COMPARSION OF TRMM PRECIPITATION SATELLITE DATA OVER CENTRAL JAVA REGION – INDONESIA

Andung Bayu Sekaranom, Emilya Nurjani, M. Pramono Hadi, Muh Aris Marfai

Faculty of Geography, University of Gadjah Mada, Yogyakarta, Indonesia

Abstract: This research aims to compare precipitation data derived from satellite observation and ground measurements through a dense station network over Central Java, Indonesia. A precipitation estimate from the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) 3B42 Version 7 are compared with precipitation data from interpolated rain gauge stations. Correlation analysis, mean bias error (MBE), and root mean square error (RMSE) were utilized in the analysis for each thee-monthly seasonal statistics. The result shows that the 3B42 products often estimate lower rainfall than observed from weather stations in the peak of the rainy season (DJF). Further, it is revealed that the 3B42 product are less robust in estimating rainfall at high elevation, especially when humid environment, which is typical during the rainy season peak, are involved.

Introduction

The atmospheric processes above the Indonesian maritime continent are considered very complex (Schott and McCreary 2001, Salahuddin and Curtis 2011, Hashiguchi et al. 2013). This condition is generated from various processes ranging from short intra-annual and inter-annual circulation, such as El-Nino Southern Oscillation (ENSO), Madden Julian Oscillation (MJO), and Dipole Mode (Ropelewski and Halpert 1987, Saji et al. 1999, Morita et al. 2006). A specific annual precipitation pattern occurs in this area, where very distinct monthly cumulative precipitation can be found between the dry season and the wet (rainy) season (Aldrian and Susanto 2003). The dry season is often associated with a period (consists of several months) with no/very small amount of monthly precipitation. In contrast,

precipitation is very high in the wet season, with accumulated monthly precipitation reaches up to 400 mm per month (Aldrian and Susanto 2003). Precipitation driving hazards, such as floods and landslides, are often occurred in the entire Indonesian maritime continent (WMO 2002, 2004, 2010). The hazards are often associated with heavy precipitation in the rainy season and drought at dry season (Naylor et al. 2001). Crucial issues related to hazards are emerging due to the predicted impact of climate change towards the region. It is believed the hazards frequencies and magnitudes are increasing due to higher extreme events generated by the climate change (Naylor 2007).

Efforts in monitoring precipitation in the study area is important due to the linkages between climate, hazards, and potential impacts of the climate change (Naylor 2007). Unfortunately, the distribution of rain gauge networks is often

DE GRUYTER OPEN

not uniform due to complex terrains, which in turn produces obstacles to the monitoring efforts (Liu and Zipser 2014). Rainfall data obtained from rain gauges/weather stations ideally produce an accurate measurement from point coordinates. However, it is often found that rain gauges spatial distribution is very dense in flat-lowland areas but sparse in mountainous areas, such as in the Global Precipitation Climatology Centre (GPCC) dataset (Schneider et al. 2008). Uneven distribution of the weather stations potentially produces a large bias in precipitation interpolation (Yatagai et al. 2012). Moreover, due to orographic factors in the study area, precipitation variability in the mountainous areas is considered higher than the lowlands (Mair and Fares 2010).

Measuring precipitation in the mountainous regions is important due to the role of mountains as an upper area of the watershed. Precipitation occurred in this area has a very significant contribution to groundwater and surface water systems, which are very important in the hydrological cycle (Collischonn et al. 2008, Chen et al. 2011). Since the 1960s, there are emerging techniques in the utilization of satellite technologies to estimate precipitation (Michaelides et al. 2009). Satellite data can provide important information due to its ability to measure precipitation on a regular basis. Further, satellite precipitation data have several advantages compared to ground-based measurements. It can measure precipitation on the earth surface uniformly. Thus, it can overcome various physical barriers which are often found from the ground measurements, for example, due to topographical obstacles (Liu and Zipser 2014). The low latency of satellite precipitation data provide ability in supporting real-time monitoring, which is often lacked by the ground observations (Kuligowski et al. 2013). Broad ranges of utilization of satellite precipitation estimate now exist at global and regional scales, for example in identifying land-ocean-atmosphere interaction, climate variability analysis, and as an input in hydrological cycle and the early warning system (Morita et al. 2006, Mair and Fares 2010, Chen et al. 2011, Kuligowski et al. 2013).

Several satellite precipitation estimates with a very high spatial resolution now exist. One of the examples provided by Tropical Rainfall Measuring Mission (TRMM) satellite observation. The main objective of this project is to measure the precipitation in the tropics and its variation based on the radar satellite data. TRMM is equipped with five sensors in measuring the precipitation in the tropics (Kummerow et al. 1998, JAXA 2005). Although precipitation data derived from satellite estimates have higher advantages in measuring precipitation than weather station networks, there are some variation about the quality of the TRMM precipitation estimation.

Various comparisons have been conducted for the TRMM datasets surrounding the study area, especially for the for the Multisatellite Precipitation Analysis (TMPA) products. Comparison of the TMPA estimates with rain-gauges over Bali has been conducted by As-Syakur et al. (2011), indicates that TMPA estimation often underestimate rainfall compared to weather station observation. Prasetia et al. (2013) compare the TMPA-rain gauge biases as function of different climate zones over Indonesia. The result shows that there are variation in the biases as the effect of different climatological condition. Comparison of TMPA data with weather station over Celebes Islands conducted by Giarno et al. (2018) shows that there are possible effect of mountainous area towards the rainfall underestimation in TMPA. Comparison with other satellite precipitation estimation conducted by Vernimmen et al. (2012), particularly for the Climate Prediction Center (CPC) MORPHing technique (CMORPH) and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) shows that all of the satellite precipitation datasets tend to estimate lower rainfall than observed from rain-gauges. However, there are several location in Indonesia where the TMPA estimate higher rainfall. However, compared to rainfall estimation using reanalysis datasets indicates that satellite estimation performs better, especially during westerlies monsoon over South East Asia (Peña-Arancibia et al. 2013). From the above results, it can be concluded from the above researches that the product quality varied among regions. Various variables can be involved in the estimation biases, including climatological properties and elevation.

From the previous researches, it is also known that over Indonesian monsoonal region, the quality of precipitation estimation is higher than the other climate regions (Prasetia et al. 2012). However, a more detailed comparison is required,
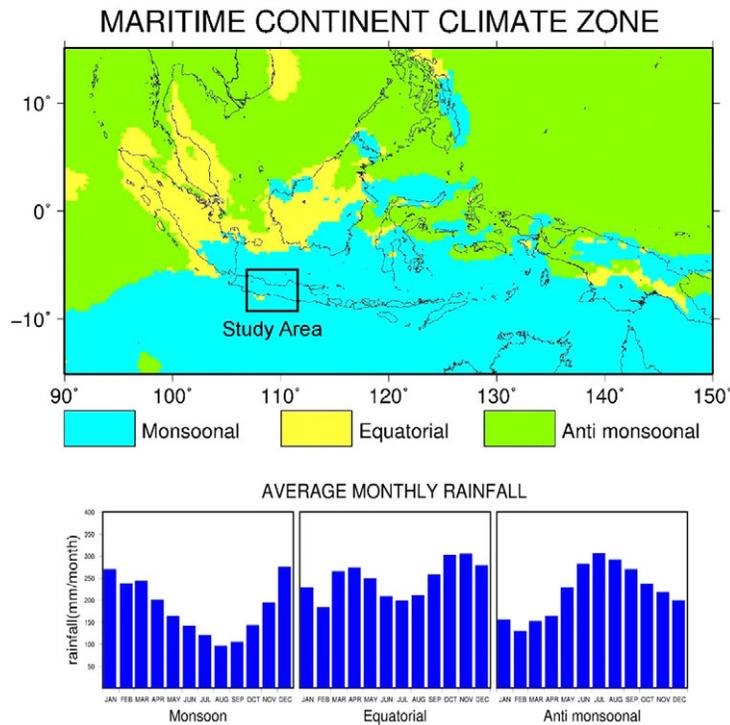
Fig. 1. Three different major climate zones in Indonesia based on Aldrian and Susanto (2013). Precipitation data were derived from TRMM 3B42.

for example, by taking the effect of topography as the function of biases. This paper is intended to compare precipitation data determined by TRMM 3B42[1] product with a dense weather station networks across Central Java-Indonesia. The main objective of this paper is to measure the capability of the satellite product, especially related to large biases and stochastic errors, compared to the ground observation. By identifying the where the large biases occur, we could learn what aspect that could contribute to the differences between the two datasets. The more detailed analysis in this research is given in this research by focusing the biases characteristics as function of topographical condition. This is because the study area consists of complex topographical condition from lowland to mountainous area with high elevation.

