Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012)

O. P. Prat[1,2] and B. R. Nelson[3]

[1]Cooperative Institute for Climate and Satellites-NC (CICS-NC), North Carolina State University, Asheville, USA
[2]NOAA/National Centers for Environmental Information, Asheville, NC, USA
[3]Center for Weather and Climate, NOAA/National Centers for Environmental Information, Asheville, NC, USA

Correspondence to: O. P. Prat (olivier.prat@noaa.gov)

Abstract. We use a suite of quantitative precipitation estimates (QPEs) derived from satellite, radar, and surface observations to derive precipitation characteristics over the contiguous United States (CONUS) for the period 2002–2012. This comparison effort includes satellite multi-sensor data sets (bias-adjusted TMPA 3B42, near-real-time 3B42RT), radar estimates (NCEP Stage IV), and rain gauge observations. Remotely sensed precipitation data sets are compared with surface observations from the Global Historical Climatology Network-Daily (GHCN-D) and from the PRISM (Parameter-elevation Regressions on Independent Slopes Model). The comparisons are performed at the annual, seasonal, and daily scales over the River Forecast Centers (RFCs) for CONUS. Annual average rain rates present a satisfying agreement with GHCN-D for all products over CONUS ($\pm 6\%$). However, differences at the RFC are more important in particular for near-real-time 3B42RT precipitation estimates (-33 to $+49\%$). At annual and seasonal scales, the bias-adjusted 3B42 presented important improvement when compared to its near-real-time counterpart 3B42RT. However, large biases remained for 3B42 over the western USA for higher average accumulation ($\geq 5\,\mathrm{mm\,day^{-1}}$) with respect to GHCN-D surface observations. At the daily scale, 3B42RT performed poorly in capturing extreme daily precipitation ($> 4\,\mathrm{in.\,day^{-1}}$) over the Pacific Northwest. Furthermore, the conditional analysis and a contingency analysis conducted illustrated the challenge in retrieving extreme precipitation from remote sensing estimates.

1 Introduction

Over the last decades, numerous long-term rainfall data sets have been developed using rain gauge (RG) precipitation measurements, remotely sensed (ground-based radars, satellites) quantitative precipitation estimates (QPEs), or combining different sensors, each of which have specific characteristics and limitations. Extensive information on precipitation measurement methodologies and available precipitation products can be found in Michaelides et al. (2009), Kidd et al. (2010), and Tapiador et al. (2012) among others. One of the limitations in using rain-gauge-based precipitation data sets lies in the fact that the geographical coverage is not spatially homogeneous. By contrast, multisensor satellite-based products – PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks; Sorooshian et al., 2000) and variants PERSIANN-CDR (Climate Data Record; Ashouri et al., 2015), CMORPH (CPC MORPHing technique; Joyce et al., 2004), and TMPA (TRMM (Tropical Rainfall Measuring Mission) Multisatellite Precipitation Analysis; Huffman et al., 2007) – or ground-based radar rainfall estimates – NCEP (National Centers for Environmental Prediction) Stage IV (Lin and Mitchell, 2005) or, more recently, the National Mosaic and Multi-sensor QPE (NMQ/Q2) (Zhang et al., 2011) – provide an opportunity to broach the problem of sparse observations over land and/or ocean. Precipitation data sets at high spatial (typically 4–25 km) and temporal (1–6 h) resolution allow for assessing annual, seasonal, and daily characteristics of precipitation at local, regional, and continen-

tal scales (Huffman et al., 2001; Sorooshian et al., 2002; Nesbitt and Zipser, 2003; Liu and Zipser, 2008; Nesbitt and Anders, 2009; Sapiano and Arkin, 2009; Prat and Barros, 2010a; Sahany et al., 2010; Kidd et al., 2012; Prat and Nelson, 2013a, b, 2014 among others).

The purpose of this study is to evaluate the ability of QPE products to describe precipitation patterns and capture precipitation extremes over a multi-annual time frame. While a lot of studies are available that compared different radar/satellite products on an event-to-event basis, in this work we focus on the long-term perspective (11 years). The objective of this study is to provide a comparison of a suite of common QPEs derived from satellites, radars, and rain gauges data sets for the period 2002–2012 over the contiguous United States (CONUS). Our aim is to evaluate the ability of satellite (TMPA 3B42, 3B42RT) and ground-based remotely sensed (Stage IV) precipitation products to describe precipitation patterns. In particular, we will investigate how the different QPE products compare with respect to long-term surface observations and what are the associated uncertainties. The choice of 3B42 is guided by the fact that a monthly accumulation adjustment is performed on the near-real-time algorithm 3B42RT and thus provides bias-adjusted precipitation estimates when compared to non-adjusted versions of CMORPH and PERSIANN. Furthermore, there are a fair amount of studies available that compare the respective merit of the data sets described above either against each other or against other data sets used as a reference. Those studies often investigate isolated events such as intense precipitation or focus on a time period that is limited by day, month, or season. It is seldom that studies that deal with the long-term assessment of precipitation products (annual or multi-annual basis) are available in the scientific literature (Chen et al., 2013). The remotely sensed data sets will be compared against surface observations from the Global Historical Climatology Network-Daily (GHCN-D) and estimations from the Parameter-elevation Regressions on Independent Slopes Model (PRISM), which combines surface observations with a digital elevation model to account for the orographic enhancement of precipitation. Both GHCN-D and PRISM will be used as a baseline for QPE product evaluations. The study will analyze 11 years (2002–2012) of rainfall data over CONUS. The duration of the study will allow the assessment of systematic biases and capture year-to-year and seasonal variability. In addition to long-term average precipitation characteristics, we will investigate the ability for each of those QPE products to capture extreme events and how they compare with surface observations.

The paper is organized as follows: in the first section, we present briefly the precipitation data sets used in this study. In the second section, we present a comparison between precipitation estimates at the annual and seasonal scales. In the third part, we investigate the impact of differing spatial and temporal resolutions with respect to the data sets' ability to

capture extreme precipitation events. Finally, the paper summarizes the major results of this study.

2   Precipitation data sets and algorithms description

In this section, we provide a brief description of these different precipitation data sets used. The interested reader will refer to the references cited.

2.1   Rain gauge precipitation data sets: GHCN-Daily

Precipitation surface observations are taken from the Global Historical Climatology Network-Daily (GHCN-D). The data set gathers records from over 80 000 stations over 180 countries. About two-thirds of those stations report total daily precipitation only, and other stations include additional information such as maximum and minimum temperature, snowfall, and snow depth (Menne et al., 2012). The entire data set is routinely quality-controlled to ensure basic consistency. The GHCN-D data set incorporates surface observations from different sources (see Table 2 in Menne et al., 2012). We selected the subset from the US Cooperative Observer network (US-COOP), which represented about 9000 stations. The US-COOP network includes first-order stations (1600 manual and automatic synoptic stations) and stations from volunteer observers. Figure 1a presents the location of the 8815 surface observations in the GHCN-D database over CONUS. For the current study, only the 4075 rain gauges reporting at least 90 % of the time during the period 2002–2012 are selected to ensure stable statistics (Fig. 1b). Although there is a 50 % decrease in the total number of rain gauges, the remaining rain gauges conserved a comparable spatial distribution to the original network, and the removed gauges were evenly distributed throughout CONUS. The surface stations are compared with the nearest pixel of the gridded precipitation estimates derived from the selected data sets (PRISM, Stage IV, 3B42, 3B42RT) described below.

2.2   Rain gauge gridded precipitation data sets: PRISM

The PRISM algorithm (available at http://www.prism.oregonstate.edu/) combines point data with a digital elevation model to generate gridded estimates of precipitation along with a suite of climatological variables such as temperature, snowfall, and degree dew point among others (Daly et al., 1994). Data are available at the daily, monthly, and annual scale and at various spatial resolutions (800 m to 4 km). In this work, we use the monthly precipitation estimates at the 4 km nominal spatial resolution (data set AN81m: PRISM Technical Note 2013). The PRISM precipitation estimates incorporate surface data observations from GHCN-D among others. The systematic comparison of point surface observations from GHCN-D and gridded estimates from PRISM will be performed as a consistency check. The PRISM precipitation estimates will be used as a baseline data set to evalu-
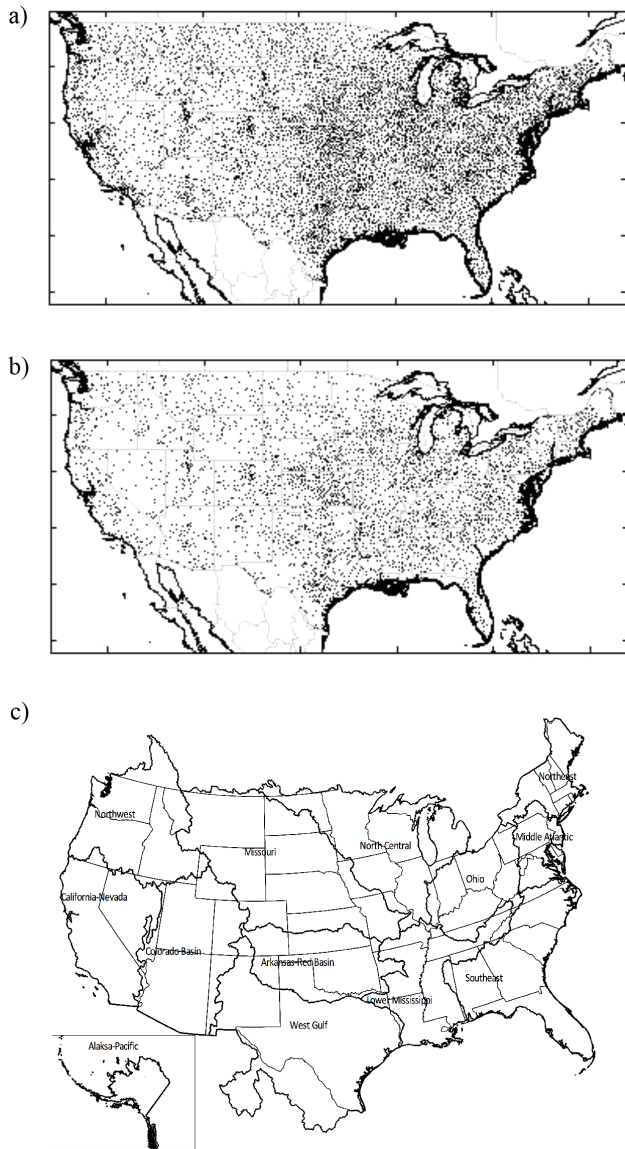
Figure 1. (a) Locations of the GHCN-D rain gauges locations over CONUS: **(a)** total 8815 rain gauges and **(b)** the 4075 rain gauges reporting at least 90 % of the time during the period 2002–2012. **(c)** National Weather Service (NWS) 12 River Forecast Centers (RFCs).

Table 1. List of the 12 NWS RFCs and corresponding number of GHCN-D rain gauges.

Nbr	ID	Name	Number of COOP RG	
			Unconditional (reporting 90 % of time)	Unconditional (total operational)
0	CONUS	CONUS	4075	8815
1	ABRFC	Arkansas-Red Basin	325	637
2	CBRFC	Colorado Basin	243	521
3	CNRFC	California–Nevada	202	537
4	LMRFC	Lower Mississippi	363	728
5	MARFC	Middle Atlantic	166	378
6	MBRFC	Missouri Basin	654	1385
7	NCRFC	North-Central	541	1225
8	NERFC	Northeast	195	462
9	NWRFC	Northwest	260	570
10	OHRFC	Ohio	378	833
11	SERFC	Southeast	387	781
12	WGRFC	West Gulf	361	758

ate remotely sensed precipitation products (Stage IV, 3B42, 3B42RT) at the annual and seasonal scale.

2.3 Radar precipitation data sets: the Stage IV analysis

The NCEP Stage IV product, herein referred to as Stage IV, is a near-real-time product that is generated at NCEP separately from the NWS Precipitation Processing System (PPS) and the NWS River Forecast Center (RFC) rainfall processing. Originally the Stage IV product was intended for assimilation into atmospheric forecast models to improve quantita-

tive precipitation forecasts (QPFs) (Lin and Mitchell, 2005). However the length of record, consistency of data availability, and ease of access has made the Stage IV product attractive for many applications. Data are available in GRIB format for hourly, 6-hourly, and daily temporal scales, and they are gridded on the Hydrologic Rainfall Analysis Projection (HRAP) (Reed and Maidment, 1995, 1999) at a nominal 4 km spatial resolution. Stage IV represents the final stage of the process that combines mosaicked estimates from the 12 RFCs. The gauges used at the RFC level for bias adjustment include available hourly rain gauges such as HADS (Hydrometeorological Automated Data System) gauges, ASOS (Automated Surface Observing System), and AWOS (Automated Airport Weather Stations) reports (Hou et al., 2014). Furthermore, some of the western RFCs (Colorado Basin: CBRFC; California–Nevada: CNRFC; northwestern: NWRFC) do not use radar estimates due to poor coverage over mountainous areas. Those RFCs use an automated analysis of rain gauge observations (Mountain Mapper) that incorporates the gridded monthly precipitation climatology from PRISM. Although changes in gauge-adjustment procedures are always possible at the RFC level, including the incorporation of the best available in situ observations, it is reasonable to assume that the in situ observations used in the Stage IV bias-adjustment procedure (HADS, ASOS, AWOS) are different from the US-COOP subset of GHCN-D used for the evaluation. For the western RFCs, however, the incorporation of the PRISM climatology that uses GHCN-D in situ data will have to be kept in mind for any analysis. Figure 1c presents the geographical extent of the 12 RFCs, and Table 1 reports the number of available rain gauges by RFC. The Stage IV precipitation data are available via http://data.eol.ucar.edu/codiac/dss/id=21.093. The reader will find a more detailed description of the Stage IV precipitation retrievals from the RFC level and

up to the final mosaicked product in Nelson et al. (2015). In addition to radar only reflectivity scanning and processing (beam blockage, hot and cold biases, bright-band contamination, anomalous propagation, cone of silence), the final mosaicked estimates present biases that are visible in the long-term averages. The fact that not all the RFCs use the same precipitation estimation algorithm generates radar-to-radar and RFC-to-RFC discontinuities (Nelson et al., 2010, 2015).

