How Well the Early 2017 California Atmospheric River Precipitation Events Were Captured by Satellite Products and Ground-based Radars?

Yixin Wen[1]*, Ali Behrangi[1], Haonan Chen[2,3], Bjorn Lambrigtsen[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

[2]Colorado State University, Fort Collins, CO, USA

[3]NOAA/Earth System Research Laboratory, Boulder, CO, USA

Revised Manuscript Submitted to *Quarterly Journal of the Royal Meteorological Society*

"Advances in Remote Sensing of Rainfall and Snowfall" Special Collection

September 2017

*Corresponding author address: Yixin Wen, 4800 Oak Grove Drive, Pasadena, CA 91109
Email: Yixin.Wen@jpl.nasa.gov

ABSTRACT

In January and February of 2017, California experienced multiple heavy storms that caused serious destruction of facilities and economic loss, although it also helped to reduce water storage deficit due to prolonged drought in previous years. These extreme precipitation events were mainly associated with Atmospheric Rivers (ARs) and brought about 174 km^3 of water to California according to ground observations. This paper evaluates the performance of six commonly used satellite-based precipitation products (IMERG, 3B42RT, PERSIANN, CCS, CMORPH, and GSMaP), as well as ground-based radar products (Radar-only and Radar-lgc) in capturing the ARs precipitation rate and distribution. It is found that precipitation maps from all products present heavy precipitation in January and February, with more consistent observations over ocean than land. Though large uncertainties exist in quantitative precipitation estimation (QPE) over land, the ensemble mean of different remote sensing precipitation products over California is consistent with gauge measurements. Among the six satellite-based products, IMERG correlates the best with gauge observations both in the detection and quantification of precipitation, but it is not the best product in terms of root mean square error (RMSE) or bias. Compared to satellite products, ground weather radar shows better precipitation detectability and estimation skill. However, neither radar nor satellite QPE products have good performances in quantifying the peak precipitation intensity during the extreme events, suggesting that further advancement in quantification of extremely intense precipitation associated with AR in the Western United States is needed.

KEYWORDS: Atmospheric River, QPE, water resources, remote sensing, satellite, ground weather radar, extreme events

Remote Sensing of 2017 California Atmospheric River Precipitation

1. Introduction

In January and February of 2017, excessive precipitation with local amounts exceeding 1000 mm, fell in Pacific coast and Western United States. The extreme precipitation alleviated ongoing drought conditions in California, but also produced catastrophic flooding and landslides in the Bay Area, wrecked Oroville Dam's spillway, and closed Interstate 80 in the Sierra Nevada under record-breaking blizzards (Taylor, 2017). These extreme precipitation events were predominantly fueled by long and narrow channels of large integrated water vapor transport commonly referred to as Atmospheric Rivers (ARs) (Zhu and Newell, 1994; Ralph et al., 2006).

ARs generally start from mid-latitude oceanic regions, but it can stall as they move onshore, leading to prolonged rainfall and flooding. While ARs occur globally, their impacts are most significant when they make landfall and interact with the topography (Gimeno et al. 2014). A number of studies have examined the importance of ARs in producing flooding in California and other western states (e.g., Ralph et al., 2004, 2006). It is concluded that ARs, combined with orographic enhancement and intense wind are key factors determining the extent of heavy rain and flood (Ralph et al., 2003; Waliser et al., 2017). In addition, Neiman et al. (2008) used a combination of satellite and ground based data and showed that, at least in California, ARs produce twice as much precipitation as all other storms. Guan et al. (2010, 2013) cited about 40% of the annual snow accumulation in California's Sierra Nevada was during ARs over the period of water years 2004-2010. Therefore, accurate measurement of extreme precipitation associated with ARs at a range of spatial and temporal resolutions is invaluable for a variety of scientific applications, ranging from real-time flood forecast to the evaluation of regional application of weather and water models.

For such large-scale atmospheric systems, only remote sensors can provide good coverage of comprehensive precipitation observations at relatively high spatiotemporal resolution. Over ocean, and before ARs make landfall, satellite can be used to retrieve precipitation associated with AR systems. After ARs move onshore, ground weather radar network provides intensive observations of ARs. However, accurate measurement of precipitation from ARs over western U.S. remains challenging due to complex precipitation microphysics caused by land-ocean interaction in the coastal zone and complex terrains in the mountainous region. First, the lower tropospheric air temperature during AR is warmer than other winter storms. In addition, once such systems make landfall over the mountainous west, they generate substantial orographic precipitation. Moreover, a large fraction of land surface in the western U.S. during AR in winter is snow or ice. These features make it difficult for spaceborne passive microwave (PMW) or infrared (IR) sensors, or even the active ground-based radars to estimate precipitation. IR-based techniques are indirect and incline to underestimate heavy precipitation from shallow clouds and false detect precipitation over ice and snow surface (Kidd et al., 2003; Behrangi et al., 2009). PMW-based retrieval has better physics than IR method. At low frequency band, PMW sensors are able to sense the thermal emission of rain, whereas at higher frequency band the PMW sensors can detect scattering properties of ice particles in the precipitation layer and on tops of convective systems. However, PMW-based techniques also have difficulties in capturing warm rains (Neiman et al. 2005). In addition, the ice and snow surface adds more uncertainties to PMW-based precipitation retrievals. Ground weather radar dramatically increases the ability of observing precipitation in high space and time resolutions through measuring reflectivity from reflected precipitation echoes. The ice and snow surface does not affect radar precipitation measurements. Despite these advances, reliable ground radar based precipitation measurements are difficult to obtain over mountainous regions, due to the uncertainties associated with empirical $Z-R$ relations and inadequate coverage induced by terrain blockages (Maddox et al. 2002; Willie et al. 2017). In order to quantify the uncertainties in observing AR extreme precipitations, the gauge measurements, which are relatively dense over the U.S., can be used to assess the weaknesses and strengths of various space and ground radar quantitative precipitation estimation (QPE) products. The insights gained from these analyses can help algorithm developers design more robust retrieval methods. Furthermore, it can provide users with a better quantitative understanding of the range of uncertainties that the current remote sensing products offer. While the outcome of the present study over the Western US may not be directly transferable to many other regions of the world, it can provide an overall insight on the range of uncertainties that one may expect over similar conditions.

Behrangi et al. (2016) investigated a broader set of ARs over western United States using various space-based precipitation products. The objective of this paper is to assess the performance of several popular multi-satellite precipitation products and ground radar network precipitation products in capturing extreme precipitation brought by ARs in January and February 2017 at finer temporal resolution (3-hourly scale). This study was inspired by Behrangi et al. (2016) and also motivated by a series of natural disaster related events (e.g., the epic 2017 California floods and mudslides) caused by excessive precipitation brought by ARs. Accurate estimation of rainfall during such extreme events is critical for California water resource management and flood protection (Cifelli et al. 2017). This study is also motivated by the recent development of satellite based precipitation retrievals. Two new products included in this study, namely, IMERG and GSMaP are probably the two most used products in the GPM era. Hence, we take this opportunity to explore if the current remote sensing technology would have improved the near-real-time QPE in challenging circumstances (i.e., extreme events over complex terrain, coastal region, complex cloud microphysics processes, and cold season when snowfall and snow on the surface add other dimensions to the listed challenges). The remainder of this paper is organized as follows. In section 2 we describe the datasets used in the study. Section 3 presents differences among different remote sensing precipitation products. Section 4 investigates the performance of various precipitation products focusing on one extreme event, and the paper is concluded in Section 5.

2. Precipitation dataset, study area and methods

In this study, AR events were first defined by the Integrated Water Vapor (IWV) from NASA Modern-Era Retrospective Analysis for Research Application, version 2 (MERRA-2) (Gelaro et al., 2017) greater than 20mm. Consequently, all precipitation occurring in this area in January and February of 2017.

a. Satellite Products

A number of multi-satellite precipitation products have been developed and available to the public, such as 1) Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG) (Huffman and Bolvin, 2015; Huffman et al, 2016), 2) the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42 real-time, version 7 (3B42RT; Huffman et al. 2007), 3) Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Sorooshian et al., 2000), 4) PERSIANN-Cloud Classification System (CCS) (Hong et al., 2004), 5) CPC MORPHing technique (CMORPH) (Joyce et al., 2004; Xie et al., 2017), and 6) Global Satellite Mapping of Precipitation (GSMaP) (Kubota et al., 2007). All these products provide near-real-time (NRT) precipitation estimates without infusion of rain gauge information, as well as post-real-time (PRT) estimates with gauge adjustment. PRT products are more accurate than NRT products in most cases (e.g., Behrangi et al., 2011), but the latency of PRT products generally is up to months. For real-time flood warning operations, NRT is the only choice to provide timely information. Therefore, it is important to understand how well NRT products can capture the rainstorms in January and February of 2017 in California.

