPE evaluation studies which reported low bias levels of the MPE products (e.g., Westcott et al., 2008; Wang et al., 2008). In addition to aggregating the bias over each year, we also assessed the bias on an event scale (Fig. 8), where an event is defined as continuous raining period interrupted by no longer than 6 h of no rain and with a rainfall depth of at least 5 mm. Overall, about 50% of the events have a bias within ±25%, 90% of them have a bias within ±50%, and only 10% of the events have a bias exceeding 50% with some as high as 100%. The histogram of event-scale bias indicates a skewed distribution with 65% of events having a negative underestimation bias. Events with high average rain rate (>8 mm/h) are mostly characterized with negative bias.

As described earlier, the total bias consists of three components; hit bias, missed-rain bias, and false-rain bias, the values of which may cancel each other when added up to form the total bias. The three bias components were calculated at the hourly scale and are shown for the two four-gauge pixels (Fig. 9); note that these volumes are combined totals from the two individual pixels. The percentages shown in the figure represent bias values relative to the total rainfall depth. It is clear that the three bias components

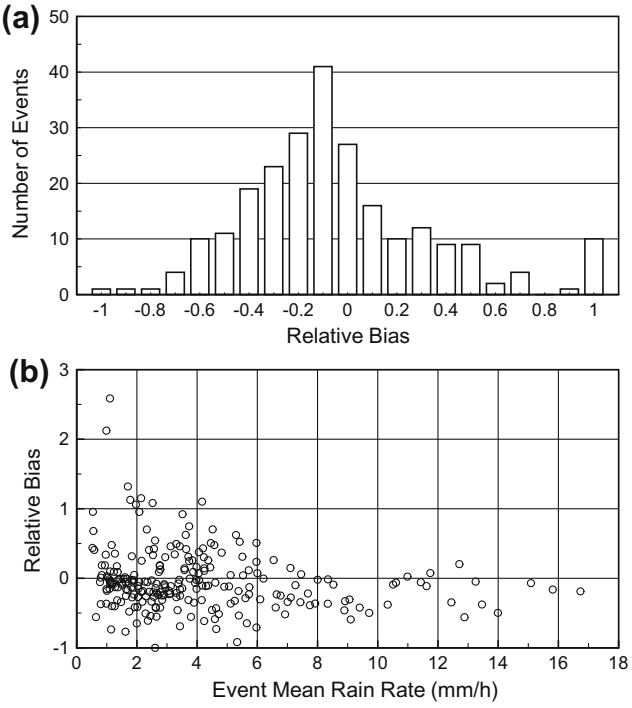

Fig. 8. (a) Distribution of the event-scale bias of the MPE estimates for all rainfall events in 2004–2006. (b) Event-relative bias as a function of the event mean rain rate.

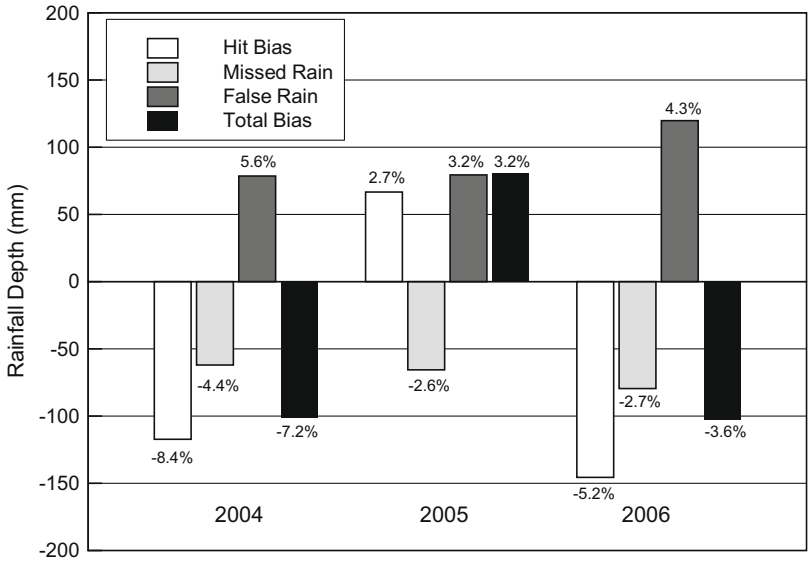

Fig. 9. Decomposition of the total bias of the MPE into its three components, hit bias, missed rain, and false rain for each year. The percentage shown above each bar indicates the ratio of the bias component to the total annual rainfall amount.

have comparable magnitudes (~2% to 8%) to each other and to the overall bias. While the hit-bias explains a significant portion of the overall MPE bias, both the false rain and the missed rain have comparable contributions. It is also interesting to observe that the value of each of the three bias components can be cancelled or diminished by the other components due to their respective signs; therefore, examining only the total overall bias can be rather misleading.

To examine whether the MPE bias is dependent on the rain rate magnitude, the full sample was divided into sub-samples based on different ranges of the reference rain rate and the conditional bias was calculated for every range (Fig. 10). The relative bias is calculated by normalizing with the mean rain rate of each range. To avoid deterioration of the sample size, the conditioning was done after pooling together data from the 3 years. It is clear that the MPE bias depends on the magnitude of the rain rate. The bias tends to be positive for the lower range of reference rates and decreases gradually to become negative for higher rates. While the unconditional bias was in the range of 3–7%, the conditional bias has rather high levels (60–90%) at small intensities (<0.5 mm/h) and decreases to about −20% at large rainfall rates (>10 mm/h). This overall behavior of the conditional bias, which is consistent across the three analyzed years, indicates that the MPE estimates tend to overestimate small rain rates and underestimate large rain rates. Similar observations of this behavior were also reported in other MPE evaluation studies (e.g., Westcott et al., 2008).