2.4 Satellite precipitation QPE data sets: TMPA 3B42 and 3B42RT

The satellite QPE TMPA 3B42 Version 7 (V7) blends optimally different microwave data sets from low earth orbit satellites (TMI: The Microwave Imager; SSM/I: Special Sensor Microwave Imager; AMSR-E: Advanced Microwave Scanning Radiometer–Earth Observing System; AMSU-B: Advanced Microwave Sounding Unit-B), along with calibrated IR estimates of rain-gauge-corrected monthly accumulation (Huffman et al., 2007). TMPA 3B42 provides precipitation estimates for the domain 50° S–50° N at a 3-hourly and quarter-degree resolution ($0.25° \times 0.25°$) from which seasonal, daily, and sub-daily precipitation characteristics can be derived. The quality of the blended precipitation estimates depends on the number of satellite estimates available at a given time stamp and on the sensor characteristics. Over the years, the retrieval algorithms of the different products incorporated within 3B42 were modified. The algorithm 3B42 itself had several versions, and a major improvement of the precipitation estimates was provided in 2007 to correct for low biases (Huffman et al., 2007). The 3B42 V7 represents substantial improvement when compared to the previous version (V6). The version 7 incorporates additional satellite products along with the reprocessed versions of the merged algorithms (TMI, SSM/I, AMSR-E, AMSU-B). However, the major upgrade consists of the use of a single, uniformly processed surface precipitation gauge analysis from the Global Precipitation Climatology Centre (GPCC) (Huffman and Bolvin, 2013). The gauge analysis used is the GPCC Monitoring Product at 1° grid resolution (Schneider et al., 2010, 2011). This specific analysis uses SYNOP (synoptic weather observation reports) and CLIMAT reports that are received in near-real time from 7000 to 8000 automated stations worldwide. While it is possible that some of the first-order automated synoptic stations included in GHCN-D are also used in the GPCC gauge analysis (SYNOP), most of the US-COOP subset of the GHCN-D stations used for evaluation are not a part of the GPCC Monitoring Product used in the 3B42. Since it is virtually impossible to track down and identify the automated stations that are or are not used in the bias-adjustment procedure for 3B42, we are confident that this number remaining relatively low will not compromise the independent assessment of the 3B42 data set. The use of the GPCC rain gauge analysis

explains most of the differences observed between V6 and V7 over land and over coastal areas. A brief comparison of both versions was provided for the period 1998–2009 over North and Central America, which encompasses the current CONUS domain, in Prat and Nelson (2013b, their Fig. B1). In this work, we also use the near-real-time version of the product (3B42RT), which is produced operationally and does not use the monthly rain gauge correction (GPCC) but incorporates an a priori climatological correction (Huffman et al., 2007). In addition to products relying on gauge measurements (GHCN-D, PRISM) and incorporating gauge information for bias-adjustment purposes (Stage IV, TMPA 3B42), the use of the near-real-time data set 3B42RT has a double objective. First it provides a quantification of the systematic biases and the adjustment performed with respect to surface observations. Second, it aims to examine the suitability of satellite precipitation products to capture precipitation patterns and extremes in near-real time.

3 Annual precipitation: differences between data sets

3.1 Annual average precipitation

Figure 2 displays the annual average precipitation derived from PRISM (Fig. 2a), Stage IV (Fig. 2b), 3B42 (Fig. 2c), and 3B42RT (Fig. 2d) for the period 2002–2012. All data sets present comparable precipitation patterns with higher rainfall east of 97° W, over the southeast (SE), and over the Pacific Northwest (NW). Precipitation derived from Stage IV displays a closer agreement with PRISM, with comparable rainfall over the Pacific Northwest and over the Rockies. The adjusted 3B42 presents a better visual agreement with PRISM and Stage IV than the near-real-time version 3B42RT. However, rainfall over the Pacific Northwest is noticeably lower than that retrieved from PRISM and Stage IV. The effect of monthly accumulation correction between 3B42 and 3B42RT is particularly conspicuous over the Pacific Northwest, the Rockies, and the northeast (NE). Over NE, the annual average precipitation differences between 3B42 and 3B42RT are above $+2\,\mathrm{mm\,day^{-1}}$. Annual average precipitation differences between bias-adjusted and non-adjusted data sets are about $+1\,\mathrm{mm\,day^{-1}}$ over NE. Those differences are about $-1.5\,\mathrm{mm\,day^{-1}}$ over the Rockies. CONUS-wide the mean average annual precipitation for the unadjusted 3B42RT is $2.62\,\mathrm{mm\,day^{-1}}$; for 3B42 it is $2.54\,\mathrm{mm\,day^{-1}}$ (3 % difference).

3.2 Comparison with surface observations

To compare the different estimates for the annual average precipitation, we make the assumption that each rain gauge represents with sufficient accuracy the area-averaged rainfall over the native resolution of the different products evaluated: PRISM and Stage IV ($4 \times 4\,\mathrm{km^2}$) and 3B42 and 3B42RT ($4 \times 4\,\mathrm{km^2}$). While there are well-known limitations of us-
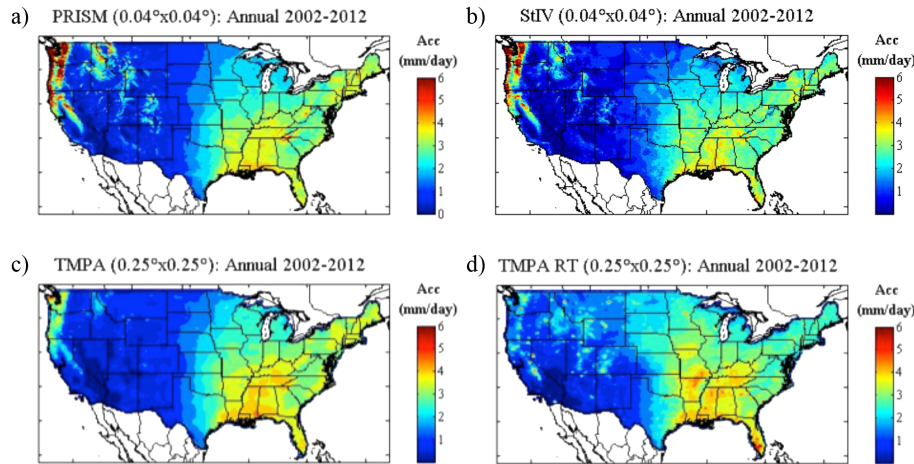
Figure 2. Annual average precipitation derived from **(a)** PRISM, **(b)** Stage IV, **(c)** TMPA 3B42, and **(d)** TMPA 3B42RT for the period 2002–2012.

ing rain gauge point measurements to evaluate area-averaged rainfall retrieved from sensors with coarser spatial resolution (Ciach and Krajewski, 1999; Ciach et al., 2003, 2007; Habib et al., 2004), the random sampling errors due to differing resolutions are mostly dominant at the sub-daily scales (Ciach and Krajewski, 1999). For accumulation period of several days the correlation distance (maximum distance between stations beyond which the correlations become insignificant) is on the order of several hundred kilometers (Gutowski Jr. et al., 2003). Those distances are several orders of magnitude greater than the sensors' spatial resolution.

Figure 3 displays the scatterplots along with the Q–Q (quantile–quantile) plots for annual average precipitation derived from PRISM, Stage IV, 3B42, and 3B42RT when compared to GHCN-D. Over CONUS (Fig. 3a), we observe a very good agreement between GHCN-D surface observations and PRISM ($a = 0.98$; $R^2 = 0.98$) as expected due to the fact that PRISM gridded precipitation estimates incorporate GHCN-D stations. Values for the mean annual average precipitation and the associated standard deviation (σ) are relatively close at $2.42\,\mathrm{mm\,day^{-1}}$ ($\sigma = 1.11\,\mathrm{mm\,day^{-1}}$) for PRISM and $2.47\,\mathrm{mm\,day^{-1}}$ ($\sigma = 1.14\,\mathrm{mm\,day^{-1}}$) for GHCN-D. The differences observed toward higher rain rates ($R > 6\,\mathrm{mm\,day^{-1}}$) are due to the algorithm that uses a digital elevation model and incorporates complex precipitation processes such as rain shadows and coastal effects among others (Daly et al., 1994). Comparison of Stage IV estimates with GHCN-D displays an overall satisfying agreement ($a = 0.93$; $R^2 = 0.93$) with lower precipitation estimates for Stage IV for rain rates greater than $4\,\mathrm{mm\,day^{-1}}$. The satellite QPEs (3B42, 3B42RT) display the highest mean annual average precipitation over CONUS when compared to other precipitation estimates (GHCN-D, PRISM, Stage IV), with $2.54\,\mathrm{mm\,day^{-1}}$ for 3B42 and $2.62\,\mathrm{mm\,day^{-1}}$ for 3B42RT, along with a lower correlation coefficient (Fig. 3a).

However, while the mean annual average precipitation is higher than surface observations, 3B42 and 3B42RT display negative biases in the upper part of the distribution ($R > 4\,\mathrm{mm\,day^{-1}}$) as revealed by the Q–Q plots. In addition, the bias-adjusted 3B42 presents a better agreement with surface observations ($a = 1.00$; $R^2 = 0.83$) than the near-real-time precipitation estimates from 3B42RT ($a = 0.99$; $R^2 = 0.36$). Overall, a better agreement is found for Stage IV than for the satellite estimates (3B42, 3B42RT) in the upper part of the distribution. The differences between surface observations (GHCN-D) and the precipitation data sets (PRISM, Stage IV, 3B42, 3B42RT) vary greatly when considering RFCs separately. For instance, the Lower Mississippi (LM) displays a good agreement regardless of the data set considered (Fig. 3b). PRISM ($3.64\,\mathrm{mm\,day^{-1}}$) presents the best agreement with GHCN-D ($3.75\,\mathrm{mm\,day^{-1}}$). Little differences are found between 3B42 and 3B42RT in terms of average rain rate of 3.87 and $3.90\,\mathrm{mm\,day^{-1}}$, respectively, albeit with a narrower distribution (lower σ) for the bias-adjusted 3B42 than for 3B42RT. Over the Missouri Basin (MB), important differences are observed between 3B42 and 3B42RT with respect to GHCN-D. The bias-adjusted 3B42 displays a satisfying agreement with surface observations that contrasts with the overestimation displayed by the near-real-time 3B42RT observations (Fig. 3c). Over NW, the bias-adjusted 3B42 presents a substantial improvement when compared to 3B42RT, but severe underestimation remains for precipitation above $4\,\mathrm{mm\,day^{-1}}$ (Fig. 3d). Although closer to surface observations, Stage IV displays a similar underestimation at higher rain rates. Table 2 summarizes the differences between GHCN-D and the other data sets. For PRISM, the linear regression coefficient when compared to surface observation (GHCN-D) remains within a narrow range (0.97 to 1.03) for the different RFCs considered. The differences are statistically significant at the 5 % significance for about
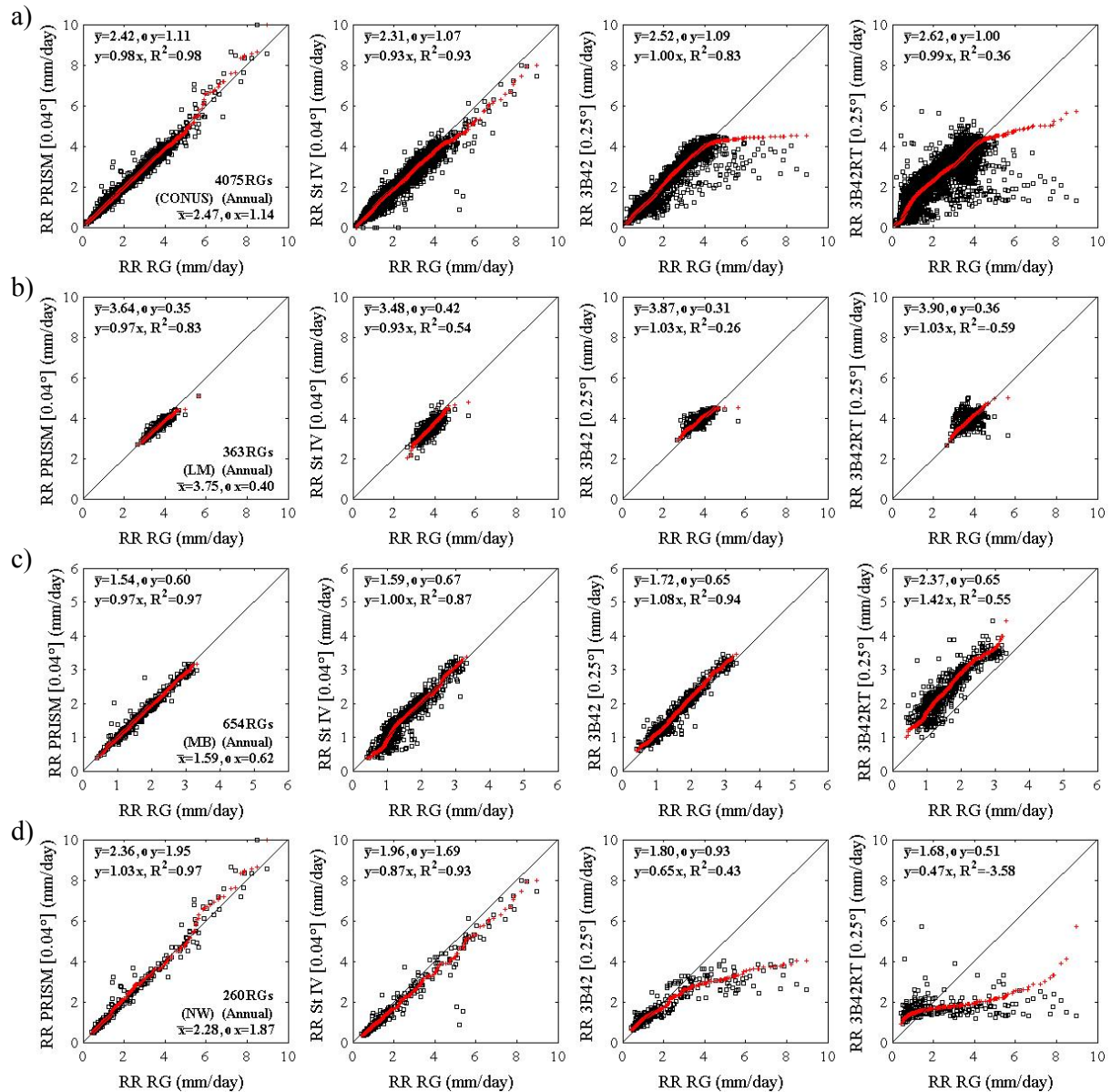
Figure 3. Scatterplots (black) and quantile–quantile (red) plots for annual precipitation derived from PRISM, Stage IV, TMPA 3B42, and TMPA 3B42RT when compared to GHCN-D for **(a)** CONUS, **(b)** Lower Mississippi (LM), **(c)** Missouri Basin (MB), and **(d)** Northwest (NW) for the period 2002–2012. Please note that for row **(c)** the scale is different than for rows **(a)**, **(b)**, and **(d)**.