The IMERG Late run product is designed to combine data from all satellites in the Global Precipitation Measurement (GPM) constellation. The new version V04A of IMERG Late run is used in this paper. TRMM 3B42RT combines various PMW-derived precipitation estimates with PMW-calibrated IR-based estimates, and relies on climatology for bias adjustment. GSMaP takes advantage of precipitation retrievals from TRMM and other low Earth orbit satellites and interpolates them via spatial propagation information obtained from IR data. The PERSIANN and CCS derive precipitation purely from a single IR channel (~11 um). PERSIANN is a pixel-based approach and PMW precipitation retrievals are used to update the parameters that relate IR to precipitation intensity, whereas CCS is a patch-based approach in which the relation between IR and precipitation rate is established for each class of cloud patches. CMORPH produces a temporally and spatially complete precipitation field by morphing the PMW precipitation data using motion vectors derived from geostationary satellite IR data. In another word, CMORPH uses precipitation estimates exclusively from PMW retrievals. The spatial and temporal resolution of different remote sensing precipitation products are summarized in Table 1. For the sake of evaluation, we mapped all the products onto common 0.25° × 0.25° spatial and 3-h temporal resolution grids in this study. For all products with original spatial resolution higher than 0.25° × 0.25°, the remapping was performed by averaging all fine resolution grids that fall inside the grid with coarser resolutions. The degradation of hourly products to 3-h temporal resolution was achieved by averaging all the times steps falling within the coarser time window. It is important to note that satellite and surface instruments measure fundamentally different quantities. The remote sensing observations, the TRMM for example, measures the volume-integrated MW emission within the instrument's instantaneous field of view, while the gauge measurements can only represent one location point. Interpolating to the common grids cannot solve the mismatch among different products with different resolutions.

b. Multi-Radar/Multi-Sensor (MRMS) precipitation products

MRMS is a quantitative precipitation estimation (QPE) system integrating radar, rain gauge, and numerical weather prediction data. It generates automated, seamless national 3D radar mosaic and multisensory precipitation estimates (Zhang et al. 2016). Currently, MRMS mainly produces four types of QPE products: 1) radar-based QPE, 2) gauge-based QPE, 3) local gauge bias-corrected radar QPE, and 4) gauge-and-precipitation-climatology-merged QPE. In this study, we used the first three QPE products since the last one is still under test.

Radar-based QPE (hereafter referred to as Radar-only) is derived using different empirical $Z-R$ relationships for different surface precipitation types, such as warm or cold stratiform rain, convective rain, tropical-stratiform or tropical-convective rain mix (Zhang et al. 2016). Polarimetric variables are not used in the operational version because various polarimetric radar QPE schemes are still under evaluation across CONUS. Radar QPE provides a high-resolution and rapid update of spatial precipitation distributions, but also carries uncertainties in the estimates because of imperfect empirical relationships between radar reflectivity and precipitation rate, as well as discrepancies between radar measurements aloft and rainfall near the ground caused by precipitation changes in the vertical. Radar QPE is challenging over this study domain characterized by complex terrain, but it is commonly used to validate satellite precipitation products (e.g., Kirstetter et al. 2012).

The gauge-only QPE (hereafter referred to Gauge) products archived in MRMS system are interpolated based on hourly rainfall records mainly from approximately 7,000 rain gauges from the Hydrometeorological Automated Data System (HADS; https://hads.ncep.noaa.gov/index.shtml). These gauge data are quality controlled through an automated scheme that compares each gauge report with collocated hourly radar QPE values. Gauge measurements outside a predefined range around the hourly radar QPE are filtered out as bad data. The quality-controlled gauge data are then interpolated onto MRMS 0.01° × 0.01° grid via an inverse-distance-weighting (IDW) scheme. In January and February over high elevation region, the phase of precipitation is mainly snow, which is unable to be measured by tipping-bucket gauges. Fortunately, some HADS gauges installed in California are with antifreeze solution or with heater, which are capable of measuring frozen precipitation.

The local gauge bias-corrected radar QPE (hereafter referred to Radar-lgc) adopts a bias correction method described in Ware (2015). First, the hourly rainfall differences between radar and gauges at each gauge station are interpolated onto MRMS grids via an IDW scheme. The interpolated difference field is then subtracted from the hourly radar QPE field. Generally, the local gauge bias correction provides consistent improvements over the radar-only QPE (Zhang et al., 2016). But a study presented by Willie et al (2017), who evaluated various MRMS products at different time scales over this region, showed that although radar rainfall performance would be enhanced after VPR and gauge correction, the improvement was not really significant. Note that the Radar-lgc product is not purely independent from Gauge product in this study. The comparisons between radar-only and gauge adjusted products can help understand the dependencies in the validation analysis.

c. Study region

The study region (Fig. 1j) includes the Eastern North Pacific Ocean before AR landfall and mountainous areas with high elevations that can induce orographic enhancement. ARs can be observed before, during, and after they make landfall and hit mountainous areas with high elevations. For comparison of different precipitation products, precipitation total, intensity and distribution are investigated both over ocean and over land. Over ocean, validation of satellite product is challenging due to the lack of in situ measurement, so cross validation of satellite QPE products among themselves is conducted. Over land, because the western United States is fairly well instrumented with rain gauges, the evaluation of various remote sensors' capability of observing extreme precipitation is conducted in a smaller region (see red box in Fig. 1j) in California using MRMS gauge QPE product as reference. This area was selected because 1) it received the largest volume of precipitation in early 2017 compared to other areas; 2) the excessive precipitation had significant impacts on the State of California, including catastrophic flooding and landslides, also alleviating drought conditions.

d. Verification statistics

Two major aspects to address for the difference between remote sensing QPE products and gauge measurements are (1) the capability to detect precipitation and (2) the accuracy in quantifying precipitation rate. Simple contingency table statistics are applied to answer the first question. The contingency table statistics describing the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) are used to evaluate remote sensing QPE products. These indexes are computed based on the number of hits (H), false alarms (F), and misses (M).

$$\mathrm{POD} = \mathrm{H}/(\mathrm{H}+\mathrm{M}) \tag{1a}$$

$$\mathrm{CSI} = \mathrm{H}/(\mathrm{H}+\mathrm{F}+\mathrm{M}) \tag{1b}$$

$$\mathrm{FAR} = \mathrm{F}/(\mathrm{H}+\mathrm{F}) \tag{1c}$$

To answer the second question, four statistical indices for evaluating remote sensing QPE products are selected. The Relative Bias (RB) is used to assess the systematic bias of products. Spearman's rank correlation coefficient (CC) is used to assess the agreement between remote sensing products and gauge observations. The mean absolute error (MAE) measures the average magnitude of the error while the root-mean-squared error (RMSE) weights more to larger errors.

$$RB = \frac{\sum P(i) - \sum G(i)}{\sum G(i)} \times 100\%, \tag{2a}$$

$$CC = 1 - \frac{6\sum\left(Rank_{P(i)} - Rank_{G(i)}\right)^2}{N(N^2-1)}, \tag{2b}$$

$$MAE = \frac{\sum|P(i) - G(i)|}{N}, and \tag{2c}$$

$$RMSE = \sqrt{\frac{\sum(P(i) - G(i))^2}{N}} \tag{2d}$$

Here, $P(i)$ and $G(i)$ represent the i^{th} matching pair of rainfall amounts estimated by remote sensing products and observed by gauges, respectively. And N represents the total number of matching pairs. In (2b), $Rank_{P(i)}$ and $Rank_{G(i)}$ represent the assigned rank value in the ascending order of the remote sensing products and gauge observations, respectively. Only data pairs with nonzero values from both gauge and remote sensing sources are considered as the four indices are focused on quantitative measurement rather than detection.

3. Uncertainty of remote sensing QPE products

Figure 1(a-h) shows maps of total accumulation of AR-driven precipitation estimated by satellites and ground radars in January and February of 2017. Gauge data from MRMS (Fig. 1h) are used as a reference for total precipitation comparison over land. Corresponding digital elevation map is shown in Fig. 1j. Because there is no gauge measurement of precipitation, the relative performance of satellite QPE products is not evaluated over ocean. Fig. 1a-f show that satellite products are fairly consistent in capturing the precipitation pattern over ocean, except large discrepancy to the west of southern British Columbia (Lat: 50°, Long: -130°), where the IR-based products (i.e., PERSIANN and CCS) show much lower precipitation than other products in this region. IR-based QPE retrieval algorithms are mainly based on the general assumption that colder or higher clouds statistically produce more intense rainfall, so they are prone to underestimate heavy precipitation from ARs with the bulk of the water vapor flux generally below 850 hPa (Rahph et al., 2005). Over land, gauge product shows the highest precipitation occurred over Sierra Nevada, likely due to the orographic lifting of precipitation on the windward side of the mountains. Among all the eight remote-sensing QPE products, GSMaP generates the highest precipitation amount and largest precipitation area over and in the vicinity of the Sierra Nevada. CMORPH, purely based on PMW sensors, barely captures precipitation over snow and frozen lands, which is in agreement with the study presented in Behrangi et al. (2016).

In capturing the precipitation observed by Gauge, CCS captures the orographic precipitation pattern over Sierra Nevada but underestimates the precipitation amount compared to gauge measurements. Also, CCS shows the largest precipitation coverage overall. Especially in the interior western United States region, CCS presents the highest precipitation among all products including MRMS local gauge bias-corrected radar QPE and gauge QPE products. The MRMS Radar Only product shows underestimation compared to gauge measurements. The artifact precipitation circles in the precipitation map indicate the radar beam might be too high and sampling the mixed-phase precipitation above the freezing level. As expected, Radar-lgc product is fairly consistent with gauge-based product since it is calibrated with gauge measurements.