Analysis of MPE detection

To further characterize the detection capability of the MPE product, we examine two probabilities: probability of detection (POD), and probability of false detection (POFD); Fig. 11. To examine the significance of detection limitations in MPE estimates, we also kept track of the volume of rain that is either missed due to lack of detection (POD < 1) or falsely detected (i.e., POFD > 0). This analysis is reported for the hourly scale only. Consider first the POD results. Conditioned on SRR larger than zero, the MPE estimates show low to medium POD values (0.5–0.6). The undetected surface rainfall exceeded 1 mm/h for less than 0.6% of the time (largest value is 6.5 mm/h). When conditioned on surface rain rates larger than a certain threshold, the POD increases which indicates that the rather low POD values are caused by lack of detection of small rainfall intensities (less than 0.13 mm/h). The POD increases to about 0.9 when conditioning on surface rain rates larger than 0.13 mm/h. As shown earlier in the bias decomposition results, the volume missed due to lack of detection is smaller than 5% of the total rainfall volume. The POD continues to improve with the increase of the surface threshold rain rate and reaches 0.9 and higher when SRR is larger than 1 mm/h. The volume of missed rain diminishes and approaches zero at a threshold of 7 mm/h. Results on the POFD (Fig. 11) show that false detection of rainfall by MPE occurs less than 2% of the time. This translates into a falsely detected volume of rainfall of 5% or lower. Falsely detected intensities had a maximum of 8 mm/h and exceeded 1 mm/h for less than 1% of the time. Similar to the POD, the PODF values decline rapidly with the increase of the rainfall threshold which indicates that most false detection cases were associated with low MPE intensities. It should be noted that POD and PODF values, especially at low thresholds, are sensitive to differences in the minimum detectable rainfall by MPE and SRR. The minimum SRR non-zero value is controlled by the average of observations from individual rain gauges, each of which has a detectable threshold of 0.254 mm/h. While the first stage of the MPE product (Stage I or DPA) has a threshold of 0.25 mm/h, the final MPE product can have rainfall hourly values that are lower than 0.25 mm/h (over the area of this study, the minimum MPE non-zero rainfall amount was found to be 0.13 mm/h). This particular value is not enforced by the MPE algorithm; instead, it is possibly a result of various data processing effects such as spatial interpolation between a zero gauge report and a nonzero radar report, bias adjustment of radar data, or mosaicking of zero and no-zero regional radar analyses into the national Stage IV product (personal correspondence with David Kitzmiller, NWS, and Ying Lin, NCEP).

Analysis of agreement and disagreement statistics

The linear association between MPE and the reference rainfall is assessed using the Pearson's correlation coefficient (Table 1). Overall, the MPE hourly estimates show strong correlation values (around 0.8 for 2004 and 2005 and above 0.9 for 2006), which reflects a reasonable skill for the MPE product in reproducing the temporal fluctuations of the reference rainfall. The correlation ex-

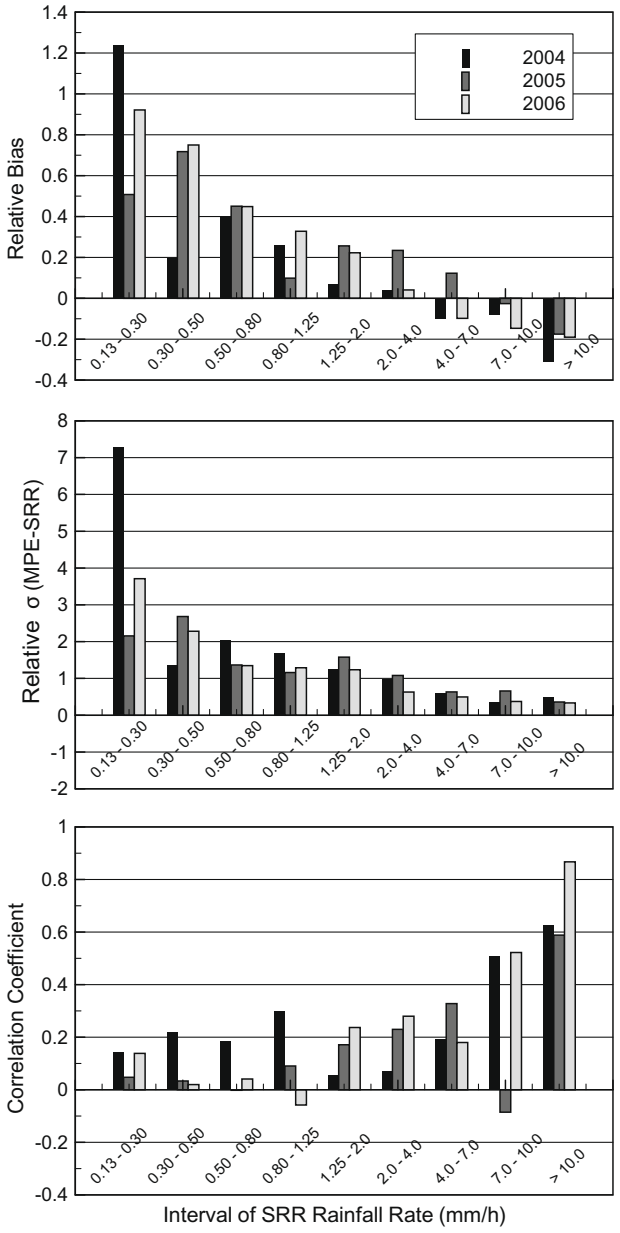

Fig. 10. Statistical performance measures of MPE versus SRR conditioned on various ranges of SRR.

ceeds 0.9 when the MPE is accumulated to daily and monthly scales. Similar to the conditional bias, the correlation level depends on the magnitude of surface rainfall intensity (Fig. 10). Very low correlation values are reported for the small to medium rainfall intensities, which indicates a poor association between the MPE estimates and the reference rainfall at such intensities but also may be partly attributed to data quantization and limited resolutions. High correlations are obtained only for large rainfall intensities (>10 mm/h), which reflects the ability of MPE estimates to better track surface rainfall during the intense part of the storms. However, such high correlation levels might be artificially inflated to some extent due to the sensitivity of the Pearson's correlation statistic to extreme values.

Now we turn to statistics that focus on assessing disagreement between MPE and reference rainfall using the standard deviation of differences between MPE and SRR hourly intensities. The standard deviation of differences is calculated unconditionally Table 1 for each year and conditionally on different intervals of SRR after combining data from the 3 years (Fig. 10). The results are presented in absolute units (mm/h) and also as a ratio relative to the mean rainfall rate of each SRR interval. The unconditional standard deviation of differences exceeds 100% in each of the 3 years. When conditioned on the magnitude of rainfall intensity, the relative standard deviation attains relatively high values (200–400%) for small intensities (<0.5 mm/h), remains around 100% for intermediate intensities (0.5–2 mm/h) and decreases to less than 50% for large rainfall intensities (>5–10 mm/h).

