Journal of the Iowa Academy of Science: JIAS

Volume 124 | Number 1-4 Article 3

2017

A statistical, spatial, and hydrologic comparison of gauge-based and MPE-based rainfall measurements

Richard Bernatz
Luther College

Jour. Iowa Acad. Sci. 124(1-4):11–23, 2017

A statistical, spatial, and hydrologic comparison of gauge-based and MPE-based rainfall measurements

RICHARD BERNATZ

Department of Mathematics & Environmental Studies Program
Luther College, 700 College Drive, Decorah, Iowa 52101

Gauge-based and multi-sensor precipitation estimation (MPE) data are compared on hourly, daily, monthly and event time scales at site locations over a 12-year period. Gauge data is collected at 16 sites within a 950 km^2 portion of the Upper Iowa River in northeast Iowa. Average relative MPE bias is positive for all but the event time scale, and has a magnitude of less than 0.10 for all scales. Gauge and MPE average correlation coefficients range from 0.73 on the hourly scale to 0.92 on the event and monthly scales. The MPE relative bias standard deviation decreases from 1.70 mm on the hourly scale to 0.27 mm on the monthly scale. Decomposition of hourly bias reveals that the false positive portion is the most significant component. Seventy percent of MPE accumulation have a relative bias of 0.5 or less when hourly accumulations are 7 mm or greater. Pearson product-moment coefficient analysis reveals strong similarities in spatial correlations as a function of site separation. Rainfall time series for the basin are constructed from the two data sources and used as input to a Blocked Topmodel rainfall runoff scheme to provide another means of comparison on a basin-wide spatial scale. Five goodness-of-fit measures are used for quantifying the viability of simulated flows. No statistically significant difference in annual means using the difference sources is found for any of the measures.

INTRODUCTION

Accurate measurement of rainfall accumulation (depth) and intensity (accumulation per time interval) is important for purposes such as climate analysis and hydrological forecasting. Two means of providing estimates for these quantities are a network of tipping bucket (TB) rain gauges and multisensor precipitation estimation (MPE) using radar, satellite and gauge data. Gauge-based data generally are considered to be an accurate measurement of ground surface rain at specific sites (Villarini and Krajewski 2008). MPE-based estimates for rainfall intensities and accumulation are calculated for relatively small surface regions called "pixels." MPE-related measurement provides near real-time and more comprehensive spatial information lacking in gauge network data (Wood et al. 2000). This paper compares these two methods of rainfall estimation in three ways. First, comparisons of rainfall statistics (mean accumulation, bias, relative bias and others) on various time scales is done using MPE data and gauge data recorded at 16 sites within the Upper Iowa River basin in northeast Iowa. Next, spatial characteristics of the two measurements are analyzed using the Pearson product-moment (PPM) coefficient. Finally, the validity of rainfall runoff simulations using the two data sources as input to a hydrologic model will be compared using six goodness-of-fit measures based on simulated river discharge versus observed river discharge.

Cole and Moore (2008) derived three gridded rainfall estimators from gauge and MPE data within the years 2002 to 2004: gauge-only data (networks of eight or nine telemetry tipping-bucket rain gauges) using a multi-quadratic interpolation scheme, gauge-adjusted radar, and unadjusted radar data. These estimators were assessed from a rainfall perspective as well as input to hydrological models for the Darwen (136 km^2) and Kent (212 km^2) river catchments. Hydrological models performed best when rain gauge-only data were used as input. However, this is attributed to model calibration using gauge-only data. The model simulations give very compelling evidence supporting the need for frequent and spatially varying gauge-adjustment of radar rainfall.

MPE (4×4 km^2 pixels) data were compared to gauge data collected on a dense rain gauge network in south Louisiana for years 2004 to 2006 (Habib, et al. 2009). Two MPE pixels had four gauges situated within them, and three had a single rain gauge. Comparisons using MPE pixels with multiple gauges show bias, defined by MPE accumulation-gauge accumulation, on an event-scale basis reaches up to $\pm 25\%$ of total event depth for half of the events, and exceeds 50% for 10% of the events. Negative bias occurred on 65% of the events. Conditionally, MPE overestimates rainfall rates by 60-90% for rates lower than 0.5 mm h^{-1}, and underestimate larger rates by as much as 20% for rates greater than 10 mm h^{-1}. A single gauge within an MPE pixel gives an artificially high MPE estimation error by as much as 120-180%.

Data from 49 gauges over the 135 km^2 Brue catchment in Somerset, England for September 1993 through June 1996 were used by Wood et al. (2000) to examine the accuracy of rainfall data recorded by the gauges against estimates derived from radar. Some radar pixels had a "super-dense" eight-gauge distribution within the 2 km^2 pixel. At rainfall intensities of 16 mm h^{-1}, a single gauge estimate for rainfall within a 2 km square, and over the catchment, has a standard error of 33% and 65% respectively. Radar data for the same resolutions give corresponding errors of 50% and 55%.

Price et al. (2011) compared the accuracy of daily streamflow simulation based on daily rain gauge and higher-resolution MPE data on two sub-basins of the Neuse River in North Carolina with different drainage scales (21 km^2 and 203 km^2) over an eight-year simulation period (1 January 2002 to 31 August 2010). MPE flow

bernatzr@luther.edu

simulations produce R^2 values of 0.64 and 0.54 for the larger and smaller sub-basins, respectively. Gauge-data simulations resulted in R^2 scores of 0.19 for both sub-basins, thereby indicating that MPE data generate more accurate stream flows than gauge data. They attribute the discrepancy to the improved spatial resolution of MPE data, and not to the error associated with rain gauge data.

A study on the Severn Uplands (UK), with an area of 2,065 km^2 (Biggs and Atkinson 2011) lead researchers to conclude that six rain gauges could be used to predict flows with similar accuracy to radar data during an extreme hydrological event. The study covered the intense rainfall and large river flows over the time period of November 2006 and December 2006.

Although this study is similar in nature to studies cited above, it is unique in that the 12-year length (years 2004 through 2015) of the gauge network data is longer by four years than the study period of Price et al. (2011), and longer by eight years or more than those used in the remaining studies referenced above. The longer interval provides the ability to compare the rainfall measures for a range of wet and dry seasons. The longer time frame significantly increases the sample size, allowing for greater statistical validity. For example, the hourly scale sample size in this study is 48,071. The corresponding sample size for Habib et al (2009) is 3,645. In terms of hydrologic input, the longer time interval includes four significant flood events of the Upper Iowa River within the study region. Consequently, hydrological efficacy of the two measures for simulating river discharge may be compared using multiple flood events.

STUDY REGION

The Upper Iowa River (UIR) originates in southeastern Minnesota, and the majority of its 250-km path to the Mississippi River winds through northeast Iowa, entering the Mississippi River on the Iowa side of the Iowa–Minnesota border. Paleozoic-age bedrock units consisting of carbonates (limestones and dolomites), sandstones, and shales make up the lower depths of the UIR basin. The sandstone and carbonate strata typically act as aquifers, and the shales as aquitards. Overlying the bedrock is a variable thickness of unconsolidated materials, which are typically thicker in the western part of the basin and thinner to the east (Wolter, et al. 2011). Soil in the western two-thirds of the basin upstream from Decorah is classified as silty and loamy mantled firm till plain, characterized by gently sloping to very steep dissected till plain. Land use is primarily crop and grazing land on ridge tops and valley bottoms, with a mix of dairy, beef and cash grain agricultural enterprises. Deciduous forests populate the side slopes. The remaining portion of the Decorah upstream basin is identified as driftless loess hill and is further characterized by highly dissected hills and valleys. It is comprised by well- to moderately well-drained silty soils over bedrock residuum.

The United States Geological Survey (USGS) maintains volume flow gauging stations near Bluffton, Iowa (station #05387440) and Decorah, Iowa (station #05387500) where discharge and gauge height are recorded on a 15-minute interval basis. The basin upstream from Bluffton has a drainage area of 950 km^2. The annual average discharge ranges from 3.33 $m^3\ s^{-1}$ to 19.23 $m^3\ s^{-1}$. The maximum peak flow for the 13-year recording period is 470 $m^3\ s^{-1}$. The basin upstream from the Decorah station has a drainage area of 1323 km^2. The annual average discharge ranges from 2.73 $m^3\ s^{-1}$ to 24.68 $m^3\ s^{-1}$. The maximum peak flow over the 64-year recording period is 965.6 $m^3\ s^{-1}$.

Rain Gauge Data

A recording rain gauge network for the UIR basin upstream from Decorah was set up for May through October 2004, with six sites distributed on the perimeter of the basin (gauges 1 to 6 as shown in Fig. 1). As many as nine additional gauges were placed at various sites in subsequent years. Sixteen different sites, as shown in Fig. 1, have data for at least six seasons (May through October).

Rain gauges used to collect data are the Onset RG2 data tipping bucket (TB) logging gauges that record the time, within 0.5 second, that the bucket tips. Each tip represents 0.254 mm of precipitation. The system has a calibration accuracy of $\pm 1.0\%$ (up to 25.4 mm per hour), a time accuracy of ± 100 ppm at 20° C, and a capacity of 8,000 tips (2032 mm). Gauges were set up annually in mid-May and removed from the sites at the end of October. Gauges were calibrated annually, either in the field or during the launching process prior to returning them to the field.

The magnitude of the rain gauge errors is highly dependent on the local rainfall intensity and the timescale. At moderate rainfall intensities of 10 mm h^{-1}, gauges have been determined to have relative standard errors of 6.4% for the 5-minute time scale and 2.3% for 15-minute timescale (Habib et al. 2001). Ciach (2003) determined relative standard errors of 4.9% and 2.9% respectively for similar conditions.

Using gauge data to give areal rainfall leads to spatial error. Villarini and Krajewski (2008) show that the standard deviation of the spatial sampling error decreases with increasing rainfall intensity and accumulation time, and increases with increasing pixel size.

