

# **Evaluating the Benefits of Merging Near-Real-Time Satellite Precipitation Products: A Case Study in the Kinu Basin Region, Japan**

NIKOLAOS MASTRANTONAS

*IHE Delft Institute for Water Education, Delft, Netherlands, and Centre for Ecology & Hydrology, Wallingford, United Kingdom*

BISWA BHATTACHARYA

*IHE Delft Institute for Water Education, Delft, Netherlands*

YOSHIHIRO SHIBUO AND MOHAMED RASMY

*International Centre for Water Hazard and Risk Management, Tsukuba, Ibaraki, Japan*

GONZALO ESPINOZA-DÁVALOS

*IHE Delft Institute for Water Education, Delft, Netherlands*

DIMITRI SOLOMATINE

*IHE Delft Institute for Water Education, and Water Resources Section, Delft University of Technology, Delft, Netherlands, and Water Problems Institute of RAS, Moscow, Russia*

(Manuscript received 14 September 2018, in final form 5 April 2019)


## ABSTRACT

After the launch of the Global Precipitation Measurement (GPM) mission in 2014, many satellite precipitation products (SPPs) are available at finer spatiotemporal resolution and/or with reduced latency, potentially increasing the applicability of SPPs for near-real-time (NRT) applications. Therefore, there is a need to evaluate the NRT SPPs in the GPM era and investigate whether bias-correction techniques or merging of the individual products can increase the accuracy of these SPPs for NRT applications. This study utilizes five commonly used NRT SPPs, namely, CMOPRH RT, GSMaP NRT, IMERG EARLY, IMERG LATE, and PERSIANN-CCS. The evaluation is done for the Kinu basin region in Japan, an area that provides observed rainfall data with high accuracy in space and time. The selected bias correction techniques are the ratio bias correction and cumulative distribution function matching, while the merged products are derived with the error variance, inverse error variance weighting, and simple average merging techniques. Based on the results, all SPPs perform best for lower-intensity rainfall events and have challenges in providing accurate estimates for typhoon-induced rainfall (generally more than 50% underestimation) and at very fine temporal scales. Although the bias correction techniques successfully reduce the bias and improve the performance of the SPPs for coarse temporal scales, it is found that for shorter than 6-hourly temporal resolutions, both techniques are in general unable to bring improvements. Finally, the merging results in increased accuracy for all temporal scales, giving new perspectives in utilizing SPPs for NRT applications, such as flood and drought monitoring and early warning systems.

## 1. Introduction

Precipitation is a major component of the global water cycle and the main forcing in hydrological processes. Its accurate estimation in space and time is of immense

---

 Denotes content that is immediately available upon publication as open access.

---

Corresponding author: Nikolaos Mastrantonas, nikmas@ceh.ac.uk



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

importance for decision-making and planning for a broad range of applications. Lately, due to the limited availability of adequate ground-based observations in many areas and the advances in remote sensing, there is an increasing interest in satellite precipitation products (SPPs). These products have near-global coverage, are freely available, and provide rainfall estimates at reasonably fine spatial and temporal resolution. There exists an extensive literature related to the evaluation of these products and/or the possibility of using them in hydrological applications in different areas, as, for example, for catchments in Asia (Xue et al. 2013; Long et al. 2016; Kim et al. 2017), South America (Collischonn et al. 2008; Dinku et al. 2010), North America (Yilmaz et al. 2005), Africa (Stisen and Sandholt 2010), Australia (Woldemeskel et al. 2013), and Europe (Lo Conti et al. 2014; Duan et al. 2016).

Research has shown that SPPs come with limitations and their performance varies across different areas. For example, Hughes (2006), Brown (2006), and Asadullah et al. (2008) found that the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN), a commonly used SPP, overestimated the rainfall in South Africa, Indian subcontinent and high elevations of Uganda, whereas according to Hirpa et al. (2010) it severely underestimated the rainfall at high altitudes in Ethiopia.

Xie and Xiong (2011) opinioned that all SPPs have spatially varying, temporally changing, and range-dependent biases. Many studies applied bias correction for improving the quality of the products. Such techniques are, for example, the mean field bias (Smith and Krajewski 1991; Borga et al. 2002) and the ratio bias correction (Arias-Hidalgo et al. 2013). Probabilistic methods such as the quantile mapping are also used (Xie and Xiong 2011).

Nevertheless, after bias correction, errors have still been found. These errors are associated with limitations on the sensors of the satellites, the processing algorithms, and the selected bias-correction techniques (e.g., Madadgar et al. 2014; Xie et al. 2017). A merging of the individual SPPs though, may produce a dataset with possibly fewer errors than the original products, as it is already identified not only for SPPs (Shen et al. 2014; Khairul et al. 2018), but also rainfall estimates in general (Beck et al. 2017). There exists a variety of available techniques, each of them with its own merits (Hasan et al. 2016).

Despite the availability of near-real-time (NRT) SPPs, most of the research has focused on the gauge-corrected SPPs that have prolonged latency periods. In contrast, the information provided by NRT SPPs has latency from a couple of hours to a couple of days and can be very useful for applications such as early warning systems. NRT SPPs have already been used for

assessment of the flood extent and consequent early warning systems, as, for example, for the Beneu River in Nigeria (Haile et al. 2016), for the city of Riyadh in Saudi Arabia (Tekeli and Fouli 2016), for the Awash River in Ethiopia (Koriche and Rientjes 2016), as well as in global scale (Wu et al. 2014).

After the launch of the Global Precipitation Measurement (GPM) mission in February 2014, some of the NRT SPPs make use of the new satellites and have improved algorithms, resulting in potentially increased accuracy and greater applicability to NRT operations. Therefore, there is a need to evaluate the NRT SPPs in the GPM era, identify their errors, and quantify their accuracy at different rainfall intensities. Moreover, there is a necessity to investigate whether bias-correction techniques, and merging of the individual sets, can increase the accuracy for NRT applications and improve the rainfall estimates, as it is already found for other rainfall products (e.g., Adler et al. 1993; Mitra et al. 2003; Xie and Xiong 2011; Shen et al. 2014; Khairul et al. 2018). This is more crucial, since the majority of the published works is associated with analysis for daily up to monthly scales (Nikolopoulos et al. 2013), and thus the merits and limitations of the techniques are not well recorded for fine temporal scales. This is the main motivation of the current study.

In this work five NRT SPPs (Table 1) are evaluated: 1) Climate Prediction Center morphing technique in real time (CMORPH RT); 2) Global Satellite Mapping of Precipitation in near-real time (GSMaP NRT); 3) Integrated Multi-Satellite Retrievals for GPM (IMERG) EARLY; 4) IMERG LATE; and 5) PERSIANN-Cloud Classification System (PERSIANN-CCS). The products are referred to as CMORPH, GSMaP, EARLY, LATE, and PERSIANN respectively (the bold part of the full names on Table 1). The SPPs are evaluated through a case study in the Kinu region in Japan for the period February 2015 to December 2016, a period that is common for all selected SPPs (the period February–December 2015, is referred as year 2015 in the rest of this study).

## 2. Case study and data

### a. Study area

The Kinu basin (1761 km<sup>2</sup>) and its surrounding area that are used for this study (15116 km<sup>2</sup>) lies between 35°42' and 37°12'N and between 139°06' and 140°18'E on the Honshu island of Japan (Fig. 1). The Kinu River has a length of about 177 km and is the longest tributary of the Tone River (Niroshinie et al. 2016). The variability in the topographic features of the area results in significant differentiation of the meteorological characteristics. In

TABLE 1. Summary of the selected SPPs. The bold text indicates how these SPPs are referred to in this study.

Product and version used	Provider	Spatial/temporal resolution	Latency	Available data from	Sensors used	Methods used
<b>CMORPH</b> RT V0.X_RT_RAW	NOAA CPC	$0.0727^{\circ} \times 0.0727^{\circ}$ , 30 min	2 h	23 Jan 2015	LEO-PMW	Kalman filter
<b>GSMaP</b> NRT v6	JAXA	$0.10^{\circ} \times 0.10^{\circ}$ , 1 h	4 h	10 Oct 2008	LEO-PMW GEO-IR	Look up tables, Kalman filter
IMERG <b>EARLY</b> v04A	NASA	$0.10^{\circ} \times 0.10^{\circ}$ , 30 min	~5 h	31 Dec 2014	LEO-PMW GEO-IR	Kalman filter, intersatellite calibration, Cloud Classification System
IMERG <b>LATE</b> v04A	NASA	$0.10^{\circ} \times 0.10^{\circ}$ , 30 min	15 h	31 Dec 2014	LEO-PMW GEO-IR	Kalman filter, intersatellite calibration, Cloud Classification System
<b>PERSIANN</b> CCS	University of California	$0.04^{\circ} \times 0.04^{\circ}$ , 1 h	1 h	1 Jan 2003	GEO-IR	Artificial neural network

the headwaters of the catchment the elevation is more than 2000m and the rainfall varies between 1600 and 2100mm, while downstream the topography is quite flat with elevation below 200m and rainfall around 1400 mm (Yasuda et al. 2016). There are substantial seasonal differences in rainfall between the cloudy and rainy period of June–July (plum rain season, called baiu in Japanese; Taniguchi 2016), the wet season (July–September) when typhoons occur, and the remaining months of the year.

Factors influencing the selection of this area include the density of the available rain gauge network and the importance of the basin to the region. The river originates in the Nikko area, with its shrines and temples being listed as UNESCO heritage monuments, while the basin has a total population of about 550 000 people with many cities serving as commuter locations feeding Tokyo. Another important reason for selecting this basin is the occurrence of three severe typhoon events (Etaui, Nangka, and Midulle) within the study period that

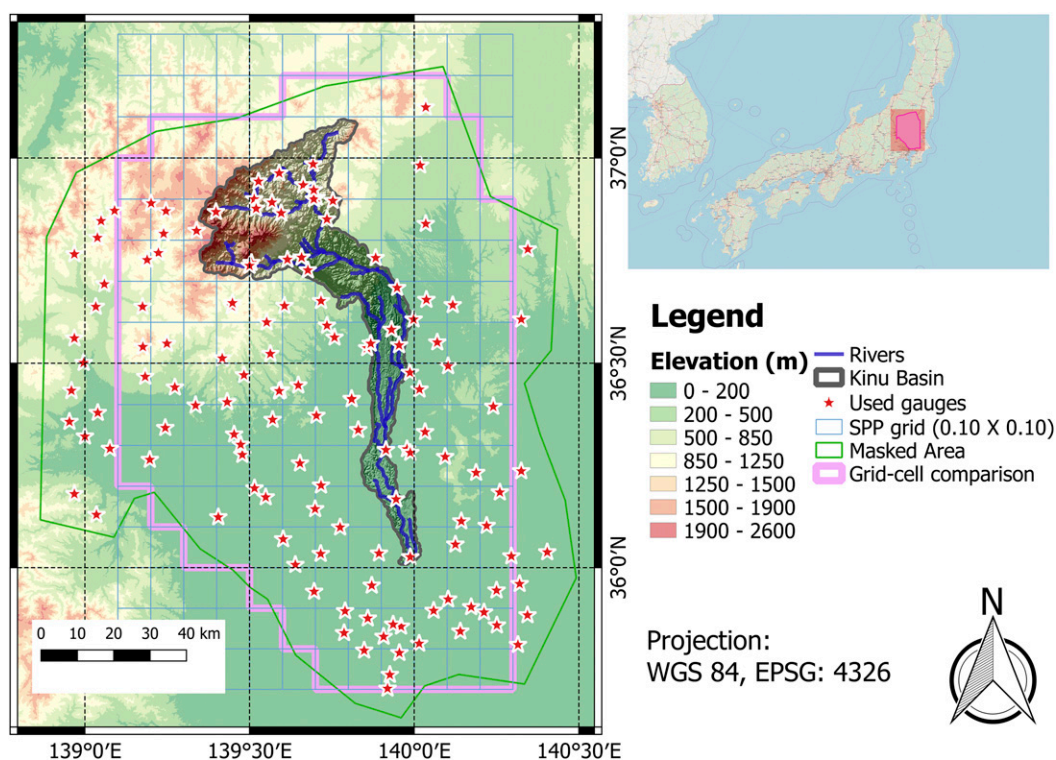


FIG. 1. Location of the Kinu catchment region, Japan, as well as the study rain gauges.

provides the opportunity to analyze the accuracy of the SPPs at high rainfall intensities. In particular, Typhoon Etau brought an accumulated rainfall of up to 600 mm locally in less than 48 h, leading to severe flooding and substantial damage, with more than 7000 affected buildings (Yasuda et al. 2016). Finally, there was an interest to explore the applicability of SPPs in small basins.