Rainfall self-correlation structure

We now examine how the MPE product can reproduce the underlying spatial and temporal organization present in the surface rainfall. Consider first the spatial auto-correlation. Ideally, one should establish correlation functions or variograms from the two datasets at various spatial lags. However, based on the gauge network spatial setup, estimation of surface rainfall is available only over the two 4-gauge pixels. Therefore, we limited the auto spatial correlation analysis to one lag only of 4-km, which is the separation distance between these two neighboring pixels. Correlation coefficients between hourly rainfall intensities at the two pixels were calculated for both the reference rainfall and the MPE product. The results represent correlation coefficients of hourly intensities stratified by the month (Fig. 12). The overall correlation trends of the MPE are very similar to those of the reference rainfall, especially during cold months. The agreement is still reasonable during warm months (except June) with MPE having overall lower self-correlation levels, which is likely due to the dominance of highly variable localized storms during this time of the year. Unlike the spatial correlation, the temporal auto-correlation can be calculated for various time lags; however, results for 1-h lag only are presented (Fig. 12) since the correlation becomes practically negligible for longer lags. As expected, temporal auto-correlations are generally low, but it is interesting to see that the MPE data produces similar temporal 1-h self-correlation values to those of the surface rainfall.

Effect of pixel sub-variability and gauge sampling errors on MPE evaluation

In most studies that are concerned with evaluation of radar-based rainfall products, single-gauge measurements are usually used as the surface rainfall reference. As discussed earlier, the near-point sampling of a single gauge and its limited representation of area-average rainfall over the size of a 4×4 km^2 MPE pixel can contaminate the information sought on actual MPE estimated errors. The experimental setup available in this study makes it possible to further investigate this issue and provide a quantitative assessment of such contamination. To do this, we follow a simple data-based approach and recalculate the MPE error assuming single-gauge observations as the reference rainfall. Then, we compare the results to those obtained earlier when average observations from four gauges within one pixel were used as the reference rainfall. To be consistent, the single-gauge analysis was based only on individual gauges from those located within the two 4-gauge pixels. The impact of using single-gauge observations as a reference is expected to depend on how variable rainfall is over the scale of an MPE pixel. Therefore, we divided the data into the same three sub-samples used earlier (Section "Surface reference rainfall") which represent uniform rain ($CV \leqslant 0.2$), medially variable rain ($0.2 < CV \leqslant 0.5$) and highly variable rain ($CV > 0.5$). Statistics of the MPE performance using both sets of reference rainfall (single gauge and average of gauges) are compared in terms of bias, standard deviation of differences and the correlation coefficient

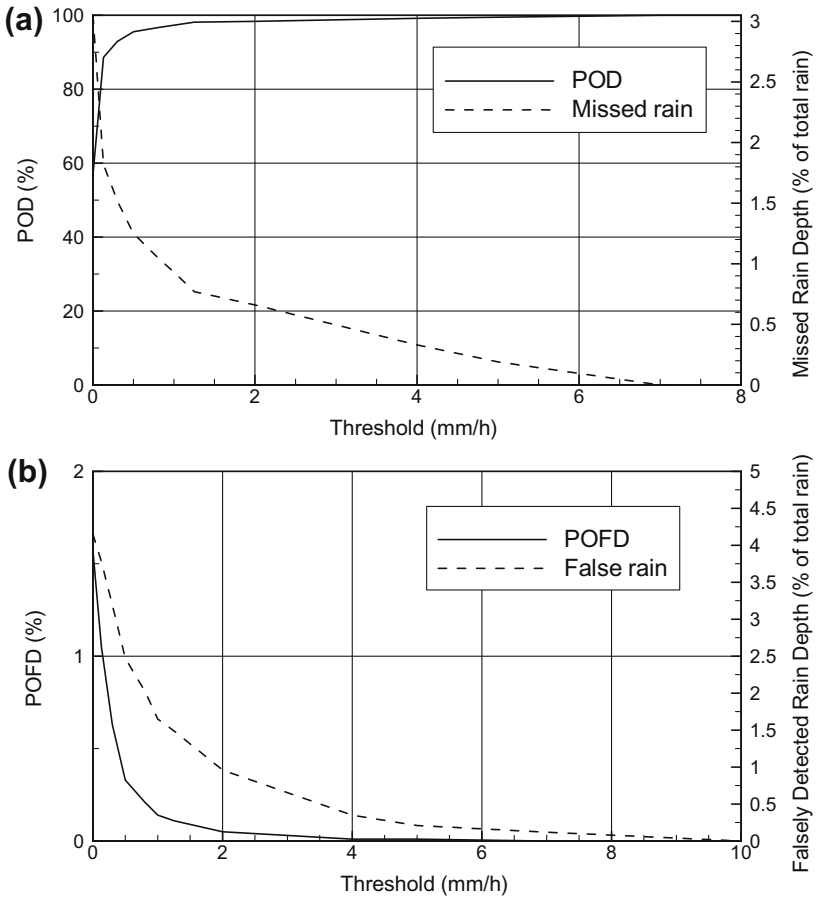

Fig. 11. (a) Probability of detection (POD) of MPE conditioned on various SRR thresholds (left *y*-axis) and the corresponding percentage of missed rain (right *y*-axis). (b) Probability of false detection (POFD) of MPE condition on various SRR thresholds (left *y*-axis) and the corresponding percentage of falsely detected rain (right *y*-axis).