half of the RFCs (5 over 12). For Stage IV, the variations are greater and indicate a general underestimation, with a varying between 0.87 and 1 and statistically significant differences for 9 of the RFCs. The bias-adjusted 3B42 presents a wider variation range ($0.63 < a < 1.11$), which is noticeably narrower than the coefficient obtained with the near-real-time precipitation estimates 3B42RT ($0.52 < a < 1.42$).

Figure 4a displays the average annual precipitation derived from all data sets for the different RFCs. The Lower Mississippi exhibits the highest average annual rain rate regardless of the data set. The Colorado Basin displays the lower av-

erage annual rain rate for GHCN-D, PRISM, Stage IV, and 3B42. For 3B42RT the minimum is found for California–Nevada (CN). The differences with respect to GHCN-D are presented in Fig. 4b. Over CONUS differences are found between −6.4 % (St. IV) and +6.1 % (3B42RT). For PRISM differences are below 4 % regardless of the RFC considered. CONUS-wide, the precipitation estimates derived from 3B42 and 3B42RT are relatively close with a slightly lower rain rate for 3B42 (−4 %). The magnitude of the bias adjustment (difference between 3B42RT and 3B42) remains below 7 % over most of the basins (Arkansas-Red Basin (AB): +5 %;

Table 2. Average rain rate (mm day^{-1}) and comparisons with surface observations (GHCN-D) for annual precipitation estimations derived from PRISM, Stage IV, 3B42, and 3B42RT. The comparison (%|a|R^2) includes the differences (%) and the linear regression coefficients (a; R^2) over CONUS and each RFC. For each QPE data set, the numbers in bold and italic-bold indicate the upper and lower limits when compared to GHCN-D. The asterisk indicates that the data sets are statistically different at the 5 % significance level with respect to surface observations.

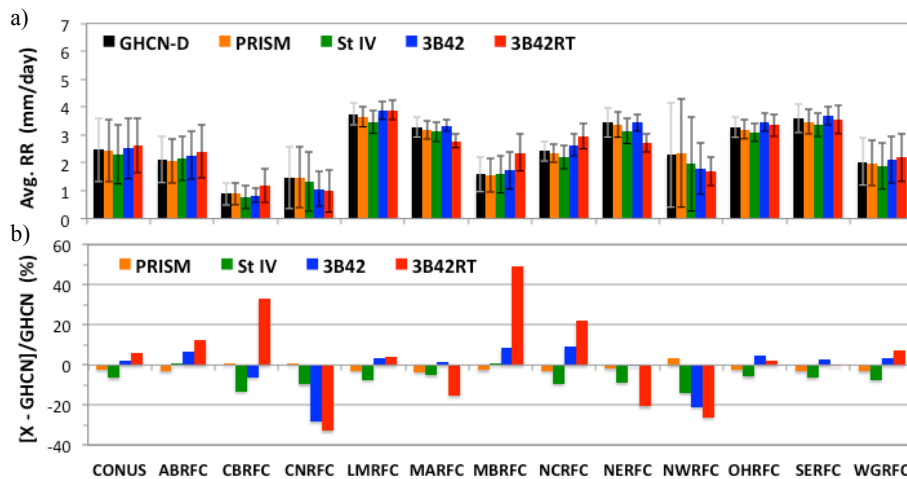| ID	GHCN-D	PRISM	Stage IV	TMPA	
				3B42	3B42RT
CONUS	2.47	2.42 [−2.2|0.98|0.98]*	2.31 [−6.4|0.93|0.93]*	2.52 [+2.1|1.00|0.83]*	2.62 [+6.1|0.99|0.36]*
ABRFC	2.13	2.06 [−3.0|0.97|0.99]	**2.14[+0.7|1.00|0.96]**	**2.27[+6.9|1.06|0.96]***	2.40 [+12.5|1.12|0.94]*
CBRFC	0.89	0.90 [+0.7|0.99|0.87]	0.77 [−13.0|0.89|0.86]*	0.84 [−6.1|0.86|−0.02]	1.18 [+33.2|1.22|0.02]*
CNRFC	1.46	1.46 [+0.3|1.00|0.98]	1.32 [−9.7|0.91|0.92]	*1.05[−28.0|0.63|0.54]**	*0.98[−32.2|0.52|−0.10]**
LMRFC	3.75	3.64 [−2.8|0.97|0.84]*	3.48 [−7.3|0.93|0.54]*	3.87 [+3.2|1.03|0.26]*	3.90 [+3.9|1.0|−0.59]*
MARFC	3.28	*3.17[−3.4|0.97|0.89]**	3.12 [−5.0|0.95|0.63]*	3.34 [+1.6|1.01|−0.21]	2.78 [−15.2|0.84|−1.49]*
MBRFC	1.59	1.54 [−2.6|0.97|0.97]	1.59 [+0.4|1.00|0.87]	1.72 [+8.4|1.08|0.94]*	**2.37[+49.3|1.42|0.56]***
NCRFC	2.42	2.35 [−2.9|0.97|0.90]*	2.19 [−9.3|0.91|0.65]*	2.63 [+9.0|1.09|0.74]*	2.95 [+21.9|1.21|0.17]*
NERFC	3.44	3.38 [−1.8|0.98|0.89]	3.14 [−8.7|0.91|0.43]*	3.43 [−0.1|0.99|−0.88]	2.73 [−20.7|0.78|−1.42]*
NWRFC	2.28	**2.36[+3.6|1.03|0.97]**	*1.96[−13.9|0.87|0.93]**	1.80 [−20.8|0.65|0.43]*	1.68 [−26.1|0.47−3.58]*
OHRFC	3.28	3.20 [−2.3|0.98|0.89]*	3.09 [−5.8|0.94|0.45]*	3.44 [+4.9|1.05|0.42]*	3.35 [+2.2|1.02|0.13]
SERFC	3.58	3.47 [−3.1|0.97|0.84]*	3.36 [−6.2|0.93|0.45]*	3.68 [+2.7|1.02|−0.08]*	3.56 [−0.7|0.98|−0.08]
WGRFC	2.04	1.98 [−3.0|0.97|0.98]	1.89 [−7.4|0.93|0.94]*	2.11 [+3.2|1.03|0.96]	2.19 [+7.4|1.06|0.87]*

Figure 4. Average annual precipitation derived from GHCN-D, PRISM, Stage IV, TMPA 3B42, and TMPA 3B42RT for the different RFCs. Differences (%) with respect to GHCN-D surface measurements.

CN: −7 %; LM: +0.7 %; NW: −7 %; OH: −3 %; SE: −3 %; WG: +4 %). This can be explained by the fact that 3B42RT uses an a priori bias adjustment based on climatological for the near-real-time algorithm (Huffman et al., 2007; Huffman and Bolvin, 2013). Important bias correction is performed over the Midwest (MB) with a +38 % difference between 3B42RT and 3B42, reducing the differences with GHCN-D from +49 % down to +9 % for 3B42RT and 3B42, respectively. This correction is to account for the overestimation of summertime convection by passive microwave retrieval that tends to associate important sub-cloud evaporation with precipitation as we will see in the next section. The CBRFC is the domain that displays the most important difference be-

tween 3B42RT and 3B42 (+42 %). For the remaining RFCs, the differences between 3B42RT and 3B42 remains moderate between −21 and +12 % (MA: −17 %; North-Central (NC): +12 %; NE: −21 %). Stage IV presents globally lower differences with GHCN-D (−14 to +1 %) and PRISM (−17 to +4 %) than the satellite estimates 3B42 (−28 to +7 %) and 3B42RT (−32 to +49 %). At the RFC level, Stage IV almost systematically underestimates precipitation except for two RFCs (AB, MB) with a lower rainfall of −7 % when compared to GHCN-D CONUS-wide (Table 2). For the western RFCs (CB, NW), Stage IV presents a better agreement with surface observations (−12 %) than 3B42 (−23 %) and 3B42RT (−25 %). The lower differences can be explained by
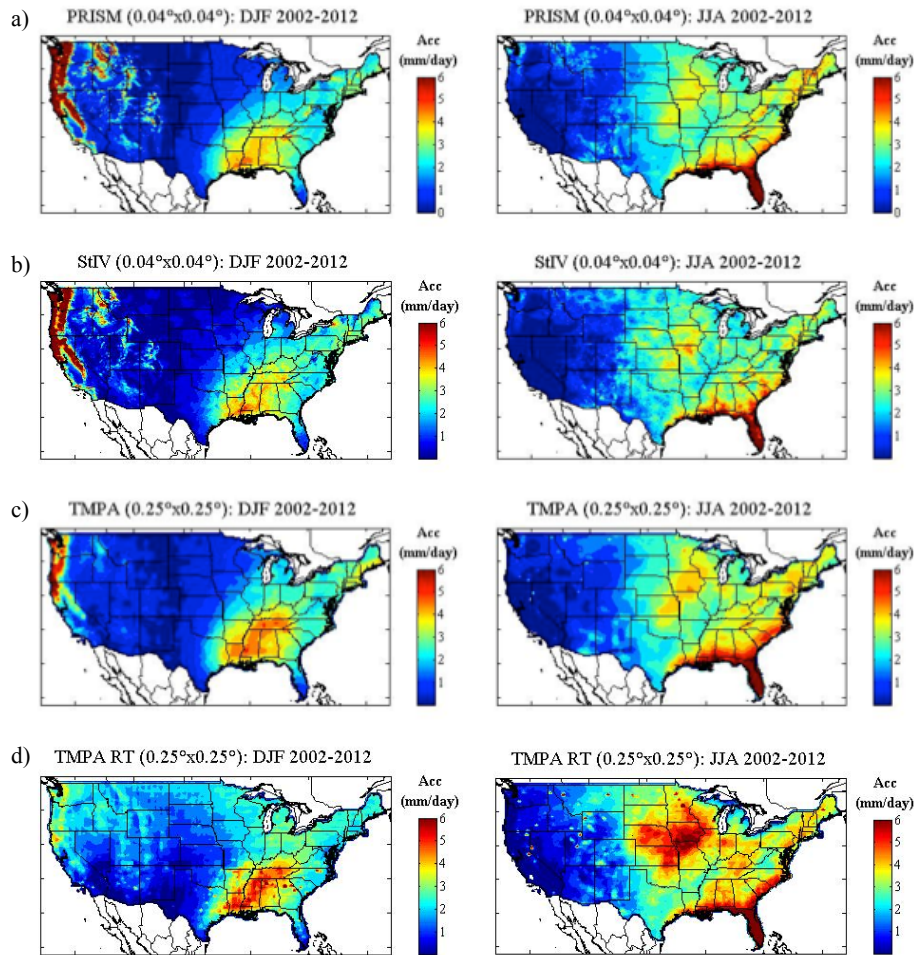
Figure 5. Winter (DJF: left column) and summer (JJA: right column) precipitation derived from **(a)** PRISM, **(b)** Stage IV, **(c)** TMPA 3B42, and **(d)** TMPA 3B42RT for the period 2002–2012.

the fact that the western RFCs use the Mountain Mapper approach and gauge-only estimates (Hou et al., 2014; Nelson et al., 2015).