Study area

The research is conducted in Central Java Region, as a part of Indonesian maritime

continent. In general, the Maritime Continent is located over the South East Asian tropics, where more than two-thirds of the area consists of the ocean (Aerts et al. 2009). There are five large islands over the area, namely Sumatra, Borneo, Java, Celebes, and Papua Islands. This archipelago has three different climate regions (Aldrian and Susanto 2003), including monsoonal type, semi-monsoonal type, and anti-monsoonal type (Fig. 1). The monsoonal region is highly influenced by a monsoonal circulation, consisting of the northwest monsoon (in October–March) and the southeast monsoon (in April–September). Large amount of moistures during the northwest monsoon produce heavy precipitation in the wet season (Qian 2008). Therefore, a single rain peak is found around December to February. A different condition can be found in the semi-monsoonal region since it has two peaks or rainfall. Those two peaks (occurred in October–November and March–May), are influenced by the northward and southward movement of the Inter-tropical Convergence Zone (ITCZ). The anti-monsoonal region has a reversed characteristic with the monsoonal region. In this area, a peak of precipitation often occurs in April to September. It is possible that the annual rainfall pattern in this

---

1 Explanation of the 3B42 can be found on the webb page: https://pmm.nasa.gov/data-access/downloads/trmm.
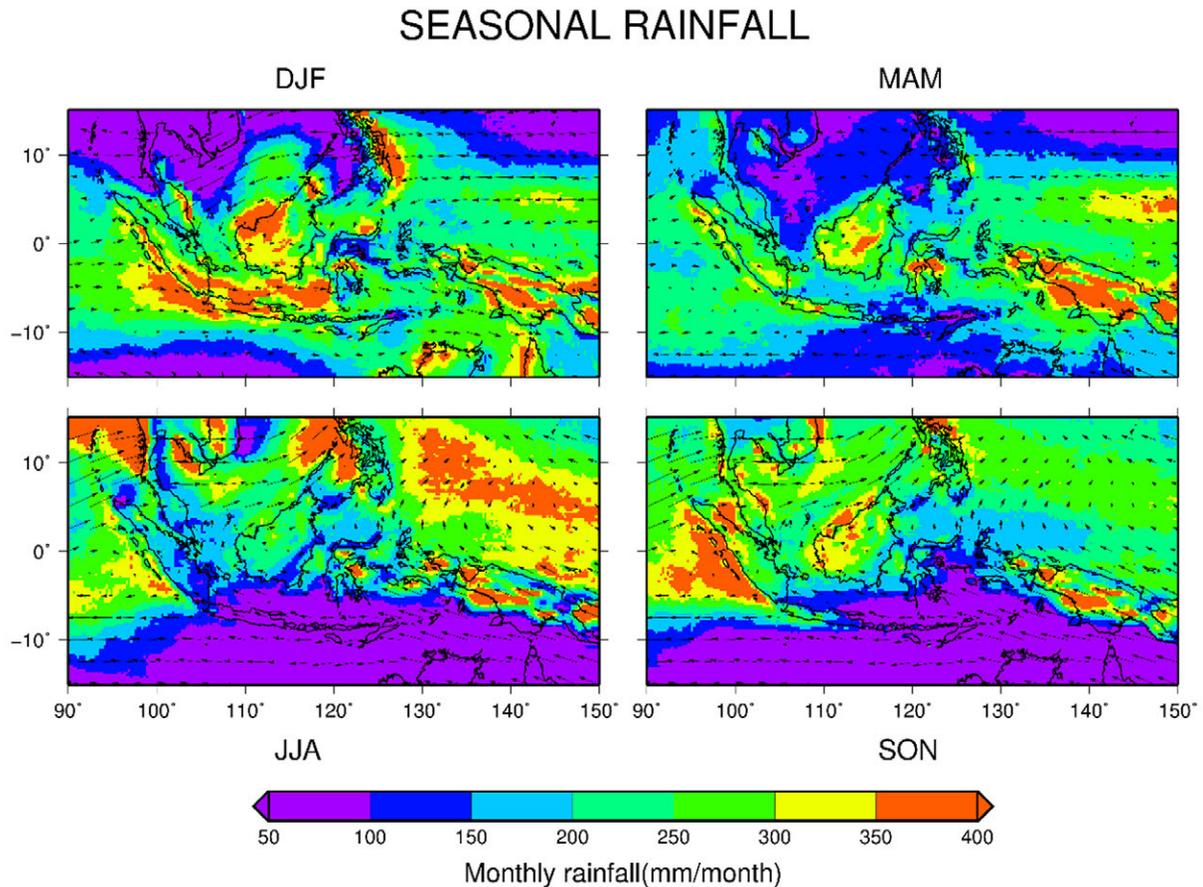
SEASONAL RAINFALL

Fig. 2. TRMM 3B42 composite seasonal rainfall and the corresponding NCEP reanalysis horizontal wind vectors from 1998–2010.

region is associated with the Indonesian Ocean Through flow which moves across small islands in the eastern part (Aldrian and Susanto 2003). It is also highly possibly that the maximum monthly rainfall in this region are closely related to distribution of maximum convective when ITCZ passes this region (Chang et al. 2005). Moreover, precipitation in this region are known to be highly enhanced by El-Nino Modoki during JJA (As-Syakur et al. 2016)

The Central Java Region is located at the central part of Java Island, 108–111° E and 7–8.5° S. This area is characterized by monsoonal climate region. The peak of the wet season occurred in December to February (DJF) (Fig. 2). The figure also shows that the monthly precipitation during this season reaches up to 400 mm per month. In the dry season, a very distinct condition can be found, in which very small amount of precipitation occurred. Based on the obtained weather station data, no precipitation is often received by the area In June–August (JJA). This is likely due to changing the direction of the monsoonal

wind (Fig. 2). Precipitation is increasing at the mountainous areas in the central part of the island. The maximum elevation of the mountainous area in the central part reaches about 3000 m in height. The Northern and Southern part of the study area are characterized by lowland-flat area with elevation less than 500 m. The mountainous area is known to trigger an orographic process that forces moist air masses to be condensed and converted to rainwater (Qian 2008). Based on the weather station data, the annual precipitation in the mountainous area reached up to 2,700 mm a^{-1}, while in the lowland area (at north and south part of the island), the annual precipitation is about 1,500–2,000 mm a^{-1}.

Research methodology

Datasets

Monthly rainfall dataset obtained from some weather stations over Central Java, recorded by the
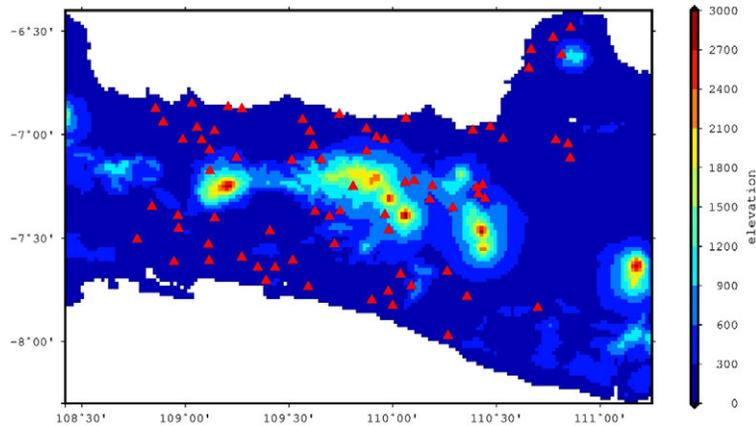
Fig. 3. Fig. 3. Rain gauge stations used for the analysis, represented as red triangle. Shading color represent elevation. (Precipitation data are from BMKG; elevation data are based on ETOPO).