### *b. Data used*

#### 1) RAIN GAUGE OBSERVATIONS

The area has an extensive network of rain gauges operated by two authorities [Automated Meteorological Data Acquisition System (JMA 2019) and Hydrological and Water Quality Database (Water Information System 2019)] with a total of 199 gauges. Although the gauges record every 10 min, the available data are preprocessed and have hourly resolution. After the necessary quality control (maximum of 20% missing data allowed), 137 gauges are used (Fig. 1), with each point of the comparison area being less than 30 km distance from at least one rain gauge.

#### 2) SATELLITE PRECIPITATION PRODUCTS

There are two categories of meteorological satellites that provide data for rainfall estimates: the low-Earth-orbiting (LEO) satellites that mainly use passive microwave sensors (PMW) and the geostationary (GEO) satellites that mainly use infrared (IR) sensors. The SPPs use data from one or more of these types of satellites.

CMORPH-RT was developed by the National Oceanic and Atmospheric Administration (NOAA; Joyce et al. 2004). It provides 30-min mean precipitation on a  $0.0727^\circ \times 0.0727^\circ$  grid over the globe ( $60^\circ\text{S}$ – $60^\circ\text{N}$ ) with 2-h latency. The data are available since the end of January 2015, while from November 2017 onward the data come with 3 h of latency and improved algorithms. CMORPH uses the rainfall estimates derived from the LEO-PMW sensors and propagates them in time, by information provided by the GEO-IR sensors.

GSMaP NRT is provided by the Japanese Aerospace Exploration Agency (JAXA; Kubota et al. 2007). It gives precipitation information on a global longitude scale and  $60^\circ\text{N}$ – $60^\circ\text{S}$  latitude from October 2008 and has a spatial and temporal resolution of  $0.10^\circ \times 0.10^\circ$  and 1 h, respectively. The latency of the product is 4 h. It should be noted that the GSMaP NOW version (0-h latency) was undergoing an update for data correction at the time of this study, and thus it was not used. The GSMaP algorithm combines GEO-IR and LEO-PMW data and employs the Kalman filter to update the rain rate produced by the forward propagation of the precipitation obtained from the microwave sensors (Ushio et al. 2009). Version 6 of the product is used for this study.

The two IMERG products used in this study, are provided by NASA (Huffman et al. 2015, 2017), as part of the GPM mission, a joint project of NASA and JAXA. The products come at a 30-min temporal resolution and a grid of  $0.10^\circ \times 0.10^\circ$  covering the globe ( $60^\circ\text{S}$ – $60^\circ\text{N}$ ). Version 4 (V4A, latest available at the time of the analysis) is used, which provides data from the end of December 2014 onward. EARLY has  $\sim 5$ -h latency, and LATE has 15-h latency (in the latest versions EARLY and LATE have 4- and 12-h latency respectively). The products use a unified algorithm; the data are derived from a constellation of LEO satellites and are merged and intercalibrated, while data obtained from the GEO satellites are used for rainfall propagation. These products come with a range of embedded precipitation subsets: the subset used in this study is the recommended one, multisatellite precipitation estimate with gauge calibration (precipitationCal).

PERSIANN-CCS was developed at the University of California, Irvine (Hsu et al. 1997; Hong et al. 2004). It provides hourly precipitation data at  $0.04^\circ \times 0.04^\circ$  resolution grid and global coverage ( $60^\circ\text{S}$ – $60^\circ\text{N}$ ) with a latency of 1 h and temporal coverage from January 2003. An algorithm is used for extracting cloud features from GEO-IR data that are consequently used for providing rainfall estimates employing an artificial neural network.

### 3. Methodology

The methodology consists of three steps and is illustrated in Fig. 2. For the steps two and three the data are divided into calibration (year 2015) and validation (year 2016) sets.

#### *a. Step 1: Initial data processing*

The selected SPPs have different temporal and spatial resolutions. Some products have subhourly resolution and are aggregated to hourly. Moreover, when necessary, SPPs are resampled from their original resolution to  $0.10^\circ \times 0.10^\circ$ . In particular, the resampling is done on PERSIANN and CMORPH because their resolution is  $0.04^\circ \times 0.04^\circ$  and  $\sim 0.07^\circ \times 0.07^\circ$ , respectively. This resampling is done by area weighted averaging of the old grid cells (e.g.,  $0.04^\circ \times 0.04^\circ$ ) to the  $0.10^\circ \times 0.10^\circ$  resolution. The gauges are not used during the resampling of the SPPs. The resolution  $0.10^\circ \times 0.10^\circ$  is selected because it is the finest resolution at which all five SPPs could be resampled, without introducing additional uncertainty and possible errors by downscaling any of the products (Table 1). The few missing values (less than 2.3% for each grid cell for all SPPs) are infilled by linear interpolation for up to 3-hourly gaps, and with 0 mm for longer gaps, since this value corresponds to the vast majority of the rainfall data; note that the gauge-derived



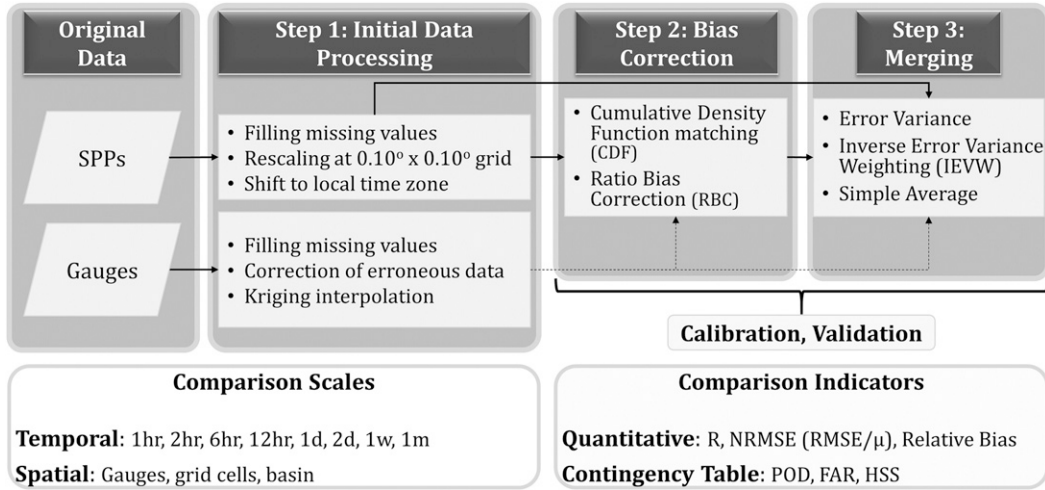


FIG. 2. Flowchart presenting this study's methodology

accumulated rainfall at the Kinu basin during the gaps of the SPPs was less than 1.5% of the total rainfall. This filling technique is selected due to the very fine temporal resolution of the SPPs. Also, spatial interpolation would introduce more uncertainty, because most of the missing values are clustered in nearby grid cells. Finally, the time zone of all SPPs is shifted from UTC to the Kinu basin's local time (UTC + 9 h).

The missing values that exist in the rain gauges measurements (the mean percentage of missing data across all gauges is less than 1%) are infilled. The inverse distance weighting (IDW) interpolation method (Xu et al. 2015) is selected because of the dense gauge network and the fine temporal resolution of the data. Finally, the ordinary kriging method is used for creating a gridded dataset from the gauge measurements (at the same  $0.10^\circ \times 0.10^\circ$  resolution grid as the SPPs).

### b. Step 2: Bias correction

For this study the selected bias correction techniques are the ratio bias correction (RBC) and the cumulative distribution function (CDF) matching. The SPPs have varying errors for different types or rainfall events, intensities, seasons, and areas (Dinku et al. 2008; Xie et al. 2011; Gao and Liu 2013); thus, both techniques are applied on the gauge locations and interpolated on the  $0.10^\circ \times 0.10^\circ$  resolution grid. The corresponding rainfall of each SPP at each gauge location is derived from the original spatial resolution of the product before any necessary upscaling (e.g., CMORPH, PERSIANN). Moreover, a temporal classification is applied since the precipitation in the study area is seasonal. The data are divided into rainy/wet (June–September) and nonrainy/dry (October–May) seasons. The record length is not sufficient to consider finer temporal scales for calculating

the bias correction factors. If a long period of data (rain gauge and SPPs) is available, monthly scale should be considered (Arias-Hidalgo et al. 2013).

#### 1) RATIO BIAS CORRECTION

The RBC technique for improving the SPPs has been widely used due to its simplicity and good performance (e.g., Adler et al. 2000; Arias-Hidalgo et al. 2013; Bhatti et al. 2016). The particular steps followed are presented below:

- (i) Calculation of the correction factor  $f_{i,e}$  for each gauge  $i$  and season  $e$ , by dividing the accumulated rainfall of the gauge with the accumulated rainfall of the corresponding grid cell for each SPP:

$$f_{i,e} = \frac{\sum_{m=1}^N P_{i,m,e}^s}{\sum_{m=1}^N P_{j,m,e}^s} \quad (1)$$

- (ii) Interpolation of the correction factors with the IDW technique to create the final gridded correction factor  $f_{j,e}$ .
- (iii) Multiplication of the original SPP at each time step and grid cell with the corresponding gridded factor for the specific season, leading to the bias corrected version of the product:

$$P_j^{\text{RBC}} = f_{j,e} \times P_{j,e}^s \quad (2)$$

In Eqs. (1) and (2),  $P_{i,e}^s$  and  $P_{j,e}^s$  are the hourly rainfall recorded at the rain gauge location  $i$  and the season  $e$  from the rain gauge and the corresponding grid cell of the SPP, respectively. Parameter  $N$  is the total number of rainfall measurements in season  $e$ ,  $P_j^{\text{RBC}}$  is the RBC corrected rainfall of the selected SPP at the grid cell  $j$ ,  $f_{j,e}$  is the bias correction factor at the grid

cell  $j$  for the season  $e$ , and  $P_{j,e}^s$  is the original rainfall of the SPP at the grid cell  $j$  for season  $e$ .

## 2) CUMULATIVE DISTRIBUTION FUNCTION MATCHING

The CDF method is already successfully implemented for various rainfall products (e.g., [Huffman et al. 2004](#); [Ines and Hansen 2006](#); [Xie and Xiong 2011](#); [Serrat-Capdevila et al. 2016](#)). This technique transforms the rainfall estimates in order to achieve a similar CDF with the observed data. For this study, the method is applied at hourly time step, for maintaining the low latency of the original products and being able to utilize them for NRT applications. The steps are as follows:

- (i) CDF matching for every gauge and corresponding grid cell at an hourly time step:

$$P_i^{\text{cor}} = F_i^{-1g}[F_i^s(P_i^s)]. \quad (3)$$

- (ii) Calculation of the difference between the original and CDF-corrected SPP at each gauge location and each time step ( $\text{Dif}_i$ ):

$$\text{Dif}_i = P_i^{\text{cor}} - P_i^s. \quad (4)$$

- (iii) IDW interpolation of the differences at  $0.10^\circ \times 0.10^\circ$  resolution grid ( $\text{Dif}_j$ ).
- (iv) Addition of this gridded difference to each grid  $j$  of original satellite data  $P_j^s$ , resulting in the bias-corrected version of the product  $P_j^{\text{CDF}}$ :

$$P_j^{\text{CDF}} = P_j^s + \text{Dif}_j. \quad (5)$$

In Eqs. (3)–(5),  $F_i^s$  is the CDF of the SPP rainfall estimation of the grid cell corresponding to the rain gauge  $i$ ,  $F_i^{-1g}$  is the inverse CDF rainfall measurement of the rain gauge  $i$ , and  $P_i^{\text{cor}}$  and  $P_i^s$  are the corrected and actual rainfall estimation by the SPP of the grid cell corresponding to the rain gauge  $i$ .

It should be noted that due to the extreme skewness of the data ( $\sim 90\%$  of time steps rainfall is zero) and the inability to fit a theoretical distribution, the nonparametric empirical cumulative distribution and the subsequent quantile mapping are used.

### c. Step 3: Merging

Following the recommendation of previous studies ([Hasan et al. 2016](#); [Khairul et al. 2018](#)), this study evaluates the potential benefits of merging NRT SPPs. For merging, a linear combination of the SPPs is considered. LATE is excluded because its latency (15 h) is significantly

larger than the highest latency of the remaining four products ( $\sim 5$  h for EARLY).

The merging is performed for a range of temporal scales. For merging, the optimal version of each SPP at the chosen temporal scale is selected, choosing between the original SPP, and the two bias-corrected versions. The optimal version is defined by the calibration results, based on the lowest normalized root-mean-square error (NRMSE) on the Kinu basin calculated by the corresponding grid cells, and not based on the lumped basin's rainfall statistics. This is preferred for considering in a direct way the spatial variation of the performance of the SPPs.

The variance of the errors (denoted var) is the selected indicator for the merging of the SPPs. For each SPP the error between the gauge observation and the corresponding grid cell is calculated for each time step. The var is derived for the wet and dry season at the gauges and interpolated with the IDW method on the  $0.10^\circ \times 0.10^\circ$  resolution grid.

