Title
Optimal merging of multi-satellite precipitation data in urban areas

Journal
Theoretical and Applied Climatology, 147(3-4)

ISSN
0177-798X

Authors
Oliazadeh, Arman
Bozorg-Haddad, Omid
Pakdaman, Morteza
et al.

Publication Date
2022-02-01

DOI
10.1007/s00704-021-03895-4

Peer reviewed

Theoretical and Applied Climatology (2022) 147:1697–1712
https://doi.org/10.1007/s00704-021-03895-4

ORIGINAL PAPER

Optimal merging of multi-satellite precipitation data in urban areas

Arman Oliazadeh[1] · Omid Bozorg-Haddad[1] · Morteza Pakdaman[2] · Ramin Baghbani[3] · Hugo A. Loáiciga[4]

Received: 8 January 2020 / Accepted: 7 December 2021 / Published online: 23 January 2022

Abstract

This paper develops and applies algorithms for optimally merging satellite precipitation products with rain-gauge precipitation for accurate rainfall estimation. The satellite-based precipitation products (SBPs) PERSIANN-CDR, TMPA-3B42, GPM-IMERG, and GSMaP MKV are combined and evaluated to generate accurate rainfall estimates. Daily satellite precipitation data are compared with corresponding station data to calculate the bias for the period 2014–2019. Three different algorithms are proposed whose adjustable parameters are optimally determined by solving constrained optimization algorithms to produce combinations of satellite-based precipitation products. The optimal combination is named optimally merged satellite-based precipitation (OMSBP). The root mean square error (RMSE), coefficient correlation (CC), and the Nash–Sutcliffe error (NSE) are employed to test the proposed method with precipitation data for the Tehran urban region, Iran. The spatially resolved results over the studied urban area establish that TMPA-3B42, with an average value MAE, MBE, and RMSE equal to 0.68 mm, − 0.31 mm, and 2.94 mm, leads to better estimates of precipitation than those of PERSIANN-CDR, IMERG, and GSMaP MKV. Moreover, algorithms alg7 and alg8 yielded better results with respect to the MAE and MBE, respectively. Lastly, algorithm alg3 produced better results than alg7 and alg8 based on the RMSE, NSE, and CC corresponding to precipitation predictions.

1 Introduction

Precipitation is a significant component of the water cycle and the principal driving hydrologic flux. Remote sensing is a powerful tool for determining hydrologic indexes that are used in various fields of water resources (Kim et al. 2014; Zhang et al. 2019a, 2019b, 2019c; Chawla et al. 2020; Isnain and Ghaffar 2021; Duan et al. 2021). Estimation of precipitation at the local and global scales is essential for quantifying hydrologic balances and for accurate hydrologic modeling (Sun et al. 2018; Mahmood et al. 2019; Foufoula-Georgiou et al. 2020). Many developing countries have low-density precipitation monitoring networks and have limited local-scale monitoring (Sharifi et al. 2016; Tiwari et al. 2020). For instance, high populated metropolitans are threatened by natural disasters, and their inhabitants are vulnerable to flooding by heavy rainfall. Satellites can provide useful datasets for precipitation monitoring and prediction in areas where there are few ground stations (Smith and Rodriguez 2017; Mahtab et al. 2018; Yang et al. 2019; Ogato et al. 2020; Oliazadeh et al. 2021).

There are multiple studies of satellite-based precipitation and its applications to hydrologic modeling. Those studies have encompassed catchments in Africa (Dembélé and Zwart 2016; Guilloteau et al. 2016), Asia (Kim et al. 2017; Vu et al. 2018), Australia (Khan et al. 2018), Europe (Duan et al. 2016), North

✉ Omid Bozorg-Haddad
OBHaddad@ut.ac.ir

Arman Oliazadeh
Arman.Oliazadeh@ut.ac.ir

Morteza Pakdaman
pakdaman@cri.ac.ir

Ramin Baghbani
rb2132@msstate.edu

Hugo A. Loáiciga
Hugo.Loaiciga@ucsb.edu

[1] Department of Irrigation & Reclamation Engineering, College of Agriculture & Natural Resources, University of Tehran, Tehran, Iran

[2] Disasters and Climate Change Group, Climatological Research Institute (CRI), Atmospheric Science and Meteorological Research Center (ASMERC), Mashhad, Iran

[3] Department of Agriculture and Biological Engineering, Mississippi State University, Starkville, MS 39762, USA

[4] Department of Geography, University of California, Santa Barbara, CA 93016-4060, USA

America (Maggioni et al. 2016), South America (Salio et al. 2015), and the world (Derin et al. 2016).

Several satellite-based precipitation products such as TMPA 3B42RT and 3B42V7, Climate Prediction Center morphing (CMORPH) technique, PERSIANN-CDR precipitation products, and Global Satellite Mapping of Precipitation–gauge adjusted (GSMaP-Gauge) were evaluated concerning their effectiveness in detecting intense rainfall in cities of China (Huang et al. 2014; Lu et al. 2018; Jiang et al. 2018; Ren et al. 2018; Li et al. 2019). Their major results show SBPs underestimated rainfall and exhibited significant deviations from temporal rainfall variations. Moreover, the GSMaP-Gauge featured the best results at the daily temporal scale, and all other SBPs underestimated extreme precipitation. In addition, several applications of satellite-based precipitation for streamflow simulation have been reported to produce acceptable results (Jiang et al. 2017; Ma et al. 2018; Wei et al. 2018; Zhu et al. 2019; Liu et al. 2020; Chao et al. 2021).

Various combinations and merging of widely used satellite rainfall estimations were evaluated over regions with variable topographic and climate conditions (AghaKouchak et al. 2012; Golian et al. 2015; Nie et al. 2016; Hazra et al. 2019; Mastrantonas et al. 2019). A combination of the individual satellite precipitation products (SPPs) may provide a dataset with a higher correlation with gauge data than individual satellite products (Beck et al. 2017; Yang et al. 2017; Khairul et al. 2018). Also, several regression-based methods for distributed-data applications have been reported (Wang and Li 2021; Wang et al. 2021), and specialized algorithms for merging daily precipitation data from several sources have been implemented for local and global scale predictions, such as the Multi-Source Weighted-Ensemble Precipitation (MSWEP) applied in Australia and Africa (Awange et al. 2019).

Few optimization approaches have been reported for deriving satellite-based precipitation products useful for improving predictions in the field of water resources studies. The optimal merging of SPPs is helpful for applications such as early warning and flood control. The four sets of satellite-based precipitation applied in this study are useful for water resource management, particularly in poorly gauged and ungauged basins.

The main objectives of this study are (1) evaluating the performance of the SBPs relative to the available ground rainfall measurements in Tehran city and (2) further improving the SBPs' rainfall estimates by applying two fitting optimal merging techniques simultaneously. Achieving these objectives yields accurate rainfall datasets that can be applied in water resources management.

Satellite / Period	1	2	...	j	...	n
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1n}
2	x_{21}	x_{22}	...	x_{2j}	...	x_{2n}
3	x_{31}	x_{32}	...	x_{3j}	...	x_{3n}
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{in}
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$
m	x_{m1}	x_{m2}	...	x_{mj}	...	x_{mn}

Reported data from the j-th satellite for all time periods

Fig. 1 Illustration of the data table

2 Study area

Tehran, the capital of Iran, is located in northern Iran. Its total area is about 730 km^2 (Shahbazi et al. 2016), with a latitudinal range of 35° 33′ N to 35° 53′ N and a longitudinal range of 51° 6′ E to 51° 36′ E (Delfani et al. 2010), as shown in Fig. 1. Tehran's climate is mostly defined by its topographic characteristics, with the Alborz Mountains to the north, and the country's central desert to the south in which elevation varies from 1000 to 2300 m (Keikhosravi 2019). The average annual temperature ranges between 24 and 26 °C, although the highest temperatures in summer can reach 43 °C, and the lowest temperatures in winter can be as low as − 30 °C.

The average annual precipitation in Tehran equals 245.8 mm/year, according to observed precipitation data from 1951 through 2015. Regional differences in annual precipitation are considerable. The maximum average annual precipitation can reach 426 mm/year at the Shemiran station, which is located in the northern section of Tehran, while the minimum average annual precipitation is about 123.8 mm/year at the Geophysics station, which is placed in the central area of the city.

3 Data sets

3.1 Gauge stations

The gauge precipitation data utilized in the present work were provided by the Iran Water Resources Management Co. (IWRM) based on a network of daily rain gauge data. The daily ground observed rainfall data are derived from four rain gauges distributed over the study regions corresponding to the period 2014 through 2019. The characteristics of the four ground stations are listed in Table 1. Data preprocessing was performed and multi-year daily means were used to fill in the missing data.