Single-gauged sites are more susceptible to operational problems (mechanical or electrical issues) and clogging as opposed to those with a dual-gauge setup (Krajewski et al. 2003, Steiner et al. 1999). Because of the known irregularities in single-gauge site collection, gauge data will not be considered the "true" ground data in this study. Rather, the study endeavors to explore the relative differences in rainfall estimates of the two sources, as well as their relative efficacies as hydrologic model inputs–especially associated with near-flooding or flooding events in the upper reaches of the Upper Iowa River.

MPE Data

Radar-based data used in this study is hourly multisensor precipitation estimation (MPE) produced by the 12 National Weather Service (NWS) River Forecast Centers (RFCs) distributed over the continental United States. The essential component of radar-based rainfall estimates is the Next Generation Weather Radar (NEXRAD) system of 160 WSR-88D (Weather Surveillance Radar–1988 Doppler) radars. NEXRAD rainfall products have four stages (I to IV) depending on the level of preprocessing, calibration, and quality control (Fulton 2002, Jayakrishnan et al. 2004, Xie et al. 2005). MPE combines rain gauge data, NEXRAD rainfall estimates, and Geostationary Operational Environmental Satellite (GOES) products. (Note: The rain gauge data used in this study is independent of that used in the MPE product.) The MPE data are mapped to a 4 km $\times$ 4 km polar stereographic projection grid, also known as the Hydrologic Rainfall Analysis Project (HRAP) grid, and mosaicked to produce a product known as Stage IV rainfall analysis. The Iowa Flood Center (IFC) transforms these data to a 2.0 km $\times$ 2.0 km pixel grid for input to its hydrologic modeling application. The resulting grid, over-lain on the study region, is shown in Fig. 1.

Sources of uncertainty for radar-based rainfall estimates include the effect of different radar beam heights due to the difference in distances between the radar location and various points within the region monitored by the radar. Another error source is resolution loss due to beam spread at greater distances from the radar site. Other sources of uncertainty include anomalous propagation of the radar electromagnetic wave due to the deviation of the vertical temperature profile from an assumed distribution, as one example. The particular

Z-R relationship used to calculate the rainfall rate *R* as a function of the reflectivity *Z* may be an issue (Austin 1987). For example, right-band contamination from an enhanced radar return caused by melting snow can lead to errors in precipitation intensity of up to a factor of five (Illingworth and Thompson 2011).

METHODS

The "raw" MPE data used in this study estimates rainfall accumulations for each MPE cell for each hour interval (01 to 24) beginning on the integer hour. Therefore, the rain gauge time series data are aggregated to the same hourly intervals. Data from all sites are lumped under the reasonable assumption that all sites are within the same climatological region. The MPE pixel corresponding to each of the 16 gauge sites is determined based on the gauge coordinates and the MPE pixel vertex coordinates. If a gauge site was close to the border of two or more MPE pixels, the corresponding MPE estimate is determined by averaging the corresponding pixel accumulations. Hyetographs for each site and year were used to identify times for which the gauge at the site may have become clogged, and the corresponding data segments in the time series are removed from the site's time series.

Statistical Analysis

A comparative statistical analysis is developed using samples of paired gauge-MPE accumulations on various time scales. Each pair in the sample has a non-zero accumulation for at least one of the measures. Site-based time scales include hourly, daily, monthly, and site event. In this study, a site event is defined by a continuous rainy period wherein rainfall is interrupted by no more than a 6-hour interval and the accumulation total is at least 5 mm at a given site. Hourly and event time-scales are relevant for purposes such as hydrologic modeling. Daily and monthly time-scales are more relevant for climatological purposes.

Statistical measures applied on the various time scales include:

$$\text{Bias} : \bar{R}_{MPE} - \bar{R}_G \tag{1a}$$

$$\text{Relative bias} : \frac{\bar{R}_{MPE} - \bar{R}_G}{\bar{R}_G} \tag{1b}$$

$$\text{Standard deviation of difference} : \sigma_{(R_{MPE}-R_G)} \tag{1c}$$

$$\text{Relative standard deviation of difference} : \frac{\sigma_{(R_{MPE}-R_G)}}{\bar{R}_G} \tag{1d}$$

$$\text{Pearson's correlation coefficient} : \frac{\overline{(R_{MPE} - \bar{R}_{MPE})(R_G - \bar{R}_G)}}{\sigma_{MPE}\sigma_G} \tag{1e}$$

R_{MPE} is the MPE rainfall accumulation per time scale and R_G is the corresponding rain gauge value. Expressions with the over-bar represent sample means, and σ represents the sample standard deviation.

One reliability measure of MPE data is the frequency of zero MPE accumulation for a non-zero gauge accumulation on a given time scale. The relative frequency of such occurrences is defined to be the probability of false zeros (PFZ). A similar probability of false positives (PFP) is defined for cases in which MPE accumulations are non-zero and gauge accumulations are zero. These probabilities are conditioned by either the magnitude of the non-zero accumulation or the magnitude of non-zero rainfall intensity.

Spatial Analysis

Comparison of spatial characteristic of the two measurements is made using the Pearson product-moment (PPM) coefficient *r*, shown in Equation (2). The coefficient is a popular estimator (Gebremichael and Krajewski 2004) of the linear dependence between rainfall *R* at two locations (x_1, y_1) and (x_2, y_2) separated by the Euclidean distance *h*.

$$r(h) = \frac{R(x_1,y_1)R(x_2,y_2) - R(x_1,y_1)R(x_2,y_2)}{\sqrt{\left[R(x_1,y_1)^2 - R(x_1,y_1)^2\right]\left[R(x_2,y_2)^2 - R(x_2,y_2)^2\right]}} \tag{2}$$

Terms with angle braces represent the mean of the quantity, so the coefficient is the covariance of the two quantities divided by the product of their standard deviations. For non-negative values such as rainfall accumulation, a coefficient value of 1.0 implies a perfect correlation and 0.0 implies no correlation.

Hydrologic Model Input Analysis

The viability of MPE-measured rainfall relative to gauge data is analyzed by comparing the simulated discharge of the Upper Iowa River using rainfall inputs derived from the two data sources. Simulated discharges in cubic meters per second (m^3s^{-1}) are calculated for the Bluffton and Decorah sites using the block-wise Topmodel framework (Takeuchi et al. 1999). In the block-wise version of Topmodel, the Upper Iowa River basin is partitioned into blocks consisting of either a sub-basin or a river corridor entity. Hourly rainfall accumulations are used to develop rainfall time series for each Topmodel block for each year from 2004 to 2015. Digital elevation model (DEM) pixels of 30 m × 30 m are assigned to one of the 31 Topmodel blocks. Point-wise gauge accumulations are scaled up to the DEM pixels using the nearest neighbors modified Shepard's (Shepard 1968) weighting scheme. For MPE data, DEM pixels, or portions thereof, are assigned to a MPE pixel. This assignment establishes the hourly rainfall accumulation as well as the block to which the corresponding portion of the MPE pixel is assigned. Hourly DEM pixel accumulations are averaged to derive an hourly rainfall time series for each computational block for each rainfall measure.

Various goodness-of-fit measures for quantifying the accuracy of simulated streamflow relative to the observed discharge are described below. Quantities with subscript "p" correspond to predicted or simulated discharges. Those with subscript "o" represent quantities corresponding to observed discharges.

- The Gupta-Kling efficiency (GKE) (Gupta et al. 2009) is given by 1-*ED*, where *ED* is the Euclidean distance between the triple (*r*, α, β) and the "ideal" location (1, 1, 1). In the triple, *r* is the linear correlation coefficient between the simulated and observed discharges, α is the ratio of the standard deviation of the simulated values σ_p and the standard deviation of the observed values, σ_o, and β is the ratio of the mean of the simulated values μ_p and the mean of the observed values μ_o. The nature of this measure results in the tendency for underestimation of simulated flow peaks (Gupta et al. 2009).
- The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970) is 1 minus the ratio of the mean square error (MSE) and the standard deviation of the observations, σ_o. It has a range of values of $(-\infty, 1]$. Values greater than or equal to 0.9 are deemed to be "good," and values greater than or equal to 0.8 deemed to be "satisfactory" (Dawson et al. 2007). Analysis provided by Gupta et al. (2009) shows this measure has three distinctive components representing the correlation among the simulated