Figure 2 shows the ensemble mean and standard deviation of monthly precipitation accumulations calculated from five satellite products, IMERG, GSMaP, PERSIANN, CCS, and CMORPH. Consistent with that observed in Figure 1, the ensemble means of monthly precipitation (Fig. 2a and 2b) show two precipitation centers, one is over ocean, and the other is along the Sierra Nevada. However, compared to the monthly precipitation derived from MRMS gauge product (Fig. 2e and 2f), the ensemble mean of space-based products shows severe underestimation. Figure 2c and 2d show higher agreement of different products over ocean (low standard deviation) but lower agreement over land (high standard deviation), especially in the Sierra Nevada region.

Fig. 3 (a) and (b) present the histograms of precipitation intensity of different products over ocean and land, respectively. The precipitation intensity investigated here are all greater than zero. Due to the lack of ground-radar and rain gauge over ocean, only precipitation derived from satellite observations are available to show the fraction of total precipitation over ocean. Fig. 3a shows that the largest fractions of precipitation volume of 3B42RT, CMORPH, CCS, and GSMaP are all located around 1 mm hr^{-1}. IMERG has a wider distribution compared to other products. GPM Level 3 data, IMERG, has a better detectability of light precipitation, which may explain the higher fraction of intensity between 0.01 mm hr^{-1} to 0.2 mm hr^{-1}. Compared to

IMERG, 3B42RT has limitations to observe light precipitation below 0.2 mm hr^{-1}. PERSIANN and CCS are both IR-based QPE products, but their histogram curves are different. CCS adopts a patch-based approach in which the relation between IR and precipitation rate is established for different classes of cloud patch. As a result, the precipitation histogram of CCS is wider than PERSIANN, with higher fractions in both very light precipitation (below 0.03 mm hr^{-1}) and moderate/heavy precipitation (higher than 1 mm hr^{-1}). Over land, both satellite and radar products are shown in Fig. 3b with Gauge as the reference. For a fair comparison, CMORPH is excluded in this figure, because it has extremely low detectability over snow and ice surfaces. Compared to gauge measurements, 3B42RT, CCS and GSMaP place more fraction of precipitation in the high intensity range (greater than 2 mm hr^{-1}), which is consistent with previous studies. For example, Behrangi et al (2016) found CCS placed a significant fraction of precipitation in the mid-intensity range between 4 and 40 mm day^{-1} by analyzing 10-yr AR landfalling data. Tang et al (2017) found 3B42RT and GSMaP have overestimation of severe storm precipitation in the summer 2016 in South China by comparing to merged gauge QPE product. Histogram of precipitation intensity measured by ground radar is close to that measured by gauge but misses up to 50% of heavy rain with intensity greater than 2 mm hr^{-1}. There are two reasons. First, in the relatively flat environment, the radar beams are too high to sample the surface rainfall (Wen et al., 2013), causing large errors in surface rainfall estimation because of the vertical variations of reflectivity. On the other hand, in the mountainous area, it is challenging to find appropriate Z-R relations suitable to the complex rainfall processes including bright band rain with robust ice processes and subsequent melting and non-bright band rain dominated by collision and coalescence below the melting level resulting from orographic enhancement (Martner et al., 2008). After bias-calibrated by gauges, Radar-lgc shifts some extent to the gauge.

4. Case study perspective

Because of the important impacts of ARs in California, we chose a research region (shown in Fig.1j) focusing on California and look into the event with the heaviest rainfall to further investigate the performance of various remote sensing QPE products. The study area, as indicated in Figure 1j, is about 300,000 km^2, including the mountains and the Central Valley area.

Figure 4 shows time series of area-averaged precipitation rate over the study area measured by gauge. The precipitation was mainly brought by four major AR events significantly contributing to the annual water resources in California. The local maximum rainfall intensity reached 10.67 mm hr^{-1} (3-hr average) occurred at 6:00UTC, January 9 based on ground stations. Table 2 presents the total volume of precipitation brought by the four AR events to California measured by the ground gauges and remote sensors. According to the gauge products, Event 1 (Jan. 2 to Jan. 13) brings the largest precipitation volume (60 km^3) within 12 days, more than the monthly average flow of liquid water (45 km^3 per month) at the mouth of the Mississippi River (Syed et al., 2005). The water transported by the four studied AR events to California ranges from 36 km^3 to 60 km^3, totally ~174 km^3 water precipitating in California in January and February of 2017. Precipitation estimated from space and ground radar has underestimation or overestimation issues compared to that measured by gauges. However, ensemble mean of precipitation amount calculated from IMERG, 3B42, PERSIANN, CCS, GSMaP and Radar-only shows reasonable performance, covering the gauge measurements within one stand deviation (STD) range for all four events. Further investigation of remote sensing capability of capturing extreme precipitation event is conducted on Event 1, the heaviest precipitation case of the four events.

Fig. 5 presents the time series of area-averaged precipitation rate over the study area generated from eight remote sensing QPE products along with the gauge data as the reference. CMORPH has severe underestimation all the time because of PMW sensors' limitation over snow and frozen surface as discussed in previous sessions. 3B42RT has almost 3-times overestimation compared to gauge measurements on the first precipitation peak on January 8. Not only the 3B42RT has overestimation issues, but also GSMaP (two-times higher), CCS and PERSIANN are all higher than the gauge measurements of the precipitation on January 8. However, for the lower precipitation peak occurred on January 11, only GSMaP overestimates the precipitation, while other products have underestimation. IMERG generally reports less precipitation compared to that measured by gauge. Ground weather radar captures the temporal pattern of precipitation, but underestimates the total precipitation amount. Radar-lgc is consistent with gauge measurements simply because this product is corrected by quality-controlled gauge measurements.

Figure 6 shows POD, FAR and CSI as a function of 3-hourly precipitation intensity measured by gauge from Event 1. The POD generally shows a trend of improving values with increasing precipitation intensity for all products. This improvement indicates that remote sensing products likely miss the light precipitation. 3B42RT has the lowest POD at all precipitation intensities. IMERG and GSMaP show better performance of POD than others. For FAR, all products generally show decreasing trends. Note that for the precipitation intensity below 0.2 mm hr^{-1}, the high FAR values exhibited by all products are probably related to rain gauge sensitivity, which sets to approximately 0.2 mm/tip. Overall, PERERSIANN and CCS have higher FAR than other products for moderate precipitation. For heavy precipitation with intensity greater than 5 mm hr^{-1}, FARs of GSMaP and 3B42RT are the largest. IMERG shows the best FAR for precipitation intensity > 0.4 mm hr^{-1} among all satellite products. The CSI value gives a comprehensive evaluation of detectability of remote sensing products. The radar-based products have highest CSIs. The difference between radar-only and radar-lgc is negligible. Among all satellite products, IMERG shows the best CSI scores.

Table 3 presents the metrics calculated based on event scale for Event 1 to assess the accuracy of remote sensing products in quantifying precipitation rate. IMERG performs much better than other satellite QPE products in terms of CC. The RB of IMERG is down to -45.84%, while the CCS, 3B42RT, and PERSIANN are -13.12%, 28.31%, and -28.45%, respectively. However, it is worth noting that the MAE of IMERG is lower than CCS, 3B42RT, and PERSIANN, suggesting that better RB of these three products maybe resulted from the cancellation of positive and negative biases. It should be noted that we also investigated the version 3 IMERG (not shown) and it shows anomalously high RB, MAE, and RMSE, likely due to an error in the V03 algorithm. It is encouraging that V04 IMERG has resolved the problem and appears as one of the best among the other products. The statistical scores of radar-only product are slightly better than IMERG. The comparisons of all events are included in Table 4 and the results are similar to the ones based on Event 1 shown in Table 3.

The error of remote sensing products compared to gauge on 3-hourly scale is shown in Fig. 7. Figure 7 shows the median, the quartile, and the range (10% and 90%) of the error distribution based on Event 1 and all the four events. Only data pairs with nonzero values from both gauge and remote sensing sources are considered since the boxplots in Fig. 7 is focused on quantitative measurement rather than detection. Fig. 7 shows that the performances of all remote sensing products are fairly good in terms of the median value of the error. GSMaP has overestimations, while other products show underestimations. Radar product, as the ground based active MW sensor, shows the best performance in terms of the median value and the range among all remote sensing products as we expect. For satellite products, the medians of PERSIANN and CCS are closer to zero than other ones. However, the range of PERSIANN and CCS are larger than IMERG. The 3B42RT shows the largest positive error and largest range of error.

The intense rainfall could trigger floods, thus accurate estimation of extreme rainfall is always critical for flood monitoring, forecasting, and migrating (Gourley et al. 2017). Figure 8 shows the error distribution of remote sensing products focusing on heavy precipitation observations. The threshold to select heavy precipitation measurements is 3-hour-intensity greater than 5 mm hr^{-1}. Compared to Fig. 7, the performances of all products degrade. CMORPH has the largest negative error. CCS, PERSIANN, IMERG, and Radar Only have severe underestimation as well. The medians of 3B42RT and GSMaP are closer to zero than other products, which can be attributed to the severe overestimation in the precipitation peak time (indicated in Fig. 5).