Indonesian Bureau of Meteorology, Climatology, and Geophysics (BMKG) is used as a reference to validate the satellite precipitation data. The data cover a period between 1998 and 2010. The weather stations are sparsely distributed in the research area. Several areas have weather stations that are densely distributed, while in other areas, there are less/no stations (Fig. 3). This is especially related to the topographical condition. A high density of the network is found in the north and the south part of the island. A lower weather station density is found in the central part, which is mostly mountainous region. 60 stations are located in the lowland area with elevation less than 500 m. 12 stations are located at the elevation between 500–1,000 m, and only one station is located at the area with an elevation higher than 1,000 m (Table 1). Data pre-processing has been implemented to ensure the quality of ground observation data, including missing data identification, correlation checking, and consistency checking.

There are three main rain estimation product in TRMM, which are based on TRMM Precipitation Radar (PR2A25), TRMM Microwave Imager (TMI2A12), and the TRMM Multisatellite Precipitation Analysis (TMPA 3B42/3B43)

Table 1. Number of weather stations based on the elevation.

Elevation (m a.s.l.)	Number of stations
0–250	51
250–500	9
500–750	9
500–1,000	3
>1,000	1

(Kummerow et al. 1998). In this research, the satellite precipitation data is obtained from 3B42 observations by the Japan Aerospace Exploration Agency (JAXA) earth observation centre and National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) precipitation processing system (Huffman et al. 2007). The TRMM rainfall product used in this research were TRMM 3B42 Version 7. TRMM 3B42 is space-time average of some passive microwave and infrared based products that are averaged to $0.25° × 0.25°$ grid ($±28$ km^2). This product has 3 hourly temporal resolution, in which each data contains averaged rain rates during the period of recording (in mm h^{-1}). The 3B42 rain estimations are based on:

1. Level 1 TRMM TMI Brightness Temperature (Tb) (2A12),
2. Level 2 PR–TMI combined rain profile (2B31),
3. NSIDC level 2 AMSR-E precipitation,
4. NOAA/NCDC M-CLASS SSMIS Tb,
5. NOAA/NCDC CLASS SSMIS antenna temperature,
6. NESDIS Microwave Surface and Precipitation Products System (MSPPS) level 2 AMSU-B precipitation,
7. NESDIS MSPPS operational level 2 Microwave Humidity Sounder (MHS) precipitation,
8. NOAA/NCDC level 3 Gridded Satellite (GridSat-B1) IR Tb,
9. NOAA/NWS/CPC merged 4-km geostationary satellite IR Tb. The 3B42 final product is later obtained by further calibrating the data with GPCC level 3 precipitation gauge analysis on a monthly basis (Huffman and Bolvin 2013).

Although monthly precipitation data are also provided by the TRMM 3B43 product, the 3B42 product is considered more applicable for the future analysis purpose, especially in identifying the source of biases in the input datasets. This is because the biases in the 3B42 can be originated mostly from passive microwave or infrared datasets as the input algorithm. For example, the high frequency channels in passive microwave sensors are known to be sensitive to ice particles (Sekaranom and Masunaga 2017), while the low frequency channels are known to be sensitive to cloud and liquid water path (Masunaga et al. 2005). On the other hand, satellite infrared sensors are known to be sensitive to cirrus clouds (Heidinger et al. 2010). Comparison of different satellite retrieval variables obtained from different sensors or by using product intercomparison, could be utilized to reveal the source of biases (Sekaranom and Masunaga 2017). However, to conduct such a comparison, data series that close to instantaneous satellite retrieval are required rather than the monthly dataset.

To produce monthly precipitation, daily precipitation were calculated and then were accumulated in monthly basis. To convert daily precipitation from 3 hourly rainfall rates, a single dataset of mean rain rate (in mm h^{-1}) are multiplied by the temporal resolution (3 hour). Eight rain rate values from the raw dataset collected in daily basis were then summed together to produce daily rainfall (in mm d^{-1}), and then summed again based on the month. Hereafter, the monthly weather station precipitation data were compared with monthly rain accumulated satellite data.

Data Pre-Processing

Comparison of TRMM precipitation data with rain gauge stations must follow several requirements to be fulfilled. The processes for example are data quality checking and outlier identification. Quality of rain gauge data in the research area often distorted due to the abundance of missing data. This is especially related to rain gauge instruments that are broken for several months, in which no recording can be made during that period. These missing values (in daily basis) were not replaced by certain techniques, for example by interpolating to the nearest stations. Simply

explained, months with missing data were omitted and were not considered for the comparison. Data obtained from rain gauge stations also must have good quality to be interpolated. Outliers from the rain gauge data possibly generated from various source. For example due to rain gauge stations that not properly placed, inaccuracy of rain gauge instruments, or data that not properly recorded. Outliers also could be possibly generated by intense rainfall within small area especially in mountains. The existence of outliers can produce low correlation of precipitation between rain gauge stations and also inconsistency of the records over time.

To remove stations which possibly contain an excessive number of outliers, two procedures were implemented. Correlation-distance analysis and consistency analysis were utilized to remove these stations. Monthly correlation from a single station was examined from three nearest stations. This process repeatedly held to all of the 70 rain gauge stations in the study area. The correlation values were then compared with the distance relative to each other and then plotted into a graph. This correlation-distance concept are originated from a condition that two rain-gauges located at a near distance will tend to capture similar precipitation systems than farther location, so the correlation will be higher for the closer stations (Asquith and Famiglietti 2000). The result of the correlation-distance analysis is shown in Figure 4. The x-axis represents the distance (in km), and Y-axis represents the correlation. Stations with low correlation but has close distance can be identified from this graph, which then removed. Stations that are omitted from the correlation-distance analysis are shown by red color.
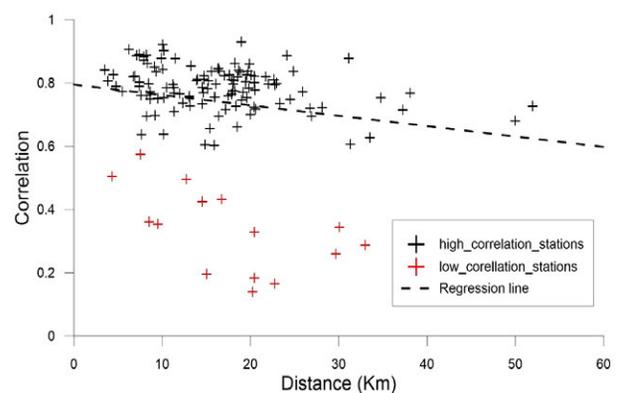