Table 1. Comparative statistics for gauge and MPE accumulations for hourly, event, daily and monthly scales.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	All
Hourly Basis													
Sample Size	1792	3764	5436	5258	4538	5263	4043	3067	3482	4115	3914	3399	48071
Gauge Mean (mm)	1.86	1.61	0.83	1.40	1.74	1.27	2.26	1.38	1.42	1.59	1.62	1.38	1.49
MPE Mean (mm)	1.86	1.63	0.99	1.70	1.65	1.29	2.10	1.33	1.46	1.79	1.75	1.62	1.57
Gauge σ (mm)	4.06	4.08	1.95	3.33	4.10	2.75	4.99	3.40	3.80	3.87	3.42	4.23	3.67
MPE σ (mm)	3.66	2.93	1.77	3.02	3.31	2.78	3.84	2.69	2.74	3.42	2.89	3.88	3.08
Average Bias (mm)	-0.01	0.02	0.16	0.30	-0.08	0.02	-0.16	-0.05	0.03	0.20	0.13	0.24	0.08
(MPE - Gauge) σ (mm)	2.24	2.44	1.24	2.58	2.71	1.77	4.03	1.81	2.59	2.72	2.78	2.58	2.53
Relative Bias	0.00	0.01	0.19	0.21	-0.05	0.02	-0.07	-0.03	0.02	0.13	0.08	0.17	0.05
(MPE - Gauge) Relative σ (mm)	1.20	1.52	1.49	1.84	1.56	1.40	1.79	1.31	1.82	1.71	1.71	1.87	1.70
Correlation	0.84	0.81	0.78	0.67	0.75	0.79	0.61	0.85	0.73	0.73	0.62	0.80	0.73
Event Basis													
Sample Size	352	554	667	765	803	713	785	602	624	691	565	499	7620
Gauge Mean (mm)	9.49	10.93	6.79	9.61	9.82	9.34	11.63	7.03	7.95	9.43	11.22	9.40	9.41
MPE Mean (mm)	8.62	9.44	6.26	9.99	8.28	8.65	9.80	5.99	6.90	9.36	10.32	9.16	8.58
Gauge σ (mm)	14.36	17.99	9.54	15.02	23.77	12.44	16.98	9.66	11.87	13.09	15.38	16.09	15.44
MPE σ (mm)	12.94	13.27	9.03	14.26	21.94	12.59	14.39	8.00	10.01	12.84	13.27	15.46	13.92
Average Bias (mm)	-0.87	-1.49	-0.53	0.38	-1.54	-0.69	-1.84	-1.04	-1.05	-0.06	-0.90	-0.23	-0.83
(MPE - Gauge) σ (mm)	5.18	7.04	3.52	8.00	6.18	5.19	5.33	3.77	6.15	6.25	7.91	6.32	6.10
Relative Bias	-0.09	-0.14	-0.08	0.04	-0.16	-0.07	-0.16	-0.15	-0.13	-0.01	-0.08	-0.02	-0.09
(MPE - Gauge) Relative σ (mm)	0.55	0.64	0.52	0.83	0.63	0.56	0.46	0.54	0.77	0.66	0.70	0.67	0.65
Correlation	0.93	0.94	0.93	0.85	0.97	0.91	0.96	0.93	0.86	0.88	0.86	0.92	0.92
Daily Basis													
Sample Size	453	910	1266	1189	1091	940	928	809	786	892	720	703	10687
Gauge Mean (mm)	7.38	6.66	3.58	6.18	7.23	7.09	9.84	5.23	6.31	7.34	8.81	6.79	6.72
MPE Mean (mm)	7.34	6.73	4.25	7.52	6.89	7.21	9.13	5.05	6.45	8.32	9.51	7.98	7.07
Gauge σ (mm)	12.05	14.82	6.77	11.51	16.46	10.23	15.05	8.76	10.69	12.97	12.15	13.91	12.45
MPE σ (mm)	10.77	11.78	6.61	11.22	15.23	10.26	12.75	7.52	9.39	12.64	11.66	13.77	11.45
Average Bias (mm)	-0.03	0.08	0.68	1.34	-0.34	0.12	-0.71	-0.18	0.14	0.99	0.70	1.20	0.35
(MPE - Gauge) σ (mm)	5.19	6.83	3.26	7.38	5.64	4.96	6.30	3.97	6.07	6.03	7.06	6.24	5.88
Relative Bias	0.00	0.01	0.19	0.22	-0.05	0.02	-0.07	-0.03	0.02	0.13	0.08	0.18	0.05
(MPE - Gauge) Relative σ (mm)	0.70	1.03	0.91	1.19	0.78	0.70	0.64	0.76	0.96	0.82	0.80	0.92	0.87
Correlation	0.90	0.89	0.88	0.79	0.94	0.88	0.91	0.89	0.83	0.89	0.83	0.90	0.88
Monthly Basis													
Sample Size	30	60	70	65	70	65	70	70	70	75	55	60	760
Gauge Mean (mm)	111.38	98.72	64.68	99.14	103.16	92.63	129.74	60.45	62.63	87.23	114.25	78.47	90.41
MPE Mean (mm)	110.85	98.54	76.94	124.57	98.49	95.22	120.15	58.40	62.47	96.73	122.87	87.39	94.65
Gauge σ (mm)	45.92	60.67	36.03	79.11	85.04	46.06	69.49	27.47	26.02	68.18	77.84	52.11	63.40
MPE σ (mm)	38.97	45.88	30.20	80.31	91.92	51.40	60.29	23.56	23.34	71.83	77.59	46.25	62.23
Average Bias (mm)	-0.53	-0.18	12.26	25.43	-4.67	2.58	-9.59	-2.05	-0.15	9.50	8.62	8.92	4.24
(MPE - Gauge) σ (mm)	15.92	33.46	18.71	41.74	17.73	16.83	16.90	13.43	17.23	21.15	21.02	24.99	24.73
Relative Bias	0.00	0.00	0.19	0.26	-0.05	0.03	-0.07	-0.03	0.00	0.11	0.08	0.11	0.05
(MPE - Gauge) Relative σ (mm)	0.14	0.34	0.29	0.42	0.17	0.18	0.13	0.22	0.28	0.24	0.18	0.32	0.27
Correlation	0.94	0.84	0.85	0.86	0.98	0.95	0.98	0.87	0.76	0.96	0.96	0.88	0.92

and observed flow, the "normalized" bias, and the relative variability of the simulated values versus the observed values.

- The index of agreement (IoA) measure is the ratio of the mean square error and the potential error (Willmott et al. 1985). Its sensitivity to extreme values is one of its strengths (Legates and McCabe Jr. 1999). Values range from 0 to 1, with values greater than or equal to 0.9 deemed to be "good," and values greater than or equal to 0.8 deemed to be "satisfactory" (Dawson et al. 2007).
- The percentage bias (PBias) is ratio of the bias (x_p - x_o) and the observed discharge x_o multiplied by 100/n, where n is the number of data points. Values range from (-∞, ∞), with an optimum score of 0. Negative values indicate, on average, an under-prediction of discharge, positive values indicate an over-prediction of discharge. Marechal (2004) states that absolute values of PBias less than 5% imply excellent model performance. Absolute values greater than 40% represent poor model performance.
- The root mean square error (RMSE) is a "standard" measure of accuracy with a range of values of [0, ∞). Squaring the difference of simulated and observed values implies the measure is biased in favor of large differences, making it sensitive to difference between predicted and observed flow associated with peaks.
- The coefficient of determination, otherwise known as the square of the Pearson product correlation coefficient squared (R^2), comprises the squared ratio of combined dispersion of the predicted and observed flow. It describes the portion of the total statistical variance in the observed flow that can be explained by the model. The values of R^2 range from 0.0 (poor model) to 1.0 (perfect model).

Model parameterization for each rainfall input source is accomplished using a Monte Carlo method for parameters over multiple runs through the simulation period (June through October). Each goodness-of-fit measure described above is used to determine viable parameter sets within the Generalized Likelihood

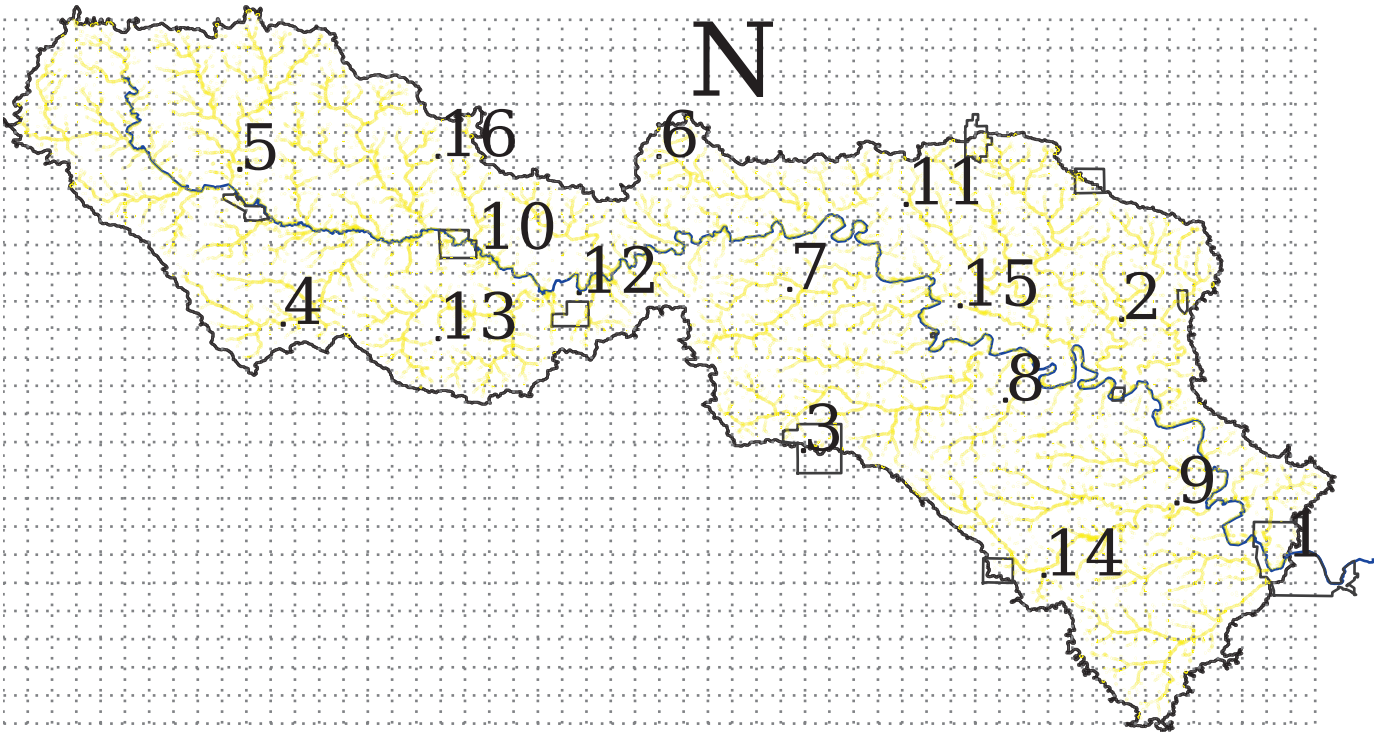

Fig. 1. The Upper Iowa watershed study region stretching from Lake Louis near Leroy, MN., in the far upper left corner, to Decorah, IA., in the lower right corner. MPE cells are indicated by the grey lines.

Uncertainty Estimation (GLUE) framework (Beven and Binley 1992). Separate viable parameter sets are determined for gauge and MPE inputs. Repeated calibration runs are made with randomly selected parameter sets until a prescribed number of "viable" sets is found for each year. Streamflow simulations are derived using a set of parameter sets randomly selected from parameter sets of a given goodness-of-fit measure from all years prior to the simulated year.

Hydrographs of observed discharge and simulated discharges derived from gauge and MPE inputs are plotted for each of the four largest events at both of the USGS gauging stations. Qualitative characteristics such as peak initiation time, peak height, and peak length of the observed volume flow are analyzed in the simulated flow hydrographs.