4   Seasonal precipitation

4.1   Seasonal precipitation patterns

Figure 5 displays the seasonal precipitation for winter (DJF: left) and summer (JJA: right) for PRISM (Fig. 5a), Stage IV (Fig. 5b), 3B42 (Fig. 5c), and 3B42RT (Fig. 5d). PRISM and Stage IV present similar precipitation patterns regardless of the season. By comparison with 3B42RT, the monthly-adjusted 3B42 displays precipitation patterns visually closer to those of PRISM. The differences between 3B42RT and 3B42 are more emphasized on a seasonal basis than observed for the annual basis (Fig. 2). For winter, the bias-adjusted 3B42 precipitation estimates are lower than 3B42RT (3B42 < 3B42RT) over the Rockies (CB), over the highest latitudes along the US–Canadian border (NC,

MB, NW), and east of the Mississippi (LM, SE). Conversely, the 3B42 estimates are found higher than the near-real-time 3B42RT (3B42 > 3B42RT) along the west coast from northern California up to the Pacific Northwest (NW, CN). For summer, the 3B42 estimates are found to be very significantly lower than 3B42RT (3B42 < 3B42RT) over the Midwest (MB, NC, AB). The rain gauge adjustment performed retrospectively corrects the possible overestimation of summertime convection by PMW sensors that mistake sub-cloud evaporation for precipitation (Dinku et al., 2010, 2011; Ochoa et al., 2014). Similarly, for the area of the LM domain located east of the Mississippi, 3B42 estimates are lower than 3B42RT.

4.2   Comparison with surface observations

Figure 6a, b present the seasonal rain rates derived from the different data sets. Between the warm and cold season, average seasonal rain rate derived from surface observations (GHCN-D, PRISM) vary from −95 % for CN to +270 %
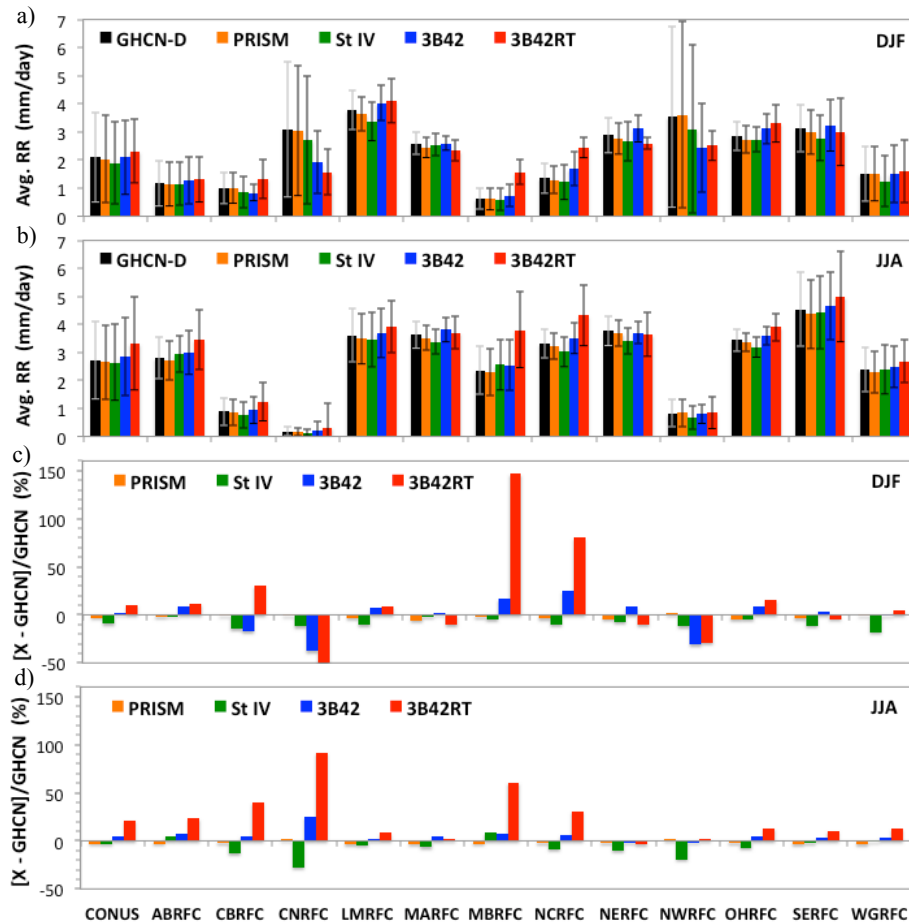
Figure 6. Seasonal rain rate derived from GHCN-D, PRISM, Stage IV, TMPA 3B42, and TMPA 3B42RT for each RFC and for **(a)** winter (DJF) and **(b)** summer (JJA). Differences between GHCN-D and PRISM, Stage IV, TMPA 3B42, and TMPA 3B42RT for **(c)** winter (DJF) and **(d)** summer (JJA).

for MB (Table 3). For winter, the minimum (maximum) average rain rate is found for MB (LM) with 0.64 mm day^{-1} (3.77 mm day^{-1}). For summer, the minimum (maximum) average rain rate is found for CN (SE) with 0.17 mm day^{-1} (4.53 mm day^{-1}). We note that the seasonal minima and maxima are observed for the same RFCs regardless of the data set except for the winter minimum for 3B42RT (Table 3). Seasonal differences between GHCN-D and PRISM remain moderate (-5.9 to $+2.1$ %) and comparable to that for the annual basis (Fig. 6c, d). For stage IV, differences with GHCN-D vary from -18 to -2 % (overall underestimation) for winter and from -28 to $+8$ % (overall underestimation) for summer. For 3B42, the differences with GHCN-D range from -38 to $+25$ % (no overall under/overestimation) in winter. In addition to fundamental limitations in radar and satellite measurement for snow and mixed precipitation events, additional uncertainties are introduced from in situ data that are used either in the adjustment (HADS, ASOS, AWOS) or in the evaluation (GHCN-D) of remotely sensed products. Among the systematic errors in measuring frozen

precipitation are evaporation, chimney effect, wind field deformation, wetting losses, delayed tips due to snow melting in the funnel, or uncertainties due to human intervention in the measurement procedure (Goodison et al., 1998; Groisman et al., 1999; Sevruk et al., 2009; Prat and Barros, 2010b; McMillan et al., 2012; Leeper et al., 2015). Although this point is beyond the scope of this study, we note that the differences observed between remotely sensed and in situ data for the higher-latitude RFCs (CB, MB, NC, NE, NW, OH) are within the range of those observed for the other RFCs experiencing cold precipitation less frequently (-4.7 to -14.8 % vs. -1.5 to -18.0 % for Stage IV and -31.1 to $+16.7$ % vs. -38.0 to $+16.7$ % for 3B42) (Table 3). For summer, differences between 3B42 and GHCN-D present a narrower range from -2 to $+25$ % (overall overestimation). Those differences represent a substantial improvement over those observed for the near-real-time 3B42RT, which vary from -49 to $+147$ % in winter and from -4 to $+92$ % in summer (Table 3). The differences between 3B42RT and 3B42 are the most important for MB and CR in winter ($+111$ and

Table 3. Average rain rate (mm day^{-1}) and differences [%] between GHCN-D and other annual precipitation estimates (PRISM, Stage IV, 3B42, 3B42RT) over CONUS and over each RFC for winter (DJF) and summer (JJA). For each QPE data sets and season, the numbers in bold and italic-bold indicate the upper and lower limits when compared to GHCN-D. The asterisk indicates that the data sets are statistically different at the 5 % significance level with respect to surface observations.

| Season | ID | GHCN-D | PRISM | Stage IV | TMPA | |
					3B42	3B42RT
DJF	CONUS	2.09	2.03 [−2.8]	1.89 [−9.5]*	2.09 [+0.1]	2.31 [+10.7]*
	ABRFC	1.17	1.14 [−2.4]	**1.15**[−**1.5**]	1.28 [+9.2]	1.31 [+11.6]
	CBRFC	1.01	1.01 [−0.2]	0.86 [−14.8]*	0.83 [−17.4]*	1.32 [+30.4]*
	CNRFC	3.09	3.06 [−1.2]	2.72 [−12.2]	*1.92*[−*38.0*]*	*1.57*[−*49.3*]*
	LMRFC	3.77	3.64 [−3.5]*	3.37 [−10.6]*	4.03 [+6.8]*	4.12 [+9.1]*
	MARFC	2.60	*2.44*[−*5.9*]*	2.54 [−2.2]	2.60 [+0.1]	2.32 [−10.4]*
	MBRFC	0.64	0.62 [−2.5]	0.60 [−4.9]	**0.74**[+**16.7**]*	**1.57**[+**146.8**]*
	NCRFC	1.35	1.30 [−4.0]	1.22 [−10.3]*	1.69 [+24.9]*	2.44 [+80.0]*
	NERFC	2.89	2.76 [−4.4]*	2.67 [−7.6]*	3.13 [+8.2]*	2.59 [−10.3]*
	NWRFC	3.54	**3.60**[+**1.8**]	3.11 [−12.2]	2.44 [−31.1]*	2.51 [−29.0]*
	OHRFC	2.86	2.73 [−4.7]*	2.73 [−4.7]*	3.11 [+8.8]*	3.31 [+15.6]*
	SERFC	3.13	3.00 [−4.0]*	2.78 [−11.2]*	3.25 [+3.7]	3.00 [−4.3]
	WGRFC	1.51	1.51 [−0.8]	*1.25*[−*18.0*]*	1.51 [+0.7]	1.59 [+4.5]
JJA	CONUS	2.73	2.65 [−3.0]*	2.64 [−3.3]*	2.85 [+4.4]*	3.32 [+21.5]*
	ABRFC	2.80	2.72 [−2.9]	2.94 [+5.2]*	3.00 [+7.3]*	3.45 [+23.2]*
	CBRFC	0.88	0.86 [−2.6]	0.77 [−13.3]*	0.93 [+5.3]	1.24 [+40.0]*
	CNRFC	0.17	0.17 [+0.6]	*0.12*[−*27.6*]*	**0.21**[+**24.7**]	**0.33**[+**91.8**]*
	LMRFC	3.62	*3.48*[−*3.8*]*	3.46 [−4.4]*	3.70 [+2.3]	3.92 [+8.2]*
	MARFC	3.63	3.51 [−3.5]*	3.38 [−6.9]*	3.81 [+4.8]*	3.71 [+2.1]
	MBRFC	2.37	2.30 [−3.0]	**2.56**[+**8.4**]*	2.55 [+7.7]*	3.81 [+60.9]*
	NCRFC	3.31	3.22 [−2.6]*	3.02 [−8.7]*	3.51 [+6.1]*	4.33 [+31.0]*
	NERFC	3.78	3.68 [−2.5]	3.39 [−10.2]*	*3.71*[−*1.9*]	*3.64*[−*3.7*]*
	NWRFC	0.82	**0.84**[+**2.1**]	0.66 [−19.1]*	0.81 [−1.7]	0.84 [+2.4]
	OHRFC	3.45	3.36 [−2.5]*	3.20 [−7.2]*	3.59 [+4.0]*	3.90 [+13.1]*
	SERFC	4.53	4.36 [−3.7]	4.41 [−2.6]	4.66 [+2.8]	5.00 [+10.3]*
	WGRFC	2.39	2.30 [−3.7]	2.39 [−0.1]	2.48 [+3.8]	2.69 [+12.5]*

+58 %, respectively), or CN and MB in summer (+54 and +49 %, respectively). The situations where the highest differences are observed correspond to significant positive biases of 3B42RT (Table 3).