It should be noted that the selection of the variance assumes that the individual products have normally distributed and unbiased errors, which is questionable for the original versions of the SPPs, increasingly so for finer time scales. In previous studies it is found that the bias correction techniques improve the results for daily and coarser temporal scales (e.g., [Xie et al. 2011](#); [Khairul et al. 2018](#)). Thus, for taking to account any potential limitations of the selected bias correction techniques for very fine temporal scales, the merging is performed not only at subdaily, but also daily and coarser scales.

There are three merging techniques used in this study, namely, (i) the error variance, (ii) the inverse error variance weighting (IEVW), and (iii) the simple average (Average). The weight of each SPP for the merging of the products is calculated based on the results of the var and the selected merging techniques. The formula for calculating the final output is given in Eq. (6):

$$P_j^{\text{Merg}} = \sum_{k=1}^n W_{k,e,j} \times P_{k,e,j}, \quad (6)$$

where  $P_j^{\text{Merg}}$  is the precipitation of merged product at the grid cell  $j$ ;  $n$  is the number of satellite products used for merging (here 4); and  $W_{k,e,j}$  and  $P_{k,e,j}$  are, respectively, the weight and rainfall of the satellite product  $k$  for the season  $e$  and the grid cell  $j$ .

### 1) ERROR VARIANCE

This method was implemented by [Hasan et al. \(2016\)](#) for combining data from radars and rain gauges, and by [Woldemeskel et al. \(2013\)](#) for combining rain gauge and TRMM data. One of the assumptions of this method is that the errors of the individual products are uncorrelated.

The formula used for calculating the weight of each product is

$$W_{k,e,j} = \frac{1}{n-1} \times \frac{\left( \sum_{k=1}^n \text{var}_{k,e,j} \right) - \text{var}_{k,e,j}}{\sum_{k=1}^n \text{var}_{k,e,j}}, \quad (7)$$

where  $n$  is the number of products used for merging,  $k$  is the satellite product for which the weight will be calculated,  $e$  and  $j$  the selected season and grid cell respectively and  $\text{var}$  is the variance of estimation errors of the product  $k$  for the season  $e$  at the grid cell  $j$ .

### 2) INVERSE ERROR VARIANCE WEIGHTING

This technique is already used for the Global Precipitation Climatology Project (GPCP) for producing a dataset of global monthly precipitation estimates (Huffman et al. 1997). The formula used for calculating the weight of each product is

$$W_{k,e,j} = \frac{1}{\text{var}_{k,e,j}^2} \bigg/ \frac{1}{\sum_{k=1}^n \text{var}_{k,e,j}^2}, \quad (8)$$

with the explanation of the symbols used being the same as the error variance method [Eq. (7)].

### 3) SIMPLE AVERAGE

The weights for each product are equal and depend only on the number of products. The formula used is given in Eq. (9):

$$W_{k,e,j} = \frac{1}{n}, \quad (9)$$

with  $n$  being the number of products used for merging.

In literature, it is reported that simple averaging of various forecasts for microeconomic time series outperformed more complicated schemes of weighting of the individual datasets, a finding called as the “forecast combination puzzle” (Stock and Watson 1999, 2003, 2004). Thus, it can be assessed whether similar findings apply also on merging rainfall estimates of SPPs.

### d. Comparison indicators

For the above mentioned three steps, the SPPs are compared with the gauges. The analysis is based on multiple indicators for a range of temporal (hourly up to monthly) and spatial (gauge location, grid cell, basin) scales.

The selected quantitative indicators are the correlation coefficient  $R$  [Eq. (10)], relative bias [Eq. (11)], and NRMSE [Eq. (12)]:

$$R = \frac{N \left( \sum_{m=1}^N P^g P^s \right) - \left( \sum_{m=1}^N P^g \right) \left( \sum_{m=1}^N P^s \right)}{\sqrt{\left[ N \sum_{m=1}^N P^{g^2} - \left( \sum_{m=1}^N P^g \right)^2 \right] \left[ N \sum_{m=1}^N P^{s^2} - \left( \sum_{m=1}^N P^s \right)^2 \right]}}, \quad (10)$$

$$\text{rel. bias (\%)} = \frac{\sum_{m=1}^N P^s - \sum_{m=1}^N P^g}{\sum_{m=1}^N P^g} \times 100, \quad (11)$$

$$\text{NRMSE} = \sqrt{\frac{\sum_{m=1}^N (P_m^g - P_m^s)^2}{N}} \bigg/ \bar{P}^g, \quad (12)$$

where  $P^g$  and  $P^s$  are the rainfall estimates by gauge and satellite rainfall product, respectively, at each gauge/grid cell or basin level, depending the spatial scale,  $\bar{P}^g$  is the average rain gauge rainfall, and  $N$  is number of observations for the various temporal scales of the analysis.

The calculations are undertaken for all time steps (unconditional) and only the time steps where rainfall is recorded according to the gauge data (conditional). The  $R$  and NRMSE are calculated for all temporal and spatial scales, while the relative bias is calculated for each spatial scale at the hourly temporal scale since it is not affected by the different temporal resolutions (for the unconditional analysis).

The contingency table classifies the satellite data based on the correct and false identification of a specific rainfall threshold (Table 2). More specifically, the data are classified as hits ( $a$ ), when both satellite rainfall (scheme) and gauge rainfall (observation) are above a preset threshold; false alarm ( $b$ ), when scheme is above the threshold but observation not; misses ( $c$ ), when scheme is below the threshold while observation is above; and correct negatives ( $d$ ), when both scheme and observation are below the threshold.

The remotely sensed products are evaluated at the grid cell spatial resolution for a range of precipitation thresholds and for two temporal scales. More specifically, for hourly temporal scale and intensities of 0.5, 1, 1.5, 2, 2.5, 5, 7.5, 10, 15, 20, and 30 mm h<sup>-1</sup>, and for 2-day temporal scale and intensities of 0.5, 2, 5, 10, 15, 20, 25, 50, 75, 100, 200, and 300 mm (2 days)<sup>-1</sup>. These two temporal scales are selected for analyzing the performance of the SPPs on the finest possible temporal scale for the particular case study, as well as on a coarser scale that can nevertheless be utilized for NRT applications (in medium/large catchments). To increase the sample

TABLE 2. Contingency table.

		Gauge $\geq$ threshold		Total
		Yes	No	
Satellite $\geq$ threshold	Yes	Hits ( $a$ )	False alarm ( $b$ )	$a + b$
	No	Misses ( $c$ )	Correct negative ( $d$ )	$c + d$
	Total	$a + c$	$b + d$	Sum = $a + b + c + d$

size, especially for high rainfall events which are rare, the data are clustered into lowlands ( $<750$  m; 114 grid cells) and highlands ( $\geq 750$  m; 38 grid cells) and each group is analyzed as one set. This results in a certain loss of information on the spatial distribution of the indicators but is used due to the limited temporal coverage of the data. Based on the contingency table (Table 2), the following indicators are used:

Probability of detection [POD; Eq. (13)]: The probability of correctly identifying a positive event across all estimations. Values range from 0 to 1, with 1 being the best score:

$$\text{POD} = \frac{a}{a + c}. \quad (13)$$

False alarm ratio [FAR; Eq. (14)]: The percentage of estimations that are incorrectly identified as positive. Values range from 0 to 1, with 0 being the best score:

$$\text{FAR} = \frac{b}{a + b}. \quad (14)$$

Heidke skill score [HSS; Eq. (15)]: It measures the fractional improvement of the scheme over the correct identification of an event due to chance. Values range from  $-\infty$  to 1. Negative scores indicate

that better results can be provided by chance, 0 means no skill, and 1 is the best score:

$$\text{HSS} = 2 \times \frac{a \times d - b \times c}{(a + c) \times (c + d) + (a + b) \times (b + d)}. \quad (15)$$

#### 4. Results and discussion

It should be noted that the results of this study refer to the specific versions of the SPPs that are used for the analysis. The subsequent versions might perform differently.

##### a. Evaluation of SPPs

The SPPs have varying performance with CMORPH, EARLY, and LATE underestimating the total rainfall for most of the area, while GSMaP and PERSIANN overestimate it (Fig. 3). All the SPPs have also challenges in depicting the spatial distribution of the accumulated rainfall.

The accumulated monthly rainfall of the SPPs is also quite variable (Fig. 4). EARLY and LATE have substantial differences on their behavior between the two years. For example, whereas the rainfall is underestimated in April, May, and July 2016, it is overestimated

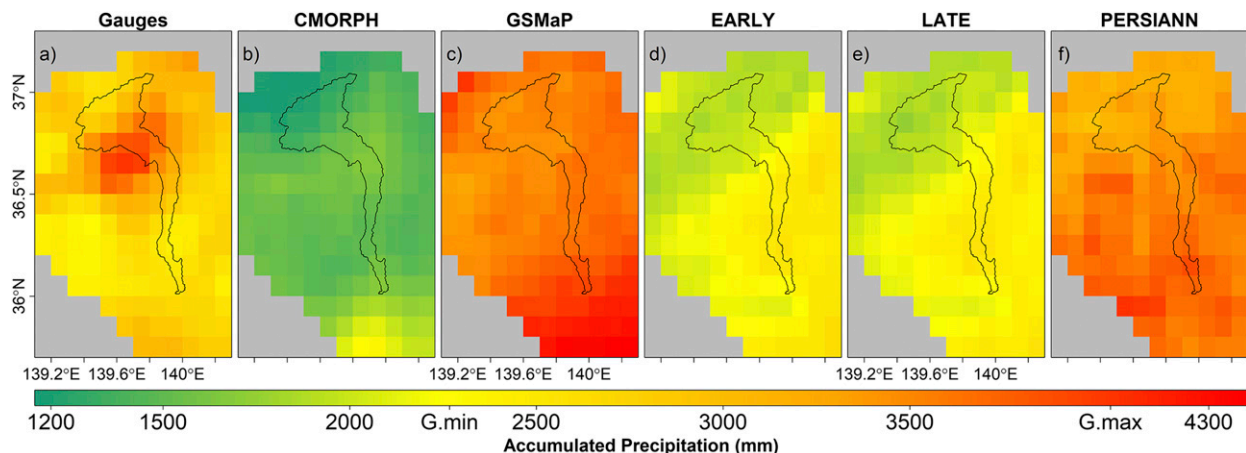


FIG. 3. Accumulated precipitation (mm) at the study area for the period February 2015–December 2016. The figure presents (a) the rain gauges' interpolated rainfall and (b)–(f) the SPPs rainfall estimates on the  $0.10^\circ \times 0.10^\circ$  resolution grid. The minimum (G.min) and maximum (G.max) values recorded from the gauges are indicated on the legend.



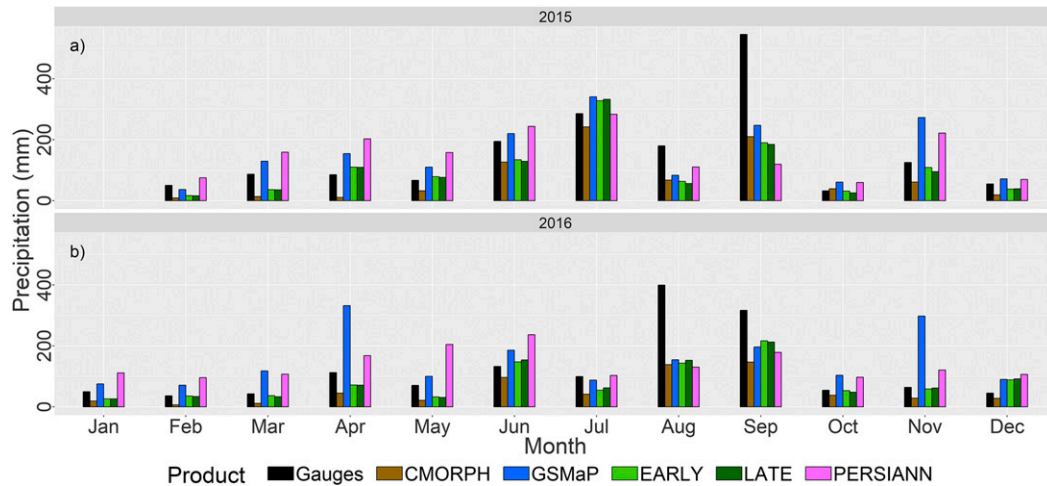


FIG. 4. Accumulated monthly rainfall for Kinu basin as it is derived by the rain gauges' interpolated data and the SPPs for the years (a) 2015 and (b) 2016.

for the same months in 2015. PERSIANN and GSMaP underestimate the rainfall during the rainy season (October–May) and overestimate it for the rest of the year, and CMORPH consistently underestimates it.