RESULTS AND DISCUSSION

Table 1 gives comparative statistics for gauge and MPE data on each of the 4 time scales for each of the 12 years using the formulas given in Equations (1a) to (1d). Results are ordered by increasing time scale. The event scale includes events as short as 1 h in duration and longer than 38 h, as shown in Fig. 2. The average event duration is approximately 9.1 h (denoted by the solid line), the median duration is approximately 7 h (dotted line), and the 90th percentile value is approximately 19 h (dotted line). The dashed lines represent locations of one standard deviation (7.84 h) on either side of the mean.

Statistical Analysis

The last column of Table 1 gives values determined by combining all years of data for the given quantity. The relative bias for each scale based on all years is positive except for the event scale. The relative bias magnitude on all scales is slightly less than 0.1. These values are similar to those reported by Habib et al. (2009), Wang et al. (2008), and Westcott et al. (2008). MPE standard deviations are slightly less than gauge standard deviations for all time scales as expected because gauge data is a point-wise measure and MPE accumulation is an areal measure subject to less variability. The standard deviation of the relative bias is largest on the hourly scale. Somewhat surprisingly, the relative bias standard deviation is larger on the daily scale than the event scale given that the event average duration is approximately 9 h. However, the event scale accumulation means are larger than those for the daily scale. The hourly scale has the weakest Pearson correlation value of 0.74, while the longer time scales have correlations around 0.9. The correlation values are slightly less

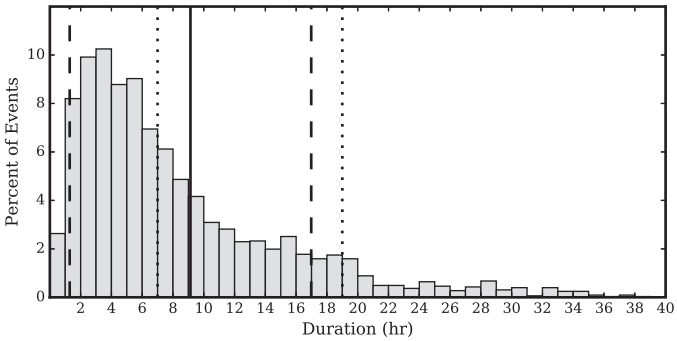

Fig. 2. Event duration relative histogram. The median event length is approximately 7 hours (dotted line), the average event duration is approximately 9 hours (solid line), the 90th –percentile event duration is 19 hours (dotted line). Lines representing one standard deviation (7.8) from the mean are plotted as dashed lines.

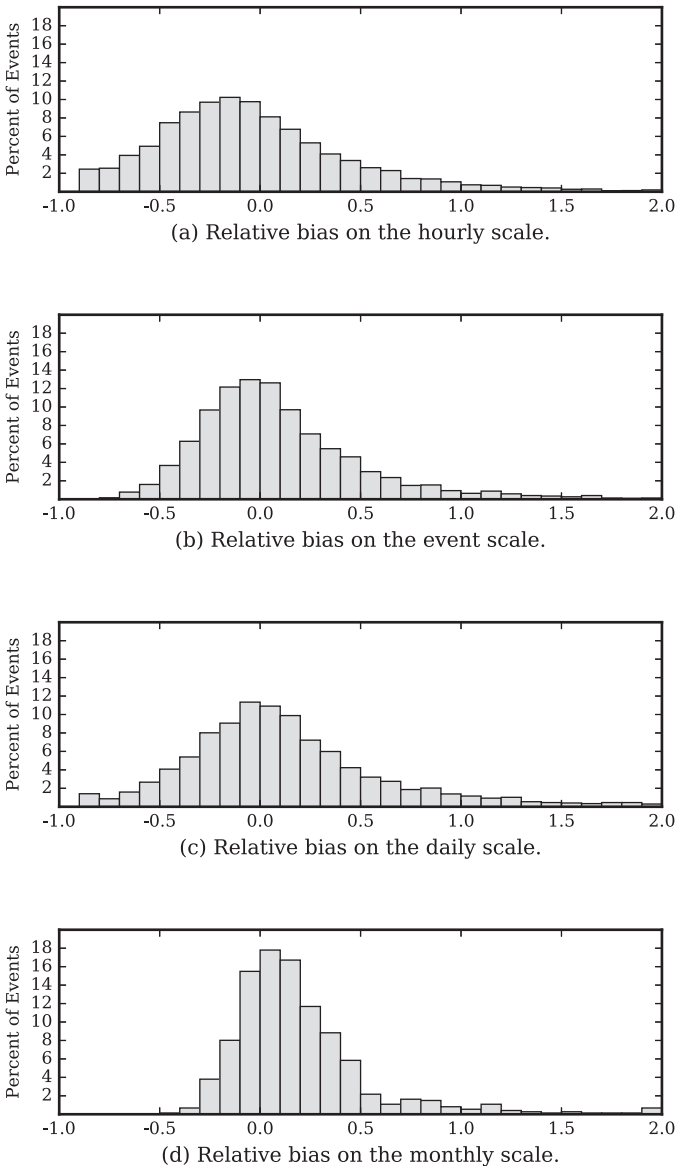

Fig. 3. Relative frequency histograms of relative bias for each of the four time scales.

than those reported by Habib et al. (2009). This is likely due to MPE pixels having as many as four gauges in that study, whereas a single gauge per MPE pixel is used in this study. On the monthly scale, there appears to be some correlation between low Pearson score and lower gauge mean values.

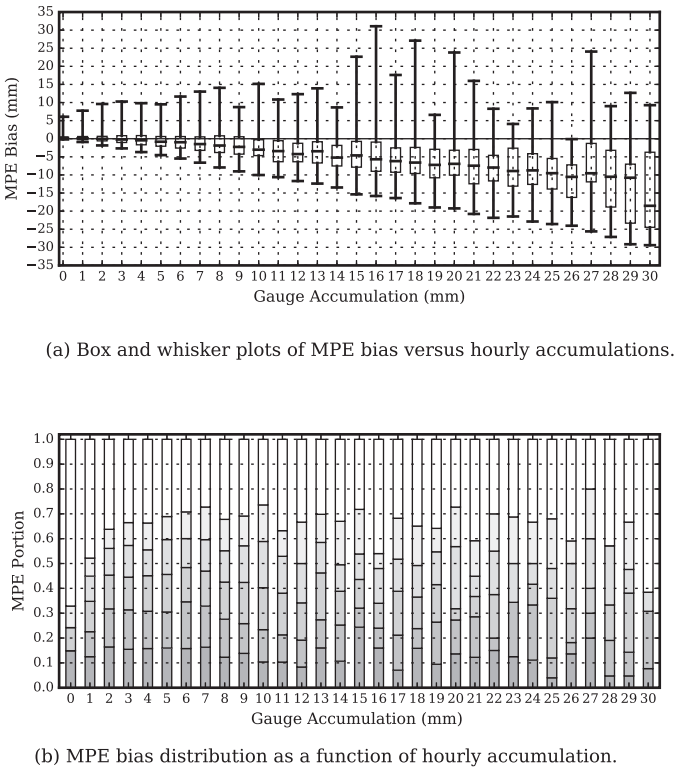

Fig. 4. MPE bias as a function of hourly gauge accumulations.

Bias Analysis

Frequency distributions of relative bias for the hourly, event, daily, and monthly scales are shown in Fig. 3. Table 2 gives the percentage of MPE values that fall below a relative bias threshold for each scale. Segregating the positive and negative occurrences quantifies the existence of any skewing. On the hourly scale, approximately 49.9% of the MPE accumulations have an absolute relative bias less than 0.3, of which 29.7% are negative and 20.2% are positive. Slightly more than 26% of the MPE accumulations had a relative bias of more than 0.5. The percentage of event accumulations to exceed a 0.5 absolute relative bias drops to 15.8%. Approximately 64.2% of the event accumulations were within 30% of the gauge reading, of which 34.8% had a negative bias and 29.4% were positive.

Conditioning the bias by the hourly gauge accumulation provides a more specific understanding of the hourly bias. MPE bias on the hourly scale is separated into bins of width 1 mm. Box and whisker plots for accumulation bins ranging from 0 to 30 mm are shown in Fig. 4(a). The median bias becomes negative for hourly gauge

Table 2. Percentage of MPE accumulations within a relative bias threshold for each time-scale.

	Less than 0.10			Less than 0.20			Less than 0.30			Less than 0.40			Less than 0.50			Greater than 0.50		
Scale	**-**	**+**	**tot**	**-**	**+**	**tot**	**-**	**+**	**tot**	**-**	**+**	**tot**	**-**	**+**	**tot**	**-**	**+**	**tot**
Hourly	9.8	8.2	17.9	20.0	14.9	34.9	29.7	20.2	49.9	38.3	24.3	62.6	45.8	27.7	73.5	13.9	12.6	26.5
Event	13.0	12.6	25.6	25.1	22.3	47.4	34.8	29.4	64.2	41.1	34.9	76.0	44.7	39.5	84.2	2.5	13.2	15.8
Daily	11.3	10.9	22.2	20.4	20.8	41.2	28.4	28.0	56.4	33.8	34.0	67.8	37.9	38.2	76.1	6.5	17.3	23.8
Monthly	15.5	17.8	33.3	23.5	34.5	58.0	27.3	46.2	73.5	28.0	55.0	83.0	28.1	60.9	89.0	0.0	11.0	11.0

Table 3. Bias decomposition on the hourly scale.

year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	All
False Zero (%)	-4.9	-2.5	-6.3	-7.2	-10.7	-5.4	-9.2	-9.3	-9.3	-3.6	-5.1	-5.5	-6.4
False Positive (%)	8.7	14.7	26.7	17.5	10.9	9.1	8.6	15.5	15.5	13.5	16.0	19.7	14.0
Hit Bias (%)	-4.3	-11.1	-1.4	11.2	-4.9	-2.0	-6.6	-4.0	-4.0	2.9	-3.0	3.0	-2.4
Net Bias (%)	-0.5	1.1	19.0	21.4	-4.8	1.8	-7.2	2.2	2.2	12.8	8.0	17.2	5.1

accumulations greater than 4 mm, and more than 75% of the MPE bias is negative for gauge accumulations greater than 9 mm.