The evaluation metrics of remote sensing QPE products at 3-hourly scale for Event 1 are shown in Table 5 and for all the four events are shown in Table 6. However, we want to note that the results based on all four events are similar to those for the first event (with largest precipitation).For all measurements including precipitation with different intensities (left part of the table), Radar-only, GSMaP and IMERG perform better than other products in terms of CC. Radar-only and IMERG are better than GSMaP in terms of RB, MAE and RMSE. As a whole, the performances of Radar-only and IMERG are the best considering the four metrics used in this study, which is consistent with Fig. 7. However, the performances of all products are seriously degraded when statistics applied to heavy precipitation only (right part in Table 5). The CC of Radar-Only is 0.24 and the best CCs of satellite products is 0.17 from 3B42RT. The RBs of all products are also deteriorated from the whole dataset to heavy precipitation. Table 5 Table 6 are consistent with Fig. 7 and Fig. 8, indicating that severe limitations exist in remote sensing products in extreme heavy precipitation estimation, though the performance is reasonable when considering the whole events.

5. Summary and conclusions

ARs are critical to the regional climate, hydrology, water resources, and socioeconomics in the semiarid western United States. The ARs precipitation events in January and February of 2017 brought totally 174 km^3 of water to California, alleviated drought but also caused floods and landslides. Accurate measurement of extreme precipitation associated with ARs is critical for flood/landslide forecasting and water resources management. The Western U. S. is fairly well instrumented and thus provides a good testbed to assess the performance of the remote sensing products under various challenging conditions such as extreme rain and snow events, orographic precipitation, and precipitation on frozen surfaces. This study assesses six satellite-based near real-time precipitation products (IMERG, 3B42RT, PERSIANN, CCS, CMORPH, and GSMaP) and two ground radar-based precipitation products (MRMS Radar-only and Radar-lgc) in capturing AR's precipitation rate and distribution, especially in extreme events. The main results are summarized as follows:

1) The precipitation map from gauge shows more than 1000 mm precipitation occurred over and in the Sierra Nevada in two months. All satellite QPE products except GSMaP underestimate the heavy precipitation. IMERG, 3B42RT, PERSIANN, and CCS are able to

capture the heavy rain pattern over the Sierra Nevada, but underestimate the total precipitation amount by -40%, -50%, -60% and -30%, respectively, compared to gauge measurements. In terms of the statistical performance over land, IMERG correlates the best with gauge observations both in the detection and quantification of precipitation, but it does not yield the best RB and RMSE. CMORPH misses the most precipitation over snow and ice surface.

2) Over ocean, different satellite products show similar precipitation patterns, except in the area close to the west of southern British Columbia where the precipitation is captured by IMERG, CMORPH, and GSMaP, but missed by IR methods (PERSIANN and CCS).

3) 3B42RT has the lowest POD at all precipitation intensities. At the same time, 3B42RT overestimates precipitation significantly at peak intensity. Both PERSIANN and CCS have false alarm issues with precipitation detection. GSMaP has fairly good detectability skill but tends to have false alarm issues at heavy precipitation. IMERG shows better performance than others in terms of POD and FAR, hence yields the best CSI. The significant improvement of IMERG compared to 3B42 is particularly encouraging.

4) The histograms of precipitation intensity show that the largest fraction of precipitation volume of 3B42RT, CMORPH, CCS, and GSMaP are all located around 1 mm hr^{-1}. IMERG has a wider distribution due to its better detectability of light precipitation. In the high intensity range (greater than 2 mm hr^{-1}), 3B42RT, CCS and GSMaP place more fraction of precipitation compared to gauge, while IMERG and PERSIANN have lower fraction.

5) Compared to satellite products, ground weather radar shows better performance in precipitation detection and estimation. However, accurate radar QPE over western U.S. remains challenging due to complex precipitation microphysics in this mountainous region. Radar shows totally 38% underestimation of rainfall compared to gauge and prone to underestimate the heavy

precipitation with intensity greater than 2 mm hr^{-1}. Radar-lgc is consistent with gauge measurements because it is bias corrected using gauge measurements.

6) For extremely heavy precipitation (3-hourly precipitation rate > 5 mm hr^{-1}), none of the products show good performance in quantifying the precipitation intensity.

The insights gained from these analyses can help algorithm developers design more robust retrieval methods. Furthermore, it can provide users with a better quantitative understanding of the range of uncertainties that the current remote sensing products offer. While the outcome of the present study over the Western US may not be directly transferable to many other regions of the world, it can provide an overall insight on the range of uncertainties that one may expect over similar conditions.

Further analysis is needed to investigate different phases of AR precipitation. The same amount of water with liquid or solid phase would have significantly different impacts on hydrological cycle, hazard forecast, and water resources management. In current study, rain gauges cannot provide snowfall information. Snow Telemetry (SNOTEL) measures snowfall over the western U. S. and thus provides an opportunity to assess remote sensing snowfall products (Serreze et al. 1999; Wen et al. 2017). Efforts are underway to investigate the liquid/frozen ratio of AR precipitation and assess the performance of commonly used remote sensing QPE products in separating solid and liquid precipitations.

Acknowledgements

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The study is partially supported by the NASA Energy and Water Cycle Study awards (NNH13ZDA001N-NEWS) and NASA WEATHER (NNH13ZDA001N-WEATHER) awards. Government sponsorship is acknowledged.

REFERENCES

Behrangi A, Hsu K, Imam B, Sorooshian S, Huffman GJ, Kuligowski RJ, 2009. PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis. *J. Hydrometeor*. **10**: 1414-1429. doi:10.1175/2009JHM1139.1.

Behrangi A, Khakbaz B, Jaw TC, AghaKouchak A, Hsu K, Sorooshian S. 2011. Hydrologic evaluation of satellite precipitation products over a mid-size basin. *J. Hydr.* **397**: 225-237

Behrangi A, Guan B, Neiman P, Schreier M, Lambrigtsen B. 2016. On the quantification of atmospheric rivers precipitation from space: composite assessments and case studies over the Eastern North Pacific Ocean and the Western United States. *J. Hydrometeor*. **17**: 369-382.

Cifelli, R, Chandrasekar V, Chen H, Lynn J. 2017. High Resolution Radar Quantitative Precipitation Estimation in the San Francisco Bay Area: Rainfall Monitoring for the Urban Environment. *Journal of the Meteorological Society of Japan*. (in press)

Gelaro, R, McCarty W, Suárez M J, Todling R, et al. 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate. 30: 5419-5454.

Gourley J, Flamig Z, Vergara H, Kirstetter PK, Clark R, Argyle E, Arthur A, Martinaitis S, Terti G, Erlingis J, Hong Y, Howard K. 2017. The Flooded Locations and Simulated Hydrographs (FLASH) project: improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc*. (In press)

Guan B, Molotch NP, Waliser DE, Fetzer EJ, Neiman PJ, 2010. Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophys. Res. Lett.* **37,** L20401, doi:10.1029/2010GL044696.

Guan B, Molotch NP, Waliser DE, Fetzer EJ, Neiman PJ. 2013. The 2010/2011 snow season in California's Sierra Nevada: Role of atmospheric rivers and modes of large-scale variability. *Water Resour. Res.* **49:** 6731–6743, doi:10.1002/wrcr.20537

Hong Y, Hsu KL, Sorooshian S, Gao XG. 2004. Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteor.* **43**: 1834–1852, doi:10.1175/JAM2173.1

Huffman GJ, Bolvin DT. 2015. TRMM and Other Data Precipitation Data Set Documentation, Mesoscale Atmospheric Processes Laboratory, NASA Global Change Master Directory Doc. 44 pp. http://pmm.nasa.gov/sites/default/files/document_files/3B42_3B43_doc_V7.pdf.

Huffman GJ, Bolvin DT, Braithwaite D, Hsu K, Joyce R, Xie P. 2016. Integrated MultisatellitE Retrievals for GPM (IMERG), version 4.5. NASA's Precipitation Processing Center, accessed July, 2016, ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/

Joyce RJ, Janowiak JE, Arkin PA, Xie P, 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.* **5**: 487–503. doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Kidd C, Kniveton DR, Todd MC, Bellerby TJ. 2003. Satellite rainfall estimation using combined passive microwave and infrared algorithms. *J. Hydrometeor*. **4:** 1088-1104, doi:10.1175/1525-7541(2003)004<1088:SREUCP>2.0.CO;2.

Kirstetter P, Hong Y, Gourley J J, Chen S, Flamig Z, Zhang J, Schwaller M, Petersen W, Amitai E. 2012. Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar–Based National Mosaic QPE. *J. Hydrometeor*. **13**: 1285-1300.

Kubota T, et al. 2007. Global precipitation map using satellite-borne microwave radiometers by the GSMaP Project: Production and validation. *IEEE T. Geoscl. Remote*. **45:** 2259-2275

Maddox R, Zhang J, Gourley JJ, Howard K. 2002. Weather radar coverage over the contiguous United States. *Wea. And Forecast*. **17:** 927-934.

Martner, BE, Yuter SE, White AB, Matrosov SY, Kingsmill DE, Ralph FM, 2008. Raindrop Size Distributions and Rain Characteristics in California Coastal Rainfall for Periods with and without a Radar Bright Band. *J. Hydrometeor.*, **9**, 408-425.

Neiman PJ, Wick GA, Ralph FM, Martner BE, White AB, Kingsmill DE. 2005. Wintertime nonbrightband rain in California and Oregon during CALJET and PACJET: Geographic, interannual, and synoptic variability. *Mon. Wea. Rev*. **133**: 1199–1223, doi:10.1175/MWR2919.1.