3.2 Satellite-based precipitation

Four satellite-based precipitation products (SBPs) (listed in Table 2) are evaluated in this study with data for the Tehran region corresponding to the period March 2014 through January 2019. The SBPs are the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PERSIANN-CDR) from the University of California, Irvine, the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA-3B42) from the National Aeronautics and Space Administration (NASA), the Global Satellite Mapping

Table 1 Calculated contingency indexes

Station	POD				FAR				CSI			
	TMPA	PERSIANN	GPM	GSMaP	TMPA	PERSIANN	GPM	GSMaP	TMPA	PERSIANN	GPM	GSMaP
Chitgar	0.48	0.88	0.83	0.60	0.59	0.71	0.66	0.62	0.28	0.28	0.32	0.31
Mehrabad	0.46	0.90	0.82	0.88	0.46	0.61	0.52	0.61	0.33	0.38	0.44	0.37
Shemiran	0.59	0.95	0.80	0.60	0.68	0.71	0.60	0.62	0.26	0.28	0.36	0.30
Geophysics	0.50	0.88	0.80	0.84	0.59	0.74	0.68	0.74	0.29	0.25	0.29	0.24

Table 2 Coverage and spatiotemporal resolutions of used SBPs in this study

Data	Full name	Spatial/temporal resolution	Source	Coverage
3B42	TRMM Multi-satellite Precipitation Analysis (TMPA) research product 3B42	3 h/0.25°	NASA	50°N–50°S
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (CDR)	Daily/0.25°	University of California	60°N–60°S
IMERG	Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) mission	0.5 h/0.1°	NASA	60°N–60°S
GSMaP-MVK	Global Satellite Mapping of Precipitation	1 h/0.1°	JAXA	60°N–60°S

of Precipitation in near real-time (GSMaP MKV) from the Japan Aerospace Exploration Agency (JAXA), and the Integrated Multi-satellite Retrievals for GPM (IMERG) from NASA.

3.2.1 TRMM

The TRMM (Tropical Rainfall Measuring Mission) is a joint satellite mission between NASA (The National Aeronautics and Space Administration of the USA, Washington, USA) and JAXA (Japan Aerospace Exploration Agency, Tokyo, Japan). It was launched in November 1997 mainly to measure tropical and subtropical precipitation. The TRMM features Precipitation Radar, Microwave Imager, and Visible Infrared Scanner as its three types of sensors. TMPA-3B42 is the multiple-adjusted daily precipitation data freely available since 1 January 1998. The product is available at relatively good spatial and temporal scales (0.25° × 0.25° and 3 h) and covers from 50° N to 50° S (Huffman et al. 2007).

3.2.2 PERSIANN

The PERSIANN data were produced by using artificial neural network algorithms to estimate the rainfall rate based on longwave IR images from geostationary earth-orbiting satellites. Data collection implemented a versatile preparing technique for refreshing the system parameters whenever independent estimates of rainfall are accessible. The PERSIANN system is improved by infrared bandwidth and daytime based on geostationary infrared imagery. Rainfall data with a spatial coverage of 60° S–N and spatial resolution of 0.25° × 0.25° are available from March 2000 through the present (Ashouri et al. 2015).

Table 3 Evaluation indexes of error statistics

Statistical indicator	Formula	Perfect value	Unit
Mean absolute error	$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \left| \left(P_{S_i} - P_{g_i} \right) \right|}{n}$	0	mm
Mean bias error	$\mathrm{MBE} = \frac{\sum_{i=1}^{n} (P_{S_i} - P_{g_i})}{n}$	0	mm
Root mean absolute error	$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(P_{S_i} - P_{g_i} \right)^2}{n}}$	0	mm
Correlation coefficient	$\mathrm{CC} = \frac{\sum_{i=1}^{n} \left(P_{s_i} - \bar{P}_s \right) \left(P_{g_i} - \bar{P}_g \right)}{\sum_{i=1}^{n} \left(P_{s_i} - \bar{P}_s \right)^2 \sum_{i=1}^{n} \left(P_{g_i} - \bar{P}_g \right)^2}$	1	NA
Nash–Sutcliffe error	$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(P_{s_i} - \bar{P}_{g_i} \right)^2}{\sum_{i=1}^{n} \left(P_{g_i} - \bar{P}_g \right)^2}$	1	NA

P_g denotes rain gauges rainfall (mm). P_s denotes estimated satellite rainfall (mm). $\overline{P}_g$ and $\overline{P}_S$ represent the mean values of gauge- and satellite-based rainfall, respectively. *n* denotes the number of data

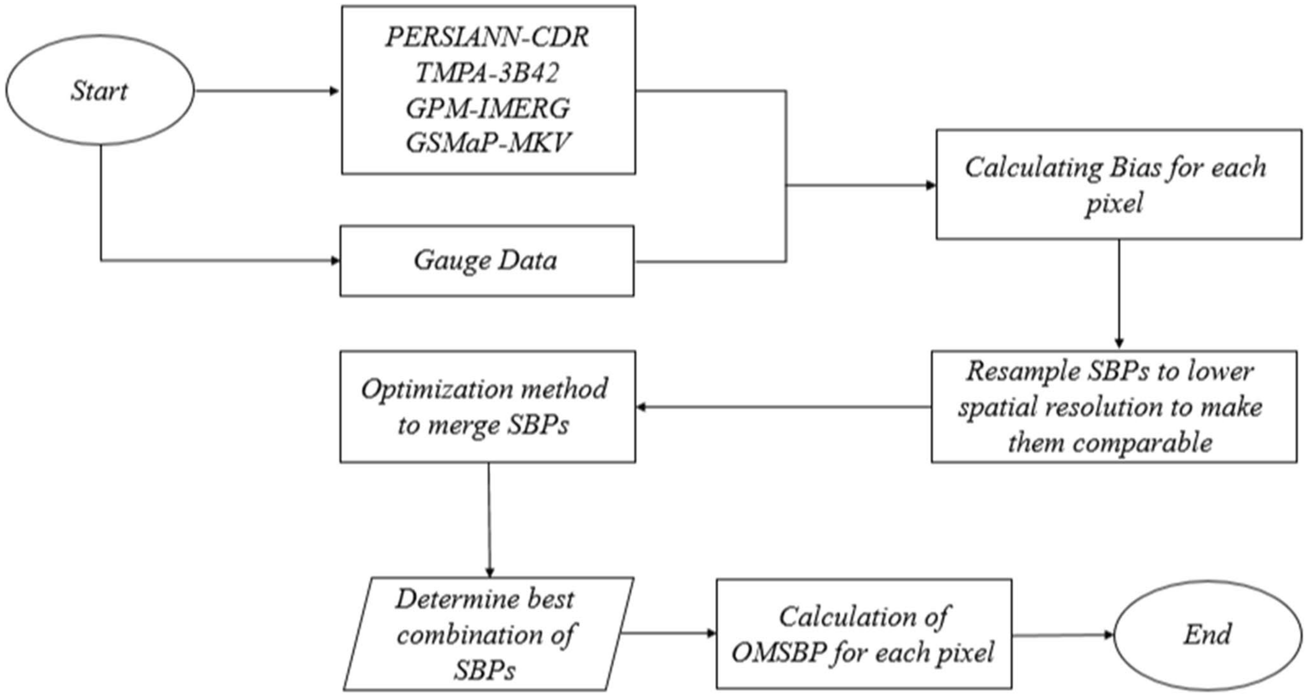

Fig. 2 Flowchart of this study's methodology

3.2.3 GPM

The GPM was launched in February 2014 and currently works in a non-sun-synchronous orbit with a tendency point of 65° to calculate light rain, snowfall, and heavy tropical rainfall as part of a NASA–JAXA cooperative program. GPM improves TRMM's worldwide coverage, providing modern satellite instrumentation, the inter-calibration of datasets from other microwave radiometers, composed combined precipitation datasets, diminished dormancy for conveying information items, simplified data access, extended global ground-validation efforts, and integrated client applications (Huffman et al. 2015).

3.2.4 GSMaP

The GPM-based GSMaP products used in this study include the standard products (MVK_V4). The GSMaP MVK product with high spatial (0.1°) and temporal (1 h) resolution is produced using the passive microwave radiometer data and infrared (IR) data with a Kalman filter to retrieve the precipitation rate generated by an atmospheric moving vector derived from two progressive IR images (Aonashi et al. 2009). The GSMaP MVK was adjusted to the observed global rainfall data from the Climate Prediction Center, evolving into GSMaP_GAU (Kubota et al. 2007). The GSMaP datasets are available in the G-Portal data service system (http://www.gportal.jaxa.jp).

4 Methods

4.1 Evaluation of satellite products

All daily satellite-based precipitation datasets were resampled to a spatial resolution of $0.1° \times 0.1°$ using the standard bilinear interpolation method suited for gridded datasets to make them comparable (Yang and Geng 2016), for evaluating the performance of the SBPs, and to compare them with gauge and OMSBPs (optimally merged satellite-based precipitations) data. Satellite precipitation products of fine resolution can be produced employing spatial interpolation techniques that are broadly used to produce a better estimation of precipitations. Yet, the pattern of precipitation is influenced by elevation in urban districts. Utilizing downscaling procedures with an emphasis on topography increases the accuracy of small-scale satellite precipitation.

The satellite pixels that contained at least one rain gauge were evaluated in this study, whereas other pixels with no rain gauges were excluded from the analysis. Several statistics herein employed (Ebert 2007) are listed in Table 3, and they are as follows:

The mean absolute error (MAE) is used to represent the first order of the discrepancies which shows the average magnitude of the error. The mean bias error (MBE) provides an estimate of the average error in the data. The closer value the MBE is to zero, the higher the accuracy.

Table 4 Contingency indices

Contingency indices	Formula	Best value
POD	$POD = \frac{\alpha}{\alpha+\beta}$	1
FAR	$FAR = \frac{\mu}{\alpha+\mu}$	0
CSI	$CSI = \frac{\alpha}{\alpha+\beta+\mu}$	1

α denotes the number of precipitation events that are correctly detected by the satellite. μ denotes the number of precipitation events detected by satellites that have no observed data at a station, and β denotes the number of precipitation events that have precipitation data at a station but were not detected by satellite-derived precipitation data

The root mean square error (RMSE) also measures the average error magnitude but gives greater weight to the larger errors (Liu et al. 2017; Ma et al. 2018). The correlation coefficient (CC) measures the degree of a linear statistical association between satellite precipitation and rain gauge observations. The value of CC ranges between − 1 and + 1. The value of + 1 indicates a perfect positive linear statistical association, while − 1 represents a negative statistical association. The CC is close to 0 whether there is no linear correlation or there is a weak linear correlation (Zhang et al. 2019a, 2019b, 2019c, 2019d). The Nash–Sutcliffe error (NSE) shows how much the observed data correspond to the satellite data. The NSE ranges from $-\infty$ to 1, and the closer it is to 1, the more accurate the predictions are (Belabid et al. 2019).