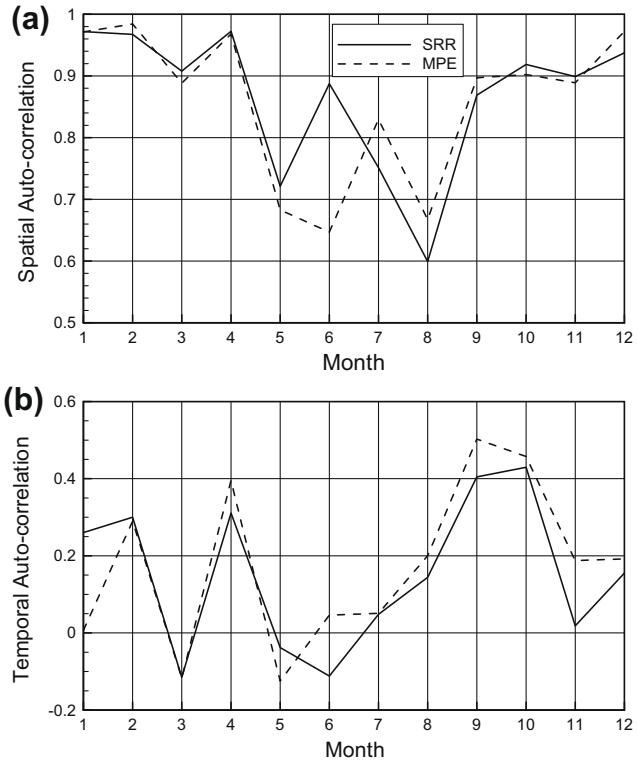

Fig. 12. Spatial (a) and temporal (b) auto-correlation of hourly rainfall for MPE and SRR stratified by the month.

(Table 2). As expected, the MPE bias is not affected since individual gauges can usually provide reasonable approximation of long-term rainfall accumulations. However, compared to the average-gauge case, the single-gauge performance statistics shows a worse than actual MPE performance as reported in the standard deviation of differences and in the correlation coefficient. These two statistics report larger differences between the single-gauge and average-gauge analysis within the samples that are characterized with higher degrees of sub-pixel variability. For the highly variable sub-sample, the change in the standard deviation of the differences increased to almost 40% and the decrease in the correlation coefficient became close to 20%. Even with the case of medium variability, the performance statistics still show significant differences (about 20% increase in the standard deviation and 6% decrease in the correlation). These results indicate that relying on observations of a single gauge within an MPE pixel will lead to an understatement of the MPE performance caused by inaccuracy that is not totally attributable to the MPE estimates.

More insight into the effect of the lack of representativeness of single gauges as a surface rainfall reference on the MPE evaluation can be gained through the following formulation:

$$(R_{MPE} - R_G) = (R_{MPE} - R_{SRR}) + (R_{SRR} - R_G) \quad (7)$$

This formulation is based on the concept of error separation (first introduced by Ciach and Krajewski, 1999a,b) in which the total difference between MPE estimates and individual gauge observations (R_G) can be decomposed into two components. The first component (first term on RHS of (6)) is the difference between MPE and the reference surface rainfall (R_{RR}) obtained by averaging

Table 2
Statistical measures of MPE versus SRR calculated for three sub-samples divided based on the degree rainfall variability (quantified by coefficient of variation CV). The comparison is shown for two cases of SRR: single gauge observations and average of four gauges.

	Uniform (CV ⩽ 0.2)		Medially variable (0.2 < CV ⩽ 0.5)		Highly variable (CV > 0.5)	
	Average of four gauges	Single gauge	Average of four gauges	Single gauge	Average of four gauges	Single gauge
Relative bias	−0.08	−0.08	−0.12	−0.12	0.08	0.08
Relative $\sigma_{(MPE-SRR)}$	0.74	0.76	0.79	0.95	2.13	2.98
Correlation coefficient	0.87	0.87	0.86	0.81	0.80	0.65

observations from the four gauges within each pixel, and represents the actual MPE error. The second component (second term on RHS of (6)) is the difference between the gauge and the reference rainfall and represents the gauge error due to lack of areal representativeness. In typical validation studies, information on R_{RR} is usually unavailable and the $(R_{MPE} - R_G)$ differences are assumed as surrogate for the actual MPE error $(R_{MPE} - R_{SRR})$.

To assess the implications of such assumption, we calculated the ratio of the standard deviation of the $(R_{MPE} - R_G)$ difference term to that of the actual MPE error $(R_{MPE} - R_{SRR})$; (Fig. 13). This ratio represents the unrealistic inflation (or deflation) of the actual MPE estimation due to using single-gauge observations as an approximation of the area-average surface rainfall. For reference, we also show the ratio of the standard deviation of $(R_{SRR} - R_G)$ to that of $(R_{MPE} - R_G)$, which reflects the significance of the gauge error caused by a lack of spatial representativeness. These ratios are calculated separately for each gauge site within the two 4-gauge MPE pixels. The reported results indicate a slight change in the standard deviation of the MPE error in the case of uniform rain sub-sample. However, a significant inflation of the error standard deviation is noticed for the case of medially and highly variable rain sub-samples. The standard deviations of the MPE error estimated by using single-gauge data as a reference are 100–140% as much as those of the best-estimate of the actual MPE error for the medially variable rain and reaches 120–180% in the case of the highly variable rain. When combining the three sub-samples, the ratio of the two standard deviations is mostly between 100% and 120%. We point out that these results are expected to be particularly valid for hourly and 4-km scales due to the significant levels of rainfall variability at such small scales (e.g., Huff and Shipp, 1969; Krajewski et al., 2003). At coarse spatial and temporal scales (daily or monthly and larger areas), and as reported by Westcott et al. (2008), the number of gauges employed to construct the surface reference rainfall do not have much impact on the results inferred regarding the performance accuracy of MPE estimates.

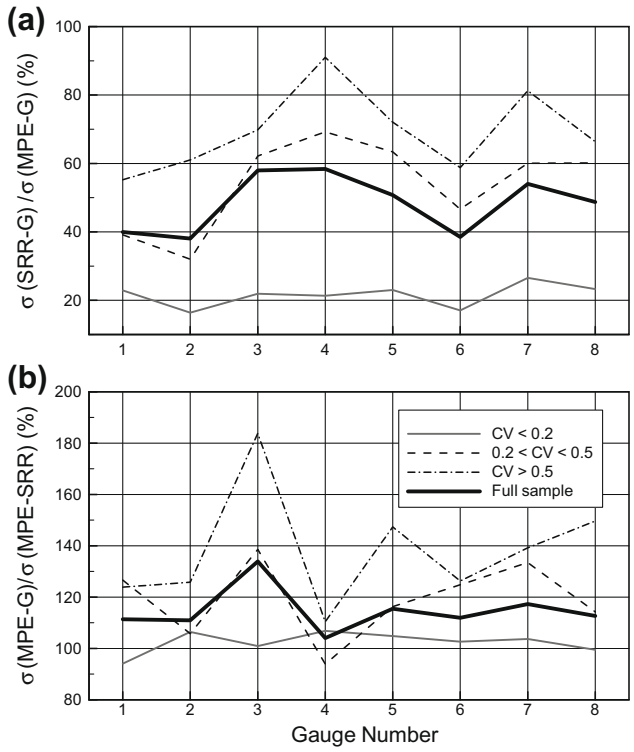

Fig. 13. (a) Ratio of the standard deviation of $(R_{SRR} - R_G)$ to that of $(R_{MPE} - R_G)$ for three sub-samples divided based on the degree of rainfall spatial variability quantified by the coefficient of variation (CV). (b) Ratio of the standard deviation of $(R_{MPE} - R_G)$ to that of $(R_{MPE} - R_{SRR})$ for each sub-sample.