Neiman PJ, Ralph FM, Wick GA, Lundquist JD, Dettinger MD. 2008. Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the west coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeor*. **9**: 22–47, doi:10.1175/2007JHM855.1

Ralph FM, Neiman PJ, Kingsmill DE, Persson PO, White AB, Strem ET, Andrews ED, Antweiler RC, 2003. The Impact of a Prominent Rain Shadow on Flooding in California's Santa Cruz Mountains: A CALJET Case Study and Sensitivity to the ENSO Cycle. *J. Hydrometeor.*, **4**, 1243-1264.

Ralph FM, Neiman PJ, Wick GA. 2004. Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Wea. Rev.* **132:** 1721–1745, doi:10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2.

Ralph FM, Neiman PJ, Wick GA, Gutman SI, Dettinger MD, Cayan DR, White AB. 2006. Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.* **33:** L13801, doi:10.1029/2006GL026689

Serreze MC, Clark MP, Armstrong RL, McGinnis DA, Pulwarty RS. 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOWTEL) data. *Water Resour. Res.* **35:** 2145–2160.

Sorooshian S, Hsu KL, Gao X, Gupta HV, Imam B, Braithwaite D. 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.* **81**: 2035–2046, doi:10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2.

Syed TH, Famiglietti JS, Chen J, Rodell M, Seneviratne SI, Viterbo P, Wilson CR. 2005. Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance, *Geophys. Res. Lett.* **32:** L24404, doi:10.1029/2005GL024851

Tang G, Zeng Z, Ma M, Liu R, Wen Y, Hong Y. 2017. Can near-real-time satellite precipitation products capture rainstorms and guide flood warning for the 2016 summer in South China? *IEEE Geosci. and Res. Lett*. (In press)

Taylor M. 2017. Managing Floods in California. Legislative Analyst's Office Report. Available at: http://www.lao.ca.gov/reports/2017/3571/managing-floods-032217.pdf

Waliser D, Guan B. 2017. Extreme winds and precipitation during landfall of atmospheric rivers. *Nature Geosci*. **10:** 179-183. doi:10.1038/ngeo2894

Ware EC. 2005. Corrections to radar-estimated precipitation using observed rain gauge data. M.S. thesis, Cornell University, 96 pp.

Wen Y, Cao Q, Kirstetter PE, Hong Y, Gourley JJ, Zhang J, Zhang G, Yong B. 2013. Incorporating NASA Spaceborne Radar Data into NOAA National Mosaic QPE System for Improved Precipitation Measurement: A Physically Based VPR Identification and Enhancement Method. *J. Hydrometeor*. **14:** 1293-1307.

WenY, Kirstetter P, Gourley JJ, Hong Y, Behrangi A, Flamig Z, 2017. Evaluation of MRMS Snowfall Products over the Western United States. *J. Hydrometeor.* **18**:1707–1713

Willie D., 2017: Evaluation of multisensor quantitative precipitation estimation in Russian river basin. *J Hydrol Eng.* 22(5), E5016002. doi:10.1061/(ASCE)HE.1943-5584.0001422

Xie P, Joyce R, Wu S, Yoo S, Yarosh Y, Sun F, Lin R. 2017. Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998. *J. Hydrometeor*. **18**: 1617-1641.

Zhang J, and Coauthors. 2016. Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc*. doi: 10.1175/BAMS-D-14-00174.1

Zhu Y, Newell RE. 1994. Atmospheric rivers and bombs. *Geophys. Res. Lett*. **21**: 1999–2002, doi:10.1029/94GL01710.

List of Tables

Table 1. A summary of QPE products used in this study.

Data	Spatial resolution	Temporal resolution	Data source
IMERG	0.1°×0.1°	0.5-hour	ftp://jsimpson.pps.eosdis.nasa.gov
3B42RT	0.25°×0.25°	3-hour	https://mirador.gsfc.nasa.gov
PERSIANN	0.25°×0.25°	1-hour	ftp://persiann.eng.uci.edu/CHRSdata/PERSIANN
CCS	0.04°×0.04°	0.5-hour	ftp://persiann.eng.uci.edu/CHRSdata/PERSIANN-CCS
CMORPH	0.25°×0.25°	1-hour	https://rda.ucar.edu/datasets
GSMaP	0.1°×0.1°	1-hour	ftp://rainmap@hokusai.eorc.jaxa.jp
MRMS Radaronly	0.01°×0.01°	2-minute	http://mrms.ncep.noaa.gov/data
MRMS Radar-lgc	0.01°×0.01°	1-hour	http://mrms.ncep.noaa.gov/data
MRMS Gauge	0.01°×0.01°	1-hour	http://mrms.ncep.noaa.gov/data

Table 2. Total amount of precipitation integrated over California for 4 major AR events in January and February of 2017. The unit is km^3.

	Event 1 1/2 - 1/13	Event 2 1/17 - 1/25	Event 3 2/2 - 2/13	Event 4 2/15 - 2/23	January	February	Total
IMERG	30.60	16.74	27.72	23.03	47.33	50.74	98.08
3B42RT	54.99	7.51	30.31	10.50	62.50	40.80	103.31
PERSIANN	33.02	13.83	18.94	20.19	46.85	39.13	85.98
CCS	46.25	26.26	31.55	32.90	72.51	64.46	136.97
CMORPH	8.43	5.91	8.87	9.20	14.34	18.07	32.418
GSMaP	69.03	62.88	64.98	65.25	131.91	130.23	262.14
Radar-only	37.81	28.15	24.70	28.01	65.96	52.71	118.67
Radar-lgc	61.04	36.95	41.27	37.70	97.99	78.98	176.97
gauge	60.64	35.58	41.11	36.54	96.22	77.65	173.87
Ensemble mean* ±STD	44.65 ± 19.10	25.98 ± 17.87	32.16 ± 15.96	29.26 ± 16.99	70.62 ± 34.51	61.42 ± 32.22	132.04 ± 66.07

*Note: The ensemble mean is calculated using IMERG, 3B42RT, PERSIANN, CCS, GSMaP and Radar-only. CMORPH and Radar-lgc are not included. CMORPH has missing data over snow and frozen surfaces. Radar-lgc is not independent of gauge data.

Table 3. Metrics of remote sensing precipitation products at event scale for Event 1 over study region.

Product	CC	RB (%)	MAE	RMSE
IMERG	0.82	-45.84	17.78	25.37
3B42RT	0.69	28.31	18.93	23.48
PERSIANN	0.63	-28.45	19.37	26.02
CCS	0.51	-13.12	17.82	22.96
CMORPH	0.47	-84.72	31.91	41.45
GSMaP	0.61	58.43	26.76	34.38
Radar-only	0.88	-36.15	14.49	20.29
Radar-lgc	0.98	0.83	3.69	5.57

Table 4. Metrics of remote sensing precipitation products for total precipitation in Jan and Feb of 2017 over study region

Product	CC	RB (%)	MAE	RMSE
IMERG	0.76	-42.04	62.61	89.92
3B42RT	0.70	-27.03	63.89	81.24
PERSIANN	0.68	-41.18	76.08	100.71
CCS	0.36	-2.81	70.79	86.50
CMORPH	0.67	-83.14	101.04	132.59
GSMaP	0.71	84.13	111.28	135.92
Radar-only	0.91	-26.96	40.20	58.76
Radar-lgc	0.99	6.02	13.16	20.06

Table 5. Metrics of remote sensing precipitation products at 3-hr scale for Event 1 over study region. Extreme heavy precipitation measurements are selected under the criteria of 3-hr intensity greater than 5 mm hr^{-1}.

Product	Event 1				Extreme heavy precipitation in Event 1			
	CC	RB (%)	MAE	RMSE	CC	RB (%)	MAE	RMSE
IMERG	0.43	-42.72	1.47	2.11	0.08	-55.40	3.98	4.37
3B42RT	0.26	65.22	2.75	4.35	0.17	-10.54	3.30	4.07
PERSIANN	0.35	-31.30	1.33	1.81	0.09	-55.26	3.46	3.92
CCS	0.28	-13.41	1.50	2.08	0.11	-48.94	3.19	3.74
CMORPH	0.26	-78.38	1.90	2.51	0.11	-85.92	5.41	5.58
GSMaP	0.47	53.50	1.90	2.84	0.01	-12.01	2.82	3.53
Radar-only	0.75	-37.73	0.81	1.26	0.24	-52.16	3.25	3.53
Radar-lgc	0.93	-0.43	0.32	0.51	0.77	-1.20	0.55	0.74

Table 6. Similar to Table 5, but the results are based on precipitation from all 4 events.