Furthermore, three Contingency Statistics indexes are herein applied to evaluate the accuracy of four satellite precipitation products (Sharifi et al. 2016). They are the probability of detection (POD), a false alarm ratio (FAR), and the critical success index (CSI). The POD measures the ratio of satellite precipitation detection to observed precipitation events. The FAR represents the ratio of the precipitation events falsely identified by the satellite products. The CSI is a proportion of total precipitation events which are correctly detected by the satellite products.

These contingency indices were calculated and are listed in Table 4. The satellite precipitation products detect precipitation more accurately if POD and CSI are close to 1, and the FAR is close to 0.

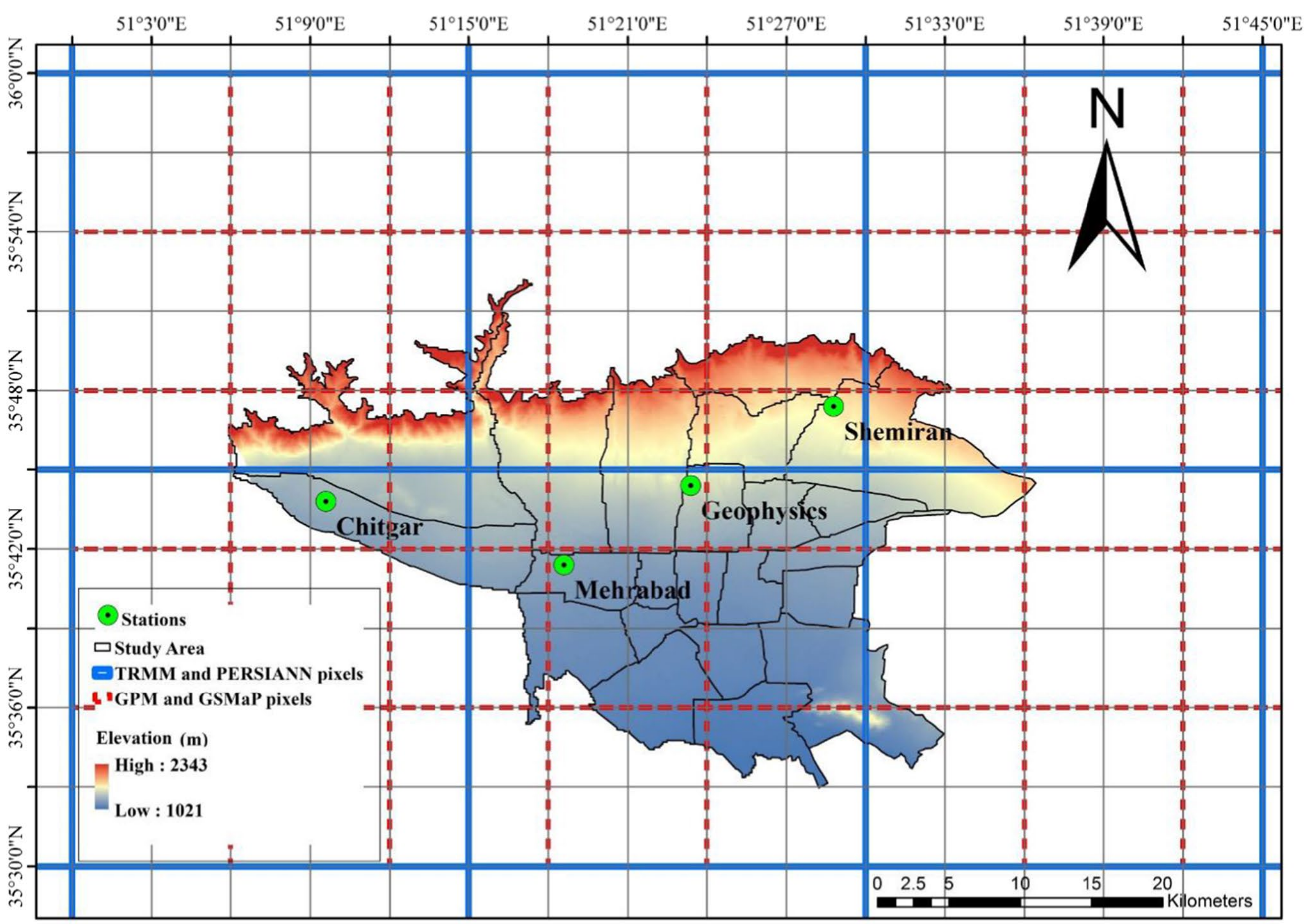

Fig. 3 Location of hydrometric stations for gauge data and the various pixels of satellite-based precipitation products in Tehran city

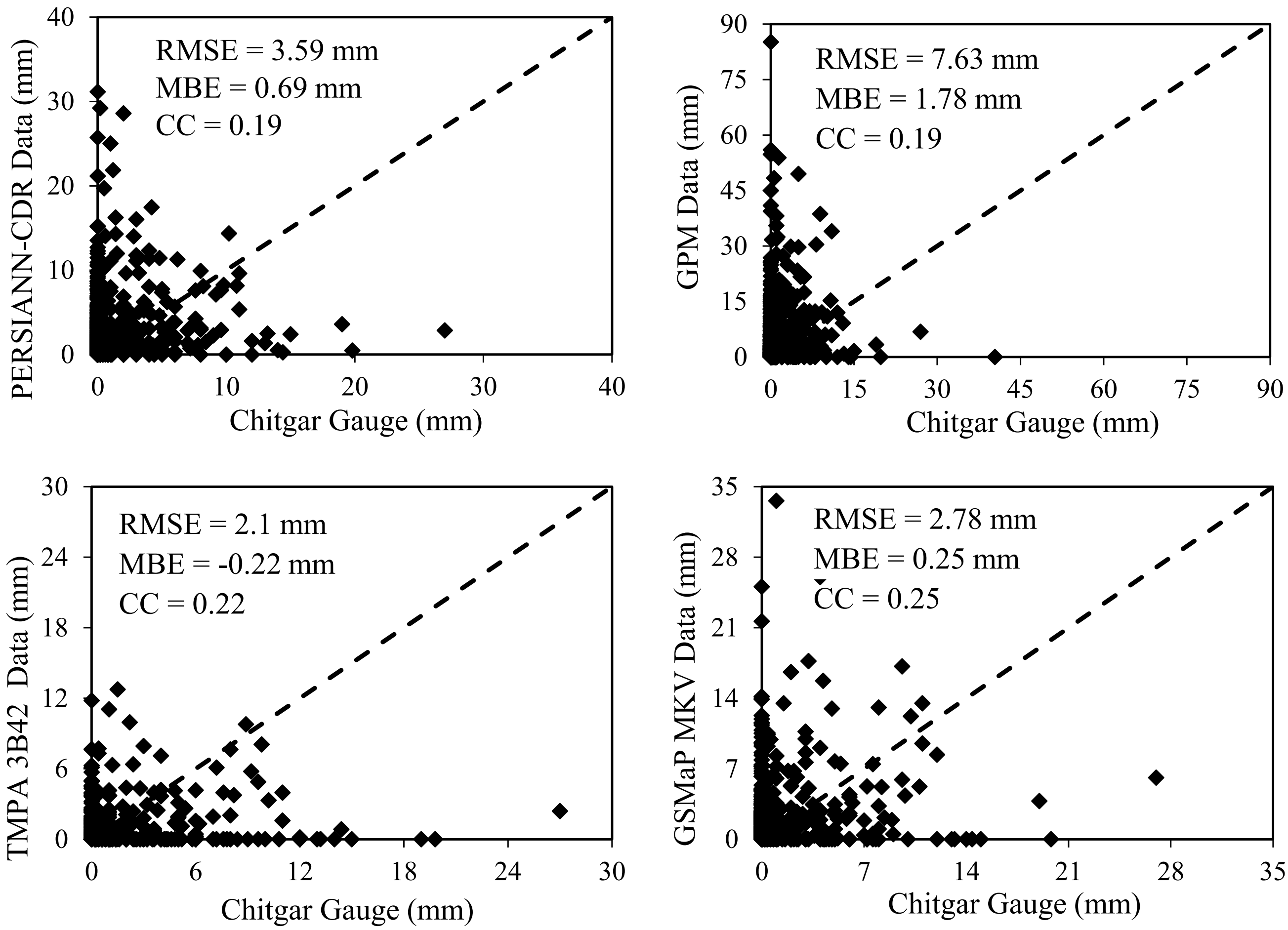

Fig. 4 Daily rainfall scatterplots of the four SBPs and Chitgar station data

4.2 Data merging

This paper's methodology combines the SBPs into an optimally merged satellite-based precipitation (OMSBP) with the least bias in each pixel. Figure 2 shows the flowchart of the optimal merging algorithm. There are four SBPs in each pixel with at least one rain gauge, namely PERSIANN-CDR, TMPA-3B42, GPM-IMERG, and GSMaP MKV. The average of daily biases (additive bias) is calculated between each SBPs and observed rainfall values from rain gauges throughout 2014–2019.