Summary, conclusions and final remarks

This paper evaluated a radar-based multisensor precipitation estimation (MPE) product with the objective of providing the user community and the algorithm developers with some insight on its accuracy in comparison to surface reference rainfall. The MPE product is produced operationally at the National Weather Service (NWS) regional River Forecast Centers and nationally mosaicked and archived as a Stage IV product at the National Center for Environmental Prediction (NCEP). The evaluation was conducted at the native resolution of the product ($4 \times 4\ km^2$ and hourly) during a 3-year period (2004–2006) over a small experimental watershed in south Louisiana, United States. A surface-rainfall dataset from a dense high-quality rain gauge network was used as independent reference for evaluating the MPE estimates. Since this gauge network was not used by the NWS in any of the MPE development stages, the results of this study can be considered as an independent validation. The dense spatial arrangement of rain gauges allowed for two MPE pixels to have four gauges located within each pixel. Such arrangement provides a reasonable approximation of area-averaged rainfall over the grid scale of the MPE product and alleviates the limitations of near-point observations typically available from sparsely located gauges. A suite of graphical and statistical techniques were implemented to characterize the differences and agreement between the MPE estimates and the reference rainfall values. Based on the results of this study, the following conclusions can be made:

1. When aggregated over long time scales (e.g., annual), the overall bias between MPE and surface rainfall is rather small. However, on an event basis, the bias reaches up to ±25% of the event total rainfall depth during 50% of the events and falls between 50% and 100% for 10% of the events. When further decomposed into its three sub-components (bias associated with successful hits, positive bias due to false detections, and negative bias due to lack of detection), it appears that these three components have comparable contributions to each other and to the total bias.
2. When conditioned on the magnitude of the rain rate, the MPE estimates tend to overestimate small rain rates (conditional bias of 60–90% for rates lower than 0.5 mm/h) and underesti-

mate large rain rates (conditional bias of −20% for rates higher than 10 mm/h). Similarly, the relative standard deviation of MPE differences from surface rain rates is quite high (200–400% of mean rainfall intensity) for small intensities (<0.5 mm/h), remains around 100% for intermediate intensities (0.5–2 mm/h) and decreases to less than 50% for large rainfall intensities (>5–10 mm/h).

3. Based on the detection analysis, MPE has an overall probability of detection of 0.6–0.82; however, most of the undetected rain is in the low range of rain rates (<0.13 mm/h) and a very good detection (90% and higher) is achieved for high rain rates. The poor detection of small rain rates led to a missed rain amount of no more than 5% of the total rainfall volume. False detection by MPE occurs less than 3% of the time and is associated with low rain rates, which indicates that evaporation and rainfall displacement by horizontal wind may be a likely reason. To put these results in a proper perspective, the corresponding POD of a single gauge over the size of an MPE pixel was calculated and was found to be comparable to that of MPE during highly and medially uniform rain, but lower by 2% approximately during highly variable rain. However, it is noted that a single gauge within the experimental research network used in this study doesn't realistically represent, at least from a data quality perspective, what one should expect from typical operational rain gauges. For such gauges, the POD can be much lower than what MPE estimates provide.
4. Despite the significant scatter between MPE and surface rainfall, especially at small intensities, the MPE hourly product has a good association to surface rainfall rates as reflected by the high overall correlation (0.82, 0.85, and 0.93 in 2004, 2005, and 2006). The correlation well exceeds 0.9 at daily and monthly scales. However, conditionally the correlation shows very low values for small to medium rainfall intensities and high correlations are obtained only for large rainfall intensities (>10 mm/h).
5. The tail of the MPE probability distribution shows good agreement with that of the surface rainfall, which indicates the ability of the MPE product to capture the occurrence of intense rainfall. The MPE product is also successful in reproducing the underlying spatial and temporal organization of the surface rainfall as quantified by the spatial and temporal self-correlations. These features are desirable for studies that rely on radar rainfall products as a driver for modeling various hydrologic processes.
6. The experimental layout used in this study made it possible to quantify the impact of relying on observations from sparse networks to validate radar-based rainfall estimates. Using a single gauge within an MPE pixel as a reference representation of surface rainfall resulted in an unrealistic inflation of the actual MPE estimation error by 120–180%. As expected, the most impact was obtained during highly variable rainfall periods.
7. An overall improvement in the performance of MPE estimates was reported over the course of the three analyzed years (2004–2006), which was reflected in all of the statistical performance metrics. The improved performance is likely attributed to factors such as: continuous improvements in the MPE algorithm (e.g., mosaicking of overlapping radars, use of climatology-based effective coverage delineation of individual radars, interactive capabilities for quality control of gauge and radar data) and increased experience by the LMRFC forecasters in using the MPE algorithm.
8. Compared to previous evaluation studies, the statistics reported in this study show better performance by the MPE product. For example, Young et al. (2008) and Grassotti et al. (2003) show correlations in the range of 0.6–0.8 for daily rainfall while higher correlations are obtained in the current study (0.8–0.9 for hourly scale and >0.9 for daily scale). We believe that the reason for the observed better performance by MPE is the enhanced quality and accuracy of the rain gauge dataset used by this study as a validation reference, which ensured that gauge-related errors are not wrongly assigned to radar estimation uncertainties.