Product	Event 1				Extreme heavy precipitation in Event 1			
	CC	RB (%)	MAE	RMSE	CC	RB (%)	MAE	RMSE
IMERG	0.35	-39.93	1.29	1.99	0.12	-57.25	4.46	4.96
3B42RT	0.28	25.01	1.94	3.17	0.05	-30.93	3.65	4.38
PERSIANN	0.20	-47.21	1.15	1.67	0.08	-70.30	4.46	4.88
CCS	0.15	-8.56	1.45	2.19	0.03	-66.24	4.29	4.81
CMORPH	0.27	-76.05	1.43	2.02	0.16	-84.87	5.42	5.65
GSMaP	0.39	73.99	1.93	4.04	0.09	-12.22	3.35	4.44
Radar-only	0.76	-30.48	0.61	1.01	0.22	-54.21	3.44	3.77
Radar-lgc	0.92	3.05	0.27	0.44	0.76	-1.72	0.61	0.85

List of Figures

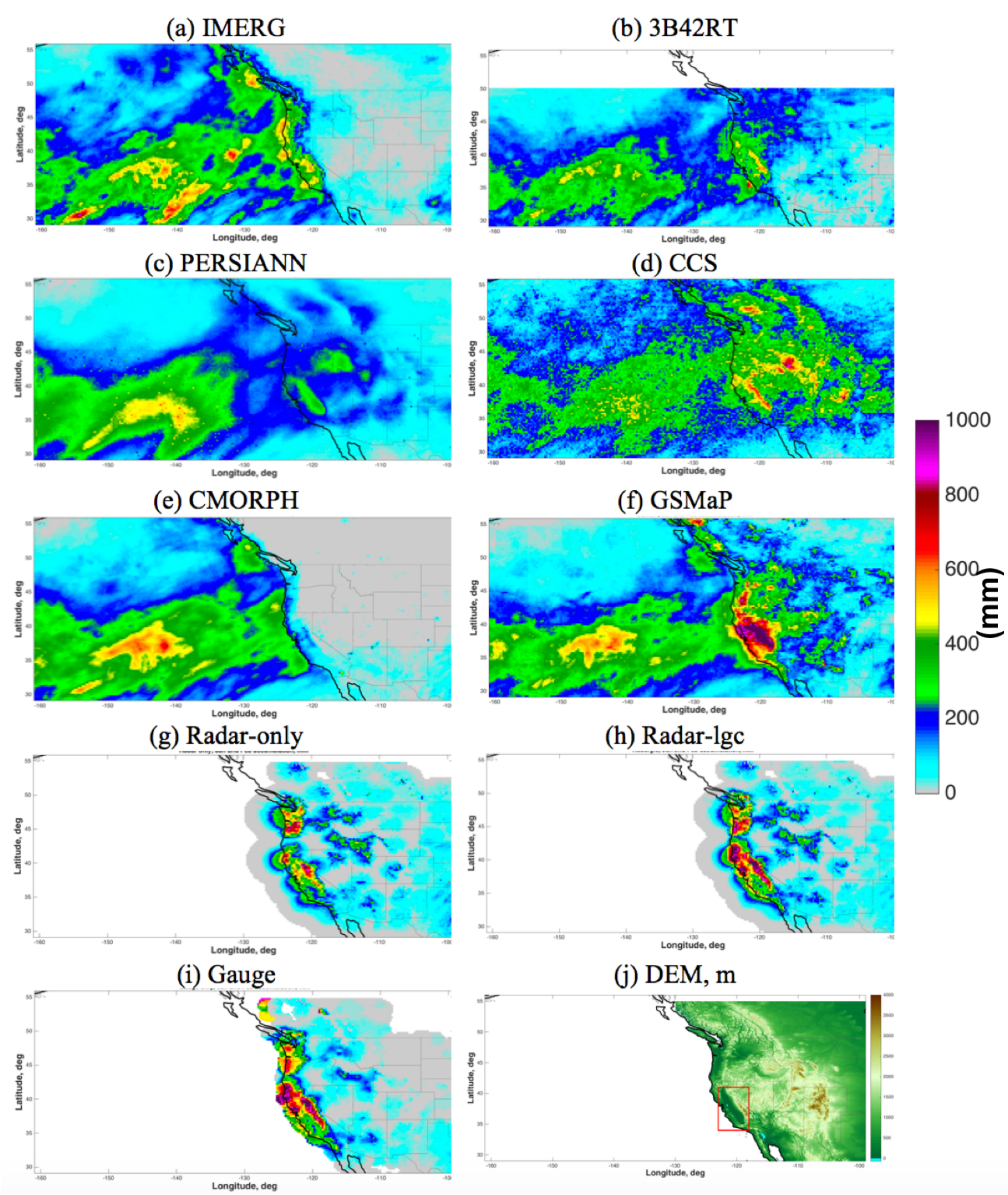
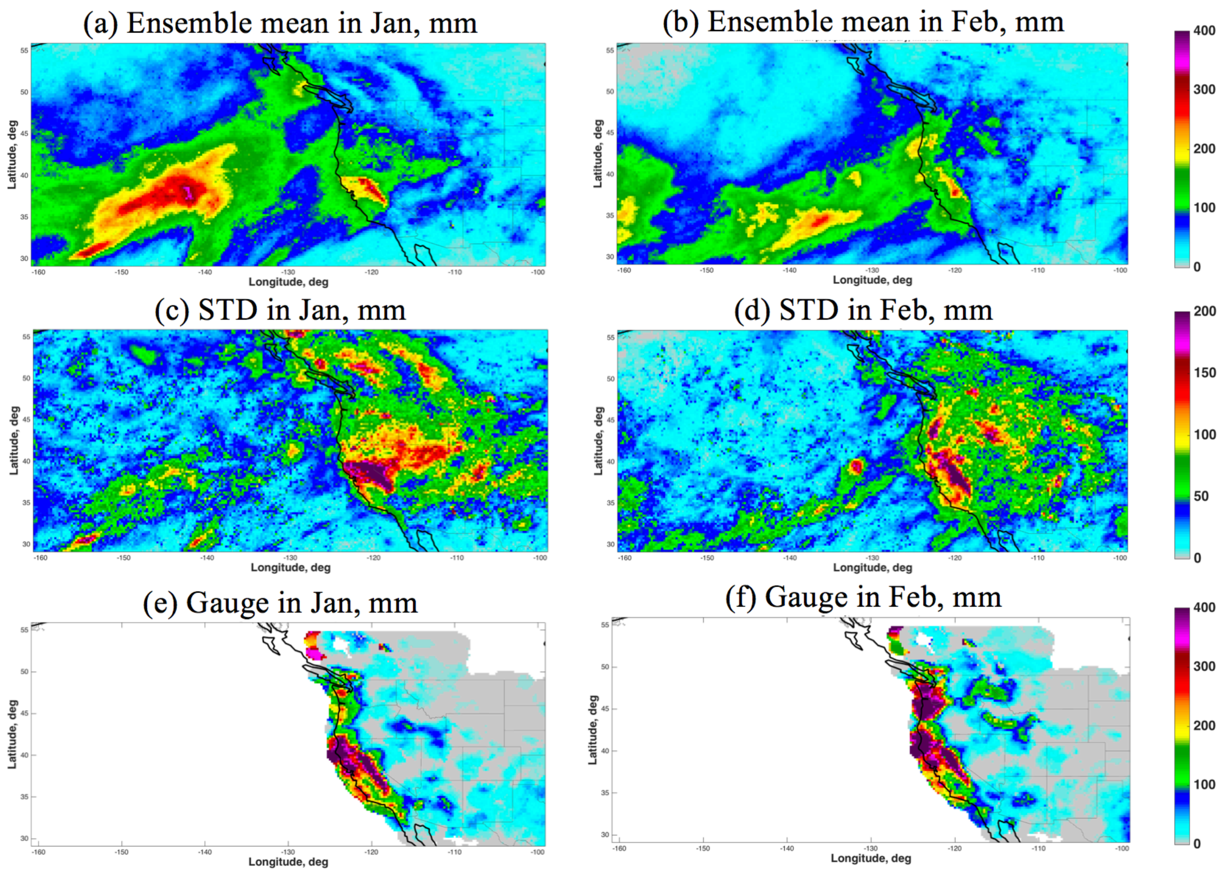


Figure 1. January and February accumulation. (a) IMERG, (b) 3B42RT, (c) PERSIANN, (d) CCS, (e) CMORPH, (f) GSMaP, (g) Radar-only, (h) Radar-lgc, (i) Gauge only, and (j) DEM. The red rectangle box in (j) DEM is the research area for case study perspective.