A closer insight into seasonal differences can be seen in Fig. 7, which displays scatterplots and Q–Q plots for the seasonal rain rate over NW. Regardless of the season (winter: Fig. 7a; summer: Fig. 7b), PRISM presents a very good agreement with surface observations, with differences of 2–3 % for the average rain rate regardless of the season. There is a four-fold difference between the maximum rain rates for winter ($R \approx 12$ mm day^{-1}) and summer ($R \approx 3$ mm day^{-1}). For winter, Stage IV displays a moderate underestimation (−12 %) when compared to GHCN-D. TMPA 3B42 presents a rainfall distribution heavily skewed toward lower rain rates ($R < 6$ mm day^{-1}) when compared with GHCN-D ($R > 12$ mm day^{-1}). For surface rain rates greater than 4 mm day^{-1}, the bias-adjusted 3B42 present a significant improvement when compared to the near-real-time 3B42RT. However, a strong negative bias remains for

3B42 that displays a comparable mean seasonal precipitation (≈ 2.5 mm day^{-1}) to 3B42RT, which is about −30 % when compared to surface observations. Summer exhibits average rain rates (≈ 0.82 mm day^{-1}) 4 times lower than winter (3.5 mm day^{-1}) (Fig. 7b). Stage IV displays a negative bias for summer when compared to GHCN-D and PRISM (≈ -19 %) and comparable with the bias observed for winter (−12 %). Differences remain significant despite the fact that Stage IV uses the PRISM/Mountain Mapper algorithm that combines automated rain gauge observations and PRISM monthly precipitation climatology (Hou et al., 2014; Nelson et al., 2015). Conversely, 3B42 presents a very good agreement with GHCN-D (−1.7 %) and PRISM (−2.4 %) and contrasts with the severe underestimation observed on the right side of the distribution during winter (Fig. 7a). The real-time 3B42RT displays small biases when compared to GHCN-D (+2.4 %) and PRISM (+0.4 %). However, rather than the indication of a good performance, results show that the locations with overestimation are compensated by those with underestimations as can be seen with the Q–Q
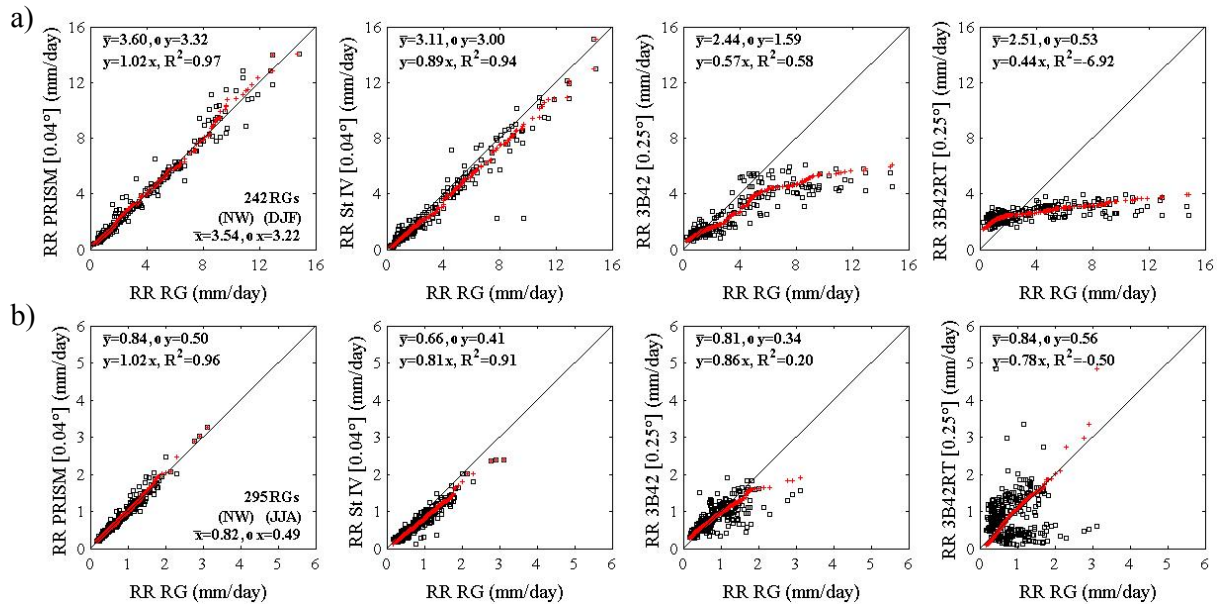
Figure 7. Scatterplots (black) and Q–Q plots (red) for the seasonal rain rate for PRISM, Stage IV, TMPA 3B42, and TMPA 3B42RT for **(a)** winter (DJF) and **(b)** summer (JJA) over Northwest (NW).

plot aligning along the diagonal. A closer look indicates that the 3B42RT pixels displaying the strongest underestimation ($< -50\%$) with respect to GHCN-D are located west of the Cascades mountain range for low to moderate elevation (< 500 m) (not shown). Comparatively, the pixels that display the strongest overestimation ($> 50\%$) are found east of the Cascades regardless of the elevation. However, while the average rain rate remains relatively constant east of the Cascades throughout the year ($R \approx 2$ mm day^{-1}), the seasonal differences are more important west of the Cascades, with average rain rates of less than 2 mm day^{-1} in summer as compared with more than 5 mm day^{-1} in winter. The important underestimation by 3B42RT west of the Cascades regardless of the season illustrates the difficulties for satellite to capture orographic and cold-season precipitation. We also note that, despite the bias adjustment, underestimation remains for 3B42 in winter due to uncertainties related to cold-season precipitation measurements mentioned earlier or by rain gauge locations that cannot fully capture orographic effects that can be observed over distances smaller than the satellite resolution (Prat and Barros, 2010a).

Further illustration of the importance of the 3B42 bias adjustment can be found in Fig. 8, which displays comparisons over MB. Again, GHCN-D and PRISM present an average rain rate difference of about 3 % regardless of the season. Similarly, Stage IV presents a good agreement with surface observations, with a small underestimation of -5 % for winter (Fig. 8a) and a moderate overestimation of $+8.4$ % for summer (Fig. 8b). For both the cold and the warm season, the improvement brought by the 3B42 bias adjustment is clearly visible. For winter, the near-real-time 3B42RT exhibits a severe overestimation of $+147$ % with respect to surface observation, which is reduced to $+16.7$ % for the bias-adjusted 3B42 (Table 3). A closer look at the wintertime precipitation (Fig. 5d) indicates higher rainfall accumulation for 3B42RT at higher latitudes and along the edges of the MBRFC when compared to the other data sets (Fig. 5a–c). These differences are certainly associated with cold season precipitation and are due to the challenge of measuring falling snow, frozen precipitation, and precipitation over snow and ice-covered areas by sensors (SSM/I, AMSU-B) used in the near-real-time 3B42RT (Huffman and Bolvin, 2013). For summer, the bias-adjusted 3B42 exhibits moderate differences of $+7.7$ % with respect to GHCN-D as compared with an overestimation of $+61$ % for 3B42RT (Table 3). The overestimation of 3B42RT is found consistently throughout the rain rate spectra (Fig. 8b). As mentioned earlier, the 3B42RT overestimation is due to uncertainties in PMW retrievals that associate summertime sub-cloud evaporation with precipitation (Dinku et al., 2010, 2011; Ochoa et al., 2014). The monthly bias adjustment (3B42) corrects efficiently for both the cold- and warm-season 3B42RT rainfall overestimation, with a reduction of the average rain rate of about 110 and 50 % for winter and summer, respectively. Overall, the bias-adjusted 3B42 performed very well over the Great Plains (MB, NC), correcting for the overestimation of summertime convection (Table 3) with comparable results to Stage IV (Table 3). Important differences remained, however, for low daily rainfall (< 1 mm day^{-1}) and for the western RFCs during wintertime, mostly due to the difficulty in capturing orographic precipitation and uncertainties in retrieving cold precipitation by the satellite (Chen et al., 2013; Huffman and Bolvin,
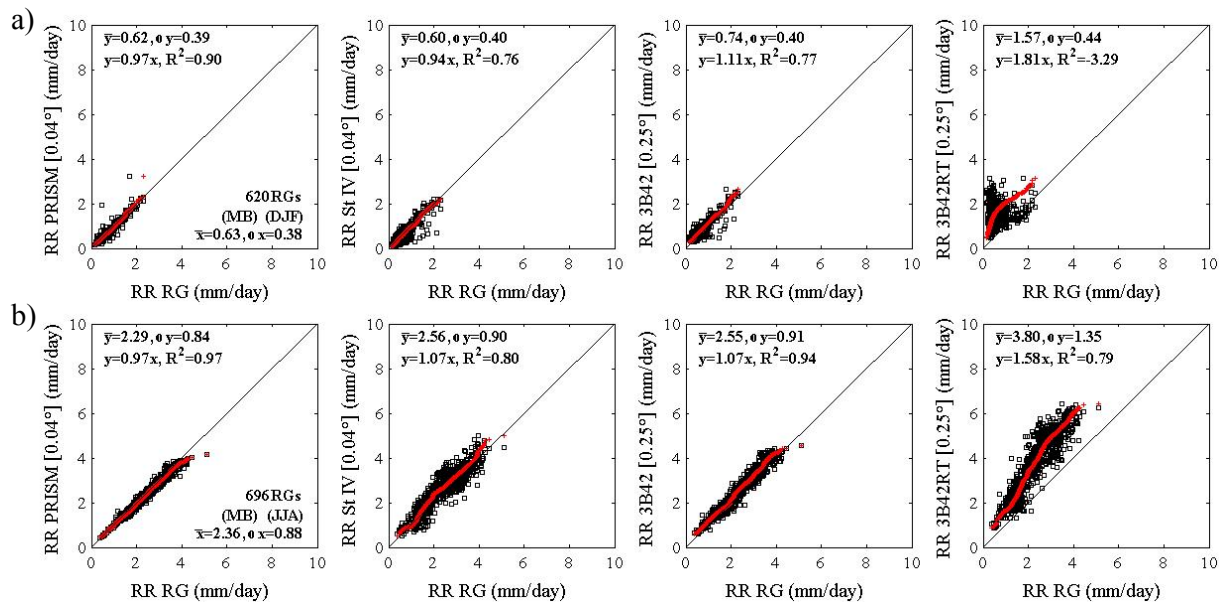
Figure 8. Scatterplots (black) and Q–Q plots (red) for the seasonal rain rate for PRISM, Stage IV, TMPA 3B42, and TMPA 3B42RT for **(a)** winter (DJF) and **(b)** summer (JJA) over Missouri Basin (MB).

2013) but also due to the rain gauges used in the bias adjustment and the evaluation (Goodison et al., 1998; Groisman et al., 1999; Leeper et al., 2015).

5 Daily precipitation

5.1 Conditional analysis and extreme precipitation

After investigating the ability of the different data sets to describe precipitation patterns, this section investigates their ability to capture intense and extreme precipitation at the daily scale. A conditional analysis was conducted using different thresholds for daily accumulation (Fig. 9). Figure 9a displays the average number of rainy days by year derived from GHCN-D (first column), Stage IV (second column), 3B42 (third column), and 3B42RT (fourth column). For Stage IV, TMPA 3B42, and TMPA 3B42RT the daily accumulation is computed 12:00–12:00 UTC. For GHCN-D, the daily accumulation computed depends on the local time and is 07:00–07:00 LST for most of the locations, which corresponds to 12:00–12:00 UTC on the eastern USA. Therefore, some uncertainties could arise from computing daily accumulation over a slightly different time period. Although the number of rainy days appears consistent in terms of magnitude for the different observation platforms, there are noticeable differences over specific areas. Despite a delicate visual comparison between point data measurements from GHCN-D and Stage IV gridded estimates due to the scarcity of station coverage, both products present a very similar pattern and a comparable number of rainy days throughout CONUS. When compared to Stage IV, 3B42 displays a lower num-