The SPPs with the best performance based on NRMSE and correlation are the EARLY and LATE (Table 3). In general, the performance of all products increases with coarser temporal scales, as also concluded by other

studies (e.g., Gaona et al. 2016). LATE outperforms EARLY, as it incorporates more information, due to its extra latency. CMORPH has very good correlation and outperforms all other products for coarser than daily temporal scales. GSMaP has moderate performance and PERSIANN very low. The drop of correlation for coarse scales for some SPPs can be attributed to the fact that for these scales there is limited sample size for robust

TABLE 3. Quantitative comparison indicators for the SPPs over Kinu basin for various temporal scales. The results for unconditional ( $P^g \geq 0$ ) and conditional ( $P^g > 0$ ) analysis are presented. The bold values correspond to the highest recorded performance for each temporal scale.

	CMORPH		GSMaP		EARLY		LATE		PERSIANN	
	$P^g \geq 0$	$P^g > 0$	$P^g \geq 0$	$P^g > 0$	$P^g \geq 0$	$P^g > 0$	$P^g \geq 0$	$P^g > 0$	$P^g \geq 0$	$P^g > 0$
<b>Correlation</b>										
Hourly	0.49	0.46	0.40	0.37	0.56	0.54	<b>0.58</b>	0.56	0.21	0.22
2 hourly	0.52	0.50	0.44	0.41	0.59	0.57	<b>0.61</b>	0.59	0.24	0.24
3 hourly	0.54	0.52	0.45	0.42	0.61	0.59	<b>0.63</b>	0.61	0.25	0.26
6 hourly	0.60	0.58	0.48	0.45	0.64	0.62	<b>0.65</b>	0.64	0.28	0.27
12 hourly	0.65	0.64	0.51	0.49	0.66	0.65	<b>0.68</b>	0.67	0.30	0.29
Daily	0.70	0.69	0.52	0.51	0.68	0.67	<b>0.70</b>	0.70	0.30	0.29
2 day	<b>0.74</b>	0.74	0.52	0.51	0.71	0.70	0.73	0.73	0.24	0.23
Weekly	<b>0.76</b>	<b>0.76</b>	0.47	0.47	0.70	0.70	0.73	0.73	0.16	0.16
Monthly	<b>0.88</b>	<b>0.88</b>	0.47	0.47	0.74	0.74	0.73	0.73	0.28	0.28
<b>NRMSE</b>										
Hourly	4.18	2.44	5.40	3.09	3.98	2.31	3.91	<b>2.28</b>	5.45	2.84
2 hourly	3.99	2.54	4.93	3.10	3.75	2.39	3.68	<b>2.35</b>	5.03	2.96
3 hourly	3.85	2.59	4.70	3.11	3.59	2.41	3.53	<b>2.38</b>	4.78	3.02
6 hourly	3.59	2.66	4.20	3.08	3.35	2.47	3.28	<b>2.43</b>	4.39	3.11
12 hourly	3.17	2.58	3.57	2.88	2.96	2.40	2.89	<b>2.34</b>	3.88	3.06
Daily	2.62	2.34	2.92	2.59	2.46	2.19	2.40	<b>2.14</b>	3.23	2.84
2 day	2.28	2.18	2.48	2.36	2.15	2.05	2.09	<b>1.99</b>	2.85	2.72
Weekly	1.36	1.36	1.48	1.48	1.28	1.28	<b>1.24</b>	<b>1.24</b>	1.71	1.71
Monthly	0.79	0.79	0.88	0.88	<b>0.74</b>	<b>0.74</b>	0.75	0.75	0.94	0.94
<b>Rel. bias (%)</b>										
Accumulated	−53.70	−58.00	13.00	<b>0.80</b>	−33.00	−41.20	−33.70	−39.70	7.30	−32.80

TABLE 4. Coefficient of determination  $R^2$  for the SPPs for 2-day accumulated rainfall at gauge locations,  $0.10^\circ \times 0.10^\circ$  grid resolution, and Kinu basin. For gauge and grid cell comparisons the data are treated as one set.

$R^2$	CMORPH	GSMaP	EARLY	LATE	PERSIANN
Gauges (point comparison)	0.37	0.24	0.38	0.41	0.07
$0.10^\circ \times 0.10^\circ$ ( $\sim 100 \text{ km}^2$ )	0.41	0.25	0.42	0.44	0.08
Kinu basin ( $\sim 1800 \text{ km}^2$ )	0.55	0.27	0.50	0.54	0.06

statistics. The results for conditional and unconditional analysis vary mainly for the NRMSE. As expected, the NRMSE is higher for the unconditional analysis, since the mean rainfall decreases substantially.

GSMaP and PERSIANN have the best performance based on relative bias (Table 3), which is nevertheless misleading. Both products underestimate the rainfall in the wet season (Fig. 4) and in the upper parts of the basin whereas they overestimate the rainfall in the dry season and in the lower parts of the basin, and these differences are mutually cancelled.

The high differences for all SPPs between the unconditional and conditional results on relative bias can be attributed to the drizzling effect and the temporal scale used to derive the indicator. Due to the drizzling effect the SPPs record very low rainfall intensities when there is no actual rainfall (Piani et al. 2010; Valdés-Pineda et al. 2016). Moreover, in the hourly scale the drizzling effect influences the results more, compared to coarser scales, where the conditional relative bias is approaching the unconditional one (e.g., the conditional relative bias at daily scale for the GSMaP and IMERG is 10.31% and  $-34.70\%$ , respectively).

Performance is also increased by aggregating in space (Table 4), which is in agreement with other studies (e.g., Bell and Kundu 2003). All SPPs apart from PERSIANN (which performs poorly in general) have increasing coefficient of determination  $R^2$  from point comparison to grid cell and basin comparison (note that the basin entries are the squares of the correlation entries for 2-day unconditional analysis in Table 3).

When the SPPs are compared for the typhoon-induced rainfall events (Table 5 for all events, and Fig. 5 for the Etai event), all of them fail to capture most of the rainfall, apart from the EARLY and LATE during Typhoon Nangka. Previous studies also showed that, in general, SPPs cannot capture high-intensity events accurately (e.g.,

Bitew and Gebremichael 2010; Nikolopoulos et al. 2013; Huang et al. 2013; Chen et al. 2014; Anjum et al. 2016).

One important reason for the underestimation seems to be the limitations of the satellites' sensors. More specifically, PMW and IR sensors struggle to depict rainfall caused by warm clouds over land (Petty and Krajewski 1996; Hobouchian et al. 2017), suggesting that typhoon-generated rainfall could be challenging. This applies especially for the Etai and Mindulle cyclones, whose centers were close to the study area, resulting in high temperatures. Moreover, LEO satellites' sensors have challenges in depicting the orographic enhancement in complex terrains (Petty and Krajewski 1996; Dinku et al. 2008; Derin and Yilmaz 2014).

One more challenge that is identified from the analysis and could influence the accuracy of the SPPs is a spatial shifting of the rainfall for all the combined PMW-IR based SPPs that use morphing techniques based on information derived from IR sensors (all besides PERSIANN). This is very clear for the Nangka event and EARLY for particular time steps [Fig. 6; e.g., at 1200 local time (LT) EARLY rainfall is shifted southwest]. For quantifying this error, the spatial correlation of EARLY with the rain gauges interpolated rainfall is calculated for each time step and a range of shifting combinations (Fig. 6c). The results show that a spatial shifting of the product substantially increases its performance (e.g., at 1200 LT the correlation increases from 0.34 up to 0.86 for shifting EARLY northwest). All SPPs (besides PERSIANN) exhibit similar behavior during all three typhoons. Thus, the shifting error, as it is already described for GSMaP (Ozawa et al. 2011; Chen et al. 2019), seems to exist for other SPPs as well.

As with the quantitative assessment, EARLY and LATE outperform the rest SPPs on the contingency table analysis (Fig. 7). Again, the performance is improved at coarser temporal resolutions. The results show

TABLE 5. Accumulated rainfall (mm) at Kinu basin during the three typhoon-induced rainfall events.

Event	Gauges	CMORPH	GSMaP	EARLY	LATE	PERSIANN
Nangka (15 Jul 2015)	112.09	57.98	67.58	127.79	121.65	31.71
Etai (9–10 Sep 2015)	400.27	74.72	77.48	82.63	92.53	7.75
Mindulle (22 Aug 2016)	107.00	30.34	47.39	46.39	51.73	24.82

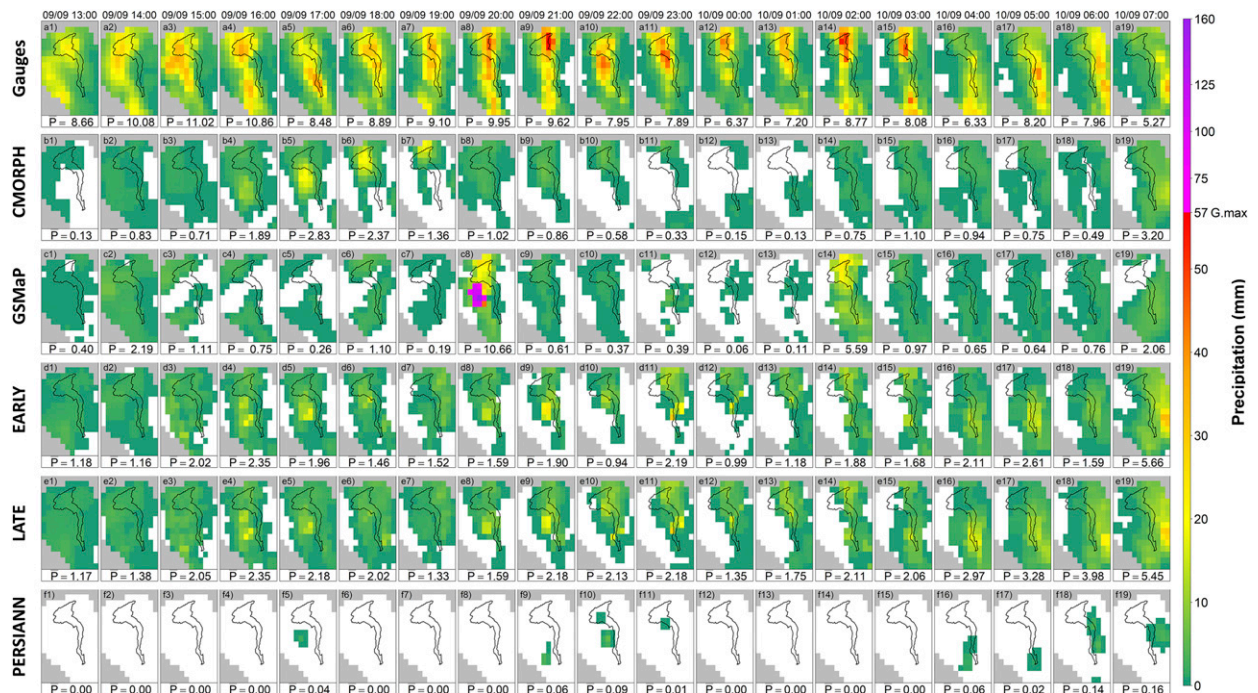


FIG. 5. Hourly precipitation during Typhoon Etou as it is depicted by (a) the rain gauges' interpolated rainfall and the SPPs [(b) CMORPH, (c) GSMaP, (d) EARLY, (e) LATE, (f) PERSIANN] on the  $0.10^{\circ} \times 0.10^{\circ}$  resolution grid. The white color indicates zero rainfall. The mean value of the precipitation from all the grid cells for each subplot is presented in the associated box. The maximum (G.max) value recorded from the gauges is indicated in the legend.

that the SPPs have two types of errors, with both errors increasing with higher rainfall intensities; an important remark given one of the potentially useful applications of SPPs is flood forecasting. First, the products fail to detect many precipitation events (Figs. 7a,b), and second, many times the SPPs wrongly estimate rainfall above the threshold for lower-intensity events (Figs. 7c,d). These errors have high impact on the HSS score (Figs. 7e,f), which deteriorates remarkably for high intensities. The elevation also affects the performance and, in general, SPPs perform better for the lowlands, possibly due to the challenges in depicting the orographic enhancement in complex terrains. It can be concluded that the errors of the SPPs not only vary spatially and temporally but also depend upon the rainfall magnitude. Similar results were also found in previous studies (e.g., Bitew and Gebremichael 2010). This finding is important when selecting bias correction techniques, since it would be advisable to select methods that take into consideration this behavior (e.g., CDF matching technique).

For high-intensity events, EARLY and LATE are considerably better than the rest in terms of POD. Their performance is in agreement with the results of Gaona et al. (2016), who evaluated a different SPP from the IMERG family (IMERG FINAL) over the Netherlands. CMORPH outperforms the other SPPs in terms of

FAR for high intensity rainfall, because it consistently underestimates the rainfall. PERSIANN's and GSMaP's poor performance in terms of FAR is associated with the overestimation of rainfall in the nonrainy season.