614 Figure 2. Ensemble Mean and STD of 6 satellite QPE products in January and February.

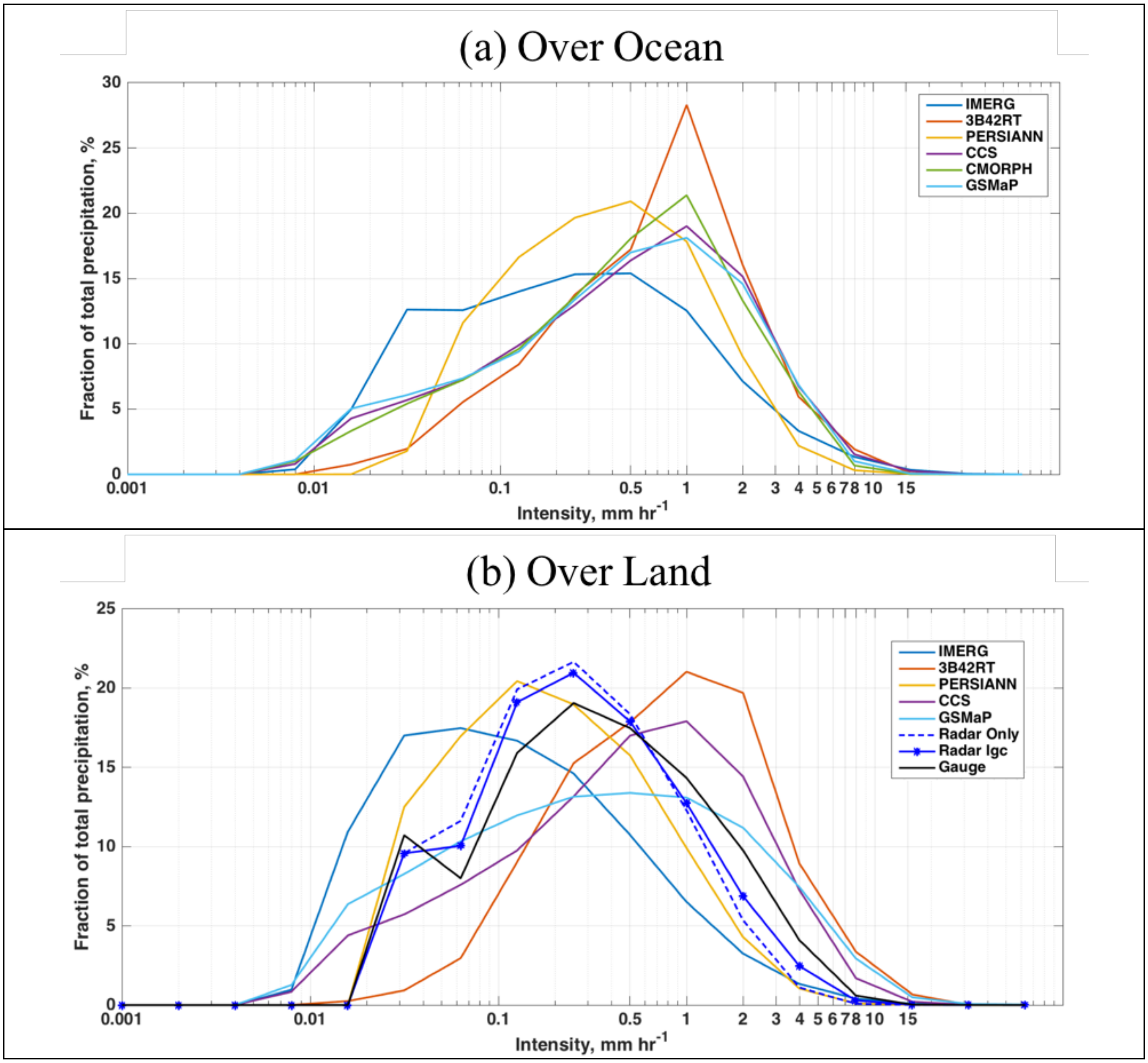


Figure 3. Histograms of remote sensing QPE products (a) over ocean; (b) over land. Histogram of gauge product is also included in (b).

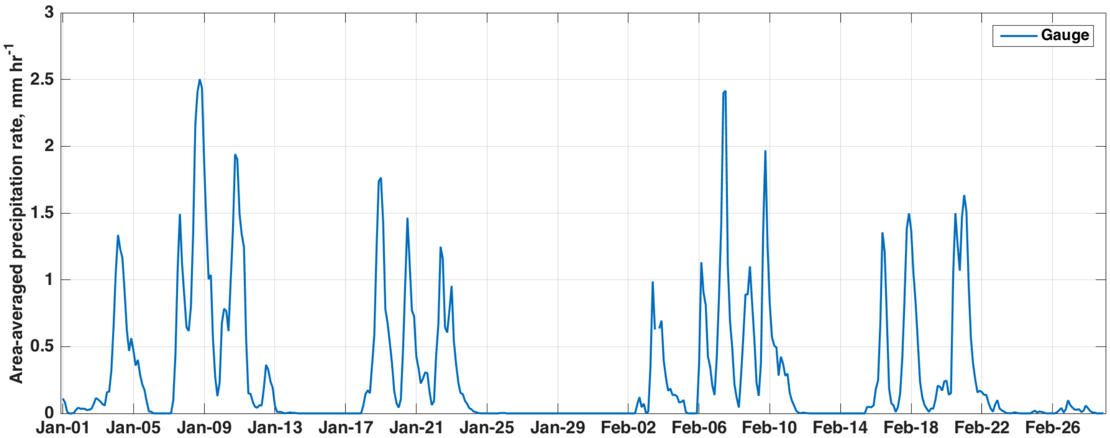


Figure 4. Time series of area-averaged precipitation rate over the study area measured by gauge.

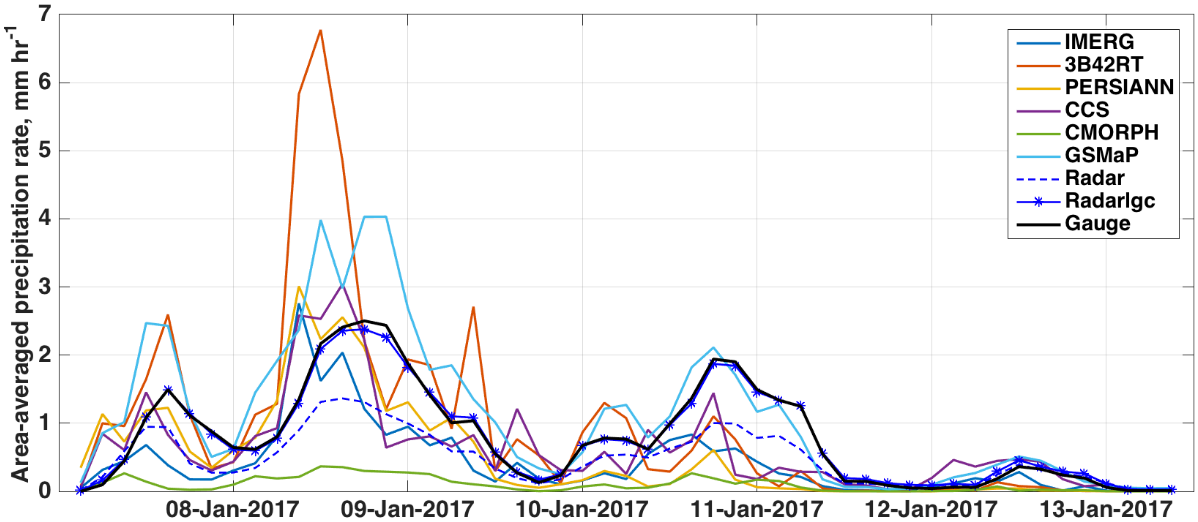
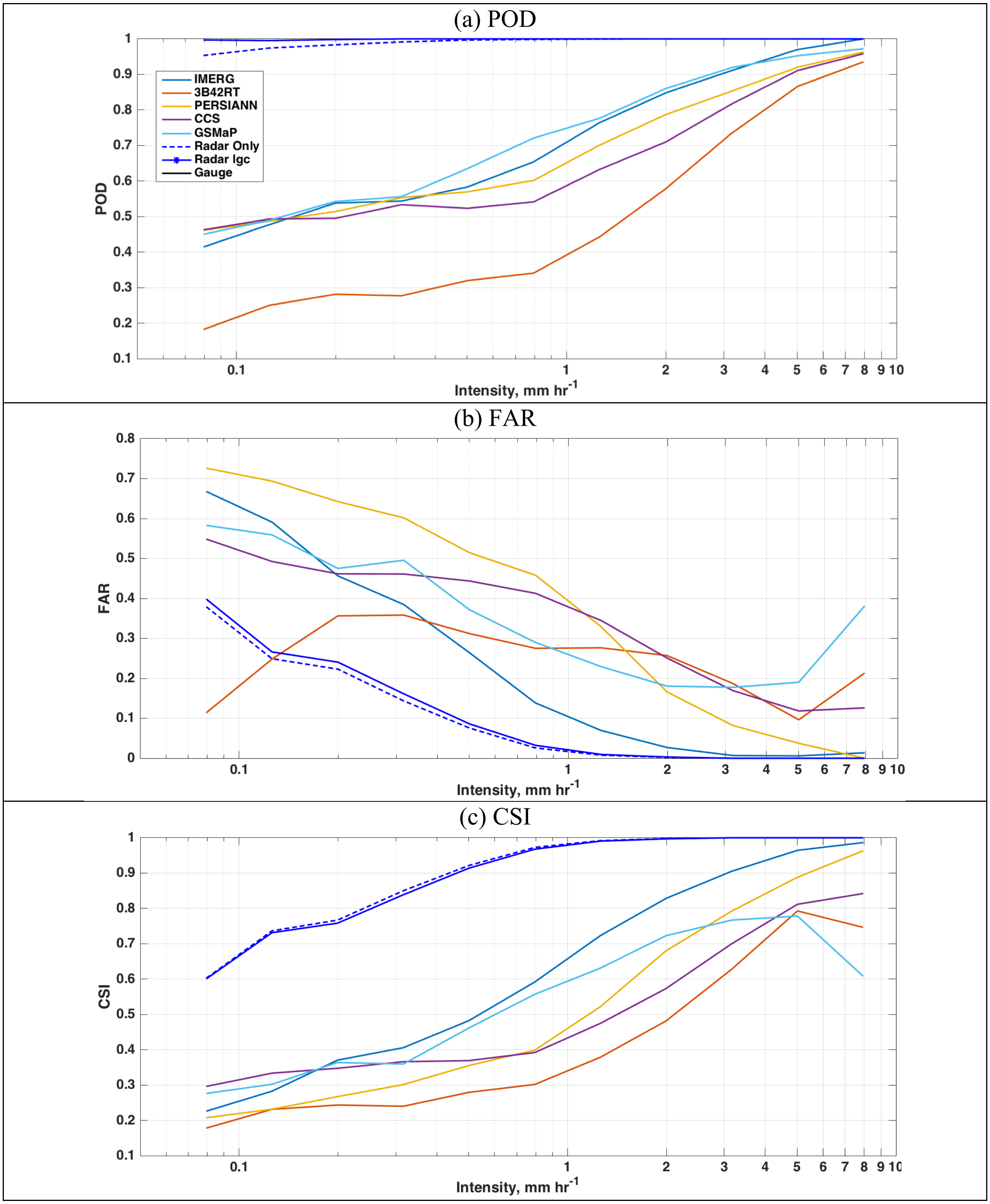


Figure 5. Time series of average precipitation rate over the study area for Event 1 generated by remote sensing QPE products.