ber of rainy days over NE, Middle Atlantic (MA), and Ohio (OH). Similarly, a lower number of rainy days are observed for 3B42 over NW when compared to Stage IV. On the other hand, 3B42 displays a higher number of rainy days over the Rockies – encompassing part or all of MB, CB, and CN – when compared to Stage IV. Different sensitivity for light rainfall detection thresholds for each sensor, the ability to retrieve snow/frozen precipitation, beam blockage over the Rockies, and/or the influence of temporal and spatial resolution can explain the differences. For instance, Stage IV higher spatial resolution could improve the detection of localized events as compared to the satellite's coarser resolution. Overall, the rain-gauge-adjusted radar (Stage IV) and satellite (3B42) data sets display a satisfying visual agreement over CONUS despite the local differences mentioned above. More important differences are observed with the real-time 3B42RT data set. Differences between the rain-gauge-adjusted satellite data set 3B42 and 3B42RT are particularly important over the western USA (Rocky Mountains) and at higher latitudes with more rainy days for 3B42RT. For daily accumulation greater than the wet millimeter days (WMMDs: $R > 17.8\,\mathrm{mm\,day^{-1}}$), significant differences are found over NW and over the southeastern USA (LM, SE) (Fig. 9b). The wet-millimeter-day threshold corresponds to the precipitation days that exceed the highest daily average over the area considered (Shepherd et al., 2007). For North America, this maximum daily average ($17.8\,\mathrm{mm\,day^{-1}}$) is recorded in Henderson Lake (British Columbia) (Source NCDC). Both Stage IV and 3B42 display similar distribution patterns of WMMDs. The most important differences are found over NW. The differences between GHCN-
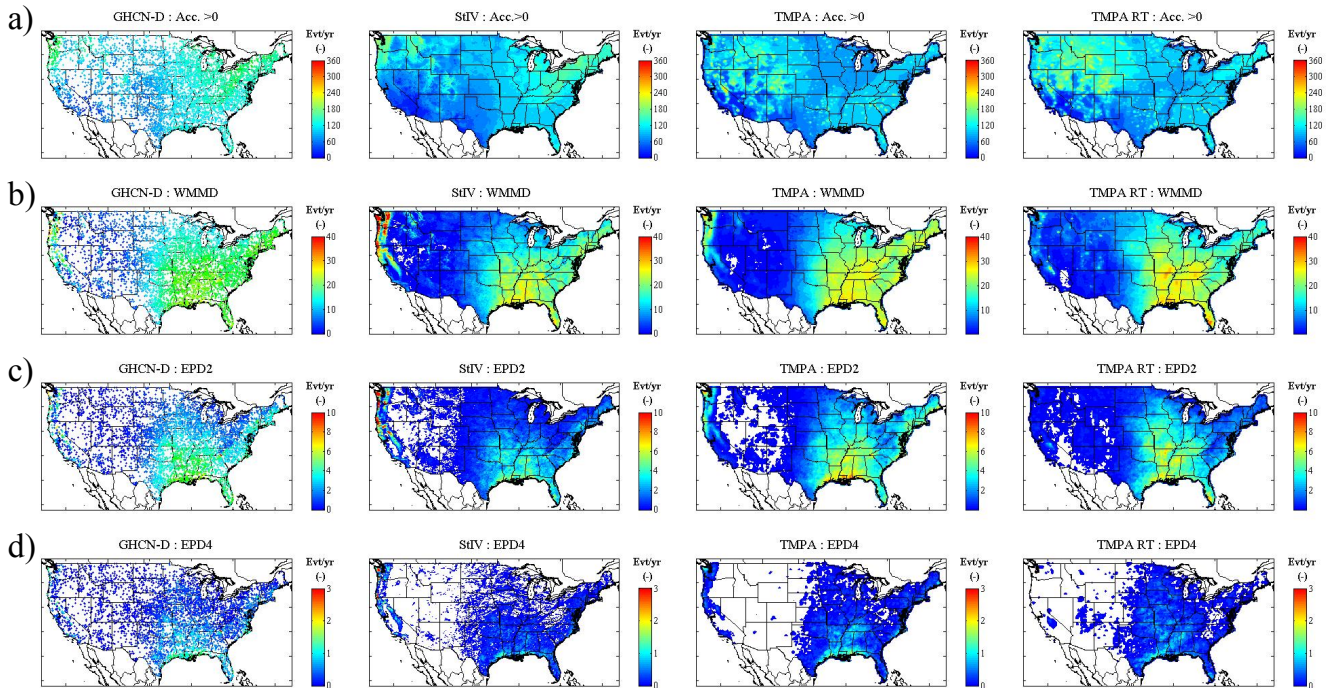
Figure 9. (a) Number of rainy days ($RR > 0$ mm day^{-1}), **(b)** wet millimeter days (WMMDs), **(c)** extreme precipitation days greater than 2 in. day^{-1} ($RR > 50.8$ mm day^{-1}: EPD2), and **(d)** extreme precipitation days greater than 4 in. day^{-1} ($RR > 101.6$ mm day^{-1}: EPD4) for Stage IV and TMPA 3B42.

D and Stage IV are due to the scarcity of station coverage over the Pacific Northwest coast. For the gridded estimates, Stage IV displays a higher number of WMMDs when compared to bias-adjusted satellite estimates 3B42. The biggest differences are observed with 3B42RT that shows a much lower number of rainy days greater than WMMDs (Fig. 9b). This is consistent with the underestimation observed for the daily averages for 3B42RT and to a lesser extent for 3B42 (Figs. 3d and 7a, b). The bias adjustment increases the number of WMMDs of 3B42 closer to Stage IV levels. Conversely, 3B42 and 3B42RT, which displayed fewer rainy days than Stage IV (Fig. 9a), present a higher occurrence of WMMDs when compared to Stage IV over NE, the upper part of the NC domain, and LM (Fig. 9b). For daily accumulation greater than 2 in. day^{-1} (> 50.8 mm day^{-1}; Karl and Plummer, 1995; hereafter EPD2), GHCN-D and Stage IV display comparable counts for EPD2 over the eastern USA, where rain gauges coverage is denser (Fig. 9c). Over NE and the southeastern USA (LM, SE), 3B42 and 3B42RT display a higher number of days with rainfall above 2 in. day^{-1} as compared to Stage IV. Daily precipitation greater than 4 in. day^{-1} (> 101.6 mm day^{-1}; Barlow, 2011; hereafter EPD4) is limited to the Pacific coast and east of 100° W, a domain regularly impacted by tropical cyclones (Prat and Nelson, 2013a, b, 2014) (Fig. 9d). These EPD4 events are relatively infrequent (three counts or less by year) and roughly correspond to the 0.1–0.5 % top daily

events regardless of the RFC considered. The bias-adjusted 3B42 and real-time 3B42RT display a comparable number of EPD4 events over the southeastern USA. The maximum occurrences are observed over the LMRFC and are higher than the daily counts for Stage IV. Over NW, the bias-adjusted 3B42 is able to better capture those extreme daily accumulation events (EPD4) with respect to 3B42RT, which displays almost no days with rainfall above 4 in. day^{-1}.

More quantitative information can be found in Fig. 10, which displays the proportion of rain gauges (GHCN-D) and the corresponding radar (Stage IV) or satellite (3B42, 3B42RT) pixels experiencing the different daily accumulation thresholds (WMMDs, EPD2, EPD4) over the 11-year period. For CONUS (central panel), we note that this proportion of stations/pixels experiencing WMMDs, EPD2, and EPD4 is comparable regardless of the platform considered. For instance, all stations/pixels experience WMMDs during the 11-year period (100 %). For EPD2, the ratio remains relatively close regardless of the sensor and varies from 88 % (Stage IV) to 95 % (3B42RT). Similarly for EPD4, the proportion is about 60 % (GHCN, 3B42, 3B42RT), with a slightly lower ratio (54 %) for Stage IV. Some interesting facts can be derived from the isolated RFC figures (border figures). A few RFCs (AB, LM, MA, SE, WG) display comparable ratios regardless of the sensor and the daily accumulation considered. Apart from a couple of RFCs (NC, OH), the ratio of Stage IV pixels experiencing extreme precipita-
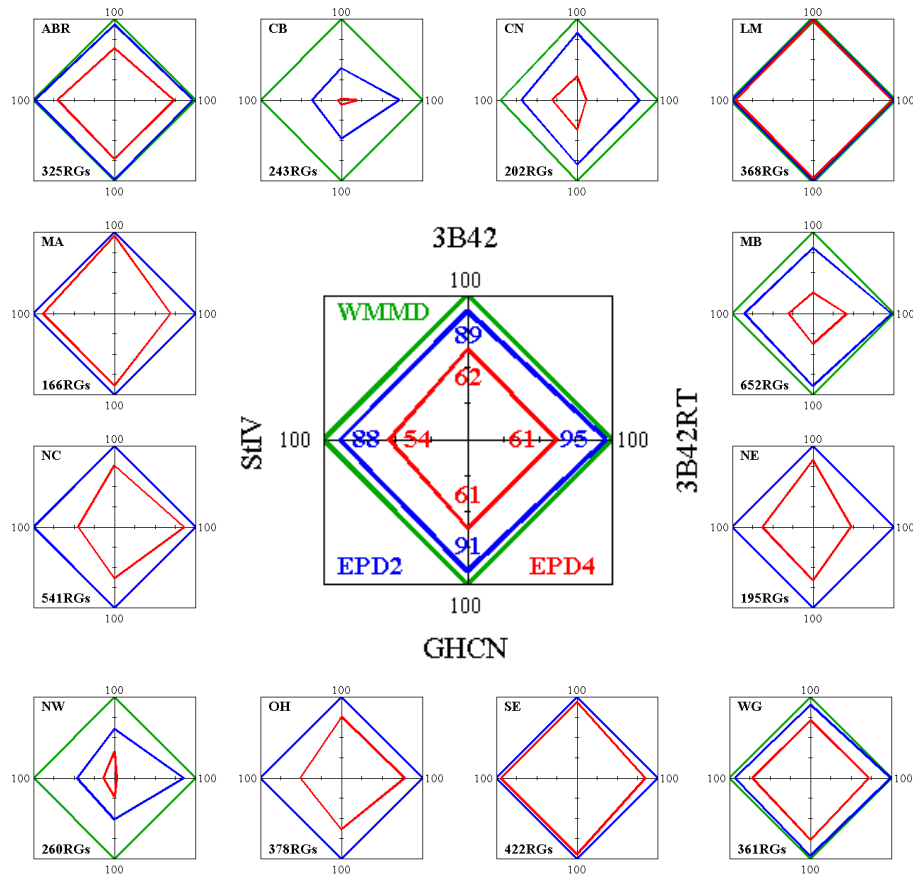
Figure 10. Proportion (%) of stations (GHCN-D) and corresponding pixel (Stage IV, TMPA 3B42, TMPA 3B42RT) experiencing WMMDs, EPD2, and EPD4 over CONUS (central figure) and for the 12 River Forecast Centers (border figures).

tion (EPD4) is relatively close to the ratio of GHCN. When looking at satellite pixels, we observe a relative symmetry for the ratio of stations experiencing EPD2. However, for western (CN, NW) and northeastern (NE) RFCs we note a strong asymmetry in the ratio of 3B42RT pixels experiencing extreme precipitation (EPD4) when compared to the other sensors. This confirms the fact that over the western USA the non-adjusted satellite QPE severely underestimates extreme daily precipitation. Interestingly, over the neighboring Colorado Basin RFC (CB) we note a higher proportion of 3B42RT pixels displaying EPD2 and EPD4 than observed for the other sensors (GHCN, Stage IV, 3B42). Furthermore, regardless of the RFC and daily accumulation considered, the ratio of pixels for 3B42 is very close to that of the GHCN stations, hence providing confidence in the bias adjustment performed. However those results have to be interpreted with caution as they present a count of the daily events over the 11-year period. The number of events decreases with increasing rain rate, and the WMMDs correspond roughly to the 90th-percentile precipitation events regardless of the RFC (Nelson et al., 2015). A test was performed to determine the interstation correlation of daily precipitation events corre-

sponding to the 90th percentile (not shown). For each station, the correlation was computed using the daily events greater than the 90th percentile regardless of the values of the other stations. Results showed that for those high-intensity events the average correlation distance was about 30–80 km, which is comparable with the satellite footprint.

Figure 11 provides a count of the total number of rainy days (Fig. 11a), WMMDs (Fig. 11b), EPD2 (Fig. 11c), and EPD4 (Fig. 11d) for GHCN-D, Stage IV, 3B42, and 3B42RT over CONUS and for each RFC at the rain gauge location. For the number of rainy days, we note that 3B42 and 3B42RT provide comparable results and display less variability across the RFCs when compared to GHCN-D and Stage IV. For GHCN and Stage IV, the RFCs over the Rockies or located partially west of 95° W (AB, CB, CN, WG) display about half of the rainy days of the eastern (MA, NC, NE, OH, SE) and NW RFCs. The Colorado Basin presents consistently the lowest average number of events by active stations regardless of the daily accumulation (WMMDs, EPD2, EPD4). On the other hand, LM presents the highest average number of events regardless of the sensor considered. For selected RFCs, the differences between 3B42 and 3B42RT are par-
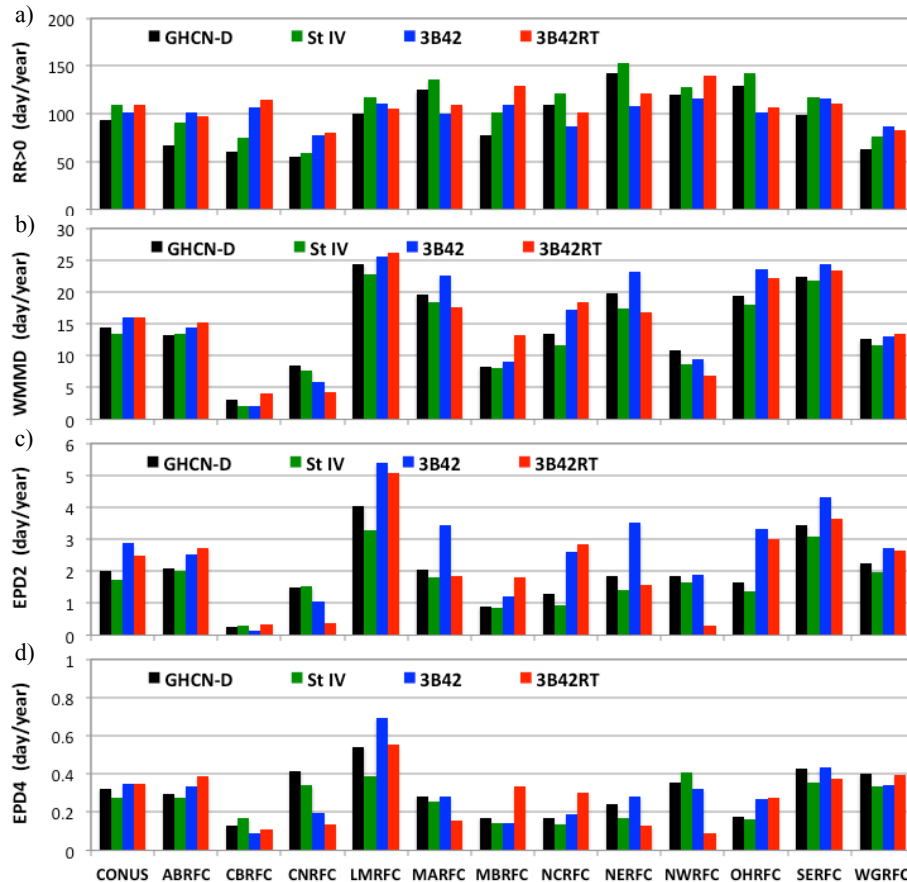
Figure 11. Average number of **(a)** rainy days; **(b)** wet millimeter days (WMMDs) i.e., precipitation days with accumulation greater than 17.8 mm day^{-1}; **(c)** precipitation days with accumulation greater than 2 in. day^{-1} (EPD2); and **(d)** precipitation days with accumulation greater than 4 in. day^{-1} (EPD4) for GHCN-D, Stage IV, TMPA 3B42, and TMPA 3B42RT over CONUS and for the 12 River Forecast Centers (RFCs). Data are for the period 2002–2012. The average number of days is normalized by the number of locations experiencing at least one event (Fig. 10).

ticularly important for EPD2 and EPD4. The biggest differences are found for NW and MB. For the latter, the number of EPD2 and EPD4 events for 3B42RT is about 50 and 130 % higher, respectively, than for 3B42 and is attributed to summertime convection and sub-cloud evaporation over the Midwest as mentioned earlier. Over NW, the number of EPD2 and EPD4 events retrieved after bias adjustment (3B42) is 6-and 3-fold the number of events indicated by 3B42RT due to the difficulty of capturing extreme precipitation in real time over the area. For the latest case, consider that those events (EPD4) correspond to only a handful of occurrences for the period 2002–2012, and caution is necessary when analyzing those results.