In view of the results reported in this paper, the most alarming factor about the MPE performance is probably related to the significant levels of bias observed at the event scale. Tracing the source of such biases in a post-product validation study is complicated by the fact that the final MPE product is a result of a multitude of processing algorithms and procedures that are difficult to trace and isolate. It is known that problems related to hardware calibration can introduce systematic errors into radar-rainfall estimates. However, in a comparative analysis versus space-borne radar observations, Anagnostou et al. (2001) found that the KLCH radar site doesn't suffer from any significant calibration problems. MPE Biases could also be the result of using improper Z–R relationships. In principle, the local weather forecast office (WFO) selects a certain Z–R relationship based on season and the prevailing rainfall regime and environmental conditions. For example, over the study site, the WFO in Lake Charles, LA, switches to a tropical Z–R relationship during tropical events. Unfortunately, it was not possible for us to trace back which Z–R relationships were used in every analyzed event. However, it is also recognized that uncertainties in Z–R selection are mitigated by gauge-based bias adjustment procedures. For example, the time-varying mean-field bias correction applied within the MPE algorithm is analogous to changing the multiplicative factor, A, in the $Z = A \cdot R^b$ relationship in real-time based on observed gauge data. Nevertheless, the efficiency of such bias corrections depends largely on the number and quality of rain gauges available to the MPE algorithm in the proximity of study site. According to the records of the LMRFC, the gauges used by the MPE algorithm for bias adjustment include one gauge at a distance of ~7 km from the study site and four more gauges within 10–20 km. We examined the actual event totals of these gauges and compared them to the corresponding totals from our independent network and found differences to be comparable in magnitude to the MPE biases. This indicates some possible data-quality problems in the gauges used for MPE bias adjustment. The effect of low-quality gauge data can be significant as shown by Marzen and Fuelberg (2005) who found a six fold increase in the hourly biases of MPE estimates when the gauge data had not been quality controlled.

Another more plausible source of the MPE biases, especially underestimation, over our study site is due to range-related effects (smith et al., 1996; Hunter, 1996). At a distance of ~120 km from the closest WSR-88D radar site, the radar beam is at an altitude of ~1.82 km above the validation rain gauge network. At such a range, beam partial filling and overshooting of lower cloud bases and shallower precipitation start to become of a concern. Beam overshooting usually results in low probability of detection; however, lack of detection by MPE over the study site was quite minimal and was associated with very light rain except for few hours (undetected intensities exceeding 5 mm/h but less than 7 mm/h were observed for only 8 h during the entire study period). Relative degradation of the beam sampling resolution over the study site (beam diameter of 1.85 km, approximately) increases the likelihood that rainfall fills only part of the beam and may result in underestimation of the rainfall intensity due to sample volume averaging of the received power. Confirming whether beam overshooting and spreading is the main source of the observed MPE biases requires event-by-event diagnosis of radar reflectivity fields and other environmental conditions, which is beyond the scope of this study. However, the fact that a major number of events were

characterized with a negative bias (i.e., underestimation), especially during heavy rainfall (Figs. 8 and 10) makes range-related factors, especially beam partial filling, the likely source of MPE biases in this study.

The large MPE biases observed at the event scale are problematic especially for hydrologic applications in which accurate prediction of rainfall volume is critical. For example, Habib et al. (2008a,b), Vieux and Bedient (2004) and Gourley and Vieux (2005) emphasized the importance of efficient bias removal in radar-based rainfall estimates before accurate model predictions can be realized. These studies indicated that once the radar bias is successfully removed at an event scale, or conditionally, the remaining random errors tend to diminish and get smoothed out by the rainfall-to-runoff transformation. However, this should not de-emphasize the importance of capturing the occurrence of intense rainfall – an aspect that the MPE data seems to be rather successful with, at least over the domain of this study. With future enhancement on the horizon for the NWS WSR-88D system and the MPE algorithm (e.g., higher resolutions in space and time; dual polarization capabilities; Seo et al., 2005; Ryzhkov et al., 2005), it is anticipated that limitations highlighted in the current study, such as magnitude-dependent biases and detection problems, will be alleviated so that future MPE products and their utility for flood forecasting and flash flood warnings (Smith et al., 2005) can be enhanced.

This study also emphasizes and reiterates the case made by several past studies (e.g., Steiner et al., 1999; Krajewski et al., 2003) on the critical need for experimental surface observation networks that can provide reliable validation datasets as a prerequisite for the assessment of current and future remote-sensing based rainfall estimates. In particular, the quantitative results reported in this study indicated that observations from single gauges within an MPE pixel should not be used in validation analysis so that reliable assessment of the product performance can be achieved. This is particularly important if the validation is performed at small temporal scales that are of significance for MPE-driven hydrologic applications.

Another attribute that is crucially important is the quality of rain gauge data used in the validation analysis. While this study was based on high-quality gauges available within an experimental research setup (e.g., regular gauge maintenance and calibration, frequent site visits and downloads, redundancy checks through the dual-gauge setup), other typical gauge networks are usually contaminated with various gauge-related errors that eventually blur the validation assessment of any MPE products (Steiner et al., 1999; Marzen and Fuelberg, 2005).

Finally, we point out that the validation analysis reported in this study was mainly driven by product performance evaluation purposes. However, information on the statistical characteristics of the MPE estimation error (e.g., error distribution, error dependence structure) are of significant interest for studies that focus on modeling and ensemble generation of error fields and their propagation in hydrologic predictions (Ciach et al., 2007; Habib et al., 2008; Villarini et al., 2009; Germann et al., 2006). The setup of the current rain gauge network, and experimental networks recently developed by other researchers, provide a valuable opportunity for a thorough analysis on error characteristics, which is recommended for further future studies.

Acknowledgments

This study was funded in part by a subaward with UCAR under the sponsorship of NOAA/DOC as part of the COMET Outreach Program, the LaSPACE Graduate Student Research Assistance program, the Louisiana Board of Regents Support Fund Research Competitiveness Subprogram, and the NASA/LEQSF(2005–2010)-LaSPACE contract (NASA grant number NNG05GH22H). The authors also acknowledge Ying Lin from NCEP for input regarding the Stage IV data and Brian R. Nelson and Ali Tokay for their valuable comments at the early stage of this study. Special thanks to Amy Henschke for assistance in data analysis and producing figures and tables. Valuable comments from two anonymous reviewers for enhancing the quality of the manuscript are greatly acknowledged.