The absolute relative bias portions for each accumulation bin in Fig. 4(a) are shown in Fig. 4(b). The portion of MPE accumulations having an absolute relative bias less than 0.1 is shown in the dark gray segment. The next segment, in lighter gray, represents the portion of the MPE accumulations having an absolute relative bias greater than 0.1 but less than 0.2, and similarly for the 0.2 to 0.3, 0.3 to 0.4, and 0.4 to 0.5 segments. The white segment of the bar corresponds the percentage of the MPE accumulations having an absolute relative bias greater than 0.5. The distribution of bias improves as the gauge accumulation increases from 0 to 7 mm, where 70% of the MPE accumulations have an absolute relative bias less than 0.5. No strong tendency or pattern is evident as the gauge accumulation increases further.

Bias is further detailed by decomposition into three categories: *FZ* = false zero ($R_{MPE} = 0$ and $R_G > 0$), *FP* = false positive ($R_{MPE} > 0$ and $R_G = 0$), and *HB* = hit bias ($R_{MPE} > 0$, $R_G > 0$, and $R_{MPE} \neq R_G$). The total bias (*TB*) is the sum of the three components: $TB = FZ + FP + HB$. The *FZ* portion is non-positive, the *FP* portion is non-negative, and the *HB* portion is negative, zero, or positive. Decomposition on the hourly scale for each year is shown in Table 3. Each bias component is summed over all hours and all sites, and its percentage relative to the total gauge accumulation is calculated. Values in the last column of the table are determined by summing over all years as well. Clearly, *FP* bias represents the greatest percentage of the total bias. It is naturally countered by the *FZ* bias (with smaller magnitude) in all years, and the *HB* bias in 9 of the 12 years. The resulting net bias is quite variable through the 12-year interval, with a range of -4.8% to 21.4%.

The probability of false zeros (PFZ) is experimentally determined by identifying hours in which the MPE accumulation is zero and the gauge accumulation is non-zero. There were 13,574 hours of zero MPE accumulation out of a total of 568,900 hours non-zero gauge accumulation, giving a probability of 0.023. The experimental probability of false positives (PFP) is 0.394, based on 16,263 hours for which the MPE accumulation was non-zero and 41,284 hours for which the corresponding gauge accumulation was zero. The relatively high PFP is examined more closely by conditioning it on the MPE accumulation, as shown in Fig. 5. For MPE accumulation less than 0.25 mm, the probability of a false positive is approximately 0.5. The probability drops quickly as the MPE accumulation value increases, dipping below 0.05 when the MPE amount is greater than 2 mm, and 0.01 for MPE values greater than 5.7 mm.

Fig. 6 shows the relative MPE event bias as a function of the average event rainfall rate in mm h^{-1}. MPE relative bias for all events is collected in bins of width 1.0 mm h^{-1}. The bin sample size is 4,107 for 1 mm h^{-1}, and decreases to 13 for the 15 mm h^{-1}. The median relative bias is negative for all event mean rates, and hovers around -0.25. The spread of MPE relative bias decreases significantly as the event mean rate increases. For event rates greater than 6 mm hr^{-1}, well over 75% of the corresponding MPE accumulations have a negative bias.

The distribution of MPE event bias is dependent on the magnitude of the event accumulation as shown in Fig. 7. Event bias is collected in 5 mm-wide bins for site event gauge accumulations ranging from 12.5 mm to 24.9 mm, 25.0 mm to 49.9 mm, 50.0 mm to 74.9 mm, and greater than 74.9 mm. For the event interval of 12.5 to 24.9 mm, 89% of the MPE accumulations were negatively biased, 85.4% of the MPE totals were negatively biased for the 25.0 to 49.9 interval, 91.6% were negatively biased for the 50.0 to 74.9 mm interval, and 91.2% were negatively biased for gauge accumulations of 75 mm or greater.

Spatial Analysis

Pearson product moment (PPM) coefficients are calculated using the hourly, daily, and monthly scales for both the gauge and MPE data. Plots of the coefficient as a function of site separation are shown in Fig. 8. The continuous distribution, shown as the solid curve for gauge data and the dashed curve for MPE data, is modeled by the formula:

$$r(h) = a * \exp\left[-\left(\frac{h}{b}\right)^{c}\right] \quad a \in [-1, 1],\ b > 0,\ c \in [0, 2] \qquad (3)$$

The site separation distance h is measured in kilometers. Small-scale variability and the measurement error are given by the nugget parameter a (Journel and Huijbregts 1978). Values closer to 1.0 imply smaller variability in the process and lesser measurement errors. The parameter b corresponds to the separation distance at which the site

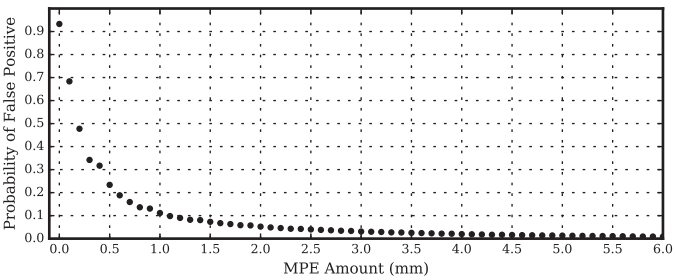

Fig. 5. The probability of a false positive (PFP) conditioned by the MPE accumulation on the hourly scale.

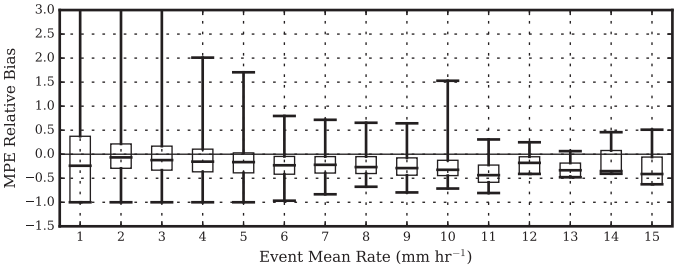

Fig. 6. Box and whisker plots of MPE event relative bias as a function of the event mean rate.

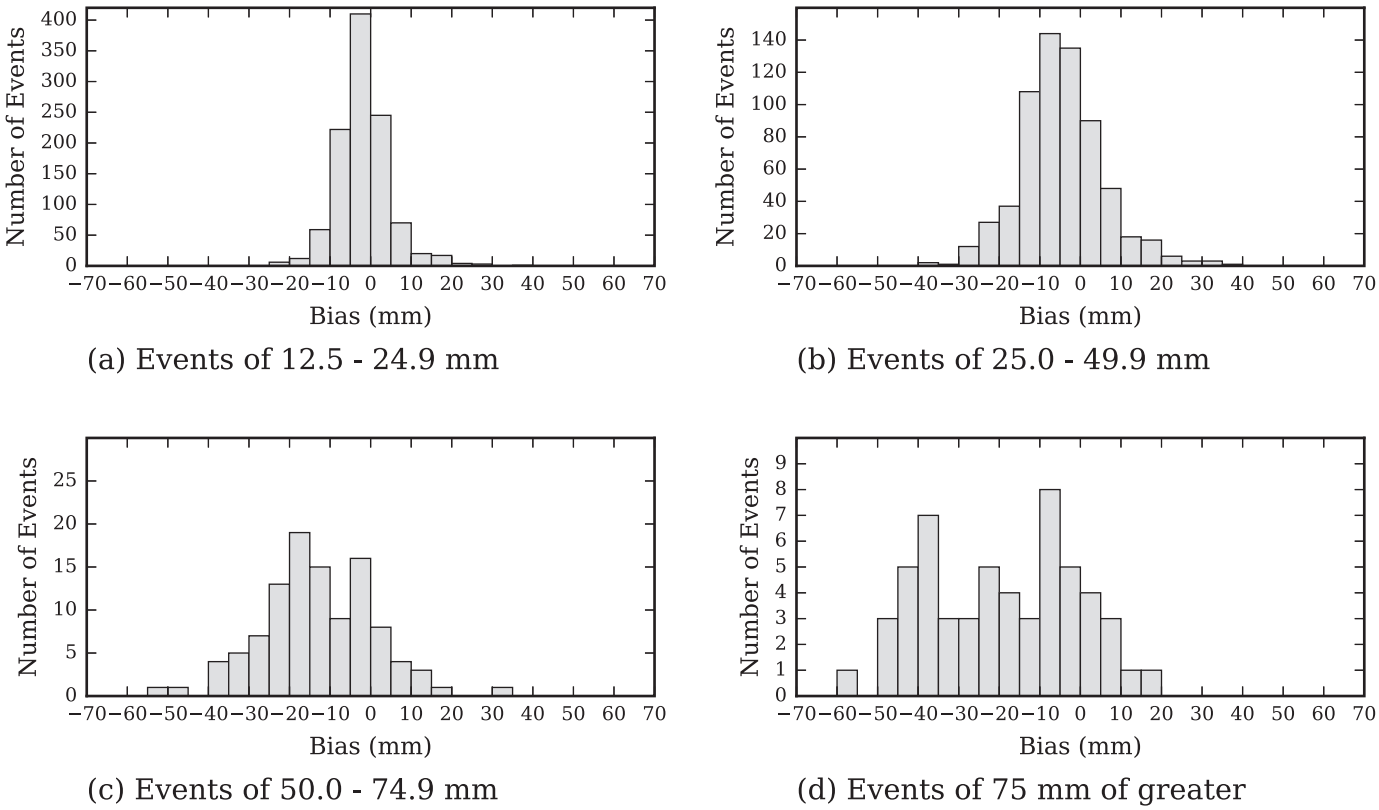

Fig. 7. Histograms of MPE event bias conditioned by gauge event accumulation.