Fig. 4. Correlation-distance plot between weather stations with deleted station in cross red marks.
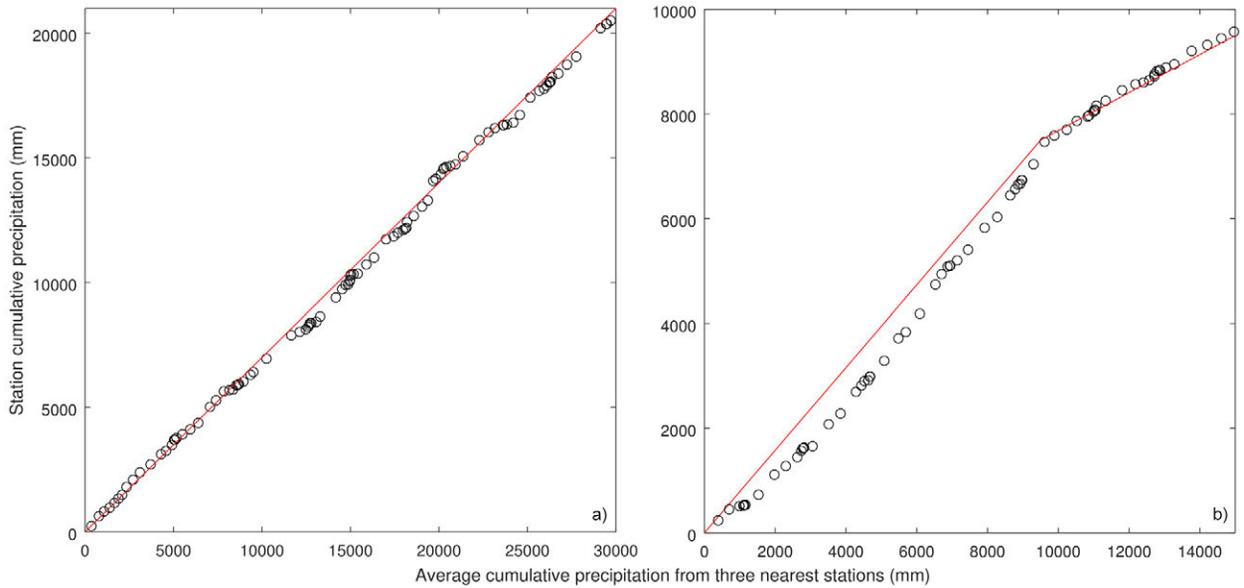
Fig. 5. a) An example of station with good consistency taken from Dracik Kramat Station, b) station with low consistency taken from Ngawen Station.

The precipitation data for each monthly time step are then compared to measure its consistency. This process employed the same stations from the correlation-distance analysis. In this analysis, monthly precipitation from examined weather stations are accumulated from the initial time step to the last time step, and compared to the three nearest stations. In general, the consistency are best compared with many station data as much as possible. However, large distance between stations could also possibly producing the biases. Three stations are selected since it often produce good result since the distance is not quite far. A linear plot indicates good consistency since the measurement is in line with the other stations. When a line break is detected, it shows low consistency which indicate a deviation between the examined station with the other station. This break indicates low consistency and are possibly generated by broken sensors or inaccuracy in the measurement. Figure 5 shows an example of station with good and low consistency. The station with good consistency shows a linear line from the initial to the end of observation, while the
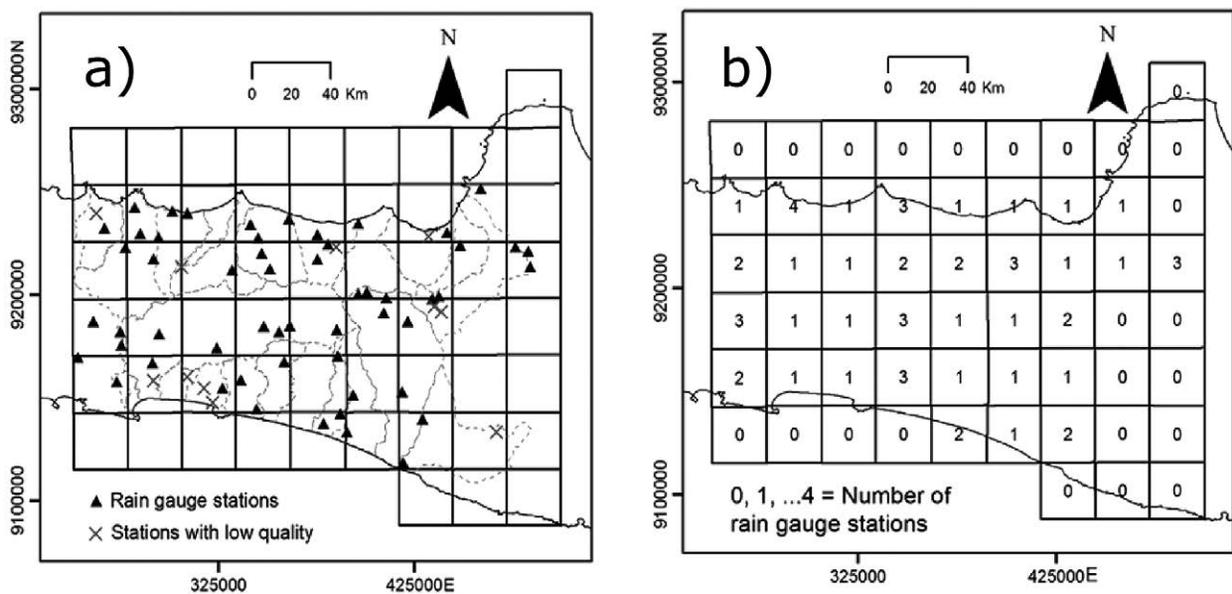

Fig. 6. Number of weather stations within each TRMM 0.25 sq. degree grid.

station with low consitency shows a line break. The weather station with this line break are removed from the analysis.

From 70 rain gauge stations data that are obtained in the research area, 12 stations were omitted due to low correlation or low consistency. Therefore, 58 rain gauge stations were utilized for the validation. The spatial distribution of these stations, as well as the omitted one, is shown in Figure 6a. There is a cluster of stations with low data quality over the Southwest of the study area. This area is a lowland so the precipitation will be more homogeneous. Therefore the main source of the low quality data, as taken from the weather station documentations, is possibly due to massive replacement of rain-gauge sensors over this area. To illustrate the station distribution inside TRMM grid, the TRMM grid is then superimposed above these stations to count the number of rain gauge stations inside each grid. From this process, about 1 to 4 rain gauge stations were accounted for each TRMM grid (Fig. 6b). This process was conducted to ensure that there is at least one rain gauge station to represent each TRMM grid. Thus, the grids which have no station representation were not accounted for the comparison.

Interpolation Methods

Comparison of weather station data with the gridded satellite data remain challenging due to different nature of both data, where weather station measures point rainfall, and satellites measure area averaged rainfall (areal rainfall). The comparison can be done directly by assuming that the precipitation is equal within the area inside each TRMM grid, and considered as a point to point analysis. Unfortunately, this method can produce a large bias due to the high variance of precipitation inside one single grid (Asquith and Famiglietti 2000). As a result, observation from the point of measurement often overestimates the areal rainfall, and therefore an areal reduction factor must be implemented (Asquith and Famiglietti 2000). Another way to compare the data is based on a grid to grid comparison, as proposed in this research. The procedure consists of a precipitation transformation from observed points to gridded (areal) data, which then compared with the TRMM grids. The transformation was conducted by interpolating unevenly distributed ground precipitation measurement to the uniform TRMM referenced grid. Several interpolation methods were examined, namely
a) Moving Average,
b) Spline,
c) IDW,
d) Kriging,
e) Nearest Neighbour,
f) Natural Neighbour.

In this research, a similar search radius of 0.5° is utilized for all of the interpolation. This search radius represent two times of the TRMM resolution and to ensure that sufficient station data are utilized for the interpolation. Since there are various parameters in the interpolation that gives influence to the interpolated value, the other interpolation parameters are set to default.

Selection of proper interpolation method is important to represent an area-averaged rainfall that matches the real condition. Improper interpolation methods might produce underestimate/overestimate results and also might produce a large bias in the data validation process. In this research, the quality of interpolated data was assessed by comparing interpolated and the observed rainfall in each station when the examined station is removed for interpolation. The comparison of interpolated statistics are then presented using 1:1 plot, correlation analysis, Mean Bias Error (MBE) and Root Mean Square Error (RMSE). The result of 1:1 plot value is shown in Figure 7, while the correlation, MBE, and RMSE are presented in Table 2. The interpolation

Table 2. Correlation value, Mean Bias Error (MBE), and Root Mean Square Error (RMSE) for several interpolation method.