Finally, the error between the SPPs and the rain gauges interpolated rainfall for each time step and grid cell is calculated, and the correlation matrix is derived. At the hourly temporal scale, the median correlation of the errors across all grid cells among the different pairs of SPPs varies from 0.40 up to 0.64, besides EARLY and LATE pair that has 0.93. For coarser temporal scales the correlation increases (e.g., for daily it ranges from 0.58 to 0.74; EARLY and LATE have 0.98). This can be attributed to the fact that all SPPs make use of similar satellites and techniques in their algorithms. The high correlation between EARLY and LATE gives one more reason for excluding LATE from merging, since their errors and spatiotemporal behavior are quite similar, and a combination of them would have a limited added value. Moreover, the fact that the errors of the SPPs are not uncorrelated could raise questions regarding the use of the error variance technique. For overcoming this problem, there are techniques for either including the covariance matrix or transforming the data for decorrelation (Hasan et al. 2016). Nevertheless, in the study of



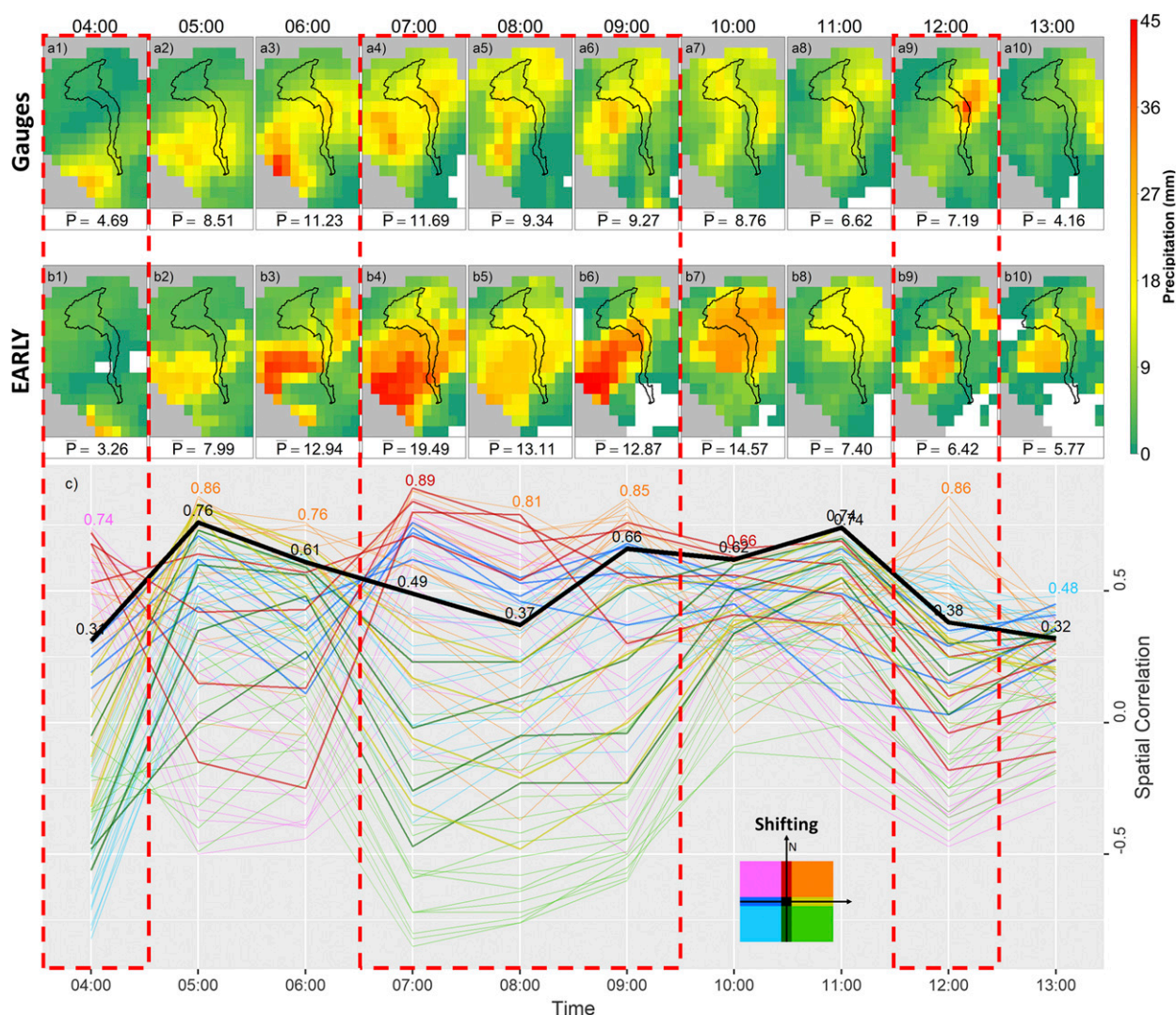


FIG. 6. Precipitation pattern during typhoon Nangka from (a1)–(a10) the rain gauges' interpolated rainfall and (b1)–(b10) EARLY on a  $0.10^\circ \times 0.10^\circ$  resolution grid. The white color indicates zero rainfall. The mean value of the precipitation for each subplot is presented on the associated box. (c) Spatial correlation of rainfall between EARLY and gauges for a range of spatial shifting combinations. EARLY is shifted from 1 up to 4 grid cells for each direction, resulting in 81 different combinations. The black line refers to the results from the original (nonshifted) location of the product, with the rest of the colors indicating a shift as per the legend. For example, the light green lines refer to the 16 combinations of shifting the product to the southeast (from 1 up to 4 grid cells in each direction). The dotted red rectangles depict the time steps when there is a high indication for shifting error for the SPP.

Smith and Wallis (2009) that examined merging techniques for forecasting of macroeconomic time series it was concluded that it is better to neglect any covariance and calculate the weights based only on the mean-squared error. Thus, the use of the error variance method is justifiable.

#### b. Bias correction

As expected, both techniques reduce the bias of the SPPs for all the study area, with the RBC method giving the best results (Fig. 8). CDF matching results in negative bias for all products, and this is related to the

inaccurate estimation of low-intensity precipitation events by the SPPs. More specifically, the SPPs at the hourly temporal scale have more time steps ( $\sim 90\%$ ) with no precipitation than the rain gauges ( $\sim 85\%$ ), leading to accumulated loss of rainfall even after CDF matching. One way of overcoming this challenge is to implement the technique at coarser scales. With longer time series, the CDF matching could, for example, be performed on a daily time step and then disaggregated to hourly, according to the ratio between hourly and corresponding daily rainfall at each time step for the original SPPs. Therefore, the accumulated rainfall of the

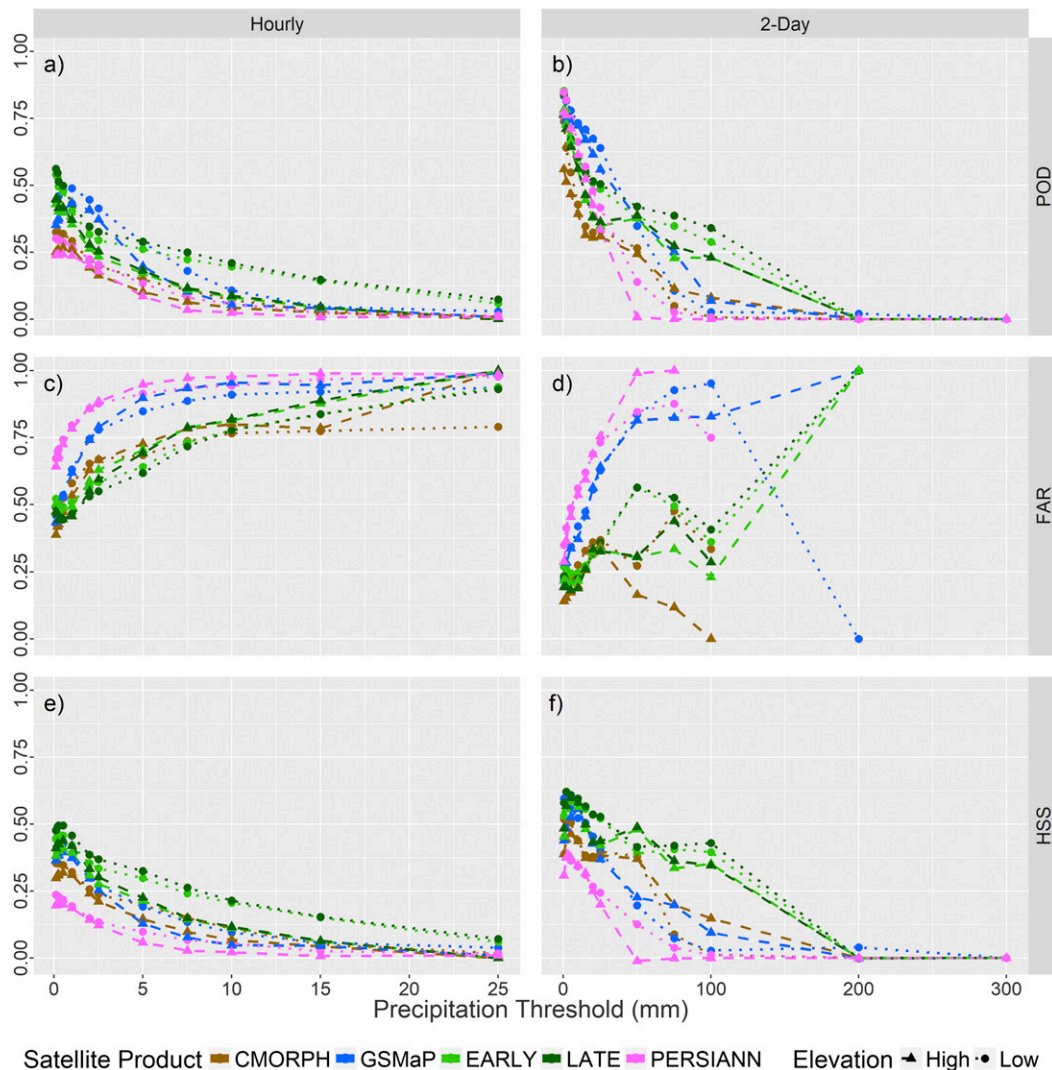


FIG. 7. (a),(b) POD, (c),(d) FAR, and (e),(f) HSS between the SPPs and the rain gauges interpolated rainfall for the  $0.10^{\circ} \times 0.10^{\circ}$  resolution grid. Two temporal scales are shown: (left) hourly and (right) 2 day. The data are clustered in lowlands ( $<750$  m) and highlands ( $\geq 750$  m).

gauges that corresponds to the additional no-rainy time steps on the SPPs will be reduced and the relative bias will be improved. Nevertheless, by working on coarser temporal scales, there will be an increase in the latency. Moreover, since the SPPs cannot capture low intensities accurately, the disaggregation would cause an overestimation of the rainfall events. It is therefore challenging to improve the performance of the SPPs on fine temporal scales by employing the CDF matching technique.

Both techniques improve the spatial pattern and magnitude of the accumulated rainfall estimates, on the validation set as well, resulting in increased accuracy, especially for coarse scales (Table 6). For fine scales, though, the accuracy decreases in both the calibration

and validation set. This can be attributed to the fact that the SPPs have very low correlation for the fine scales, thus it is very difficult to increase their performance after bias correction.