5.2 Contingency analysis between Stage IV and GHCN-D

The previous results were provided for the entire period 2002–2012. Figure 12 displays a contingency analysis between daily precipitation from the GHCN-D stations and the

corresponding Stage IV radar pixel. We will assume that the rain gauge is representative of the grid-averaged rainfall for Stage IV. The computation of the interstation correlation for daily events indicated that the correlation distance was greater than the 4 km spatial resolution of the radar (not shown). The number of rainy days ($R > 0$ mm day^{-1}) observed simultaneously at the rain gauge and the radar pixel is 62 % over CONUS (Fig. 12a). Significant differences are observed between RFCs and vary from 49 % (CB) to 71 % (OH). Events observed only by the radar are 24 % over CONUS, which is higher than the ratio for gauge-only events (14 %). A similar trend is observed regardless of the RFC considered, ranging from 18 % (NE) to 32 % (MB) for events at the radar pixel only and from 8 % (AB) to 22 % (CN) for rain-gauge-only events. With increasing rain rate the number of events observed simultaneously by the gauge and the radar decreased from 62 % ($R > 0$ mm day^{-1}; Fig. 2a), to 56 % (WMMD; Fig. 12b), to 43 % (EPD2; Fig. 12c), and to 35 % (EPD4; Fig. 12d). Furthermore, while the ratio of events

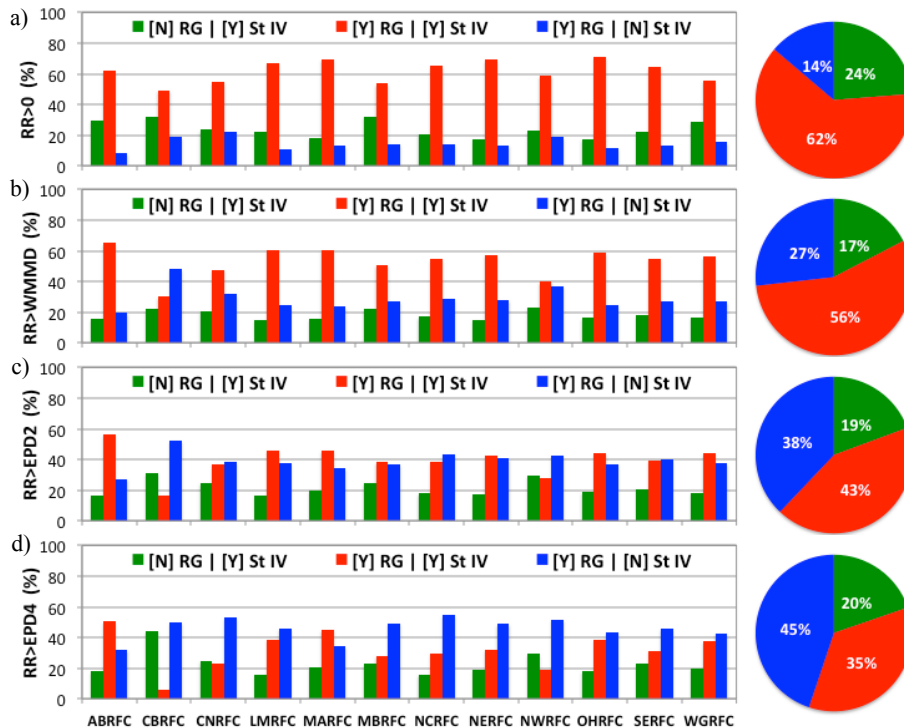**Figure 12.** Contingency as a function of the daily threshold selected – **(a)** RR > 0, **(b)** RR > WMMD, **(c)** RR > EPD2, and **(d)** RR > EPD4 – for rainfall observed simultaneously at the rain gauge and at the radar pixel (YY: red), and successively at the rain gauge only (YN: blue), or at the radar pixel only (NY: green) over CONUS (circle) and for the 12 RFCs (bars). Data are for the period 2002–2012.

observed at the radar pixel only remains relatively close around 20 % regardless of the daily rainfall threshold (i.e., between 17 % for WMMD and 24 % for $R > 0\,\mathrm{mm\,day^{-1}}$), the number of events missed by the radar increases from 14 % ($R > 0\,\mathrm{mm\,day^{-1}}$) to 45 % (EPD4). In addition for accumulation greater than $2\,\mathrm{in.\,day^{-1}}$, the number of extreme events missed by one or the other sensor is greater than the number of events observed simultaneously by both sensors (Fig. 12c). For accumulation greater than $4\,\mathrm{in.\,day^{-1}}$, the proportion of events missed by the radar becomes more important except for the ABRFC.

Figure 13 displays the contingency analysis at each station location. Regardless of the daily accumulation ($R > 0\,\mathrm{mm\,day^{-1}}$; Fig. 13a), the eastern USA and west-coast stations present a higher proportion of events observed simultaneously at the rain gauge and radar pixel (median column). There is a strong contrast between the eastern and western USA, with the eastern USA displaying a lower proportion of events observed at the gauge only. With increasing daily accumulation (Fig. 13b–d), the number of rainfall events decreases over the Rockies, the western, and the northern USA as described previously (Fig. 8). While the spatial extent of intense precipitation events becomes more and more limited to the southeastern USA with increasing daily accumulation, the ratio of events observed at the radar pixel only (right column) remains relatively constant. For concurrent rainfall

events (median column) the ratio decreases significantly for daily accumulation greater than $2\,\mathrm{in.\,day^{-1}}$ (Fig. 13c). With increasing daily accumulation, as the number of events become smaller and spatially more localized, the ratio of events missed by the radar increases importantly over the Midwest (Fig. 13b–d). However, caution is advised when looking at increasing threshold events in particular over areas where those events become more and more scarce. A closer look shows that most of the events observed at one or the other sensor (NY: left column; YN: right column) for accumulation greater than EPD2 and EPD4 are single-occurrence events. For EPD2, the single-occurrence events are located west of $-103°$ W longitude (Fig. 13e). For EPD4, apart from isolated events over the Rockies and the Pacific coast, most of the single-occurrence events are located at the edge of the southeastern USA, i.e., east of $-100°$ W and north of the $40°$ N latitude (Fig. 13f).

6   Summary and conclusion

We compared quantitative precipitation estimates from satellite (3B42 and 3B42RT) and radar (Stage IV) with surface observations (GHCN-D) and models (PRISM) over CONUS for the period 2002–2012. The comparisons were performed at the annual, seasonal, and daily scales over the major river
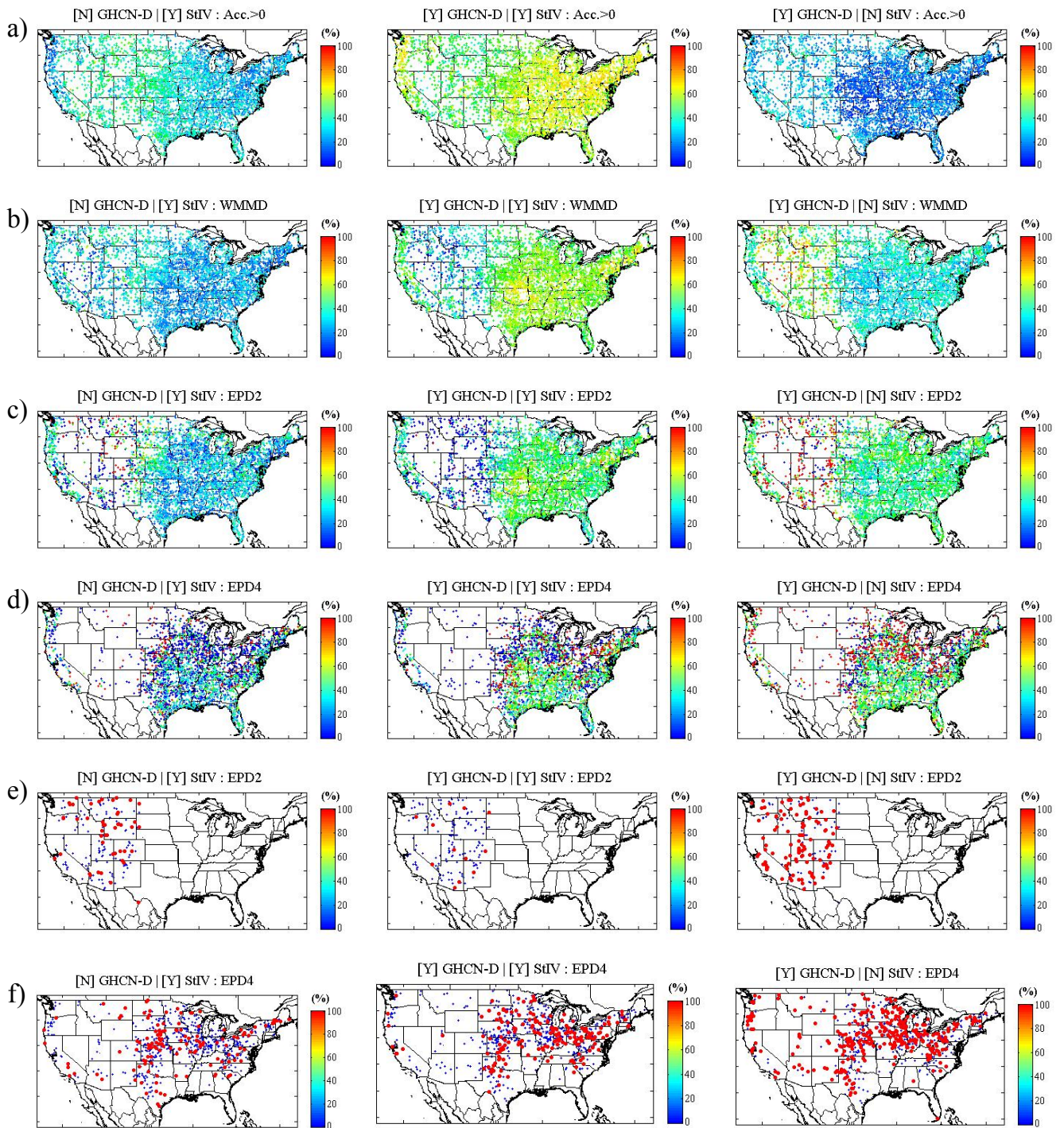
Figure 13. Contingency analysis at the rain gauge site with respect to the daily rainfall accumulation: **(a)** RR > 0, **(b)** RR > WMMD, **(c)** RR > EPD2, and **(d)** RR > EPD4 for rain observed at the radar pixel only (first column), simultaneously at the rain gauge and radar (second column), and at the rain gauge only (third column). **(e** and **f)** Same as **(c** and **d)** but only displaying single event occurrence over the period 2002–2012.

basins within CONUS. The main conclusions of this study are summarized below.

- Over CONUS the different data sets show a satisfying agreement on an annual basis, with differences ranging between −6.4 % (St. IV) and +6.1 % (3B42RT). At the RFC level, PRISM displays a difference of ±4 % with GHCN-D. Stage IV presents a tendency to underestimate when compared to surface observations (−14 to +1 %). A bigger spread of the differences is found between 3B42 and GHCN-D (−28 to +9 %). Finally, 3B42RT displays the bigger differences with GHCN-D (−33 to +49 %).