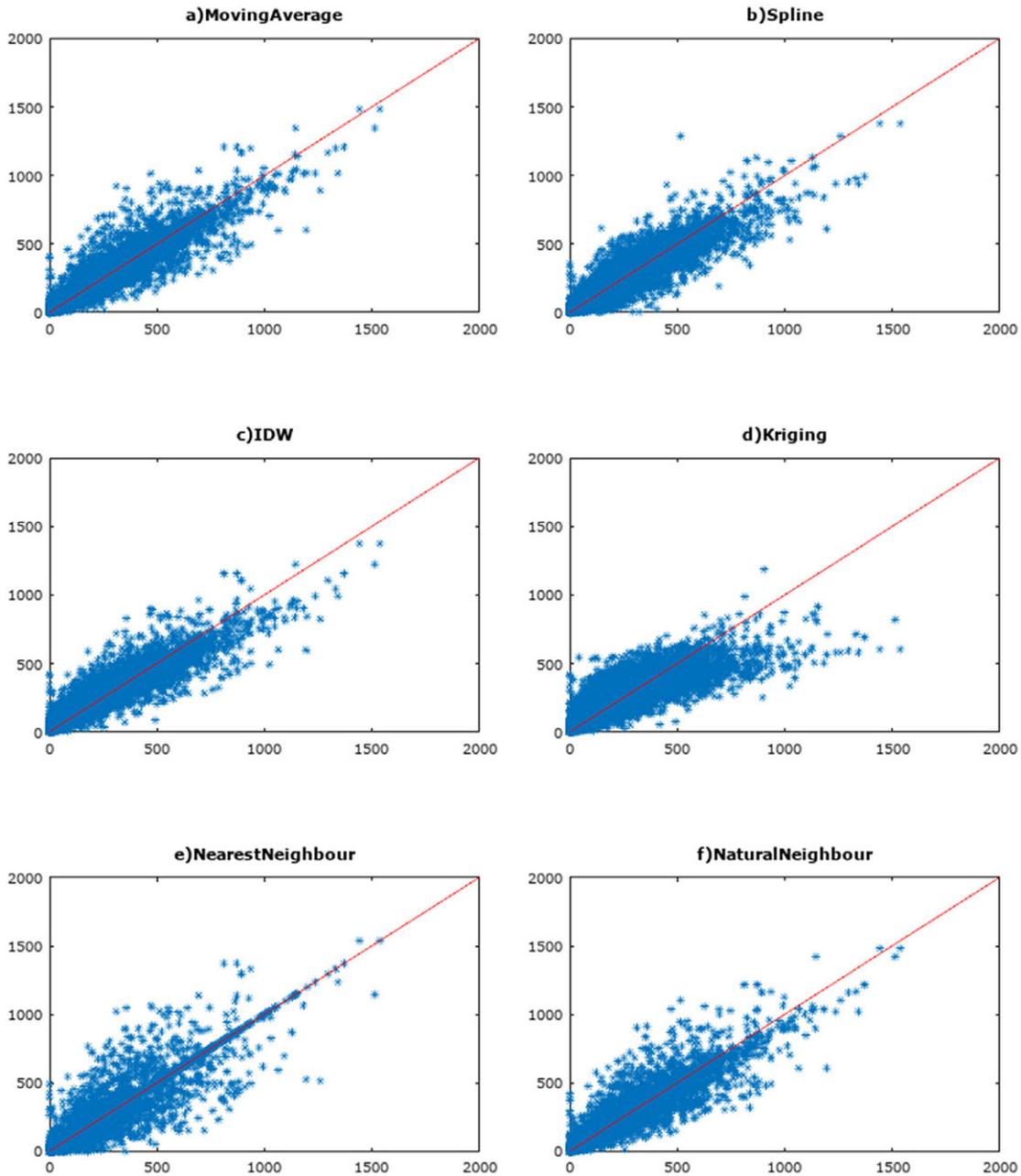| Parameter	Interpolation					
	Moving Average	Spline	IDW	Kriging	Nearest Neighbour	Natural Neighbour
Correlation	0.93	0.91	0.92	0.84	0.92	0.91
MBE	1.15	–10.61	8.01	2.18	1.02	4.21
RMSE	81.51	88.17	83.37	118.94	89.82	92.06

Fig. 7. Selection of interpolation techniques, a) Moving Average, b) Spline, c) IDW, d) Kriging, e) Nearest Neighbour, and f) Natural Neighbour.

techniques produces different results. In general, there are no large differences allo of the examined parameters. The correlation ranges from 0.91 to 0.93, which are all statistically significant at p-value <0.05, the bias ranges from –10 to 8.01 mm per month, while the RMSE ranges from 81.51 to 118.94 mm per month. The moving average interpolation is selected in this research since it has the highest correlation value and lowest RMSE value. Since it might be not sufficient to determine the best interpolation using the above parameters only, a direct comparison of TRMM grid with weather stations (non-interpolated) is also conducted to shows that the interpolated result remains objective.

Statistical Comparison

Several statistical measures were implemented to analyse the relationship between the TRMM data with rainfall estimated using the rain gauge data. The linear correlation

coefficient, mean bias error (MBE), and root mean square error (RMSE) were selected in this analysis (Feidas 2010). The equations are defined in equations 1–3:

$$Correlation = \sum_{i=1}^{n} \frac{(S_i - S)(G_i - G)}{(n-1)\ \delta S\ \delta G} \qquad (1)$$

$$MBE = \frac{1}{n} \sum_{i=1}^{n} (G_i - S_i) \qquad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (S_i - G_i)^2} \qquad (3)$$

where:
- S_i denotes the estimated TRMM 3B42 value,
- G_i denotes the reference gauge value,
- δS and δG denote standard deviation for estimated and reference values respectively,
- n is the number of paired data.

The selection of statistical measures is based on advantages and disadvantages of each method. The correlation analysis can identify the similarity of rainfall patterns from TRMM 3B42. Highly positive correlation represents the similarity in the pattern. MBE could measure the tendency, which can be positive or negative. A positive value means that satellite estimates often overestimate the precipitation regarding rainfall estimated from rain gauges as a reference and vice versa for the negative value. RMSE was used to find the absolute average error value between the reference data and the satellite data. To evaluate the reliability of the product on the monthly basis, both of the statistical errors analysis (MBE and RMSE) are expressed in mm per month differences. This analysis was conducted from 1998 to 2010 where the weather station precipitation data are available.

To measure seasonal difference within a year, comparison for the distinct season was also conducted. The seasons were grouped into December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON). This seasonal analysis was intended to identify the influence of monsoon cycle. The DJF represents the peak of the rainy season in the study area, while JJA represents the peak of the dry season. Transitions of the rainy season to the dry season and from the dry season to the rainy season is represented by MAM and SON respectively.

Results

Spatial pattern of Correlation, MBE, and RMSE

Result of the correlation analysis, MBE, and RMSE for each station with TRMM data (non-interpolated) are shown in Figure 8. In general, the correlation ranges between 0.62 to 0.88. Correlation values at the west part are slightly higher compared to the east, although no clear patterns are found in the north-south direction. Comparison of the MBE ranges from –81 to 203 mm per month (–40% to 45%). As similar to the correlation, there are no specific pattern of the biases are observed, however, over central part of the study area, which
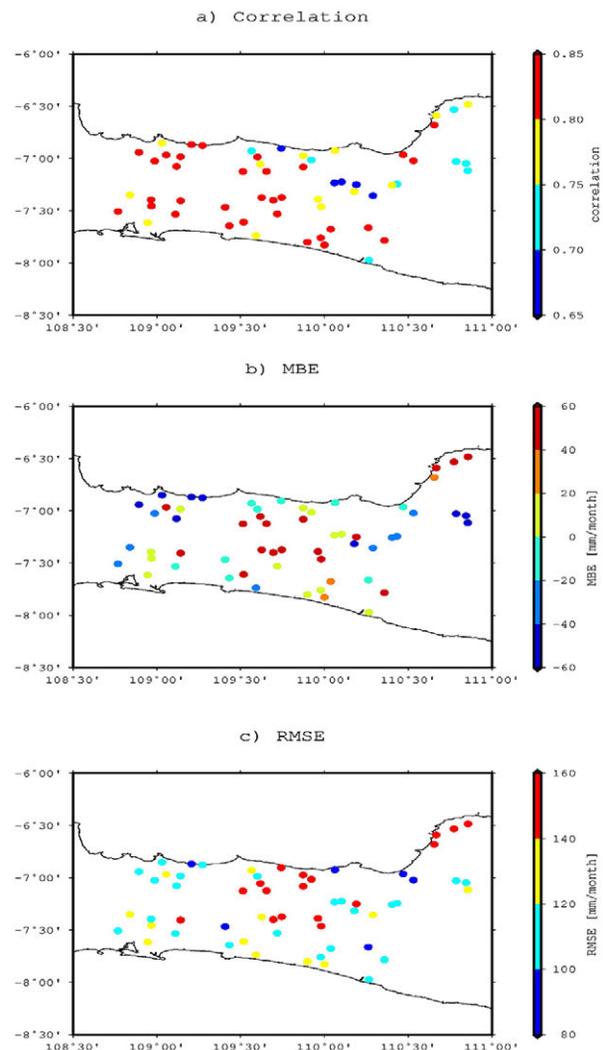