Let $x = (x_{ij})_{m \times n}$ denote the matrix data from n satellites in m observational periods, in which, x_{ij} denotes the amount of precipitation reported by the j-th satellite in the i-th period (i and j range from 1 to m, and from 1 to n, respectively). We denote the j-th column of the matrix x by x_j which indicates the reported data from the j-th satellite for all periods. Figure 3 illustrates the format of the table which contains the data. This work employs two fitting models. The first is based on the power function, and the second is a linear regression approach.

4.2.1 Power function fitting

Consider fitting the following nonlinear function to the datasets:

$$y = a_1 {x_1}^{a_{n+1}} + a_2 {x_2}^{a_{n+2}} + \ldots + a_n {x_n}^{a_{2n}} \tag{1}$$

Or equivalently

$$y = \sum_{j=1}^{n} a_j x_j^{a_{n+j}} \tag{2}$$

where the coefficients a_j and a_{n+j}, $j = 1, 2, \ldots, n$ must be determined such that Eq. (1) optimally fit the data, which consists of observations at time i, b_i. In Eq. (2) each x_j and y are vectors of length m (the number of periods), and $x_j^{a_{n+j}}$ denotes that the power function is applied elementwise. This work applies non-negative constraints $a_j \geq 0$ to avoid

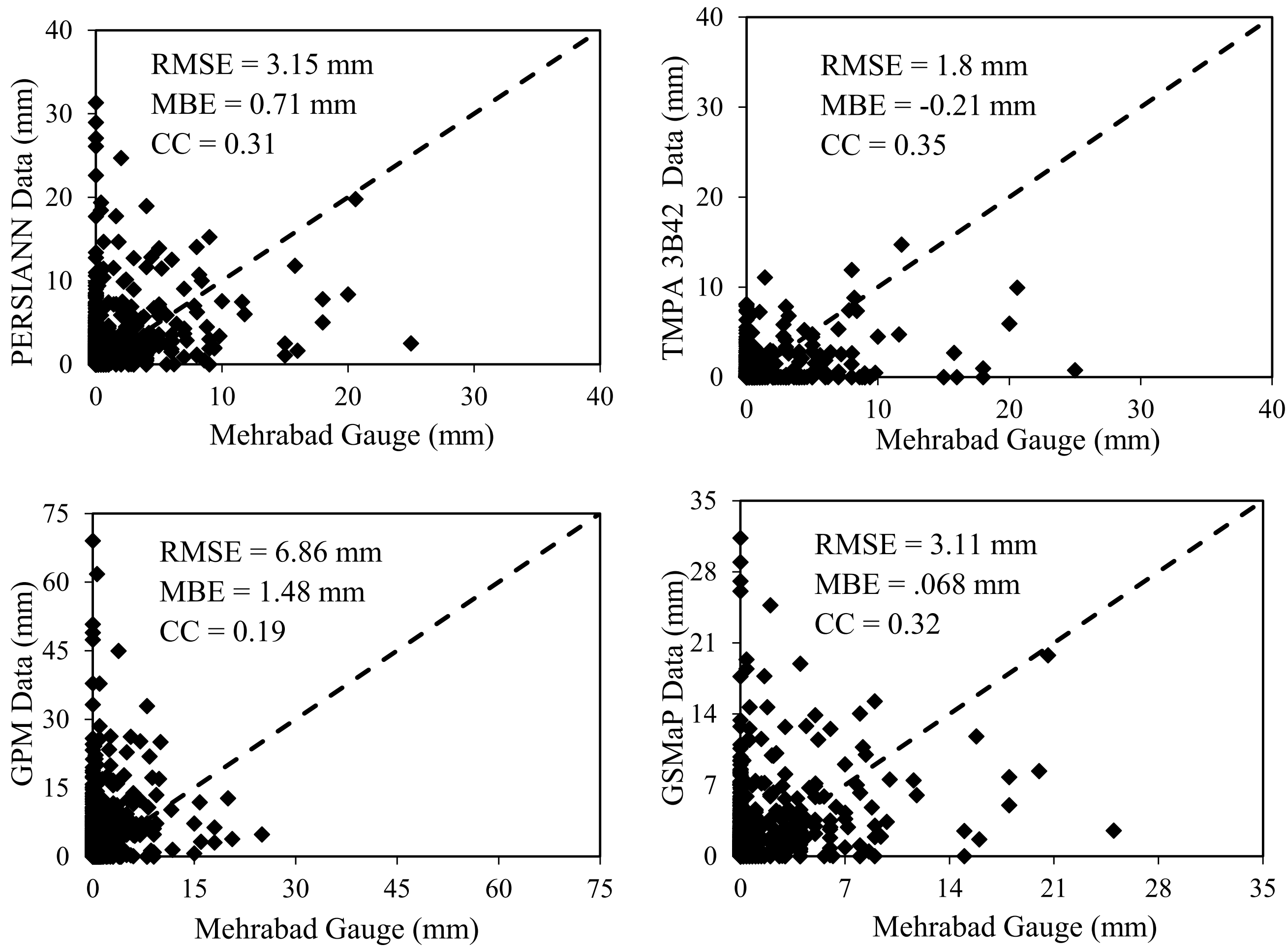

Fig. 5 Daily rainfall scatterplots of the four SBPs and Mehrabad station data

negative values of precipitation. This yields the following constrained optimization problem:

$$\mathrm{Min}\varphi(a) = \sum\nolimits_{j=1}^{n} \left(\sum\nolimits_{i=1}^{m} ((a_j x_{ij}{}^{a_{n+j}} - b_i)^2) \right) \tag{3}$$

subject to

$$a_k \geq 0, k = 1, 2, \ldots, 2n \tag{4}$$

Equations (3) and (4) define the algorithm 3 (or alg3 for simplicity) in which $\varphi(.)$ denotes the objective function, and a denotes the vector of decision variables (i.e., the unknown coefficientsa_k). As a special case suppose thatx_1,x_2,x_3, x_4 indicate the precipitation data of four satellite datasets PERSIANN, GPM, TMPA, and GSMaP, respectively. Similar to Eq. (1), and for$n = 4$, the following regression combination fits the datasets:

$$y = a_1 x_1^{a_5} + a_2 x_2^{a_6} + a_3 x_3^{a_7} + a_4 x_4^{a_8} \tag{5}$$

In this case, for the objective function of Eq. (3), we have $n = 4$ and thus $\varphi(a) = \varphi(a_1, a_2, \ldots, a_8)$. The objective function of the optimization problem (3) indicates the minimization of the distance between the observations and the estimated values of precipitation. The minimization is performed using adjustable and non-negative parameters $a_k, k = 1, 2, \ldots., 2n$. See Tang et al. (2015) and Tian et al. (2013) for details about using the power function.

4.2.2 Linear function fitting

Consider the following linear regression function:

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \sum_{j=1}^{n} a_j x_j \tag{6a}$$

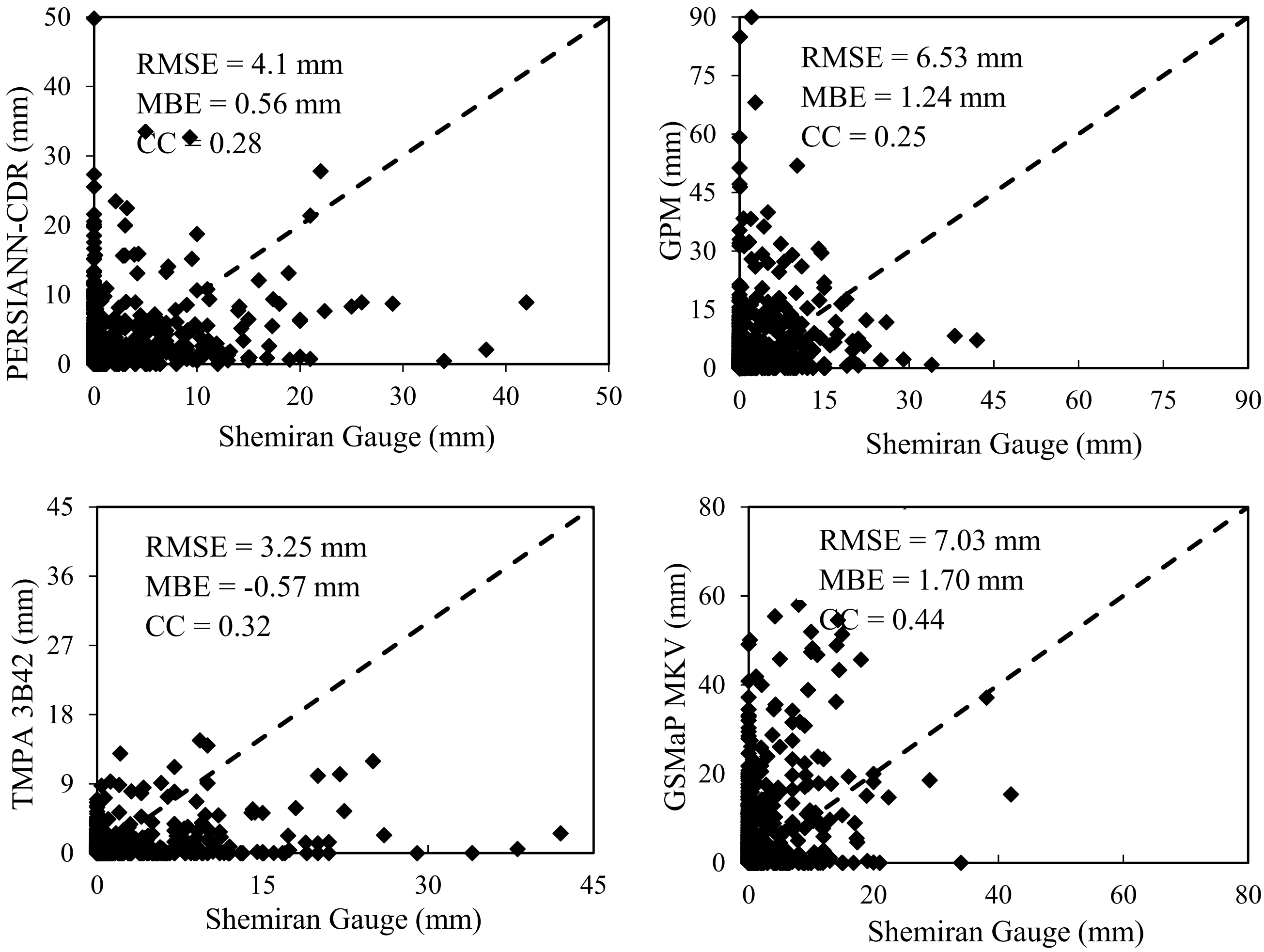

Fig. 6 Daily rainfall scatterplots of the four SBPs and Shemiran station data

Equation (6a) does not include a constant coefficient. In the case of the existence of bias, a constant coefficient is introduced in Eq. (5a) to better fit the data. Thus, add a constant to Eq. (6a) as follows:

$$y = a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_{n+1} = a_{n+1} + \sum_{j=1}^{n} a_jx_j \quad (6b)$$