In particular, one of the limitations of the CDF matching method is that the pairing is not maintained, since the CDFs of the gauges and SPPs are constructed independently, without taking into consideration the time of each observation (Madadgar et al. 2014). This limitation is highly influencing the results of the CDF matching correction for GSMaP and PERSIANN that have systematic overestimation (underestimation) of the rainfall in the dry (wet) season. Thus, the seasonal bias would require seasonal adjustment. Nevertheless, this is not selected in this research due to the limited



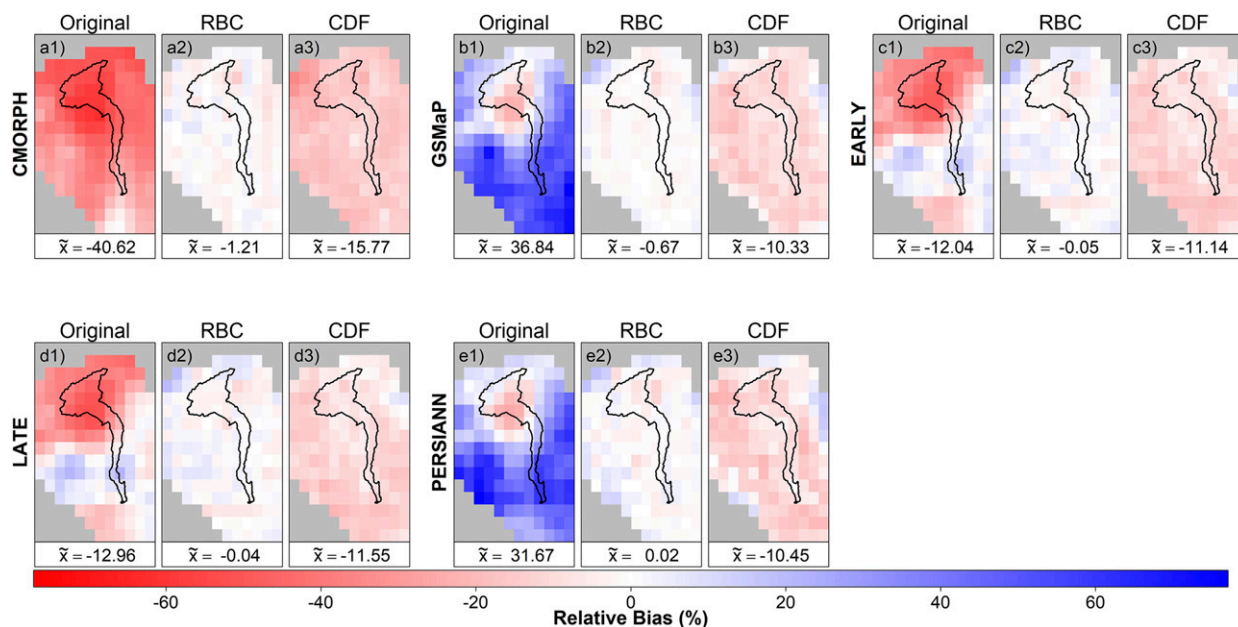


FIG. 8. Relative bias of the SPPs [(a) CMORPH, (b) GSMaP, (c) EARLY, (d) LATE, (e) PERSIANN] for the calibration set (year 2015). The figure presents the results for the original version (SPP1) and the two bias-corrected versions (SPP2 for the ratio bias correction and SPP3 for the cumulative distribution function matching) of each SPP. The median value of the relative bias from all the grid cells for each version of SPP is presented in the associated box.

record length. Similarly, the RBC is performed on seasonal scale; thus, important variations that occur in finer temporal scales (e.g., monthly), as well as intensity-induced errors, are not corrected. Due to the small temporal coverage though, it is not advisable to work on monthly or finer scale for the RBC correction, or split the data based on intensity or precipitation type (typhoon, convective, high–low intensity, etc.).

The method that performs the best for each SPP in the calibration set also performs best in the validation set, indicating that the selected bias correction technique should be decided based on the particular SPP and its characteristics.

One important observation is the mismatch of the results on the relative bias for the calibration set between the Kinu basin and the grid cells for the SPPs (Fig. 8; Table 6), something that also applies for the whole temporal coverage (Fig. 3; Table 3) as well as the validation set. It can be noticed, for example, that for the calibration set the original versions of GSMaP and PERSIANN have minor bias for the Kinu basin, although the actual bias for each grid cell is quite high, either positive or negative. These differences though, are effectively cancelled when the lumped rainfall is calculated for the basin. This result highlights once more, that the correction techniques should take into consideration the spatial behavior of the SPPs and not rely on spatially aggregated indicators that could provide misleading information.

### c. Merging

This section presents the results of merging the SPPs for two different analyses. More specifically, the merging weights are calculated at the (i) hourly temporal scale for the dry season and (ii) daily temporal scale for the whole temporal coverage. As is mentioned on the methodology, the version of each SPP with the lowest NRMSE on the calibration set is selected for the merging (Table 7). LATE is not considered as explained in the previous sections. These two analyses are selected for the following reasons:

- Hourly scale is the finest temporal scale that ensures the least possible latency of the merged products. In this scale nevertheless, the version that has the lowest NRMSE on the calibration set for all SPPs when the whole temporal coverage is analyzed, is the original version. Thus, the assumption of unbiased errors is not valid. Moreover, most of the rain of the wet season is associated with typhoons, where the SPPs have challenges. This, in combination with the small temporal coverage, results in high errors for very fine temporal scales (Table 3) and instability for coarser scales (Table 8). For the dry season though, there are no limitations related to typhoon-induced rainfall and, moreover, two SPPs have increased performance after bias correction. Although the assumption of unbiased errors is not

TABLE 6. NRMSE and relative bias (%) of the SPPs in comparison to the rain gauges interpolated rainfall for Kinu basin for the calibration (year 2015) and validation (year 2016) sets. The results refer to the unconditional analysis ( $P^g \approx 0$ ). The best product for each temporal scale (bold) and the best version of each SPP on each temporal scale (italic) are indicated.

	CMORPH			GSMaP			EARLY			LATE			PERSIANN		
	Original	RBC	CDF	Original	RBC	CDF	Original	RBC	CDF	Original	RBC	CDF	Original	RBC	CDF
Calibration															
NRMSE															
Hourly	<i>4.31</i>	4.83	4.78	4.79	4.62	5.39	4.02	4.51	4.66	<b>3.88</b>	4.34	4.56	5.48	5.68	5.68
2 hourly	<i>4.15</i>	4.46	4.38	4.51	4.27	4.83	3.87	4.23	4.37	<b>3.74</b>	4.08	4.28	5.17	5.30	5.24
3 hourly	<i>4.06</i>	4.24	4.12	4.35	4.09	4.49	3.75	4.05	4.16	<b>3.64</b>	3.93	4.11	5.03	5.11	5.05
6 hourly	3.87	3.75	3.61	4.07	3.79	4.01	3.59	3.80	3.90	<b>3.48</b>	3.66	3.82	4.72	4.75	4.70
12 hourly	3.54	3.26	<b>3.09</b>	3.66	3.43	3.52	3.30	3.31	3.36	3.21	3.19	3.30	4.29	4.31	4.27
Daily	2.96	2.61	<b>2.51</b>	3.07	2.85	3.01	2.81	2.74	2.82	2.73	2.62	2.73	3.61	3.64	3.58
2 day	2.64	2.26	<b>2.19</b>	2.70	2.49	2.54	2.50	2.31	2.39	2.43	2.20	2.30	3.26	3.29	3.23
Weekly	1.52	1.25	<b>1.23</b>	1.57	1.42	1.52	1.46	1.34	1.40	1.42	1.27	1.34	1.94	1.97	1.90
Monthly	0.76	0.51	<b>0.51</b>	0.71	0.59	0.77	0.75	0.72	0.75	0.77	0.76	0.79	0.93	0.84	0.92
Rel. bias (%)	-51.4	-1.3	-15.3	1.1	-1.1	-10.4	-33.6	-1.8	-8.9	-35.6	-1.3	-8.5	-0.2	-1.2	-10.1
Validation															
NRMSE															
Hourly	3.83	5.03	5.49	6.14	5.05	9.03	<b>3.81</b>	5.04	5.83	3.88	5.41	6.16	5.18	5.25	5.34
2 hourly	3.58	4.56	4.93	5.43	4.48	7.91	<b>3.46</b>	4.45	5.10	3.49	4.78	5.36	4.60	4.68	4.64
3 hourly	3.37	4.12	4.35	5.10	4.11	7.30	<b>3.23</b>	4.12	4.67	3.26	4.42	4.93	4.20	4.21	4.19
6 hourly	<i>3.01</i>	3.35	3.38	4.28	3.43	5.84	<b>2.83</b>	3.40	3.84	2.84	3.66	4.04	3.66	3.67	3.62
12 hourly	2.40	2.55	2.59	3.35	2.59	4.45	2.23	2.82	3.13	<b>2.21</b>	3.01	3.27	2.97	3.00	2.91
Daily	1.89	1.85	<i>1.83</i>	2.59	1.92	3.46	1.70	2.00	2.22	<b>1.66</b>	2.12	2.31	2.38	2.36	2.32
2 day	1.47	1.23	<b>1.14</b>	2.00	1.47	2.60	1.36	1.46	1.61	1.31	1.51	1.64	1.94	1.85	1.87
Weekly	1.04	0.80	<b>0.71</b>	1.30	0.93	1.60	0.93	0.92	1.00	0.89	0.91	0.98	1.25	1.23	1.20
Monthly	0.82	0.43	<b>0.38</b>	1.07	0.67	1.34	0.70	0.56	0.60	0.69	0.55	0.58	0.92	0.81	0.84
Rel. bias (%)	-56.5	-7.3	-18.0	27.3	9.5	39.3	-32.2	1.6	<b>0.2</b>	-31.4	7.8	5.0	16.7	10.8	10.1

TABLE 7. Average NRMSE for the Kinu basin based on the results of gridcell unconditional analysis. The table presents the results for all versions of the SPPs used for merging, for the two selected analysis (daily scale and whole temporal coverage and hourly scale and dry season), and both the calibration and validation set. The best version of each SPP for each analysis and set is indicated with bold.

	CMORPH			GSMaP			EARLY			PERSIANN		
	Original	RBC	CDF	Original	RBC	CDF	Original	RBC	CDF	Original	RBC	CDF
1. Hourly—dry												
Calibration	<b>5.22</b>	8.45	6.99	8.49	<b>5.66</b>	11.78	<b>5.04</b>	5.49	6.70	9.84	<b>6.39</b>	11.05
Validation	<b>5.21</b>	9.82	8.19	16.95	<b>9.97</b>	24.88	<b>6.54</b>	7.34	11.04	13.05	<b>7.76</b>	13.42
2. Daily—whole												
Calibration	3.10	2.88	<b>2.82</b>	3.26	<b>3.00</b>	3.22	2.93	<b>2.91</b>	2.98	3.83	3.81	<b>3.78</b>
Validation	<b>2.20</b>	2.37	2.37	3.11	<b>2.39</b>	4.01	<b>2.09</b>	2.57	2.82	2.84	<b>2.83</b>	2.84

valid for all SPPs, it is still useful to present the results in this temporal scale.

- (ii) Daily scale results in 24-h latency of the merged products, which is still very useful for NRT applications on medium/large basins. In this scale all selected versions are bias corrected (Table 7) and thus comply with the assumption of unbiased errors of the individual products. It should be noticed that the best version on the validation set is not the same as the calibration, with possible reasons being the short temporal coverage and the inclusion of the typhoon events that the SPPs have challenges in depicting (Table 5).

The performance of the merged products on the hourly temporal scale and the dry season is improved compared to the individual SPPs. Especially for the Kinu basin spatial scale, the merging improves the results for fine temporal scales at both calibration and validation sets

(Table 8). This is very important, because for such scales the bias correction techniques could not bring any additional improvement on the SPPs. On the calibration set, the best individual SPP for the fine scales is the original version of LATE. The IEVW product outperforms this SPP, resulting not only in improved accuracy, but also with reduced latency (LATE has 15 h and IEVW has ~5 h). Note that for coarse temporal scales, the merged products do not outperform the individual SPPs on the calibration set. This is because the version of each SPP used, as well as the variance of the errors for calculating the weights, are determined on the hourly scale, thus the weights are not the optimal ones for coarser scales.

Similarly, the merging at daily level improves the results on both calibration and validation sets (Fig. 9) for the gridcell comparison, not only at the daily temporal scale, but also for all the coarser scales. All merging

TABLE 8. NRMSE for the Kinu basin unconditional analysis at a range of temporal scales for the dry season of the calibration (year 2015) and the validation (year 2016) sets. The results refer to the best individual SPP and the merged products (merging weights calculated at hourly level for the dry season). The best value for each temporal scale is indicated with bold.