Fig. 8. a) Correlation value, b) Mean Bias Error (MBE), and c) Root Mean Square Error (RMSE) of monthly precipitation between each weather stations and TRMM 3B42.
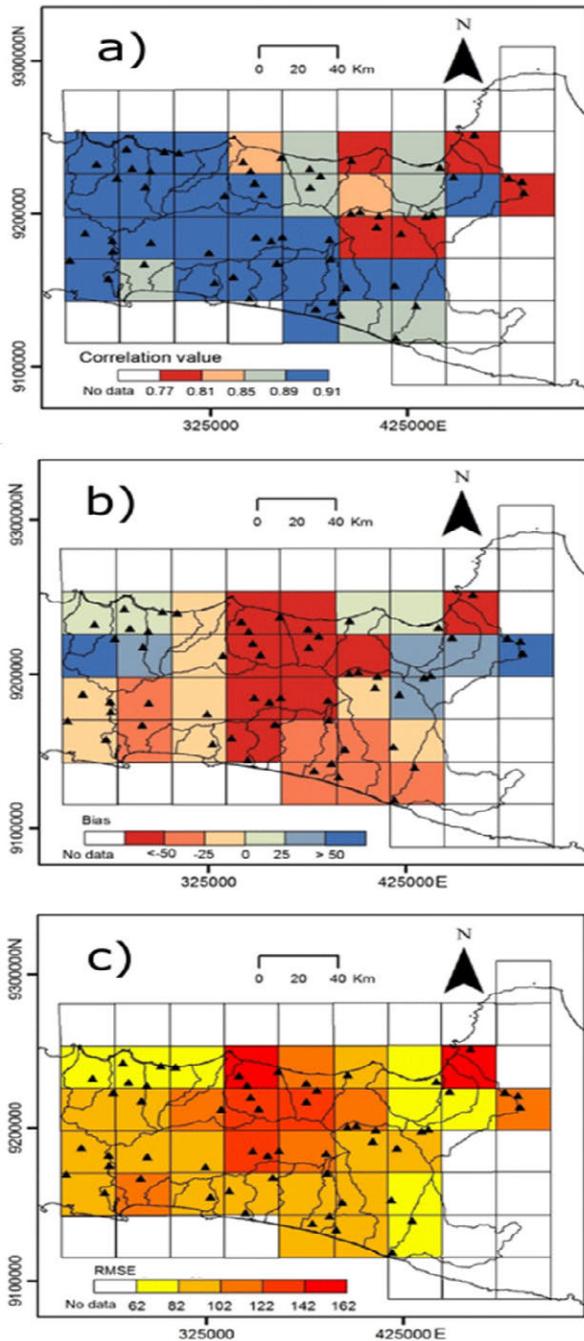
Fig. 9. a) Correlation value, b) Mean Bias Error (MBE), and c) Root Mean Square Error (RMSE) of monthly precipitation between weather stations (interpolated using Moving Average) and TRMM 3B42.

are mountainous regions, the TRMM tends to underestimate precipitation from several weather stations. Comparison of the RMSE shows that the error ranges from 85 to 265 mm per month (38% to 94%). It also appears that station with high biases at the central part also have higher RMSE.

Result of the correlation analysis, MBE, and RMSE for the interpolated data are shown in Figure 9. Overall, the interpolated product shows higher correlation compared to the station data, while the MBE and RMSE become lower. Figure 9a shows a correlation between TRMM and interpolated gauge data in the entire research area. The correlation values range from 0.77–0.91. The result of MBE (Fig. 9b) also reveals that no specific tendencies are found in the entire research area, as similar to the station data plot. At several parts, positive MBE values are identified, while negative MBE values also can be observed at the other parts. It could be observed that average monthly bias ranges from –68 to 71 mm per month (–36% to 17%). The biases tend to be negative over the central part, while positive over the west and east parts. Thus, it can be concluded that the TRMM estimation is more likely underestimate or overestimate data more locally in term of east, central, and west part of the study area. Monthly absolute precipitation differences, measured using RMSE (Fig. 9c), have a lower range than the non-interpolated data, which is about 62 to 162 mm per month (31% to 56%). These values indicate that although interpolation has been implemented, the large difference between the ground observation and the satellite estimate. From the spatial distribution of RMSE, it can be seen that the differences are higher at the central part. However, up until this point, no clear indication that the topographical factor gives an influence to this difference.

Seasonal difference of correlation, MBE, and RMSE

To initially characterize the seasonal difference, the comparison of monthly precipitation from rain gauge-TRMM for all temporal records is shown in Figure 10. The figure represents areal averaged precipitation from both of the data. Only TRMM grids with at least one rain gauge station were averaged. A similar precipitation pattern between ground observation and TRMM can be observed from this figure. It can be identified that precipitation in the dry season is quite similar to both of the data. However, the precipitation difference in the rainy season reached up to 150 mm per month lower than observed from the rain gauges. From the average of the research area, it is also shown in the figure that satellite estimates in rainy season often underestimates the precipitation.
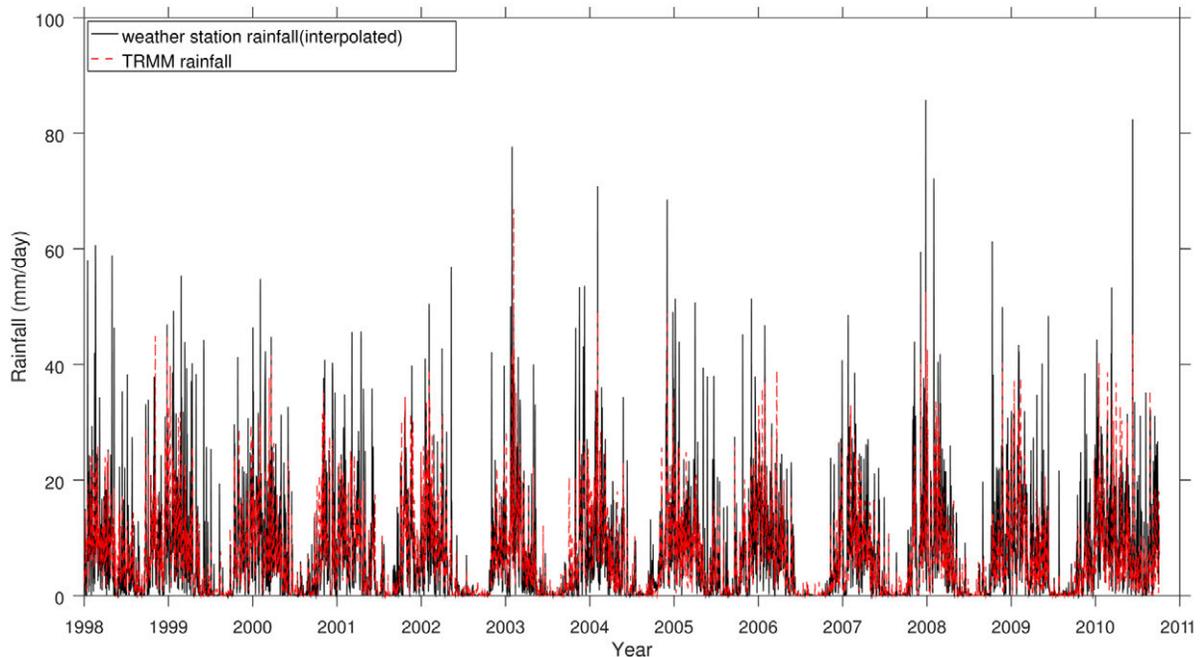
Fig. 10. Long term plot of precipitation from TRMM and interpolated weather station precipitation data.