The following least-square optimization problem must be solved to obtain the optimal values of a_j, in Eq. (6a), which for non-negativity of precipitation requires non-negative constraints:

$$\underset{a_j,\ j=1\ldots n}{\text{Min}} \quad \sum_{j=1}^{n}\left(\sum_{i=1}^{m}\left(\left(a_jx_{ij} - b_i\right)^2\right)\right) \quad (7a)$$

subject to:

$$a_j \geq 0, j = 1, 2, \ldots, n \quad (7b)$$

Problem (7a, b) is a constrained optimization problem that minimizes the distance between the observed amount of precipitation at period i (which is denoted by b_i) and it's an approximated value that is calculated using Eq. (6a). The minimization procedure is performed using the adjustable parameters $a_j, j = 1, 2, \ldots., n$.

Similarly, solve the following problem to obtain the optimal values of a_j in Eq. (6b):

$$\underset{a_j,\ j=1\ldots n+1}{\text{Min}} \quad \sum_{j=1}^{n}\left(\sum_{i=1}^{m}\left(a_jx_{ij} + a_{n+1} - b_i\right)^2\right) \quad (8a)$$

subject to:

$$a_j \geq 0, j = 1, 2, \ldots, n + 1 \quad (8b)$$

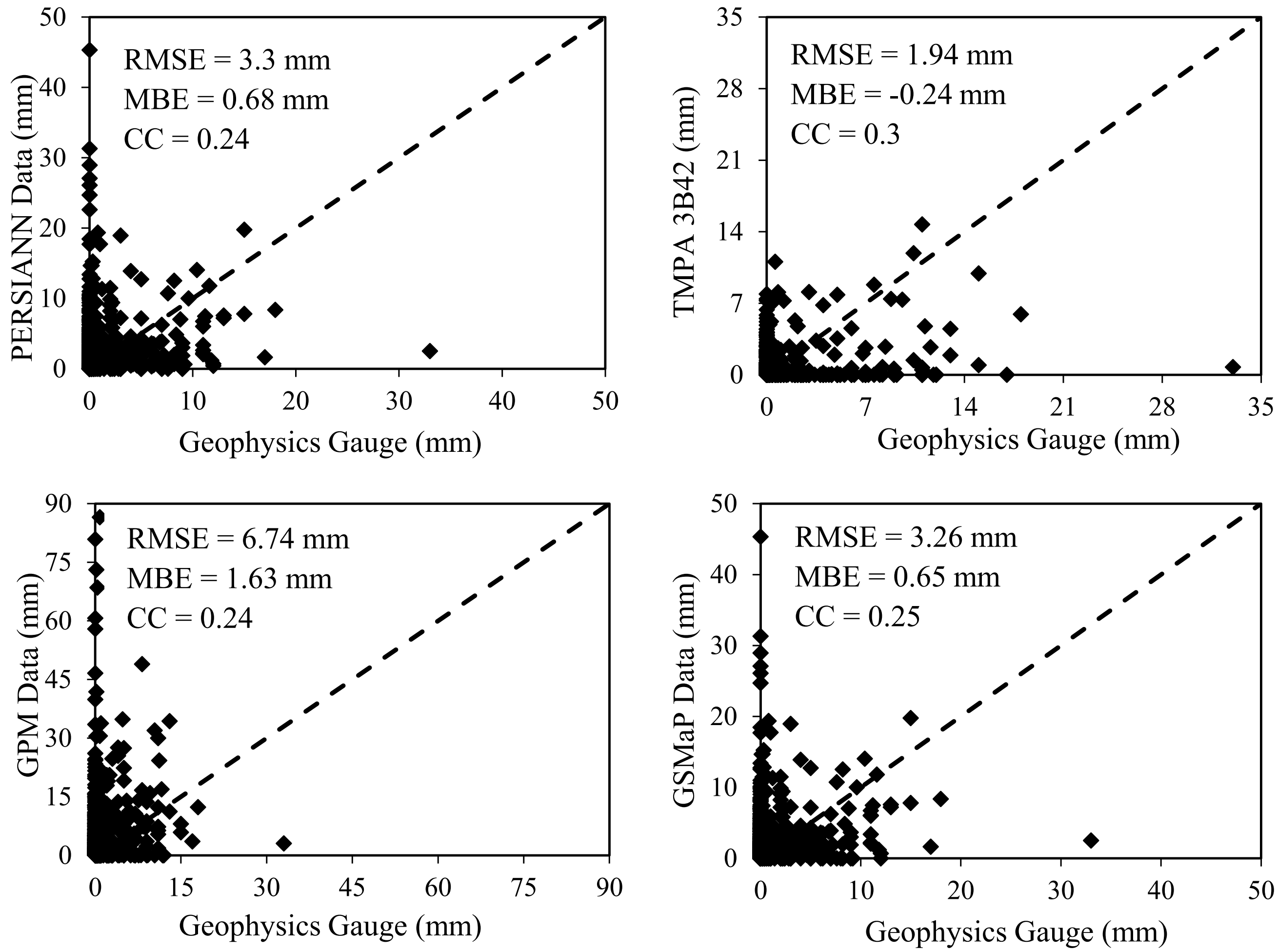

Fig. 7 Daily rainfall scatterplots of the four SBPs and Geophysics station data

Problem (8a, b) is a constrained optimization problem that minimizes the distance between the observed precipitation at period i (which is denoted by b_i) and its estimated value that is calculated with Eq. (6b). The minimization procedure is performed using the adjustable parameters $a_j, j = 1, 2, ..., n + 1$.

Equations (3)–(4), (7a, b), and (8a, b) are constrained nonlinear programming problems that correspond, respectively, to the algorithms herein named alg3, alg7, and alg8. These programming problems are solved with classical nonlinear programming algorithms such as the steepest descent, quasi-Newton methods, or with evolutionary algorithms. This paper relies on the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton method. The BFGS algorithm approximates the Hessian matrix of second derivatives to make its computation more efficient (Bazaraa et al. 2013). It is noteworthy that in our case study, the number of satellites is $n = 4$. The results obtained from the application of this paper's algorithms are discussed in the next section.

5 Results and discussion

5.1 Statistical error analysis

Three statistical indicators of errors on a daily scale were calculated (MBE, MAE, RMSE). The MBE index measures an average overestimation or underestimation. The results of Figs. 4, 5, 6 and 7 demonstrate the TMPA-3B42 data has the smallest deviation from gauge data compared to other satellite sources. Only in the Shemiran region did the PERSIANN produce the best estimate with an error deviation of 0.56 mm. In all four regions, the TMPA satellite data are underestimated as indicated by the negative ranges of the MBE. The IMERG data recorded the most deviation from terrestrial data in all stations by MBE of 1.78, 1.48, 1.24, and 1.63 mm, respectively. However, the GSMaP satellite recorded better performance than PERSIANN in other areas. It is noteworthy that the Mehrabad station is an airport station where aviation activities may affect the satellite-based precipitation estimation.

The MAE indicates that the TMPA-3B42 satellite data performed better in all four studied regions. The best results based on the MAE correspond to the Mehrabad station for TMPA-3B42 data with a value of 0.52 mm. The next best data performance was that of the GSMaP data in the Chitgar, Mehrabad, and Geophysical regions where the MAE equaled 0.8, 1, 1.2 mm, respectively, and PERSIANN at the Shemiran station with an MAE of 1.7 mm. It is concluded that the IMERG data has the poorest accuracy compared to other data sources, except in the Shemiran region according to the values of MAE.

The RMSE results show that the TMPA-3B42 data performed better than other data sources, and it equaled 0.5 mm in the Mehrabad region. The next best one was the GSMaP satellite data in the Cheitgar, Mehrabad, and Geophysical areas. The PERSIANN had better estimates than other data sources in the Shemiran region. The IMERG data had the poorest results with an RMSE ranging between 6.5 and 7.6 mm. The GSMaP satellite in the Shemiran region performed poorly concerning the three evaluation indices compared to the other areas. It is concluded that due to the higher elevation of the Shemiran region, the GSMaP satellite performs with lower accuracy of precipitation estimation in high altitudes. Also, the TMPA-3B42 data estimation is more accurate in this region than elsewhere, although it has acceptable performance in all other stations for all indices.