622 Figure 6. (a) POD, (b) FAR, and (c) CSI of remote sensing products with gauge as reference.

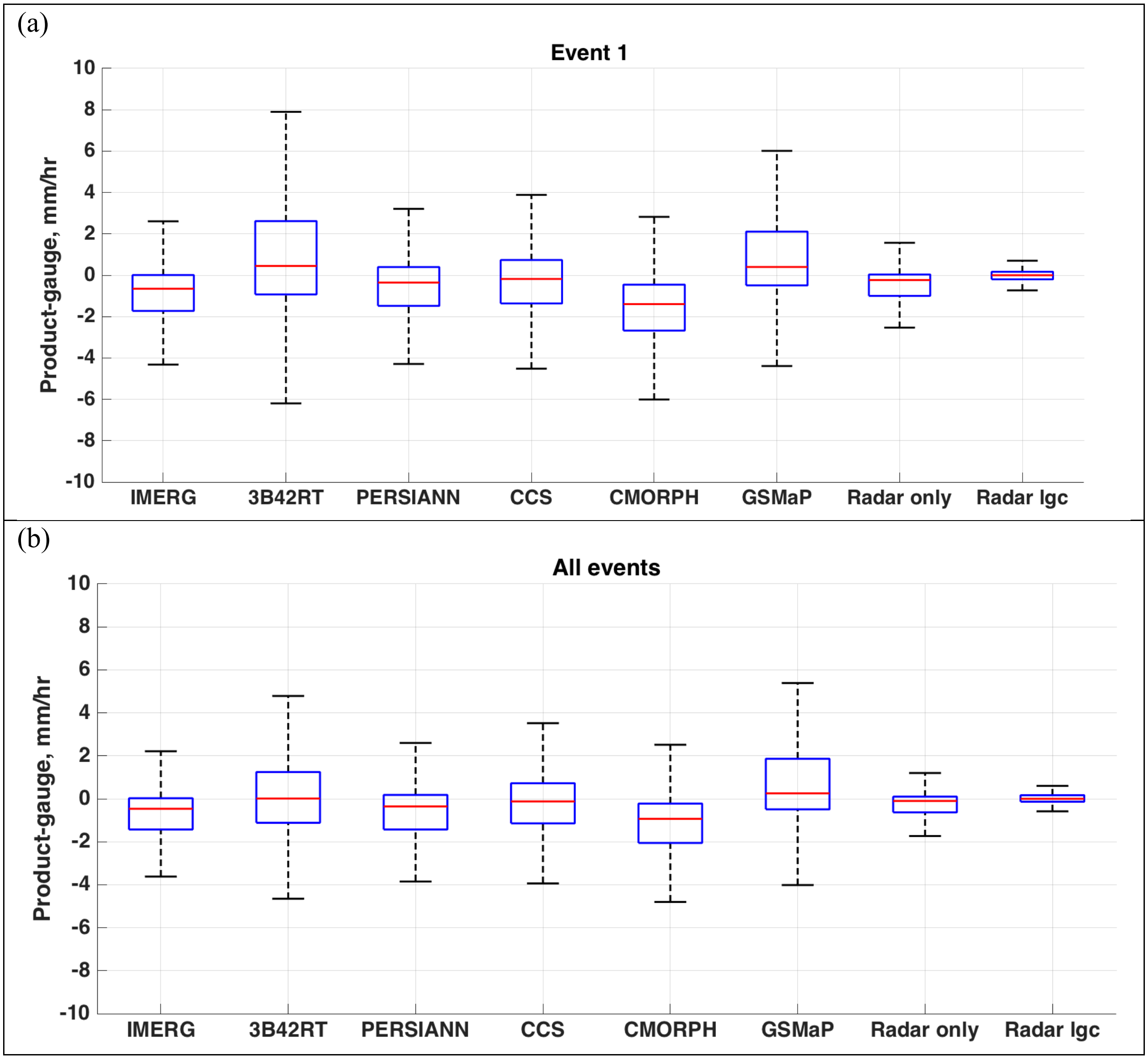


Figure 7. Error of remote sensing products compared to gauge for (a) Event 1 and (b) all four events.

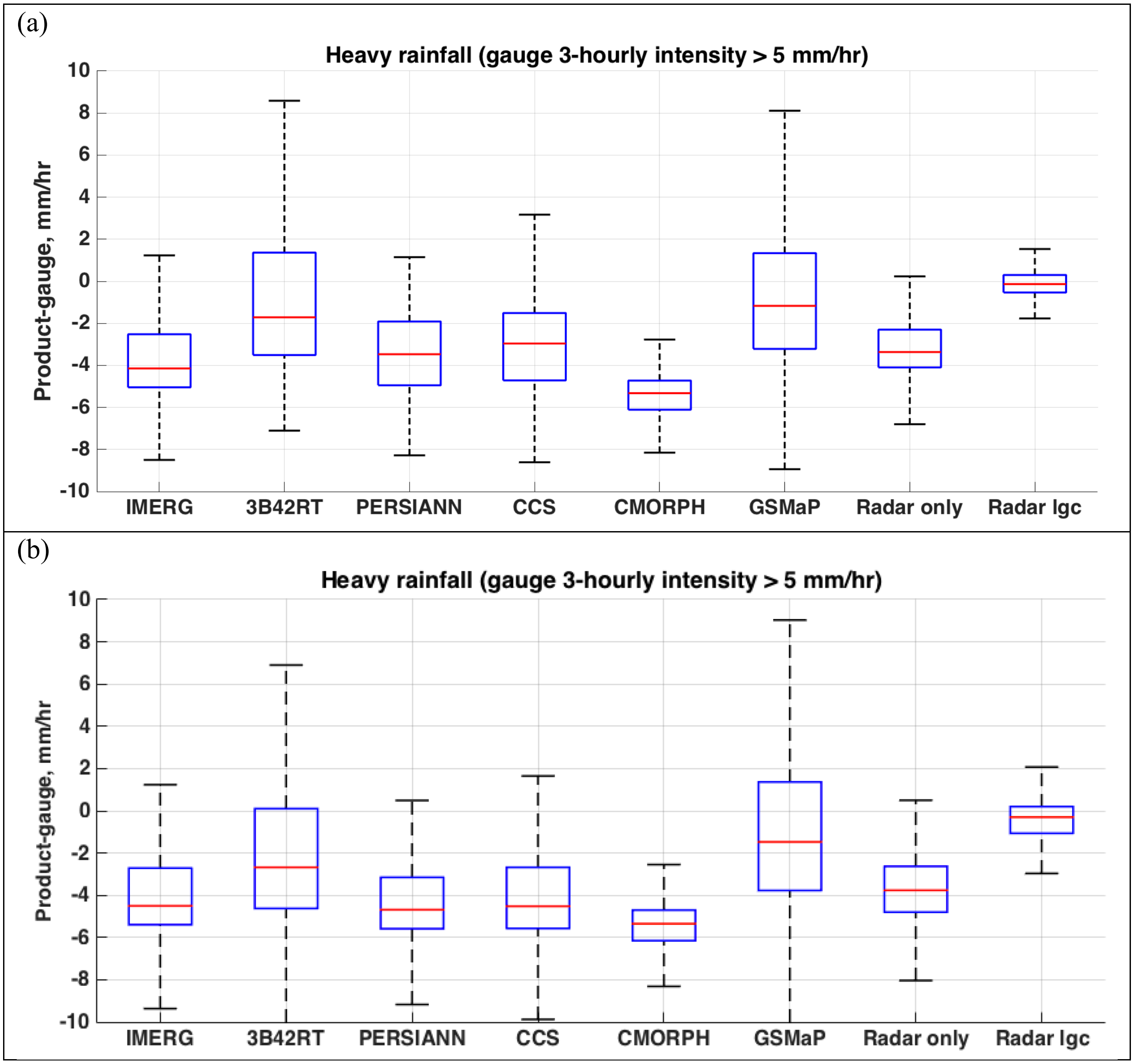


Figure 8. Similar to Figure 7, but focus on the heavy rainfall 3-hourly intensity great than 5 mm hr^{-1} measured by gauge for (a) Event 1 and (b) all four events.