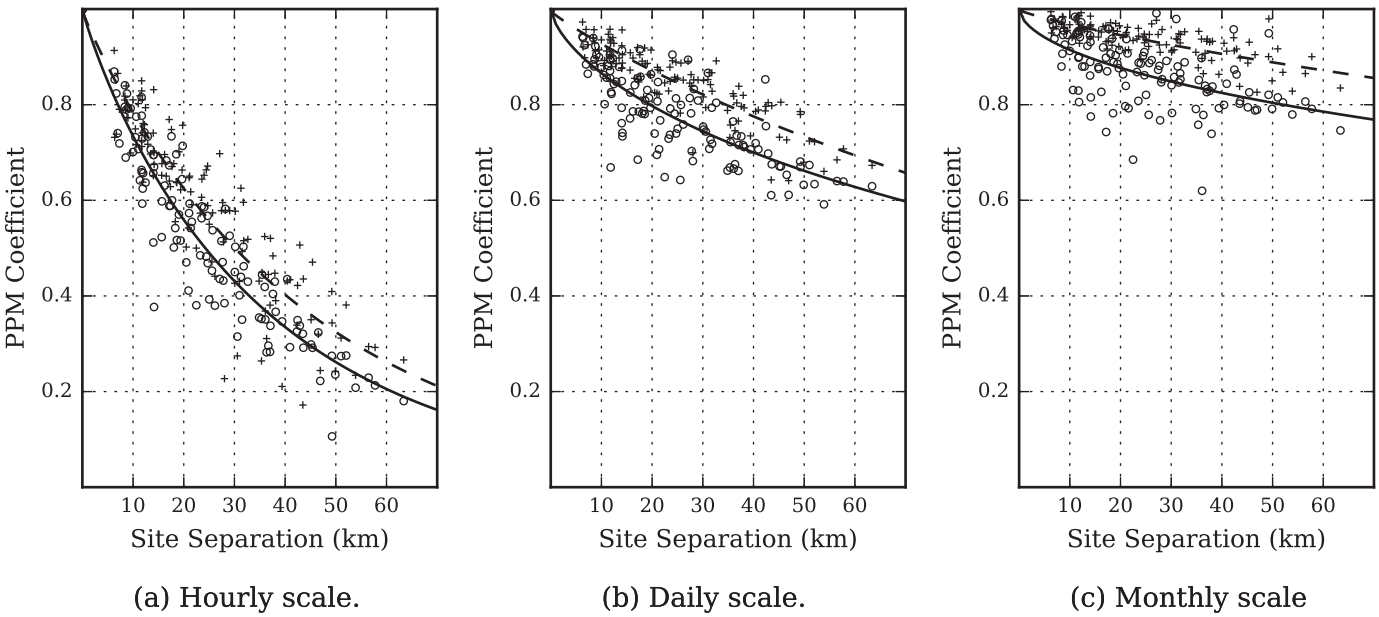

Fig. 8. Pearson product moment (PPM) coefficients as a function of site separation (km) for hourly, daily and monthly time scales. Discrete observations are shown with "o" for gauge data and "+" for MPE data. Modeled continuous distributions are shown for gauge data (solid lines) and MPE data (dashed lines).

Table 4. Gauge and MPE Pearson product moment modeling parameters for various time scales.

Scale	Hourly		Daily		Monthly	
Measure	Gauge	MPE	Gauge	MPE	Gauge	MPE
a	0.9999	0.9955	1.0	0.9957	1.0	0.9955
b	36.25	44.37	192.9	184.8	775.8	665.2
c	0.9124	0.9586	0.6549	0.9058	0.5556	0.8311

accumulations lose correlation. Parameter *c* controls the shape of the correlation function in the range of small-scale separation. The closer *c* is to unity, the more the correlation follows a true exponential decay at low separation distances. Parameter values for several time scales are shown in Table 4.

Generally speaking, parameter values for gauge and MPE data are similar for all scales. The two possible exceptions are the values for *c* for the daily and monthly time scales, with the gauge cases showing less true exponential behavior (the value of *c* further from 1.0). Loss of correlation separation is less for the gauge on the hourly scale, but slightly greater on the other two time scales. The relative increases are only 4% and 16%, respectively. The values for *a* and *b* in this study are similar to those reported by Villarini and Krajewski (2008). However, the shape factor values for gauge data on the daily and monthly scales are smaller (meaning the correlation decay is less exponential) than reported in that article.

Hydrological Input Analysis

Comparison of gauge and MPE on the basin spatial scale is examined by using the different inputs to the blocked Topmodel framework. Discharge volumes at the gauged sites of Bluffton and Decorah are calculated on an hourly interval for approximately the same June through October interval for each year (2004 to 2015). Observed discharges are used to determine the viability of the simulated flows using various goodness-of-fit measures. Tables 5 and 6 show six goodness-of-fit measures for gauge-based (G) and MPE-based (M) Upper Iowa River stream flow simulations at the Bluffton and Decorah gauge stations, respectively. Measures include the Gupta-Kling Efficiency (GKE), the Nash-Sutcliffe Efficiency (NSE), the Index of Agreement (IoA), the Percentage Bias (PBias), the Root Mean Square Error (RMSE), and the Pearson product correlation coefficient squared (R^2). The mean and standard deviation values for each measure over the set of results, shown in the last two columns, are calculated after removing results from year 2012–an extremely dry summer with very low flow volume.

Box plots for each measure using the seasonal results shown in Tables 5 and 6 are shown in Fig. 9. The outliers evident in the NSE-measure data for both the gauge and MPE cases correspond to the extreme low-flow year of 2012. The same is true for the gauge-based outlier in percentage bias (PBias). Paired t-tests on each measure indicate that the null hypothesis (equal mean values for the different rainfall input sources) cannot be rejected in any measure. PBias is negative for all years for both stations for both inputs, with the exceptions being 2012 for both measures and both stations, and gauge input for the Decorah station for 2005. The one measure that may indicate a statistically significant difference in means is PBias, where it is generally more negative for MPE based projections. The gauge data PBias outlier for the Decorah station and year 2012 is notable. The standard deviation for MPE results is generally higher for the gauge based simulations, suggesting more season-to-season variance. The gauge-based R^2 values are higher for both stations with less standard deviation, which is contrary to the findings of Price et al. (2011). The R^2 values for both gauge and MPE simulations are considerably better than those reported in the same study.

Hydrographs for the four events for both stations are shown in Fig. 10. The observed hydrographs show multiple relative maximums for each event, although some are rather subtle, such as in the cases of Bluffton 2007 and 2008.

For the 2007 event, both gauge- and MPE-based hydrographs show a "false" peak prior to the major peak at Bluffton. The MPE case is more drastic. However, the MPE hydrograph does show some waviness after the major peak, as in the observed hydrograph, although the gauge hydrograph shows no such characteristic. Both input sources result in an under-prediction of the major peak magnitude. The Decorah 2007 hydrographs show similar results, except simulations result in extremely close agreement with the magnitude and the time of the major peak.

Gauge and MPE hydrographs both greatly overestimate the Bluffton peak, gauge-based by 67% and MPE-base by 33% for the June 2008 event. MPE picked up on the small "after-peak" at Bluffton, but the peak occurred sooner than observed and did not match the observed peak in duration. The gauge hydrograph showed no after-peak. The major Decorah peak for this event was under-predicted by both gauge- and MPE-based simulations, more so by the MPE case with a relative error of 22% compared to the gauge-based relative error of 10%. The simulated peaks are less sharp than the observed peak and occur several hours later. The MPE hydrograph shows a trailing minor peak and matches the time of

Table 5. Goodness-of-fit measures for gauge-based (G) and MPE-based (M) rainfall runoff results for the Bluffton Station.

Input	Measure	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Mean*	SD*
G	GKE	0.671	0.731	0.574	0.226	0.494	0.581	0.747	0.629	0.23	0.070	0.638	0.103	0.497	0.247
M	GKE	0.594	0.245	0.510	0.538	0.649	0.442	0.472	0.045	0.425	0.422	0.676	0.060	0.423	0.218
G	NSE	0.696	0.476	-0.172	0.037	0.578	0.613	0.756	0.579	-1.05	0.283	0.733	0.335	0.447	0.299
M	NSE	0.711	-0.975	-0.03	0.068	0.798	-0.097	0.484	-0.339	-0.346	0.603	0.662	0.270	0.196	0.539
G	IoA	0.901	0.882	0.727	0.699	0.928	0.876	0.947	0.867	0.644	0.645	0.920	0.615	0.819	0.122
M	IoA	0.905	0.530	0.704	0.792	0.958	0.744	0.838	0.553	0.700	0.845	0.902	0.567	0.758	0.152
G	PBias	-21.1	-8.64	-36.6	-51.9	-41.4	-30.5	-23.3	-22.8	36.3	-48.9	-32.3	-43.4	-32.8	13.2
M	PBias	-29.2	-30.4	-10.7	-24.4	-53.4	-33.6	-44.5	-55.6	4.3	-34.8	-26.4	-46.1	-35.4	13.5
G	RMSE	8.41	2.40	1.58	16.4	30.4	4.76	8.84	3.03	1.83	26.6	6.93	13.2	11.1	9.97
M	RMSE	8.20	4.67	1.48	16.2	21.0	8.01	12.9	5.40	1.48	19.8	7.81	13.8	10.8	6.36
G	R^2	0.752	0.643	0.772	0.391	0.863	0.734	0.846	0.715	0.320	0.640	0.840	0.849	0.731	0.138
M	R^2	0.838	0.155	0.291	0.455	0.886	0.379	0.671	0.636	0.309	0.718	0.738	0.765	0.594	0.238

*- excluding year 2012

Table 6. Goodness-of-fit measures for gauge-based (G) and MPE-based (M) rainfall runoff results for the Decorah Station.