5.2 Contingency index analysis

It is seen in Table 5 that PERSIANN's precipitation observations are more accurate than other observations. This source correctly detected at least 88% of rainfall events and performed the best detection at the Shemiran station with a POD = 0.95. The TMPA-3B42 data performed more poorly in detecting daily rainfall events than the other three sources, with a POD = 0.46 for the Mehrabad station, and the index variations were negligible in the other three regions.

Concerning the FAR index, the results indicate the TMPA-3B42 data has a lower FAR index than other Mehrabad regions, which means that it recorded less false precipitation than other sources. Also, FAR has the highest percentage for the Geophysics station concerning PERSIANN and GSMaP by 74%. The PERSIANN data performs better than ground-based data detection, yet, it recorded more false precipitation than other satellites and has the weakest performance for this index. This indicator is better for GPM than the GSMaP satellite in the Mehrabad, Shemiran, and Geophysics regions. Only the GSMaP performance is better in the Chitgar region.

Table 5 Characteristics of the four ground observation stations in Tehran city

Station No	Station Name	Elevation (m)	Latitude (degree)	Longitude (degree)	Average annual precipitation (mm/year)
1	Chitgar	1305	35.73	51.16	241.3
2	Mehrabad	1191	35.69	51.31	232.8
3	Shemiran	1549	35.79	51.48	426.2
4	Geophysics	1415	35.74	51.39	165.8

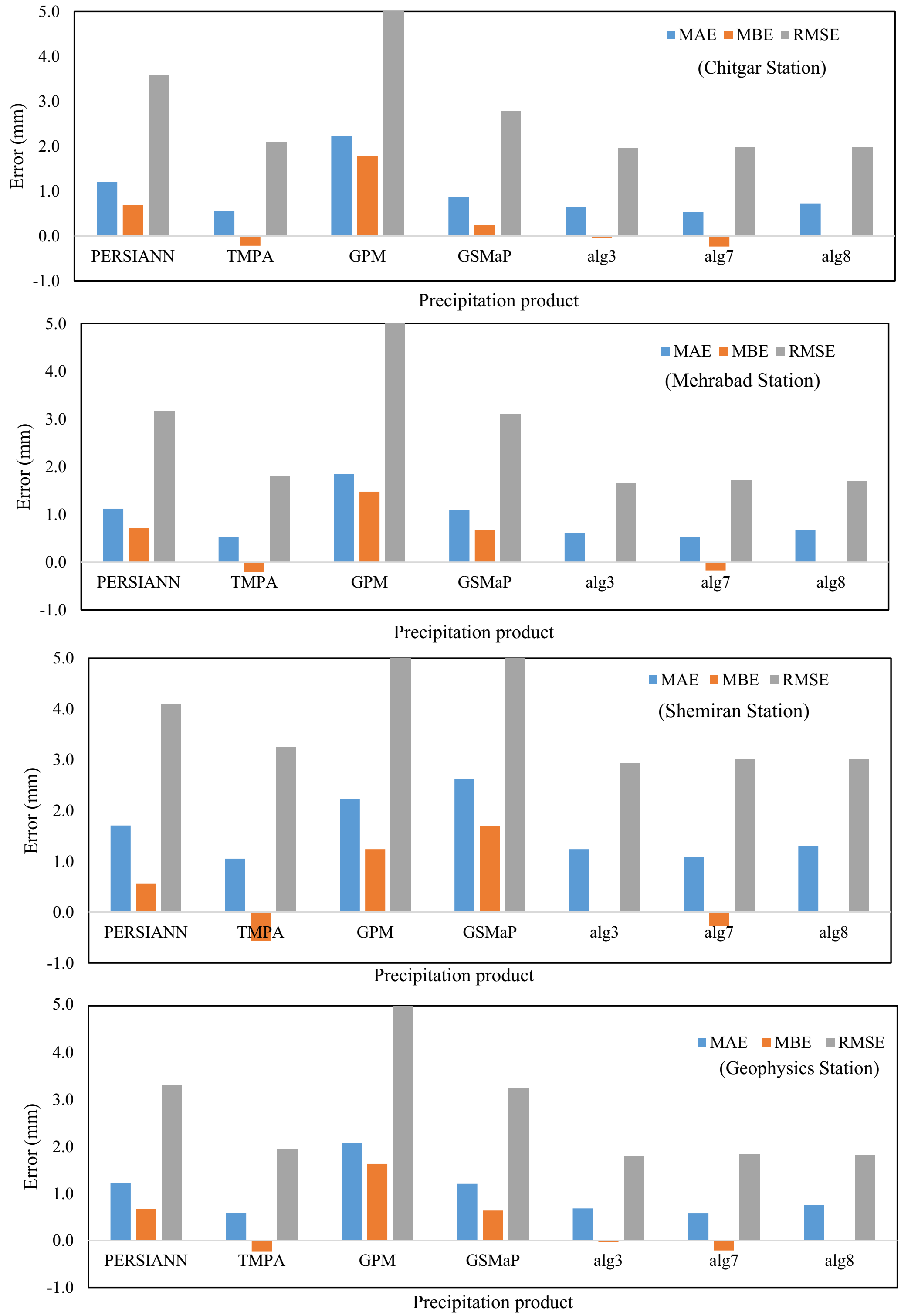

Fig. 8 Results for models involving algorithms alg3, alg7, and alg 8 at all stations

These results establish that the PERSIANN recorded more precipitation events than the other satellites, many of which were false and did not occur in reality.

The CSI, which is complementary to the POD index, shows the percentage of success in precipitation event detection for all recorded events. According to CSI

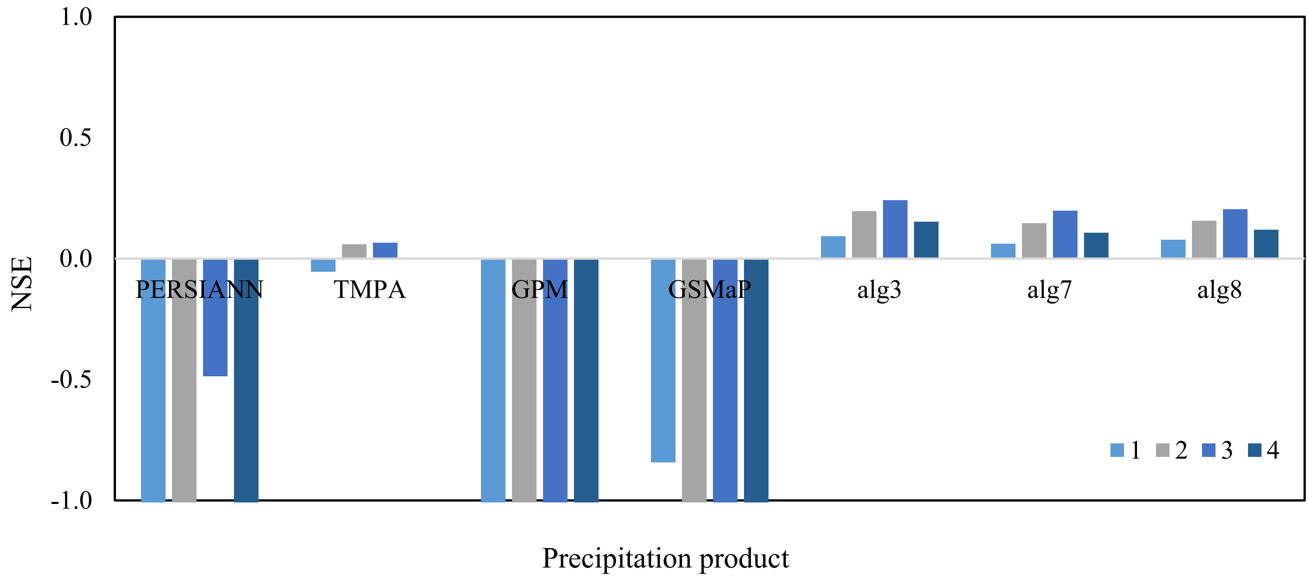

Fig. 9 Results corresponding to the NSE index for models involving algorithms alg3, alg7, and alg8 at stations based on numbers listed in Table 4

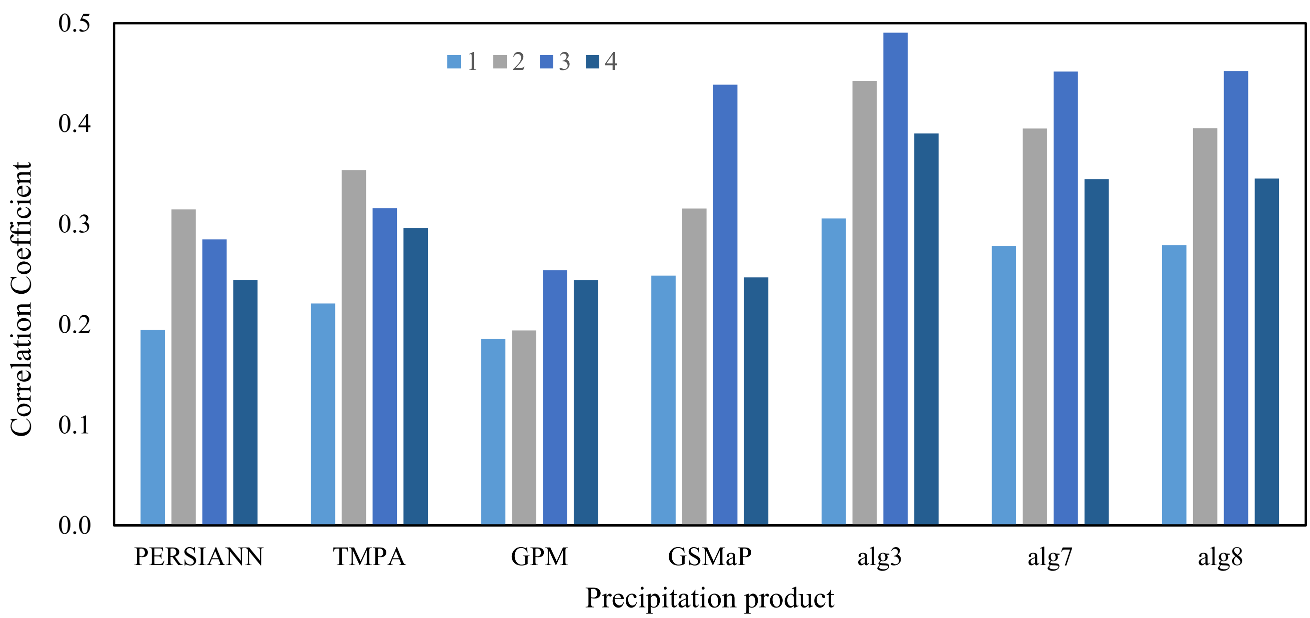

Fig. 10 Results corresponding to the CC index for models involving algorithms alg3, alg7, and alg8 at stations based on numbers listed in Table 4