References

Anagnostou, E.N., Krajewski, W.F., Smith, J., 1999. Uncertainty quantification of mean-field radar-rainfall estimates. J. Atmos. Ocean. Technol. 16, 206–215.
Anagnostou, E.N., Morales, Carlos A., Dinku, T., 2001. The use of TRMM precipitation radar observations in determining ground radar calibration biases. J. Atmos. Ocean. Technol. 18 (4), 616–628.
Austin, P.M., 1987. Relation between measured radar reflectivity and surface rainfall. Mon. Wea. Rev. 115, 1053–1070.
Boyles, R., Raman, S., Sims, A., Schwab, S., Horgan, K., Brooks, M., Frazier, A., 2006. Evaluation and applications of NCEP stage II and stage IV gage-corrected radar precipitation estimates over the Carolinas. In: Preprints, 20th Hydrol. Conf., Atlanta, GA. Amer. Meteor. Soc. (CD-ROM, 1.1).
Bras, R.L., Rodriguez-Iturbe, I., 1993. Random Functions and Hydrology. Dover, New York, USA.
Breidenbach, J.P., Bradberry, J.S., 2001. Multisensor precipitation estimates produced by National Weather Service River Forecast Centers for hydrologic applications. In: Proceedings of the 2001 Georgia Water Resources Conf., March 26–27, 2001. Institute of Ecology, University of Georgia, Athens.
Brock, F.V., Crawford, K.C., Elliott, R.L., Cuperus, G.W., Stadler, S.J., Johnson, H.L., Eilts, M.D., 1995. The Oklahoma Mesonet: a technical overview. J. Atmos. Ocean. Technol. 12, 5–19.
Ciach, G.J., Krajewski, W.F., 1999a. Radar-rain gauge comparisons under observational uncertainties. J. Appl. Meteorol. 38, 519–525.
Ciach, G.J., Krajewski, W.F., 1999b. Radar-rain gauge comparisons under observational uncertainties. J. Appl. Meteorol. 38, 1519–1525.
Ciach, G.J., Krajewski, W.F., Villarini, G., 2007. Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. J. Hydrometeorol. 8 (6), 1325–1347.
Ciach, G., 2003. Local random errors in tipping-bucket rain gauge measurements. J. Atmos. Ocean. Technol. 20 (5), 752–759.
Dyer, J.L., Garza, R., 2004. A comparison of precipitation estimation techniques over Lake Okeechobee, Florida. Weather Forecast. 19, 1029–1043.
Ebert, E.E., McBride, J.L., 2000. Verification of precipitation in weather systems: determination of systematic errors. J. Hydrol. 239, 179–202.
Fulton, R.A., Breidenbach, J.P., Seo, D., Miller, D.A., O'Bannon, T., 1998. The WSR-88D rainfall algorithm. Weather Forecast. 13 (2), 380–391.
Fulton, R.A., 2002. Activities to improve WSR-88D radar rainfall estimation in the National Weather Service. In: 2nd Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV, July 28–August 1, 2002.
Gebremichael, M., Krajewski, W.F., 2004. Assessment of the statistical characterization of small-scale rainfall variability from radar: analysis of TRMM GV data sets. J. Appl. Meteorol. 43 (8), 1180–1199.
Germann, U., Joss, J., 2001. Variograms of radar reflectivity to describe the spatial continuity of Alpine precipitation. J. Appl. Meteorol. 40, 10421059.
Germann, U., Berenguer, M., Sempere-Torres, D., Salvade, G., 2006. Ensemble radar precipitation estimation—a new topic on the radar horizon. In: Proc. ERAD Fourth European Conference on Radar in Meteorology and Hydrology, Barcelona, Spain, 18–22 September 2006, pp. 559–562.
Gourley, J.J., Vieux, B.E., 2005. Evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective. J. Hydrometeorol. 2, 115–133.
Grassotti, C., Hoffman, R.N., Vivoni, E.R., Entekhabi, D., 2003. Multiple timescale intercomparison of two radar products and rain gauge observations over the Arkansas-Red River Basin. Weather Forecast. 18 (6), 1207–1229.
Habib, E., Krajewski, W.F., 2002. Uncertainty analysis of the trmm ground-validation radar-rainfall products: application to the TEFLUN-B field campaign. J. Appl. Meteorol. 41 (5), 558–572.
Habib, E., Meselhe, E.A., 2006. Stage-discharge relations for low-gradient tidal streams using data-driven models. J. Hydrol. Eng. 132 (5), 482–492.
Habib, E., Krajewski, W.F., Ciach, G.J., 2001a. Estimation of inter-station correlation coefficient in rainfall data. J. Hydrometeorol. 2, 621–629.
Habib, E., Krajewski, W.F., Kruger, A., 2001b. Sampling errors of tipping-bucket rain gauge measurements. J. Hydrol. Eng. 6 (2), 159–166.
Habib, E., Aduvala, A., Meselhe, E.A., 2008. Analysis of radar-rainfall error characteristics and implications for streamflow simulations uncertainty. J. Hydrol. Sci. 53 (3), 568–587.
Habib, E., Aduvala, A., Meselhe, E.A., 2008a. Analysis of radar-rainfall error characteristics and implications for streamflow simulations uncertainty. J. Hydrol. Sci. 53 (3), 568–587.
Habib, E., Malakpet, C.G., Tokay, A., Kucera, P.A., 2008b. Sensitivity of streamflow simulations to temporal variability and estimation of Z–R relationships. ASCE J. Hydrol. Eng. 13 (12), 1177.
Harris, D., Foufoula-Georgiou, E., Droegemeier, K.K., Levit, J.J., 2001. Multiscale statistical properties of a high-resolution precipitation forecast. J. Hyrdol. Meteorol. 2, 406–418.