Input	Measure	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Mean*	SD*
G	GKE	0.715	-0.052	0.480	0.323	0.802	0.775	0.520	0.680	-0.186	0.259	0.802	0.183	0.499	0.289
M	GKE	0.566	0.700	0.441	0.516	0.739	0.445	0.612	0.044	0.190	0.578	0.767	0.064	0.497	0.244
G	NSE	0.700	-0.234	0.256	0.04	0.965	0.634	0.535	0.666	-3.633	0.509	0.788	0.472	0.485	0.346
M	NSE	0.626	0.649	0.156	0.128	0.953	-0.191	0.475	-0.303	-1.099	0.793	0.687	0.290	0.388	0.406
G	IoA	0.901	0.860	0.749	0.734	0.992	0.896	0.921	0.895	0.513	0.818	0.946	0.718	0.858	0.910
M	IoA	0.882	0.884	0.683	0.829	0.987	0.743	0.864	0.559	0.646	0.937	0.917	0.567	0.805	0.147
G	PBias	-23.2	18.6	-30.1	-36.1	-33.8	-3.70	-9.41	-16.8	70.5	-46.8	-16.9	-38.5	-21.5	18.7
M	PBias	-35.8	-7.48	-9.56	-16.6	-47.2	-20.3	-37.7	-53.2	24.7	-33.5	-15.1	-42.5	-29.0	15.8
G	RMSE	11.4	22.0	2.67	23.6	15.7	4.80	10.4	3.77	2.76	33.8	7.86	16.9	13.9	9.67
M	RMSE	12.7	11.8	2.85	22.5	18.2	8.66	11.0	7.44	1.86	22.0	9.56	19.6	13.3	6.41
G	R^2	0.753	0.894	0.698	0.373	0.975	0.671	0.842	0.756	0.285	0.815	0.827	0.936	0.776	0.164
M	R^2	0.789	0.653	0.305	0.558	0.936	0.372	0.678	0.719	0.319	0.871	0.737	0.801	0.677	0.200

*- excluding year 2012

initiation well, but not the duration. No such feature is shown in the gauge-based hydrograph.

The initiation of the Bluffton 2013 event was too late for both data-source hydrographs. Both do a better job with the Decorah peak initiation. Their hydrographs capture the subtle characteristics of the observed hydrograph, but very much underestimate the magnitude of the major peak and the two trailing minor peaks for the Bluffton hydrograph–the MPE estimate more so than the gauge estimate. Both input hydrographs capture the shape and timing of the Decorah major peak, but underestimate the magnitude, with a relative error of 32% and 38% for the gauge and MPE inputs, respectively. The MPE hydrograph captures some of the subtleties of the observed case, but misses on timing and magnitude. Once again, these finer features are not evident in the gauge-based hydrograph.

The 2015 event is easily the worst for both inputs in terms of major peak magnitude at both the Bluffton and Decorah stations.

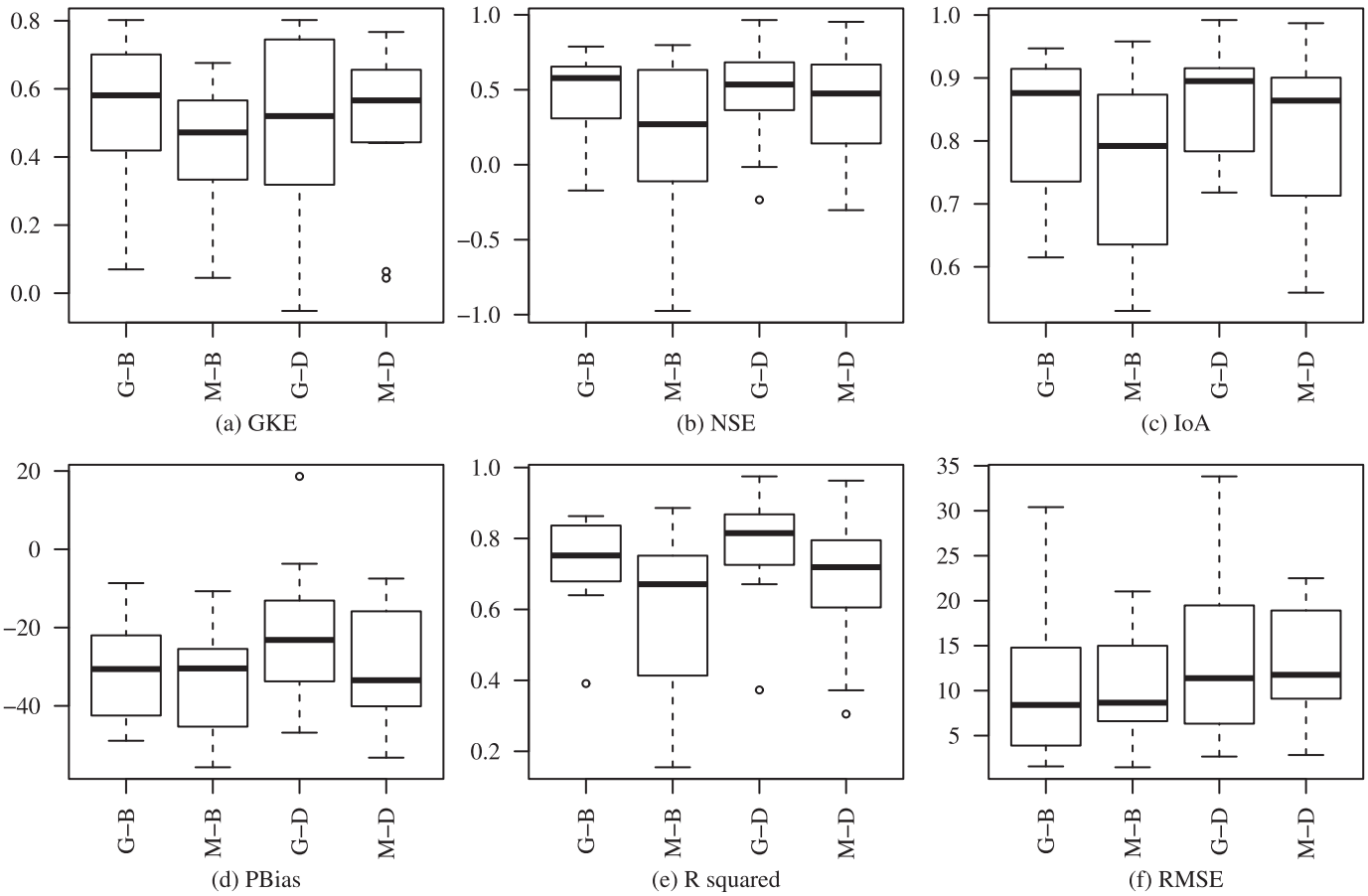

Fig. 9. Box and whisker plots for gauge-based (G) and MPE-based (M) stream flow simulation error for six goodness-of-fit measures at the Bluffton (B) and Decorah (D) stations.

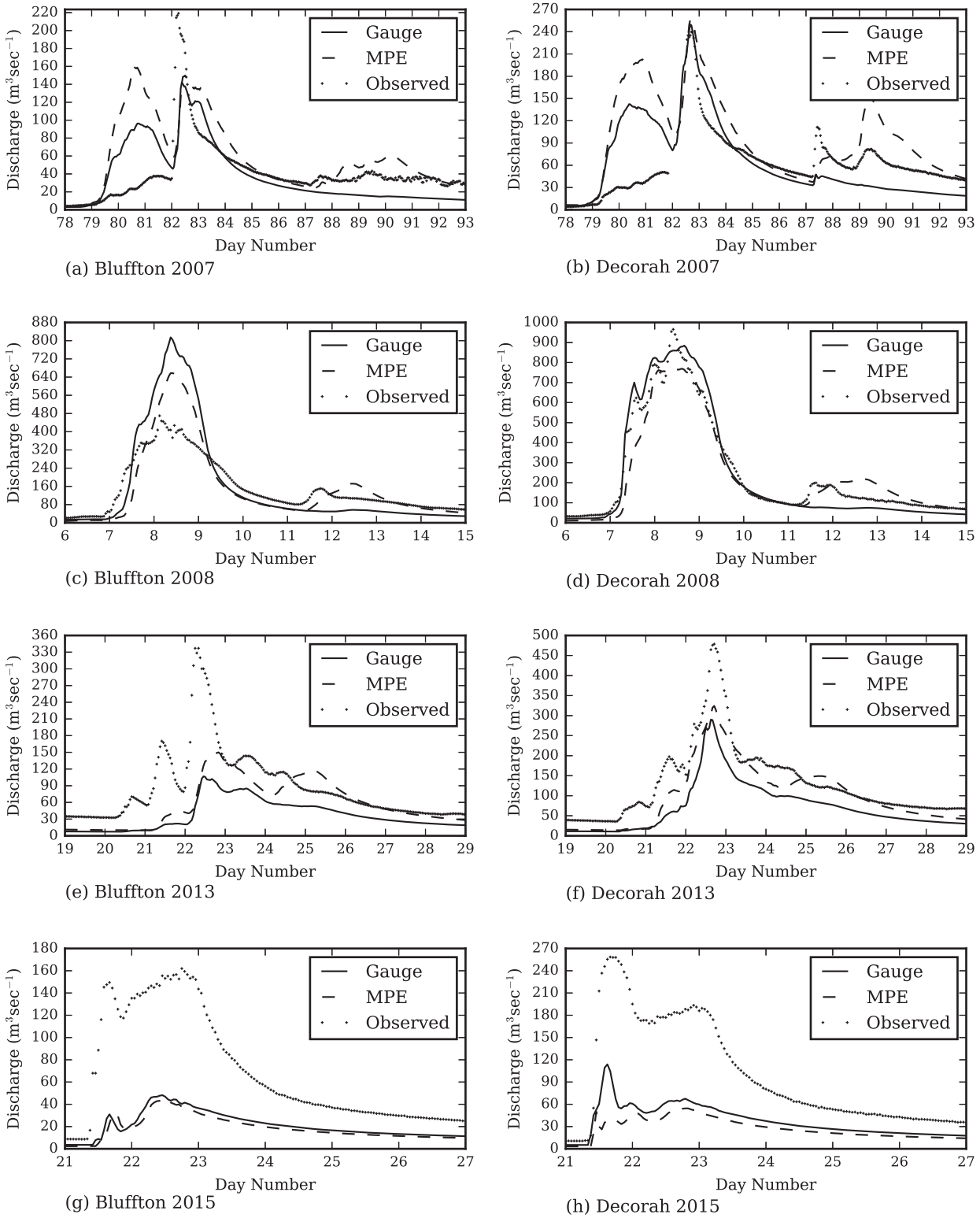

Fig. 10. Bluffton and Decorah station hydrographs for four large events.