Best individual SPP			Error variance	IEVW	Average
Calibration					
Hourly	LATE (original)	3.50	3.51	<b>3.43</b>	3.54
2 h	LATE (original)	3.21	3.26	<b>3.19</b>	3.29
3 h	LATE (original)	3.04	3.11	<b>3.03</b>	3.13
6 h	LATE (original)	<b>2.60</b>	2.73	2.66	2.75
12 h	LATE (original)	<b>2.08</b>	2.24	2.21	2.25
Daily	LATE (original)	<b>1.50</b>	1.56	1.55	1.57
2 day	LATE (original)	<b>1.05</b>	1.09	1.08	1.10
Weekly	LATE (original)	0.65	<b>0.65</b>	0.66	0.65
Monthly	PERSIANN (RBC)	<b>0.19</b>	0.28	0.32	0.27
Validation					
Hourly	CMORPH (original)	4.03	3.53	<b>3.43</b>	3.57
2 h	CMORPH (original)	3.76	3.19	<b>3.10</b>	3.22
3 h	CMORPH (original)	3.51	3.00	<b>2.93</b>	3.03
6 h	CMORPH (original)	3.02	2.46	<b>2.41</b>	2.48
12 h	EARLY (original)	2.55	2.16	<b>2.13</b>	2.17
Daily	EARLY (original)	1.92	1.63	<b>1.61</b>	1.64
2 day	LATE (original)	1.39	1.15	<b>1.15</b>	1.16
Weekly	PERSIANN (RBC)	0.78	0.62	0.63	<b>0.61</b>
Monthly	PERSIANN (RBC)	0.31	0.20	0.25	<b>0.19</b>

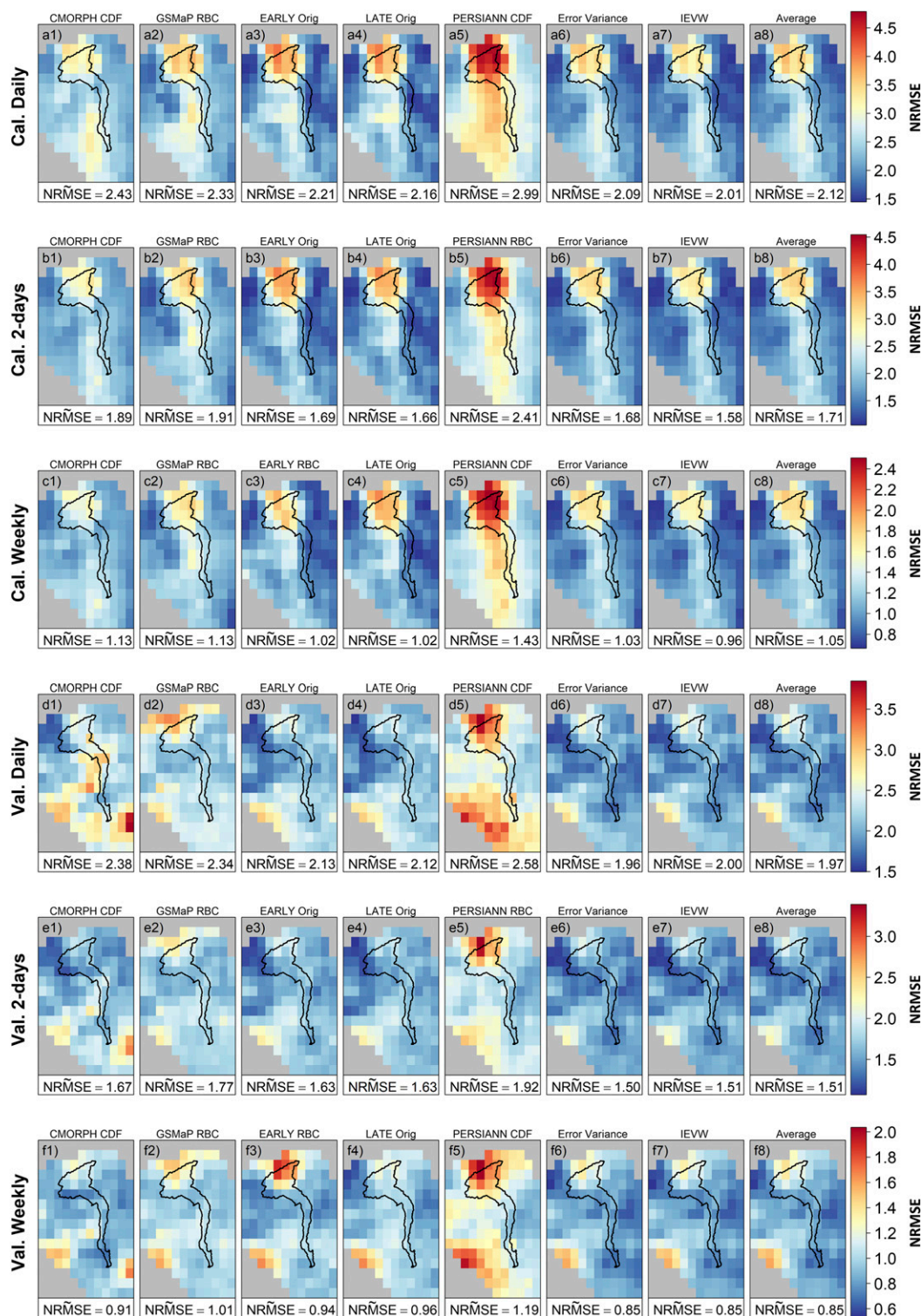


FIG. 9. NRMSE for a range of temporal scales on the  $0.10^\circ \times 0.10^\circ$  resolution grid for the calibration (2015) and validation (2016) sets. Shown for each temporal scale is the best version of each individual SPP and the merged products (merging weights calculated at daily scale). The best version of each SPP is defined based on the lowest median NRMSE between the original and the two bias corrected version of the SPP for each temporal scale at the calibration set. The median value of the NRMSE from all the grid cells for each subplot is presented in the associated box.



methods perform equally on the validation set and IEVW has the highest performance on the calibration set. This can be attributed to the used equations [Eqs. (7)–(9)]. More specifically, the IEVW method gives a wider range of values for the weights of each product at each grid cell. For the wet season, for example, the median weight of the grid cells for PERSIANN (the product with the lowest performance) is 0.11 and for EARLY (the product with the best performance) is 0.34 according to the IEVW. The error variance has weights of 0.22 and 0.26, respectively, while the average method scores 0.25 for all SPPs. Because the calibration set is quite limited and the performance of the SPPs between the calibration and validation set has differences, these more extreme values on the weights of the IEVW technique result in lower performance on the validation set. For larger datasets though, the weights of the SPPs would be more robust, and it is quite possible that the IEVW would outperform the rest of the merging methods, as already found in other studies (Khairul et al. 2018).

## 5. Conclusions and recommendations

This study evaluates the performance of five NRT SPPs in the Kinu basin region in Japan, including the use of two bias correction techniques and three merging methods that are applied taking into consideration the temporal and spatial variability on the SPPs' performance.

At this point, it is important to mention the limitations of this study. The main limitation is the short temporal coverage, which constrains the possibility for generalizing the conclusions and having robust results on the bias correction and merging steps. Moreover, this limitation creates subsequent challenges, as for example the inability to perform the bias correction techniques and merging methods with finer temporal splitting of the data. Nevertheless, this temporal coverage is chosen because it is a period that all selected SPPs have available data. One more limitation is the challenges of the SPPs in detecting typhoon-included rainfall. The three typhoon events that occurred in the area resulted in almost 20% of the total rainfall at Kinu basin. Thus, despite the small number of events, they substantially affect the evaluation of the SPPs and the performance of the selected methods on the analyzed temporal coverage and more importantly on the rainy season and the fine temporal scales. Finally, although the Kriging interpolation is quite robust and the gauge network is very extensive and of very high quality, there is still inherent uncertainty with every rainfall interpolation.

According to the results, the SPPs' performance depends on the selected temporal and spatial scale and increases with aggregation in time and/or space. Both bias-correction

techniques are able to reduce the bias and improve the spatial representation of the rainfall, as well as, increase the performance of the products for coarse temporal scales. Finally, the merging of the SPPs improves the results for all temporal and spatial scales, showing that the products could be utilized not only in lumped hydrological models, but also in distributed ones.

This study, moreover, highlights some challenges in the usage of the various SPPs and methods. In the analyzed spatial and temporal coverage, all SPPs, especially PERSIANN, have difficulties in detecting typhoon-induced rainfall, and additionally, spatial-shift error is identified for all SPPs that use data from LEO-PMW sensors and algorithms for rainfall propagation. Also, both bias-correction methods are in general unable to improve the results for finer than 6-h temporal scales. This creates an additional drawback when implementing the selected merging techniques at such temporal scales, where the original versions of the SPPs do not comply with the requirement of unbiased and normally distributed errors. Finally, the CDF matching, when applied at fine temporal scales, has challenges in eliminating the bias due to the significantly more time steps that the precipitation was estimated to be zero on the SPPs compared to the gauge observations.

Based on the results, there are avenues of future research and important research questions that need to be addressed in further studies. It is crucial to evaluate the performance of SPPs for basins of various sizes, especially in those areas affected by typhoons. The analysis can be expanded by comparing the SPPs via the simulation of the rainfall–runoff process through a hydrological model. In the future, similar studies with SPPs over extended temporal coverage should be taken up. Moreover, it is advisable to explore the performance of additional merging techniques, and the optimal method for a range of meteorological conditions and seasons. A comparison of an ensemble of SPPs together with gauge rainfall may be taken up as well. Last, it would be useful to identify the optimal combination of rain gauge and NRT satellite data to improve the accuracy of rainfall estimations for disaster risk reduction and water management tasks, such as flood and drought monitoring and early warning systems. The latter is of immense importance for countries with limited resources and poor ground observations that are nevertheless vulnerable to water-induced disasters. In such areas lies the most effective use of SPPs and any improvement in their applicability is very crucial.

*Acknowledgments.* The authors thank the data providers for distributing their products open access, as well as the three anonymous reviewers for their valuable



comments. Moreover, the first author would like to thank Mr. Jianning Ren for providing useful information regarding the analysis of the shifting error, Prof. Toshio Koike and Mr. Islam M. Khairul for the conducted discussions that contributed to the completion of this study, and Mr. Simon Parry for his fruitful suggestions for improving the quality of this manuscript.

## REFERENCES

- Adler, R. F., A. J. Negri, P. R. Keehn, and I. M. Hakkarinen, 1993: Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *J. Appl. Meteor.*, **32**, 335–356, [https://doi.org/10.1175/1520-0450\(1993\)032<0335:EOMROJ>2.0.CO;2](https://doi.org/10.1175/1520-0450(1993)032<0335:EOMROJ>2.0.CO;2).
- , G. J. Huffman, D. T. Bolvin, S. Curtis, and E. J. Nelkin, 2000: Tropical rainfall distributions determined using TRMM combined with other satellite and rain gauge information. *J. Appl. Meteor.*, **39**, 2007–2023, [https://doi.org/10.1175/1520-0450\(2001\)040<2007:TRDDUT>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<2007:TRDDUT>2.0.CO;2).
- Anjum, M. N., Y. Ding, D. Shanguan, M. W. Ijaz, and S. Zhang, 2016: Evaluation of high-resolution satellite-based real-time and post-real-time precipitation estimates during 2010 extreme flood event in Swat River Basin, Hindukush region. *Adv. Meteor.*, **2016**, 2604980, <https://doi.org/10.1155/2016/2604980>.
- Arias-Hidalgo, M., B. Bhattacharya, A. E. Mynett, and A. van Griensven, 2013: Experiences in using the TMPA-3B42R satellite data to complement rain gauge measurements in the Ecuadorian coastal foothills. *Hydrol. Earth Syst. Sci.*, **17**, 2905–2915, <https://doi.org/10.5194/hess-17-2905-2013>.
- Asadullah, A., N. McIntyre, and M. Kigobe, 2008: Evaluation of five satellite products for estimation of rainfall over Uganda. *Hydrol. Sci. J.*, **53**, 1137–1150, <https://doi.org/10.1623/hysj.53.6.1137>.
- Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo, 2017: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, <https://doi.org/10.5194/hess-21-589-2017>.
- Bell, T. L., and P. K. Kundu, 2003: Comparing satellite rainfall estimates with rain gauge data: Optimal strategies suggested by a spectral model. *J. Geophys. Res.*, **108**, 4121, <https://doi.org/10.1029/2002JD002641>.
- Bhatti, H. A., T. Rientjes, A. T. Haile, E. Habib, and W. Verhoef, 2016: Evaluation of bias correction method for satellite-based rainfall data. *Sensors*, **16**, E884, <https://doi.org/10.3390/s16060884>.
- Bitew, M. M., and M. Gebremichael, 2010: Evaluation through independent measurements: Complex terrain and humid tropical region in Ethiopia. *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer, 205–214, [https://doi.org/10.1007/978-90-481-2915-7\\_12](https://doi.org/10.1007/978-90-481-2915-7_12).
- Borga, M., F. Tonelli, R. J. Moore, and H. Andrieu, 2002: Long-term assessment of bias adjustment in radar rainfall estimation. *Water Resour. Res.*, **38**, 1226, <https://doi.org/10.1029/2001WR000555>.
- Brown, J. E. M., 2006: An analysis of the performance of hybrid infrared and microwave satellite precipitation algorithms over India and adjacent regions. *Remote Sens. Environ.*, **101**, 63–81, <https://doi.org/10.1016/j.rse.2005.12.005>.
- Chen, S., and Coauthors, 2014: Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations. *PLOS One*, **9**, e89681, <https://doi.org/10.1371/journal.pone.0089681>.
- , J. Hu, A. Zhang, C. Min, C. Huang, and Z. Liang, 2019: Performance of near real-time Global Satellite Mapping of Precipitation estimates during heavy precipitation events over northern China. *Theor. Appl. Climatol.*, **135**, 877–891, <https://doi.org/10.1007/s00704-018-2391-y>.
- Collischonn, B., W. Collischonn, and C. E. M. Tucci, 2008: Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *J. Hydrol.*, **360**, 207–216, <https://doi.org/10.1016/j.jhydrol.2008.07.032>.
- Derin, Y., and K. K. Yilmaz, 2014: Evaluation of multiple satellite-based precipitation products over complex topography. *J. Hydrometeorol.*, **15**, 1498–1516, <https://doi.org/10.1175/JHM-D-13-0191.1>.
- Dinku, T., S. Chidzambwa, P. Ceccato, S. J. Connor, and C. F. Ropelewski, 2008: Validation of high-resolution satellite rainfall products over complex terrain. *Int. J. Remote Sens.*, **29**, 4097–4110, <https://doi.org/10.1080/01431160701772526>.
- , F. Ruiz, S. J. Connor, and P. Ceccato, 2010: Validation and intercomparison of satellite rainfall estimates over Colombia. *J. Appl. Meteor. Climatol.*, **49**, 1004–1014, <https://doi.org/10.1175/2009JAMC2260.1>.
- Duan, Z., J. Liu, Y. Tuo, G. Chiogna, and M. Disse, 2016: Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. *Sci. Total Environ.*, **573**, 1536–1553, <https://doi.org/10.1016/j.scitotenv.2016.08.213>.
- Gao, Y. C., and M. F. Liu, 2013: Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. *Hydrol. Earth Syst. Sci.*, **17**, 837–849, <https://doi.org/10.5194/hess-17-837-2013>.
- Gaona, M. F. R., A. Overeem, H. Leijnse, and R. Uijlenhoet, 2016: First-year evaluation of GPM rainfall over the Netherlands: IMERG Day 1 Final Run (V03D). *J. Hydrometeorol.*, **17**, 2799–2814, <https://doi.org/10.1175/JHM-D-16-0087.1>.
- Haile, A. T., F. T. Tefera, and T. Rientjes, 2016: Flood forecasting in Niger-Benue basin using satellite and quantitative precipitation forecast data. *Int. J. Appl. Earth Obs. Geoinf.*, **52**, 475–484, <https://doi.org/10.1016/j.jag.2016.06.021>.
- Hasan, M. M., A. Sharma, F. Johnson, G. Mariethoz, and A. Seed, 2016: Merging radar and in situ rainfall measurements: An assessment of different combination algorithms. *Water Resour. Res.*, **52**, 8384–8398, <https://doi.org/10.1002/2015WR018441>.
- Hirpa, F. A., M. Gebremichael, and T. Hopson, 2010: Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *J. Appl. Meteor. Climatol.*, **49**, 1044–1051, <https://doi.org/10.1175/2009JAMC2298.1>.
- Hobouchian, M. P., P. Salio, Y. García Skabar, D. Vila, and R. Garreaud, 2017: Assessment of satellite precipitation estimates over the slopes of the subtropical Andes. *Atmos. Res.*, **190**, 43–54, <https://doi.org/10.1016/j.atmosres.2017.02.006>.
- Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteor.*, **43**, 1834–1853, <https://doi.org/10.1175/JAM2173.1>.
- Hsu, K.-L., X. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteor.*, **36**, 1176–1190, [https://doi.org/10.1175/1520-0450\(1997\)036<1176:PEFRSI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2).
- Huang, Y., and Coauthors, 2013: Evaluation of version-7 TRMM Multi-Satellite Precipitation Analysis product during the