Huff, F.A., Shipp, W.L., 1969. Spatial correlation of storm, monthly and seasonal precipitation. J. Appl. Meteorol. 8, 542–550.
Humphrey, M.D., Istok, J.D., Lee, J.Y., Hevesi, J.A., Flint, A.L., 1997. A new method for automated dynamic calibration of tipping-bucket rain gauges. J. Atmos. Ocean. Technol. 14, 1513–1519.
Hunter, S., 1996. WSR-88D radar rainfall estimation: capabilities, limitations and potential improvements. NWA Digest 20 (4), 26–36.
Jayakrishnan, R., Srinivasan, R., Arnold, J.G., 2004. Comparison of raingauge and WSR-88D stage III precipitation data over the Texas-Gulf basin. J. Hydrol. 292, 135–152.
Joss, J., Waldvogel, A., 1990. Precipitation measurement and hydrology. In: Atlas, D. (Ed.), Radar in Meteorology: Battan Memorial and 40th Anniversary. Radar Meteorology Conference, Amer. Meteor. Soc., Boston.
Kitchen, M., Blackall, R.M., 1992. Representativeness errors in comparisons between radar and gauge measurements of rainfall. J. Hydrol. 134, 13–33.
Krajewski, W.F., Ciach, G.J., Habib, E., 2003. An analysis of small-scale rainfall variability in different climatological regimes. Hydrol. Sci. J. 48, 151–162.
Lawrence, B.A., Shebsovich, M.I., Glaudemans, M.J., Tilles, P.S., 2003. Enhancing precipitation estimation capabilities at National Weather Service field offices using multisensor precipitation data mosaics. In: Preprints, 19th Conf. on Interactive Information Processing Systems, Long Beach, CA. Amer. Meteor. Soc., 8 pp.
Linsley, R.K., Franzini, J.B., Freyberg, D.L., Tochbanoglous, G., 1992. Water resources engineering, fourth ed. Irwin-McGraw-Hill, New York.
Marzban, C., Sandgathe, S., 2009. Verification with Variograms. Weather Forecast., in press.
Marzen, J., Fuelberg, H., 2005. Developing a high resolution precipitation dataset for Florida hydrologic studies. In: Preprints, 19th Hydrol. Conf., San Diego, CA. Amer. Meteor. Soc., J9.2.
Miriovsky, B., Bradley, A.A., Eichinger, W.B., Krajewski, W.F., Kruger, A., Nelson, B.R., Creutin, J.-D., Lapetite, J.-M., Lee, G.W., Zawadzkt, I., Ogden, F.L., 2004. An experimental study of small-scale variability of radar reflectivity using disdrometer observations. J. Appl. Meteorol. 43 (1), 106–118.
Over, T.M., Murphy, E.A., Ortel, T.W., Ishii, A.L., 2007. Comparisons between NEXRAD radar and tipping-bucket gage rainfall data: a case study for DuPage County, Illinois. World Environmental and Water Resources Congress.
Ryzhkov, A., Giangrande, S., Schurr, T., 2005. Rainfall estimation with a polarimetric prototype of the WSR-88D. J. Appl. Meteorol. 44, 502–515.
Seo, D.-J., Breidenbach, J., 2002. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. J. Hydrometeorol. 3, 93–111.
Seo, D.J., Breidenbach, J.P., Johnson, E.R., 1999. Real-time estimation of mean field bias in radar rainfall data. J. Hydrol. 223, 131–147.
Seo, D.J., Kondragunta, C.R., Howard, K., Vasiloff, S.V., Zhang, J., 2005. The National Mosaic and Multisensor QPE (NMQ) Project—Status and plans for a community testbed for high-resolution multisensor quantitative precipitation estimation (QPE) over the United States. In: Preprints, 19th Hydrol. Conf., San Diego, CA. Amer. Meteor. Soc. (CD-ROM, 1.3).
Smith, J.A., Seo, D.J., Baeck, M.L., Hudlow, M.D., 1996. An intercomparison study of NEXRAD precipitation estimates. Water Resour. Res. 32, 2035–2045.
Smith, J.A., Miller, A.J., Baeck, M.L., Nelson, P.A., Fisher, G.T., Meierdiercks, K.L., 2005. Extraordinary flood response of a small urban watershed to short duration convective rainfall. J. Hydrometeorol. 6 (5), 599–617.
Steiner, M., Smith, J.A., Alonso, C.V., Darden, R.W., 1999. Effect of bias adjustment and rain gauge data quality control on radar rainfall. Water Resour. Res. 35, 2487–2503.
Stellman, K., Fuelberg, H., Garza, R., Mullusky, M., 2001. An examination of radar and rain gauge-derived mean areal precipitation over Georgia watersheds. Weather Forecast. 16, 133–144.
Vieux, B.E., Bedient, P.B., 2004. Assessing urban hydrologic prediction accuracy through event reconstruction. J. Hydrol. 299 (3–4), 217–236.
Villarini, G., Krajewski, W.F., Ciach, G.J., Zimmerman, D.L., 2009. Product-error-driven generator of probable rainfall conditioned on WSR-88D precipitation estimates. Water Resour. Res., W01404. doi:10.1029/2008WR006946.
Wang, D., Smith, M., Zhang, Z., Koren, V., Reed, S., 2001. Statistical comparison of mean areal precipitation estimates from WSR-88D, operational and historical gage networks. NWS Hydrology Laboratory Technical Report, 40pp. [Available from Hydrology Laboratory, 1325 East-West Hwy, Silver Spring, MD, 20910.].
Wang, X., Xie, H., Sharif, H., Zeitler, J., 2008. Validating NEXRAD MPE and stage III precipitation products for uniform rainfall on the upper Guadalupe River Basin of the Texas Hill Country. J. Hydrol. 348, 73–86. doi:10.1016/j.jhydrol.2007.09.057.
Westcott, N.E., Knappb, H.V., Hilberg, S.D., 2008. Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. J. Hydrol. 348, 73–86. doi:10.1016/j.jhydrol.2007.10.057. 351(1), 1-12.
Wilson, J.W., Brandes, E.A., 1979. Radar measurement of rainfall – a summary. Bull. Amer. Meteor. Soc. 60, 1048–1058.
Xie, H., Hendrickx, J., Zhou, X., Vivoni, E., Guan, H., Tian, Y., Small, E., 2006. Evaluation of NEXRAD stage III precipitation data over semiarid region. J. Am. Water Resour. Assoc. 42 (1), 237–256.
Yilmaz, K., Hogue, T., Hsu, K.L., Sorooshian, S., Gupta, H., Wagener, T., 2005. Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. J. Hydrometeorol. 6, 497–517.
Young, C.B., Brunsell, N.A., 2008. Evaluating NEXRAD estimates for the Missouri River basin: analysis using daily raingauge data. J. Hydrol. Eng. 13 (7), 549.
Young, C.B., Bradley, A., Krajewski, W., Kruger, A., Morrissey, M., 2000. Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. J. Hydrometeorol. 1, 241–254.