results presented in Table 5, the GPM satellite data performs better in detecting rainfall events, and it had better accuracy in distinguishing non-rainfall events. The GPM satellite recorded better results than other data sources concerning the POD index except for PERSIANN and also had better results than other data sources for the FAR index except for TMPA.

5.3 Optimally merging of SBPs

It is seen in Fig. 8 that the TMPA data had better results for the Chitgar station judged by the best MAE value among all other satellites and methods. Also, among the three proposed algorithms, alg7 has better performance of MAE by 0.53 mm at the Chitgar and Mehrabad stations, while comparing TMPA and alg3 indicates the TMPA is slightly better for the MAE. Also, three merging techniques have the same and best MBE of about 1.7 mm at Mehrabad Station.

Figure 9 reveals the TMPA-3B42 data has better performance for the NSE among all satellites where it was in the range − 0.05 to 0.06, while alg3 generally is better among the three proposed algorithms. The TMPA-3B42 and alg3 had superior performance at the Mehrabad station.

Similar to other stations the TMPA-3B42 data outperform other satellite-based models, while alg3 performed better than alg7 and alg8 concerning RMSE, CC, and NSE. TMPA-3B42 is slightly better than alg7 with respect to MAE at the Shemiran station. Lastly, the TMPA-3B42 data are better among the satellite data sources, similar to other stations. alg3 has the best performance for RMSE, CC, and NSE at the Geophysics station.

It could be argued that the best NSE pertains to the TMPA-3B42 data at all stations among the SBPs, while the alg3 has the best NSE for the proposed algorithms where it was 0.09 at Chitgar station according to Fig. 9. Similarly, the GSMaP has the best CC of 0.44 among all other SBPs and is smaller than all implemented algorithms based on Fig. 10. This shows that all algorithms are effective in improving the CC index at all stations, whereas alg3 is the best with respect to the CC with a value of 0.5 at the Shemiran station.

This paper's results establish that algorithms 7 and 8 were the most accurate merging methods within the study area. Figure 9 establishes that all proposed algorithms changed the Nash Sutcliffe coefficient from negative values to positive values, which demonstrates the improvement achieved with the merging algorithms. It would be valuable to find the ideal combination of rain gauges and SBPs to improve the precision of precipitation estimations. These combinations can be applied to reduce the risk and uncertainties of modeling natural disasters such as flooding and droughts under future conditions.

6 Concluding remarks

This paper introduced two optimally data-merging methods and explored the accuracies of four satellite-based precipitation products: PERSIANN-CDR, TMPA-3B42, GPM-IMERG, and GSMaP MKV, for the period 2014–2019 in Tehran. Our results established that TMPA had the best performance among the satellite-based models. Also, the proposed algorithm of alg3 featured the best performance (even in comparison with individual satellite-based data) for RMSE, CC, and NSE. The algorithm alg7 has the best performance concerning the MAE and MBE.

The merging of SBPs is recommended for other urban study zones worldwide. Unquestionably, alternative combining strategies, e.g., a nonlinear mix of SBPs, may be applied to estimate precipitation with higher accuracy than that exhibited by individual data products. It is pivotal to evaluate the performance of SBPs of various temporal and spatial scales, especially in those urban areas that have low-density gauging networks. It is noteworthy that further evaluation must be performed before applying the proposed blending technique to pixels with no rain gauge data, i.e., to ungauged areas. Finally, the comparison of satellite data with the estimates from the proposed merging algorithms indicates that the merging results are more accurate than the satellite estimates and had less uncertainty. Consequently, we recommend the proposed merging procedure be applied in the validation of satellite-based precipitation data.

Future research will (1) integrate the two optimization methods herein presented coupled with an evaluation of goodness-of-fit statistics for the assessment of rainfall estimation errors and (2) incorporate combined and optimized rainfall products based on multi-source satellites to simulate urban streamflow with improved precision relative to current technologies.

Acknowledgements The authors thank Iran's National Science Foundation (INSF) for its support of this research.

Author contribution Arman Oliazadeh: software, formal analysis, writing — original draft.

Omid Bozorg-Haddad: conceptualization, supervision, project administration.

Morteza Pakdaman: software, formal analysis, writing — original draft.

Ramin Baghbani: software, formal analysis, writing — original draft.

Hugo A. Loáiciga: validation, writing, equations — review and editing.

Availability of data and material All of the required data have been presented in our article.

Code availability Any code used in this paper is available upon request.

Declarations

Ethics approval All authors accept all ethical approvals.

Consent to participate All authors consent to participate.

Consent for publication All authors consent to publish.

Conflict of interest The authors declare no competing interests.

References

AghaKouchak A, Mehran A, Norouzi H, Behrangi A (2012) Systematic and random error components in satellite precipitation data sets. Geophys Res Lett, 39(9)

Aonashi K, Awaka J, Hirose M, Kozu T, Kubota T, Liu G, ... Takayabu YN (2009) "GSMaP passive microwave precipitation retrieval algorithm: algorithm description and validation." J Meteorol Soc Japan. Ser. II, 87, 119-136

Ashouri H, Hsu KL, Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., ... & Prat, O. P. (2015). "PERSIANN-CDR: daily precipitation climate data record from multi-satellite observations for hydrological and climate studies." Bulletin of the American Meteorological Society, 96(1), 69-83

Awange JL, Hu KX, Khaki M (2019) The newly merged satellite remotely sensed, gauge and reanalysis-based multi-source weighted-ensemble precipitation: evaluation over Australia and Africa (1981–2016). Sci Total Environ 670:448–465

Bazaraa MS, Sherali HD, Shetty CM (2013) Nonlinear programming: theory and algorithms. John Wiley & Sons

Beck HE, Van Dijk AI, Levizzani V, Schellekens J, Gonzalez Miralles D, Martens B, De Roo A (2017) MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol Earth Syst Sci 21(1):589–615

Belabid N, Zhao F, Brocca L, Huang Y, Tan Y (2019) Near-real-time flood forecasting based on satellite precipitation products. Remote Sensing 11(3):252

Chao L, Zhang K, Yang Z, Wang J, Lin P, Liang J,... Gu Z (2021) Improving flood simulation capability of the WRF-Hydro-RAPID model using a multi-source precipitation merging method. J hydrol (Amsterdam), 592, 125814

Chawla I, Karthikeyan L, Mishra AK (2020) A review of remote sensing applications for water security: quantity, quality, and extremes. J Hydrol 585:124826

Delfani S, Karami M, Pasdarshahri H (2010) The effects of climate change on energy consumption of cooling systems in Tehran. Energy and Buildings 42(10):1952–1957

Dembélé M, Zwart SJ (2016) Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. Int J Remote Sens 37(17):3995–4014

Derin Y, Anagnostou E, Berne A, Borga M, Boudevillain B, Buytaert W, ... Lavado-Casimiro W (2016) "Multiregional satellite precipitation products evaluation over complex terrain." J Hydrometeorol, 17(6): 1817-1836

Duan W, Maskey S, Chaffe PLB, Luo P, He B, Wu Y, Hou J (2021) Recent advancement in remote sensing technology for hydrology analysis and water resources management. Remote Sensing 13(6):1097

Duan Z, Liu J, Tuo Y, Chiogna G, Disse M (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

Ebert EE (2007) Methods for verifying satellite precipitation estimates. In Measuring precipitation from space. Springer, Dordrecht, pp 345–356

Foufoula-Georgiou E, Guilloteau C, Nguyen P, Aghakouchak A, Hsu KL, Busalacchi A, ... Levizzani V (2020) Advancing precipitation estimation, prediction, and impact studies. Bull Am Meteorol Soc 101(9): E1584-E1592

Golian S, Moazami S, Kirstetter PE, Hong Y (2015) Evaluating the performance of merged multi-satellite precipitation products over a complex terrain. Water Resour Manage 29(13):4885–4901

Guilloteau C, Roca R, Gosset M (2016) A multiscale evaluation of the detection capabilities of high-resolution satellite precipitation products in West Africa. J Hydrometeorol 17(7):2041–2059

Hazra A, Maggioni V, Houser P, Antil H, Noonan M (2019) A Monte Carlo-based multi-objective optimization approach to merge different precipitation estimates for land surface modeling. J Hydrol 570:454–462