To further investigating the seasonal differences, the seasonal variation of correlation, MBE, and RMSE are shown in Table 3. The correlation between the two data is highest JJA (0.82), and lowest in DJF (0.54). High correlation also found in SON, but lower in MAM (0.81 and 0.68 respectively). It means that the correlation are high in the dry season and the transition from the dry season to the rainy season. Correlation are lower in the peak of the rainy season and the transition between the rainy season to the dry season. The lowest MBE value is found in JJA. In this season, the satellite estimates are quite match with ground observation data. However, MBE values for other seasons are negative. It can be concluded that the satellite data often underestimate precipitation data from the weather stations. The highest difference of MBE value is found in DJF, which is the peak of the rainy season. The average difference reaches 44 mm per month or about 12.29%. Although the correlation and

the MBE in JJA are also the lowest, The RMSE percentage value is the highest (64.25%). This is especially related to the low precipitation amount during this period, in which a slight precipitation difference can produce a large difference from the average. The RMSE differences are quite similar in MAM and SON, which are about 90 mm per month (37.56%) in MAM and 97 mm per month (54.07%) in SON. Although the actual errors are similar, it should be noted that MAM has the lowest percentages than all other seasons.

Correlation value, Mean Bias Error (MBE), and Root Mean Square Error (RMSE) as function of elevation

To identify the difference between the TRMM and weather stations over the study area, the correlation, MBE, and RMSE are plotted as a function of elevation. Comparison of the non-interpolated data and moving average interpolated data are presented in this analysis. The station data are ranked from the lowest elevation to the highest elevation from left to right. For the interpolated data, the elevation are based on average elevation within each 0.25° x 0.25° grid. As an initial step, comparison for all data are conducted, and further separated for each season to better understand the differences.

Table 3. a) Correlation value, b) Mean Bias Error (MBE), and c) Root Mean Square Error (RMSE) for seasonal data.

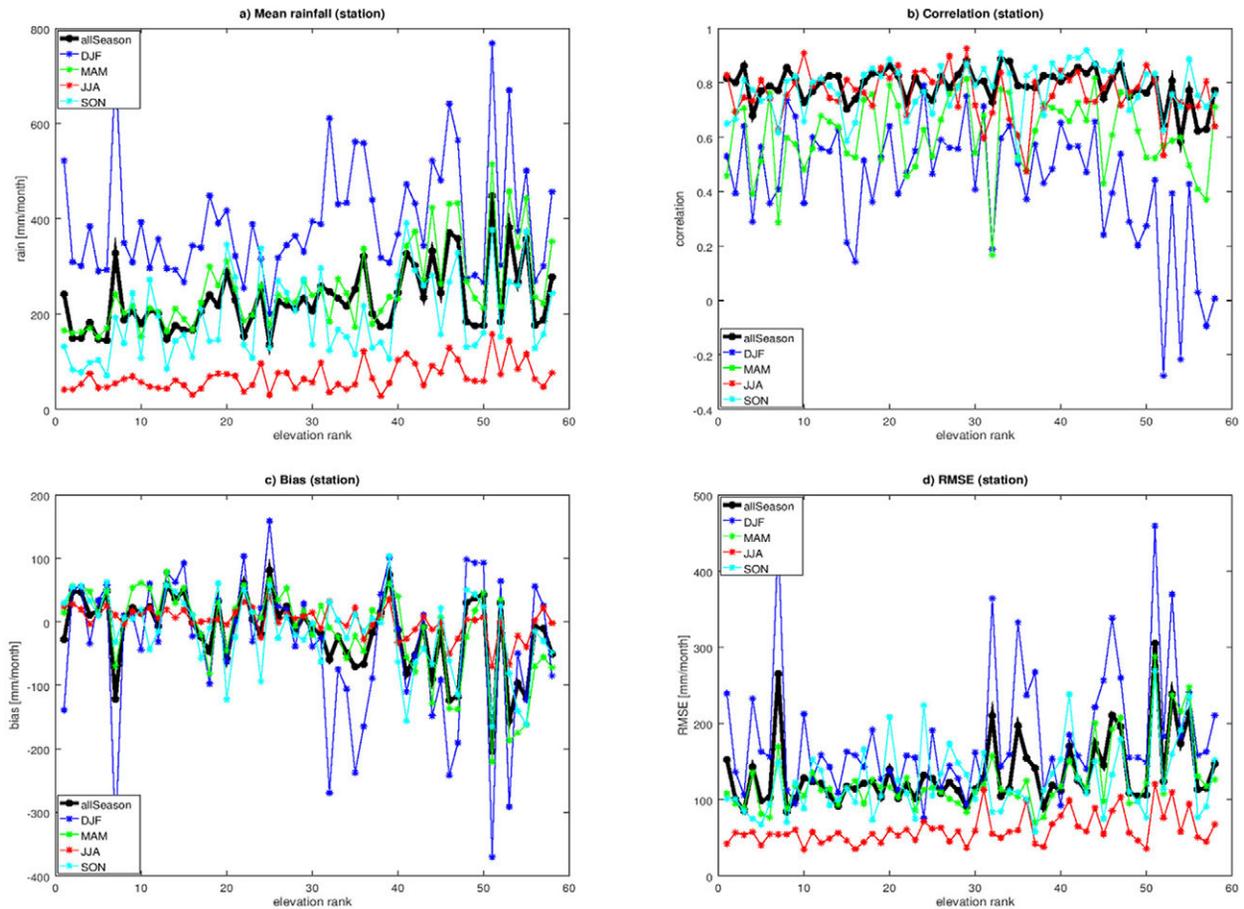| Season	Correlation	MBE	RMSE
DJF	0.54	−44 (−12.29%)	149 (41.54%)
MAM	0.68	−17 (−7.16%)	90 (37.56%)
JJA	0.82	0.94 (0.16%)	44 (64.25%)
SON	0.81	−16 (−9.09%)	97 (54.07%)

Fig. 11. a) Mean rainfall, b) correlation value, c) Mean Bias Error (MBE), and d) Root Mean Square Error (RMSE) for each station based on elevation rank from the lowest to the highest. Black-bold lines indicates average value for all season, while color lines indicate seasonal variation.

Plot of the statistical properties of the 3B42 and weather station differences as function of elevation is shown in Figure 11, contains a) mean monthly rainfall, b) correlation, c) MBE, and d) RMSE for all season and seasonal data. We first focus on the statistical result for all season. The plot of mean rainfall shows that there is a tendency of higher rainfall at higher elevation. There are also some variation of the rainfall between stations, which possibly comes from the effect of mountain shades area and differences in the localization of precipitation systems. The corresponding correlation value indicates that most of the correlation are higher than the 0.6–0.7 range, except for several stations at the highest elevation. Comparison of the MBE shows that the 3B42 data tends to estimate lower monthly rainfall for some stations. The 3B42-station biases are especially large at station with high elevation, in which almost reaching 200 mm per month. Comparison of the RMSE also shows that higher errors are often

achieved for the stations at high elevation. This is related to the higher rainfall occurred over the mountainous area compared to the lowland area.