- The bias-adjusted 3B42 represents a significant improvement when compared to the near-real-time 3B42RT. The 3B42RT biases were particularly important at the seasonal scale over the western and northwestern USA (CN, NW) and at higher latitude over the Midwest (MB, NC) during winter, with an important underestimation (−35 %) of the daily accumulation in the first case (CN, NW) and a severe overestimation (+100 %) in the second case (MB, NC). During summer, 3B42RT presents large positive biases (+45 %) over the Midwest (MB, NC). The bias adjustment (3B42) reduces those differences to moderate levels (i.e., from +100 to +22 % in winter and from +45 to +7 % in summer). Over the CNRFC, 3B42RT presents alternatively a severe underestimation for winter (−45 %) and a severe overestimation for summer (+121 %), with an overall annual difference of −23 % with surface observations GHCN-D.

- Despite the bias adjustment, large biases remained for 3B42 at higher daily average accumulation (> 5 mm day^{-1}) over CONUS. Discrepancies can be explained by the difference between point (RG) and area (satellite, radar) measurements and the difficulty in capturing localized, convective, and orographic events due to the coarser resolution. Furthermore, those differences can be more important at the seasonal scale and for selected basins in particular over the western part of CONUS (Pacific Northwest, Rocky Mountains) due to the difficulty in retrieving precipitation over mountainous areas.

- Stage IV presents an overall better agreement with surface observations than 3B42. At the seasonal level Stage IV displays the same tendency of rainfall underestimation with respect to surface observations, with differences ranging from −18 to −2 % for winter and from −28 to +8 % for summer. Comparatively, 3B42 displays a bigger spread with no particular tendency (−38 to +25 %) for winter and a tendency of rainfall overestimation (−2 to +25 %) for summer.

- At the daily scale, the conditional analysis performed using increasing daily precipitation thresholds (0–4 in. day^{-1}) showed that the sensor ability to capture intense and extreme precipitation depended on the domain considered. In particular, the near-real-time satellite QPE 3B42RT displayed poor skills in capturing intense daily precipitation over NW. The bias-adjusted 3B42 exhibited a significant improvement and level closer to surface station (GHCN-D) and radar statistics (Stage IV) over the 11-year period.

- A contingency analysis performed at the rain gauge location (GHCN-D) and the corresponding radar pixel (Stage IV), showed that with increasing daily accumulation from greater than 0 to greater than 4 in. day^{-1}, the ratio of events observed simultaneously by the gauge and the radar decreased from 62 to 45 %. Furthermore, while the ratio of events observed only by the radar remained close (around 20 %) regardless of the daily accumulation, the number of events measured at the ground but missed by the radar increased from 15 to 45 %. Although caution is required due to the fact that large rainfall events above 2 in. day^{-1} (a fortiori events greater than 4 in. day^{-1}) are infrequent and geographically limited to the Pacific Northwest and the eastern USA, results illustrate the challenge of retrieving extreme precipitation (top 1 percentile) from remote sensing.

References

Ashouri, H, Hsu, K., Sorooshian, S., Braithwaite, D., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat, O. P.: PERSIANN-CDR: Daily precipitation climate data record from multi-satellite observations for hydrological and climate studies, B. Am. Meteorol. Soc., 96, 69–83, 2015.

Barlow, M.: Influence of hurricane-related activity on North American extreme precipitation, Geophys. Res. Lett., 38, L04705, doi:10.1029/2010GL046258, 2011.

Chen, S., Hong, Y., Gourley, J. J., Huffman, G. J., Tian, Y., Cao, Q., Yong, B., Kirstetter, P.-E., Hu, J., Hardy, J., Li, Z., Khan, S. I., and Xue, X.: Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States, Water Resour. Res., 49, 8174–8186, doi:10.1002/2012WR012795, 2013.

Ciach, G. J. and Krajewski, W. F.: On the estimation of rainfall error variance, Adv. Water Resour., 2, 585–595, 1999.

Ciach, G. J., Habib, E., and Krajewski, W. F.: Zero-covariance hypothesis in the error variance separation method of radar rainfall verification, Adv. Water Resour., 26, 573–580, 2003.

Ciach, G. J., Krajewski, W. F., and Villarini, G.: Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data, J. Hydrometeorol. 8, 1325–347, 2007.

Daly, C., Neilson, R. P., and Phillips, D. L.: A statistical-topographic model for mapping climatological precipitation over mountainous terrain, J. Appl. Meteorol., 33, 140–158, 1994.

Dinku, T., Ruiz, F., Connor, S. J., and Ceccato, P.: Validation and Intercomparison of Satellite Rainfall Estimates over Colombia, J. Appl. Meteorol. Clim., 49, 1004–1014, doi:10.1175/2009JAMC2260.1, 2010.

Dinku, T., Ceccato, P., and Connor, S. J.: Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa, Int. J. Remote Sens., 30, 5965–5979, doi:10.1080/01431161.2010.499381, 2011.

Goodison, B. E., Louie, P. Y. T., and Yang, D.: WMO solid precipitation measurement intercomparison, WMO Instruments and Observing Methods Rep. WMO/TD-872, WMO, Switzerland, 212 pp., 1998.

Groisman, P. Y., Peck, E. L., and Quayle, R. G.: Intercomparison of Recording and Standard Nonrecording U.S. Gauges, J. Atmos. Ocean. Tech., 16, 602–609, 1999.

Gutowski Jr., W. J., Decker, S. G., Donavon, R. A., Pan, Z., Arritt, R. W., and Takle, E. S.: Temporal–Spatial Scales of Observed and Simulated Precipitation in Central U.S. Climate, J. Climate, 16, 3841–3847, 2003.

Habib, E., Ciach, G. J., and Krajewski, W. F.: A method for filtering out raingauge representativeness errors from the verification distributions of radar and raingauge rainfall, Adv. Water Resour., 27, 967–980, 2004.

Hou, D., Charles, M., Luo, Y., Toth, Z., Zhu, Y., Krzysztofowicz, R., Lin, Y., Xie, P., Seo, D.-J., Pena, M., and Cui, B.: Climatology-Calibrated Precipitation Analysis at Fine Scales: Statistical Adjustment of Stage IV toward CPC Gauge-Based Analysis, J. Hydrometeorol., 15, 2542–2557, 2014.

Huffman, G. J. and Bolvin, D. T.: TRMM and Other Data Precipitation Data Set Documentation, Lab. for Atmos., NASA Goddard Space Flight Cent. and Sci. Syst. and Appl. Inc., available at: ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_oc.pdf (last access: 28 January 2015), 2013.

Huffman, G. J., Adler, R. F., Morrissey, M., Bolvin, D., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multi-satellite observations, J. Hydrometeorol., 2, 36–50, 2001.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K., Hong, Y., Stocker, E. F., and Wolf, D. B.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, J. Hydrometeorol., 8, 38–55, 2007.

Joyce, R., Janowiak, J., Arkin, P., and Xie, X.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, J. Hydrometeorol., 5, 487–503, 2004.

Karl, T. R. and Plummer, N.: Trends in high-frequency climate variability in the twentieth century, Nature, 377, 217–220, 1995.

Kidd, C., Levizzani, V., and Laviola, S.: Quantitative precipitation estimation from earth observation satellites, in: Rainfall: State of the Science, edited by: Testik, F. Y. and Gebremichael, M., Geophys. Monogr. Ser., 191, 127–158, doi:10.1029/2009GM000920, 2010.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K.-L., and Braithwaite, D.: Intercomparison of high-resolution precipitation products over Northwest Europe, J. Hydrometeorol., 13, 67–83, 2012.

Leeper, R. D., Rennie, J., and Palecki, M. A.: Observational Perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and Precipitation Comparison, J. Atmos. Ocean. Tech., 32, 703–721, 2015.

Lin, Y. and Mitchell, K. E.: The NCEP stage II/IV hourly precipitation analyses: Development and applications, Preprints, Paper 1.2., 19th Conf. on Hydrology, 9–13 January, Amer. Meteor. Soc., San Diego, CA, 2005.

Liu, C. and Zipser, E. J.: Diurnal cycles of precipitation, clouds, and lightning in the tropics from 9 years of TRMM observations, Geophys. Res. Lett., 35, L04819, doi:10.1029/2007GL032437, 2008

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge, and water quality, Hydrol. Process., 26, 4078–4111, 2012.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An overview of the global historical climatology network-daily database, J. Atmos. Ocean. Tech., 29, 897–910, 2012.

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., and Lane, J. E.: Precipitation: measurement, remote sensing, climatology and modeling, Atmos. Res., 94, 512–533, 2009.

Nelson, B. R., Kim, D., and Seo, D.-J.: Multisensor precipitation reanalysis, J. Hydrometeorol., 11, 666–682, 2010.

Nelson, B. R., Prat, O. P., Seo, D.-J., and Habib, E.: Assessment and implications of NCEP Stage IV quantitative precipitation estimates, Weather Forecast., in revision, 2015.

Nesbitt, S. W. and Anders, A. M.: Very high-resolution precipitation climatologies from the Tropical Rainfall Measuring Mission precipitation radar, Geophys. Res. Lett., 36, L15815, doi:10.1029/2009GL038026, 2009.

Nesbitt, S. W. and Zipser, E. D.: The diurnal cycle of rainfall and convective intensity according to three years of TRMM measurements, J. Climate, 16, 1456–1475, 2003.

Ochoa, A., Pineda, L., Crespo, P., and Willems, P.: Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific–Andean region of Ecuador and Peru, Hydrol. Earth Syst. Sci., 18, 3179–3193, doi:10.5194/hess-18-3179-2014, 2014.

Prat, O. P. and Barros, A. P.: Assessing satellite-based precipitation estimates in the Southern Appalachian mountains using rain gauges and TRMM PR, Adv. Geosci., 25, 143–153, doi:10.5194/adgeo-25-143-2010, 2010a.

Prat, O. P. and Barros, A. P.: Ground observations to characterize the spatial gradients and vertical structure of orographic precipitation – Experiments in the inner region of the Great Smoky Mountains, J. Hydrol., 391, 141–156 doi:10.1016/j.jhydrol.2010.07.013, 2010b.

Prat, O. P. and Nelson, B. R.: Precipitation contribution of tropical cyclones in the Southeastern United States from 1998 to 2009 using TRMM precipitation data, J. Climate, 26, 1047–1062, 2013a.

Prat, O. P. and Nelson, B. R.: Mapping the world's tropical cyclone rainfall contribution over land using the TRMM Multi-satellite Precipitation Analysis, Water Resour. Res., 49, 7236–7254, doi:10.1002/wrcr.20527, 2013b.

Prat, O. P. and Nelson, B. R.: Characteristics of annual, seasonal, and diurnal precipitation in the Southeastern United States derived from long-term remotely sensed data, Atmos. Res., 144, 4–20, 2014.

PRISM Technical Note: Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States, Tech. Note, 14 pp., http://www.prism.oregonstate.edu/ (last access: 1 October 2014), 2013.

Reed, S. and Maidment, D.: A GIS procedure for merging NEXRAD precipitation data and digital elevation models to determine rainfall–runoff modeling parameters, Online Report 95-3, Center for Research in Water Resources (CRWR), University of Texas at Austin, 1995.

Reed, S. and Maidment, D.: Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling, J. Hydrol. Eng., 4, 174–182, 1999.

Sahany, S., Venugopal, V., and Nanjundiah, R.: Diurnal-scale signatures of monsoon rainfall over the Indian region from TRMM satellite observations, J. Geophys. Res., 115, D02103, doi:10.1029/2009JD012644, 2010.

Sapiano, M. R. P. and Arkin, P. A.: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data, J. Hydrometeorol., 10, 149–166, 2009.

Schneider, U., Becker, A., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: Global Precipitation Analysis Products of the GPCC, Global Precipitation Climatology Centre, DWD, 12 pp., available at: ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf (last access: 28 January 2015), 2010.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC Monitoring Product: Near Real-Time Monthly Land-Surface Precipitation from Rain-Gauges based on SYNOP and CLIMAT data, Global Precipitation Climatology Centre at Deutscher Wetterdienst, Germany, doi:10.5676/DWD_GPCC/MP_M_V4_100, 2011.

Sevruk, B., Ondras, M., and Chvila, B.: The WMO precipitation measurement intercomparison, Atmos. Res., 192, 376–380, 2009.

Shepherd, J. M., Grundstein, A., and Mote, T. L.: Quantifying the contribution of tropical cyclones to extreme rainfall along the coastal southeastern United States, Geophys. Res. Lett., 34, L23810, doi:10.1029/2007GL031694, 2007.

Sorooshian, S., Hsu, K., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of the PERSIANN system satellite-based estimates of tropical rainfall, B. Am. Meteorol. Soc., 81, 2035–2046, 2000.

Sorooshian, S., Gao, S., Hsu, K., Maddox, R. A., Hong, Y., Gupta, H. V., and Imam, B.: Diurnal variability of tropical rainfall retrieved from combined GEOS and TRMM satellite information, J. Climate, 15, 983–1001, 2002.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., and de Castro, M.: Global precipitation measurement: methods, datasets and applications, Atmos. Res., 104–105, 70–97, 2012.

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Arthur, A., Van Cooten, S., Kelleher, K., Kitzmiller, D., Ding, F., Seo, D.-J., Wells, E., and Dempsey, C.: National mosaic and multi-sensor QPE (NMQ) system, B. Am. Meteorol. Soc., 92, 1321–1338, 2011.