Huang Y, Chen S, Cao Q, Hong Y, Wu BW, Huang MY, Qiao L, Zhang ZX, Li Z, Yang XQ (2014) 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

Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, ... Stocker EF (2007) "The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales." J hydrometeorol 8(1): 38-55

Huffman G, Bolvin D, Braithwaite D, Hsu K, Joyce R, Xie P (2015) "Integrated multi-satellite retrievals for GPM (IMERG), version 4.4." NASA's Precipitation Processing Center

Isnain Z, Ghaffar SNA (2021) Using the geographical information system (gis) and remote sensing techniques for mapping the groundwater potential zones in kg Timbang Dayang, Kota Belud Sabah. Water Conserv Manag 4(1):57–60. https://doi.org/10.26480/WCM.01.2020.57.60

Jiang Q, Li W, Wen J, Qiu C, Sun W, Fang Q, ... Tan J (2018) "Accuracy evaluation of two high-resolution satellite-based rainfall products: TRMM 3B42V7 and CMORPH in Shanghai." Water 10(1), 40

Jiang S, Liu S, Ren L, Yong B, Zhang L, Wang M, ... He Y (2017) "Hydrologic evaluation of six high-resolution satellite precipitation products in capturing extreme precipitation and streamflow over a medium-sized basin in China." Water 10(1): 25

Keikhosravi Q (2019) The effect of heat waves on the intensification of the heat island of Iran's metropolises (Tehran, Mashhad, Tabriz, Ahvaz). Urban Climate 28:100453

Khairul I, Mastrantonas N, Rasmy M, Koike T, Takeuchi K (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 Sensing 10(6):828

Khan A, Koch M, Chinchilla K (2018) Evaluation of gridded multi-satellite precipitation estimation (TRMM-3B42-V7) performance in the upper Indus Basin (UIB). Climate 6(3):76

Kim K, Park J, Baik J, Choi M (2017) Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. Atmos Res 187:95–105

Kim Y, Kimball JS, Zhang K, Didan K, Velicogna I, McDonald KC (2014) Attribution of divergent northern vegetation growth responses to lengthening non-frozen seasons using satellite optical-NIR and microwave remote sensing. Int J Remote Sens 35(10):3700–3721

Kubota T, Shige S, Hashizume H, Aonashi K, Takahashi N, Seto S, ... Okamoto KI (2007) "Global precipitation map using satelliteborne microwave radiometers by the GSMaP project: production and validation." IEEE Transactions on Geosci Remote Sens 45(7): 2259-2275

Li W, He X, Sun W, Scaioni M, Yao D, Fu J, ... Cheng G (2019) "Evaluating three satellite-based precipitation products of different spatial resolutions in Shanghai based on upscaling of rain gauge." Int J Remote Sens 40(15): 5875-5891

Liu R, Ma Y, Yang Y, Han Z, Tang G, Liu Q, Hong Y (2017) Error analysis of ensemble multi-satellite precipitation datasets over the Tibetan Plateau. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 4684–4687). IEEE

Liu Y, Zhang K, Li Z, Liu Z, Wang J, Huang P (2020) A hybrid runoff generation modeling framework based on spatial combination of three runoff generation schemes for semi-humid and semi-arid watersheds. J Hydrol 590:125440

Lu X, Tang G, Wei M, Yang L, Zhang Y (2018) Evaluation of multi-satellite precipitation products in Xinjiang, China. Int J Remote Sens 39(21):7437–7462

Ma Y, Hong Y, Chen Y, Yang Y, Tang G, Yao Y, ... Liu R (2018) "Performance of optimally merged multisatellite precipitation

products using the dynamic Bayesian model averaging scheme over the Tibetan Plateau." J Geophys Res: Atmos 123(2): 814-834
Maggioni V, Meyers PC, Robinson MD (2016) A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. J Hydrometeorol 17(4):1101–1117
Mahmood GG, Rashid H, Anwar S, Nasir A (2019) Evaluation of climate change impacts on rainfall patterns in the Pothohar region of Pakistan. Water Conservation and Management 3(1):1–6. https://doi.org/10.26480/wcm.01.2019.01.06
Mahtab MH, Ohara M, Rasmy M (2018) The impact of rainfall variations on flash flooding in haor areas in Bangladesh. Water Conservation and Management 2(2):6–10. https://doi.org/10.26480/wcm.02.2018.06.10
Mastrantonas N, Bhattacharya B, Shibuo Y, Rasmy M, Espinoza-Dávalos G, Solomatine D (2019) Evaluating the benefits of merging near-real-time satellite precipitation products: a case study in the Kinu basin region, Japan. J Hydrometeorol 20(6):1213–1233
Nie S, Wu T, Luo Y, Deng X, Shi X, Wang Z, ... Huang J (2016) "A strategy for merging objective estimates of global daily precipitation from gauge observations, satellite estimates, and numerical predictions." Adv Atmos Sci 33(7): 889-904
Ogato GS, Bantider A, Abebe K, Geneletti D (2020) Geographic information system (GIS)-Based multicriteria analysis of flooding hazard and risk in Ambo Town and its watershed, West Shoa zone, Oromia Regional State Ethiopia. J Hydrol Regional Studies 27:100659
Oliazadeh A, Bozorg-Haddad O, Mani M, Chu X (2021) Developing an urban runoff management model by using satellite precipitation datasets to allocate low impact development systems under climate change conditions. Theoret Appl Climatol 146(1):675–687
Ren M, Xu Z, Pang B, Liu W, Liu J, Du L, Wang R (2018) Assessment of satellite-derived precipitation products for the Beijing region. Remote Sensing 10(12):1914
Salio P, Hobouchian MP, Skabar YG, Vila D (2015) Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. Atmos Res 163:146–161
Shahbazi H, Taghvaee S, Hosseini V, Afshin H (2016) A GIS-based emission inventory development for Tehran. Urban Climate 17:216–229
Sharifi E, Steinacker R, Saghafian B (2016) Assessment of GPM-IMERG and other precipitation products against gauge data under different topographic and climatic conditions in Iran: preliminary results. Remote Sensing 8(2):135
Smith B, Rodriguez S (2017) Spatial analysis of high-resolution radar rainfall and citizen-reported flash flood data in ultra-urban New York City. Water 9(10):736
Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu KL (2018) A review of global precipitation data sets: data sources, estimation, and intercomparisons. Rev Geophys 56(1):79–107
Tang L, Tian Y, Yan F, Habib E (2015) An improved procedure for the validation of satellite-based precipitation estimates. Atmos Res 163:61–73
Tian Y, Huffman GJ, Adler RF, Tang L, Sapiano M, Maggioni V, Wu H (2013) Modeling errors in daily precipitation measurements: additive or multiplicative? Geophys Res Lett 40(10):2060–2065
Tiwari S, Jha SK, Singh A (2020) Quantification of node importance in rain gauge network: influence of temporal resolution and rain gauge density. Sci Rep 10(1):1–17
Vu T, Li L, Jun K (2018) Evaluation of multi-satellite precipitation products for streamflow simulations: a case study for the Han River basin in the Korean Peninsula, East Asia. Water 10(5):642
Wang K, Li S (2021) Robust distributed modal regression for massive data. Comput Stat Data Anal 160:107225
Wang K, Wang H, Li S (2021) Renewable quantile regression for streaming datasets. Knowledge-Based Systems 107675
Wei G, Lü H, Crow WT, Zhu Y, Wang J, Su J (2018) Evaluation of satellite-based precipitation products from IMERG V04A and V03D, CMORPH and TMPA with gauged rainfall in three climatologic zones in China. Remote Sens 10(1):30
Yang D, Yang A, Qiu H, Zhou Y, Herrero H, Fu CS, ... Tang J (2019) A citizen-contributed GIS approach for evaluating the impacts of land use on hurricane-Harvey-induced flooding in Houston area. Land 8(2): 25
Yang XQ, Geng WJ (2016) Accuracy evaluation of TRMM-based multi-satellite precipitation in Huai river basin. Water Resources and Power 7:1–5
Yang Z, Hsu K, Sorooshian S, Xu X, Braithwaite D, Zhang Y, Verbist KM (2017) Merging high-resolution satellite-based precipitation fields and point-scale rain gauge measurements—a case study in Chile. Journal of Geophysical Research: Atmospheres 122(10):5267–5284
Zhang A, Xiao L, Min C, Chen S, Kulie M, Huang C, Liang Z (2019a) Evaluation of latest GPM-Era high-resolution satellite precipitation products during the May 2017 Guangdong extreme rainfall event. Atmos Res 216:76–85
Zhang K, Ali A, Antonarakis A, Moghaddam M, Saatchi S, Tabatabaeenejad A,... Moorcroft P (2019c) The sensitivity of north American terrestrial carbon fluxes to Spatial and temporal variation in soil moisture: an analysis using radar-derived estimates of root-zone soil moisture. J geophys res Biogeosci 124(11): 3208-3231
Zhang K, Chao L, Wang Q, Huang Y, Liu R, Hong Y,... Ye J (2019a) Using multi-satellite microwave remote sensing observations for retrieval of daily surface soil moisture across China. Water Sci Eng 12(2): 85-97 https://doi.org/10.1016/j.wse.2019.06.001
Zhang K, Wang Q, Chao L, Ye J, Li Z, Yu Z,... Ju Q (2019b) Ground observation-based analysis of soil moisture spatiotemporal variability across a humid to semi-humid transitional zone in China. J hydrol (Amsterdam), 574: 903-914
Zhu Q, Gao X, Xu YP, Tian Y (2019) Merging multi-source precipitation products or merging their simulated hydrological flows to improve streamflow simulation. Hydrol Sci J 64(8):910–920