Both simulations do a fair job of capturing the shape of the wave, and the gauged-based peak magnitude for Decorah is better than the MPE-based case, yet still has a relative error of 62%.

SUMMARY AND CONCLUSIONS

Using data collected from 16 sites over 12 years (not all sites are included in all years), the average relative bias of MPE accumulations is positive on the hourly, daily, and monthly time scales with a value of 0.05. However, the average relative bias on the event scale is -0.09. The standard deviation of the relative bias is greatest on the hourly scale, with a value of 1.70, and decreases to 0.27 on the monthly scale. MPE accumulation standard deviation is less than gauge accumulation standard deviation on all scales, probably because it represents an areal average accumulation whereas the gauge data represents a point-wise accumulation. MPE and gauge measure correlations are strongest for the event and monthly scales. This true for the event scale even though the standard deviation of relative bias is greater compared to the other scales.

The distribution of relative MPE bias is negatively skewed for all time scales. The percentage of MPE accumulations within a certain relative bias threshold increases as the time scale increases. For example, 17.9% of MPE readings had a relative bias less than or equal to 0.1 on the hourly scale. On the monthly scale, this percentage increases to 33.3%. MPE bias threshold results are better on the event scale than on the daily scale at all threshold values. For example, 84.2% of the MPE accumulations had a relative bias less than 0.50 on the event scale, whereas 76.1% met this threshold on the daily scale. Recall that the event scale has an average and median duration of 9 and 7 hours, respectively.

Hourly MPE bias conditioned by the gauge accumulations becomes more predominantly negative as the accumulation amount increases. For gauge accumulations 9 mm and greater, MPE bias is negative for 75% or more of the cases. When the absolute value of relative bias for each bin is proportioned by magnitude, there is significant improvement from 0 to 3 mm gauge accumulation. On this interval, the percentage of MPE observation with relative bias less than 0.5 improves from 32% to 67%. Thereafter, no significant improvement is seen.

Hourly bias decomposed into categories of false zero, false positive, and hit bias shows that the category of false positive represents the greatest portion of the total bias, averaging 14% of the gauge accumulation. Hit bias averaged only -2.4% of the gauge accumulation. The probability of a MPE estimation being a false positive drops to less than 0.10 for MPE accumulations of 1.0 mm and greater.

On the event scale, the mean relative bias conditioned by event rate switches from positive to negative as the rate increases from 2 to 3 mm h^{-1}. The median relative bias clusters around the value of -0.25 as the accumulation rate increase to 10 mm hr^{-1} and greater. Approximately 90% of the MPE accumulations are negatively biased on the event scale, independent of event gauge accumulation.

Gauge and MPE data exhibit similar spatial characteristics based on the Pearson product-moment coefficient values for the hourly, daily, and monthly time scales. MPE data show de-correlation occurring at long site-separation distances on the hourly scale, whereas the opposite result emerges for the daily and monthly scales.

Little if any significant differences are found in using the two rainfall measures as the basis for input to the Topmodel rainfall runoff algorithm and the GLUE methodology. Six goodness-of-fit measures were calculated for the continuous simulation interval of June through October for each year. Paired t-tests on each measure indicate that the null hypothesis (equal mean values for the measures) cannot be rejected for any of the measures.

In terms of significant basin-wide rainfall events and the corresponding river response, MPE-derived hydrographs generally do a better job of resolving subtleties in event peak characteristics such as minor peaks. However, the timing and magnitude for such features is not extremely accurate in many cases. Both inputs do well in matching the onset of the streamflow event and the overall length of the event. The one exception for initial event rise is Bluffton 2013. Both inputs either underestimated or overestimated observed major peak flows in concert. In one of the eight instances, simulated discharges based on both inputs overestimated the observed major peak magnitude–gauge-based more so than MPE by 67% versus 33% relative error, respectively. In one instance, both matched the major peak discharge extremely well (Decorah 2007). The projected discharges underestimated the major peak discharge in the six other instances with relative errors ranging from 10% (Decorah 2007) to 87% (Bluffton 2015). Both do well in the timing of major peak occurrence except for one instance. The MPE results are generally better at capturing the shape of the event wave in terms of leading or following minor peaks.

ACKNOWLEDGEMENTS

The Upper Iowa Rainfall Runoff project has received financial support (four student stipends) through the NCUR/Lancy Initiative "Humans and their environmental choices; Winneshiek County, Iowa." Support in the form of a stipend for student collaboration and funds for recording rain gauges were made possible by a United States Environmental Protection Agency Assistance Agreement (grant number X-98757301-0).

This study was made possible through the willingness of the Iowa Flood Center to provide MPE data. The author also appreciates the guidance and feedback received from Witold Krajewski, Bongchul Seo, Ricardo Mantilla, Felipe Quintero, and Gabriel Villarini.

Technical assistance was provided by Dave Pahlas and the Decorah Water Plant, Lee Bjerke and the Winneshiek County Engineer's office, Dan Wade and the Howard County Engineer's office, Terry Haindfield and John Pearson of the Iowa Department of Natural Resources, Daryl Herzmann and the Iowa Environmental Mesonet, and Adam Kiel and Paul Berland of Northeast Iowa Resource Conservation and Development.

LITERATURE CITED

AUSTIN PM. 1987. Relation between measured radar reactivity and surface rainfall. Monthly Weather Rev. 115:1053–1070.

BEVEN KJ, BINLEY A. 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrolog Proc. 6(3):279–298.

BIGGS EM, ATKINSON PM. 2011. A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the Severn Uplands, UK. Hydrolog Proc. 25(5):795–810.

CIACH G. 2003. Local random errors in tipping-bucket rain gauge measurements. J Atmos Ocean Tech. 20(5):752–759.

COLE SJ, MOORE RJ. 2008. Hydrological modelling using raingauge- and radar-based estimators of areal rainfall. J Hydrol. 358(3-4):159–181.

DAWSON CW, ABRAHART RJ, SEE LM. 2007. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Env Mod & Software. 22(7):1034–1052.

FULTON RA. 2002. Activities to improve WSR-88D radar rainfall estimation in the National Weather Service. In: Proceedings of the Second Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV. Leavesley GH (ed).

GEBREMICHAEL M, KRAJEWSKI W. 2004. Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets. J Appl Meteor. 43(8):1180–1199.

GUPTA HV, KLING H, YILMAZ K et al. 2009. Decomposition of the mean squared error and NSE performance criteria: Implication for improving hydrological modelling. J Hydrol. 377(1-2):80–91.

ILLINGWORTH A, THOMPSON R. 2011. Radar bright band correction using the linear depolarisation ratio. IAHS Publ. 3XX:1-5.
HABIB E, KRAJEWSKI W, KRUGER A. 2001. Sampling errors of tipping-bucket rain gauge measurements. J Hydrol Eng. 6(2):159–166.
HABIB E, LARSON BF, GRASCHEL J. 2009. Validation of NEXRAD multisensor precipitation estimates using an experimental dense rain gauge network in south Louisiana. J Hydrol. 373(3):463–478.
JAYAKRISHNAN RR, SRINIVASAN R, ARNOLD GJ. 2004. Comparison of raingauge and WSR-88D stage III precipitation data over the Texas-Gulf basin. J Hydrol. 292:135–152.
JOURNEL AG, HUIJBREGTS CJ. 1978. Mining Geostatistics. University of California: Academic Press.
KRAJEWSKI W, CIACH G, HABIB E. 2003. An analysis of small-scale rainfall variability in different climate regimes. Hydrol Sci J. 48:151–162.
LEGATES D, MCCABE G. JR. 1999. Evaluating the use of "goodness-of-fitness" measures in hydrologic and hydroclimatic model validation. Water Resour Res. 35(1):233–241.
MARECHAL D. 2004. A soil-based approach to rainfall-runoff modelling in ungauged catchments for England and Wales [thesis]. Silsoe (UK): Cranfield University Institute of Water and Environment.
NASH JE, SUTCLIFFE JV. 1970. River flow forecasting through conceptual models: Part 1–A discussion of principles. J Hydrol. 10(3):282–290.
PRICE K, PURUCKER ST, KRAEMER SR. 2011. Multi-scale comparison of Stage IV NEXRAD (MPE) and gauge precipitation data for watershed modeling. In: Proceedings of the 2011 Georgia Water Resources Conference. Athens (GA): US EPA Office of Research and Development.
SHEPARD D. 1968. A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM National Conference. New York (NY): Association for Computing Machinery. 517-524.
STEINER M, SMITH JA, BURGESS S et al. 1999. Effect of bias adjustment and rain gauge quality control on radar rainfall estimation. Water Resour Res. 35(8):2487–2503.
TAKEUCHI K, AO T, ISHIDAIRA H. 1999. Introduction of block-wise use of TOPMODEL and Muskingum-Cunge method for the hydro-environmental simulation of a large ungaged basin. Hydrol Sci J. 44(4):633–646.
VILLARINI G, KRAJEWSKI W. 2008. Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. Adv Water Resour. 31(7):1015–1023.
WANG X, XIE H, SHARIF H et al. 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.
WESTCOTT NE, KNAPP HV, HILBERG SD. 2008. Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. J Hydrol. 351(1):1–12.
WILLMOTT C, ACKLESON S, DAVIS R et al. 1985. Statistics for the evaluation and comparison of models. J Geophys Res. 90(C5):8995–9005.
WOLTER CF, MCKAY RM, LIU H et al. 2011. Geologic mapping for water quality projects in the Upper Iowa River Watershed. Technical Information Series No. 54. Des Moines (IA): Iowa Geological and Water Survey and Iowa Department of Natural Resources.
WOOD SJ, JONES DA, MOORE RJ. 2000. Accuracy of rainfall measurement for scales of hydrological interest. Hydrol Earth Sys Sci. 4(4):531–543.
XIE H, ZHOU X, VIVONI E et al. 2005. GIS-based NEXRAD Stage III precipitation database: Automated approaches for data processing and visualization. Comp Geosci. 31(1):65–76.