- Beijing extreme heavy rainfall event of 21 July 2012. *Water*, **6**, 32–44, <https://doi.org/10.3390/w6010032>.
- Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**, 5–20, [https://doi.org/10.1175/1520-0477\(1997\)078<0005:TGPCPG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2).
- , R. F. Adler, E. F. Stocker, D. T. Bolvin, and E. J. Nelkin, 2004: Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time. *12th Conf. on Satellite Meteorology and Oceanography*, Seattle, WA, Amer. Meteor. Soc., P4, 11, [https://ams.confex.com/ams/annual2003/techprogram/paper\\_54906.htm](https://ams.confex.com/ams/annual2003/techprogram/paper_54906.htm).
- , D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015: NASA Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.5, 30 pp., [http://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V4.5.pdf](http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf).
- , —, and E. J. Nelkin, 2017: Integrated Multi-satellite Retrievals for GPM (IMERG) technical documentation. NASA Tech. Doc., 54 pp., [https://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_technical\\_doc\\_3\\_22\\_17.pdf](https://pmm.nasa.gov/sites/default/files/document_files/IMERG_technical_doc_3_22_17.pdf).
- Hughes, D. A., 2006: Comparison of satellite rainfall data with observations from gauging station networks. *J. Hydrol.*, **327**, 399–410, <https://doi.org/10.1016/j.jhydrol.2005.11.041>.
- Ines, A. V. M., and J. W. Hansen, 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agric. For. Meteorol.*, **138**, 44–53, <https://doi.org/10.1016/j.agrformet.2006.03.009>.
- JMA, 2019: Automated Meteorological Data Acquisition System. Japan Meteorological Agency, accessed 3 April 2017, <http://www.jma.go.jp/jma/en/Activities/amedas/amedas.html>.
- Joyce, R., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: 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, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Khairul, I. M., N. Mastrantonas, M. Rasmy, T. Koike, and K. Takeuchi, 2018: Inter-comparison of gauge-corrected global satellite rainfall estimates and their applicability for effective water resource management in a transboundary river basin: The case of the Meghna River basin. *Remote Sens.*, **10**, 828, <https://doi.org/10.3390/rs10060828>.
- Kim, K., J. Park, J. Baik, and M. Choi, 2017: Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. *Atmos. Res.*, **187**, 95–105, <https://doi.org/10.1016/j.atmosres.2016.12.007>.
- Koriche, S. A., and T. Rientjes, 2016: Application of satellite products and hydrological modelling for flood early warning. *Phys. Chem. Earth*, **93**, 12–23, <https://doi.org/10.1016/j.pce.2016.03.007>.
- Kubota, T., and Coauthors, 2007: Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. *IEEE Trans. Geosci. Remote Sens.*, **45**, 2259–2275, <https://doi.org/10.1109/TGRS.2007.895337>.
- Lo Conti, F., K.-L. Hsu, L. V. Noto, and S. Sorooshian, 2014: Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea. *Atmos. Res.*, **138**, 189–204, <https://doi.org/10.1016/j.atmosres.2013.11.011>.
- Long, Y., Y. Zhang, and Q. Ma, 2016: A merging framework for rainfall estimation at high spatiotemporal resolution for distributed hydrological modeling in a data-scarce area. *Remote Sens.*, **8**, 599, <https://doi.org/10.3390/rs8070599>.
- Madadgar, S., H. Moradkhani, and D. Garen, 2014: Towards improved post-processing of hydrologic forecast ensembles. *Hydrol. Processes*, **28**, 104–122, <https://doi.org/10.1002/hyp.9562>.
- Mitra, A. K., M. D. Gupta, S. V. Singh, and T. N. Krishnamurti, 2003: Daily rainfall for the Indian monsoon region from merged satellite and rain gauge values: Large-scale analysis from real-time data. *J. Hydrometeorol.*, **4**, 769–781, [https://doi.org/10.1175/1525-7541\(2003\)004<0769:DRFTIM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<0769:DRFTIM>2.0.CO;2).
- Nikolopoulos, E., E. Anagnostou, and M. Borga, 2013: Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy. *J. Hydrometeorol.*, **14**, 171–185, <https://doi.org/10.1175/JHM-D-12-09.1>.
- Niroshinie, M. A. C., K. Ohtuski, and Y. Nihei, 2016: Effect of small rivers for the inundations due to levee failure at Kinu River in Japan. *Procedia Eng.*, **154**, 794–800, <https://doi.org/10.1016/j.proeng.2016.07.408>.
- Ozawa, G., H. Inomata, Y. Shiraishi, and K. Fukami, 2011: Applicability of GSMaP correction method to typhoon "Morakot" in Taiwan. *J. Japan Soc. Civ. Eng. Ser B1*, **67**, I\_445–I\_450, [https://doi.org/10.2208/jscejhe.67.I\\_445](https://doi.org/10.2208/jscejhe.67.I_445).
- Petty, G. W., and W. F. Krajewski, 1996: Satellite estimation of precipitation over land. *Hydrol. Sci. J.*, **41**, 433–451, <https://doi.org/10.1080/02626669609491519>.
- Piani, C., J. O. Haerter, and E. Coppola, 2010: Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.*, **99**, 187–192, <https://doi.org/10.1007/s00704-009-0134-9>.
- Serrat-Capdevila, A., M. Merino, J. B. Valdes, and M. Durcik, 2016: Evaluation of the performance of three satellite precipitation products over Africa. *Remote Sens.*, **8**, 836, <https://doi.org/10.3390/rs8100836>.
- Shen, Y., A. Xiong, Y. Hong, J. Yu, Y. Pan, Z. Chen, and M. Saharia, 2014: Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *Int. J. Remote Sens.*, **35**, 6843–6858, <https://doi.org/10.1080/01431161.2014.960612>.
- Smith, J., and K. F. Wallis, 2009: A simple explanation of the forecast combination puzzle. *Oxford Bull. Econ. Stat.*, **71**, 331–355, <https://doi.org/10.1111/j.1468-0084.2008.00541.x>.
- Smith, J. A., and W. F. Krajewski, 1991: Estimation of the mean field bias of radar rainfall estimates. *J. Appl. Meteorol.*, **30**, 397–412, [https://doi.org/10.1175/1520-0450\(1991\)030<0397:EOTMFB>2.0.CO;2](https://doi.org/10.1175/1520-0450(1991)030<0397:EOTMFB>2.0.CO;2).
- Stisen, S., and I. Sandholt, 2010: Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. *Hydrol. Processes*, **24**, 879–891, <https://doi.org/10.1002/hyp.7529>.
- Stock, J., and M. Watson, 1999: *Cointegration, Causality and Forecasting: A Festschrift for Clive W. J. Granger*. Oxford University Press, 504 pp.
- , and —, 2003: How did leading indicator forecasts do during the 2001 recession? *FRB Richmond Econ. Quart.*, **89**, 71–90.
- , and —, 2004: Combination forecasts of output growth in a seven-country data set. *J. Forecast.*, **23**, 405–430, <https://doi.org/10.1002/for.928>.
- Taniguchi, K., 2016: Future changes in precipitation and water resources for Kanto Region in Japan after application of pseudo global warming method and dynamical downscaling. *J. Hydrol. Reg. Stud.*, **8**, 287–303, <https://doi.org/10.1016/j.ejrh.2016.10.004>.
- Tekeli, A. E., and H. Fouli, 2016: Evaluation of TRMM satellite-based precipitation indexes for flood forecasting over Riyadh City, Saudi Arabia. *J. Hydrol.*, **541**, 471–479, <https://doi.org/10.1016/j.jhydrol.2016.01.014>.

- Ushio, T., and Coauthors, 2009: A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *J. Meteor. Soc. Japan*, **87A**, 137–151, <https://doi.org/10.2151/jmsj.87A.137>.
- Valdés-Pineda, R., E. M. C. Demaría, J. B. Valdés, S. Wi, and A. Serrat-Capdevilla, 2016: Bias correction of daily satellite-based rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa. *Hydrol. Earth Syst. Sci. Discuss.*, **2016**, 1–28, <https://doi.org/10.5194/hess-2016-473>.
- Water Information System, 2019: Hydrological and Water Quality Database, accessed 3 April 2017, <http://www1.river.go.jp/>.
- Woldemeskel, F. M., B. Sivakumar, and A. Sharma, 2013: Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *J. Hydrol.*, **499**, 167–176, <https://doi.org/10.1016/j.jhydrol.2013.06.039>.
- Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang, 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.*, **50**, 2693–2717, <https://doi.org/10.1002/2013WR014710>.
- Xie, P., and A.-Y. Xiong, 2011: A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res.*, **116**, D21106, <https://doi.org/10.1029/2011JD016118>.
- , S.-H. Yoo, R. Joyce, and Y. Yarosh, 2011: Bias-corrected CMORPH: A 13-year analysis of high-resolution global precipitation. *Geophysical Research Abstracts*, Vol. 13, Abstract EGU2011-1809, <https://meetingorganizer.copernicus.org/EGU2011/EGU2011-1809.pdf>.
- , R. J. Joyce, S. Wu, S.-H. Yoo, Y. Yarosh, F. Sun, and R. Lin, 2017: Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J. Hydrometeor.*, **18**, 1617–1641, <https://doi.org/10.1175/JHM-D-16-0168.1>.
- Xu, W., Y. Zou, G. Zhang, and M. Linderman, 2015: A comparison among spatial interpolation techniques for daily rainfall data in Sichuan Province, China. *Int. J. Climatol.*, **35**, 2898–2907, <https://doi.org/10.1002/joc.4180>.
- Xue, X., Y. Hong, A. S. Limaye, J. J. Gourley, G. J. Huffman, S. I. Khan, C. Dorji, and S. Chen, 2013: Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *J. Hydrol.*, **499**, 91–99, <https://doi.org/10.1016/j.jhydrol.2013.06.042>.
- Yasuda, S., Y. Shimizu, and K. Deguchi, 2016: Investigation of the mechanism of the 2015 failure of a dike on Kinu River. *Soil Found.*, **56**, 581–592, <https://doi.org/10.1016/j.sandf.2016.07.001>.
- Yilmaz, K. K., T. S. Hogue, K.-L. Hsu, H. V. Gupta, S. Sorooshian, and T. Wagener, 2005: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. *J. Hydrometeor.*, **6**, 497–517, <https://doi.org/10.1175/JHM431.1>.