Comparison of the statistics for seasonal data are further observed to characterize the 3B42-station differences between seasons. It can be observed from the mean monthly rainfall that the precipitation are highest in DJF, and reach the lowest in JJA as typical to monsoonal circulation. Comparison of the correlation values show that the correlation are extremely low for the peak of rainy season (DJF), particularly for the station located at highest elevation. However, comparison of the correlation for other seasons still show relatively high correlation, especially for JJA and SON. Comparison of the MBE shows that the stations located at high elevation also have higher negative bias in DJF than the station located at lower elevation in general. Comparison of the RMSE also shows that station at high elevation have higher errors in DJF, while lowest in JJA.
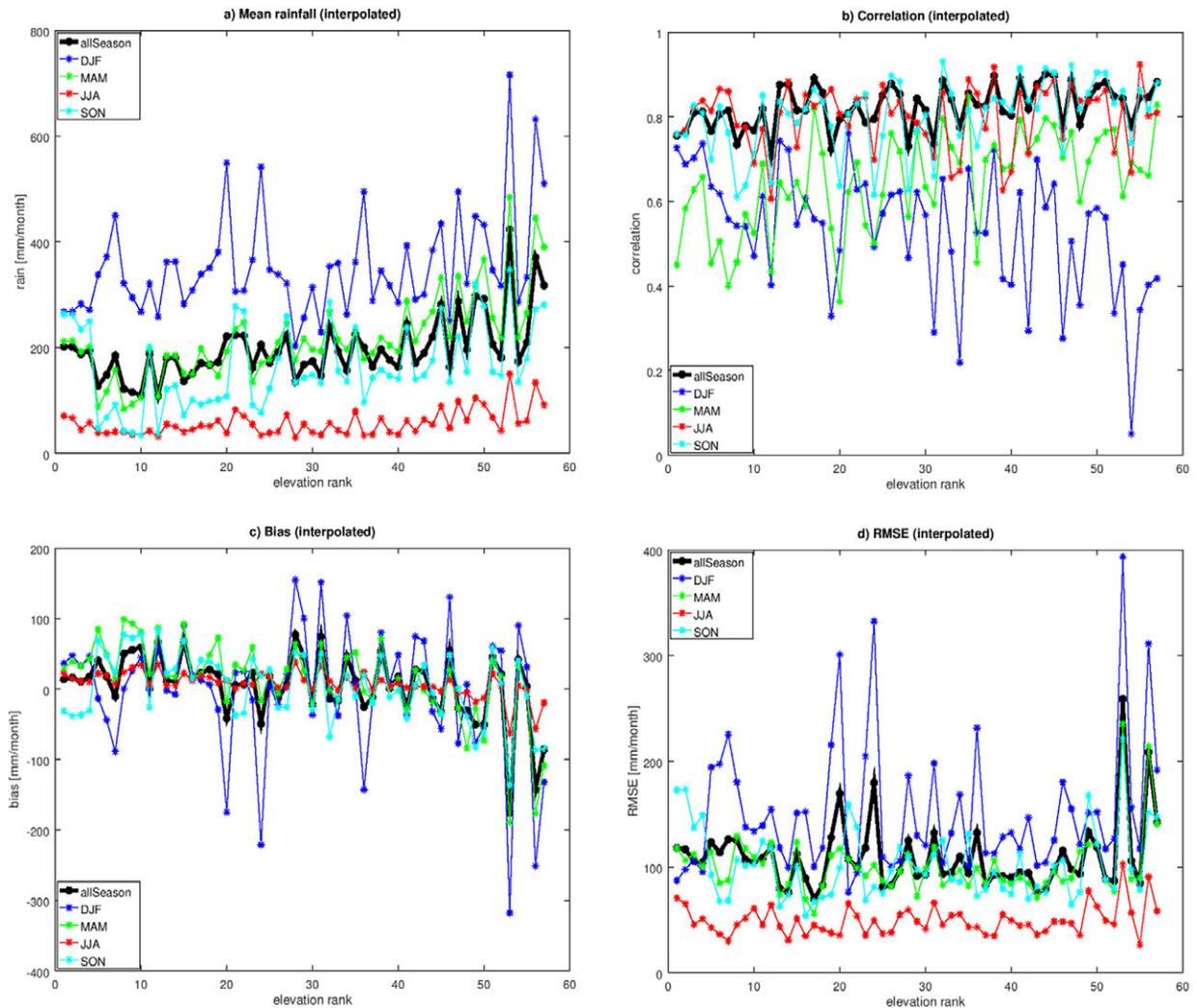
Fig. 12. Similar to previous figure, but for moving average interpolated rainfall data. The plot are based on grid elevation rank from the lowest to the highest.

Statistical comparison for the moving average interpolated data is shown in Figure 12. The result, in general shows similar pattern to what has been identified by the non-interpolated comparison. However, we could observe that the differences are much clearer than the non-interpolated comparison. The tendency of increasing precipitation as function of height is clearly observed from the graph, particularly for the DJF rainfall. Comparison of the correlation also shows that the DJF correlation tend to decrease with increasing elevation. The MBE and RMSE differences are comparable to the non-interpolated result. There are variation in the MBE that could be positive or negative around zero, except for rain grid at high elevation. The RMSE are also show that the error are large for DJF rainfall.

Discussions

From the results obtained in this research, we could summarize the statistical difference between the 3B42 estimates with weather station precipitation data that spread over the study area. First, it is important to note that the monthly precipitation are centralized at the centre part of the island, which are mountainous region. On the other hand, precipitation are lower over the lowland areas at the north and south part of the island. It is well known that the higher precipitation over the central part of the study area are generated mainly by land-ocean diurnal interaction (Qian 2008). In more detail, large instability occurred over land during afternoon takes up moisture at the surrounding ocean, which then condensed due to orographic uplifting at the

central part (Qian 2008). It has been shown in this study that the 3B42 shows a weak correlation between 3B42 estimates and weather station precipitation data. Comparison of the MBE also shows that 3B42 tends to estimate lower monthly rainfall for the mountainous area, as well as the high errors compared to the lowlands area.

It is also important to mention that DJF rainfall statistically lower compared to the other seasons. It has been shown that higher monthly rainfall in DJF are mostly sourced from humid air originated from the Indian Ocean (Fig. 2). The humid air that transported towards the central part of mountainous area somehow produce rainfall that exhibit weak correlation with 3B42 in DJF. MBE and RMSE of the DJF rainfall are also larger than the other seasons, especially over mountainous region. It is important to note that the DJF statistics are different compared to other seasons. For example, even though JJA still have relative large error in term of RMSE percentage (Table 3), but the correlation remains high. Therefore, the lower correlation in DJF might indicate that the 3B42 missed or false in identifying precipitation over the mountainous region.

To this end, it has been shown that the 3B42 estimates often produce lower rainfall estimation over the mountainous region in the study area, especially during the peak of rainy season. It is therefore important to identify possible source of the bias from the above result. First, it is known that there are large variation of the mountainous precipitation, which are tends to be inhomogeneous compared to the lowland. It is therefore possible that the weather station data measured local precipitation system that occupy small areas as compared to the 3B42 resolution. This kind of small system might be not well identified by the 3B42.

Second, it is also known that estimations from the passive microwave sensors tend to underestimate orographic rain (Shige et al. 2013). This can also be one important factor since 3B42 estimation partially derived from passive microwave estimation. Over ocean, the passive microwave low frequency channels detect emission from liquid phase hydrometeors, while the high frequency channels detect scattering from ice particles (Kummerow et al. 1996, Masunaga et al. 2005). In general, the rainfall estimation using passive microwave sensors depends mostly on high frequency channels due to land surface noises at lower frequency (Kummerow et al. 2001, Gopalan et al. 2010). The high frequency channels are more sensitive to ice-particles, which are mostly associated with deep convective clouds (Sohn et al. 2015). Moreover, signals received by satellite infrared sensors are highly correlated with cloud top temperature (Berg et al. 2006), and therefore will also be more sensitive to tall-deep convective clouds. However, in orographic precipitation, the humid air are associated with forced upward vertical motion. Thus different than the general instability applied to deep convective clouds (Shige et al. 2013). The orographic rainfall therefore might be associated with less ice-particles rather than deep convective rainfall, so the microwave sensors estimate lower rainfall.

Third, it is also worth to mention that mountainous precipitation possibly generated more likely by warm rain mechanism than cold rain mechanism. The cold rain mechanism are associated with large surface instability that important in deep convective development, while warm rain mechanism could occur in more stable environment, but with large amount of moisture (Song et al. 2017). The most important process in warm rain mechanism is intense collision and coalescence process of rain particles in very humid environment (Hamada et al. 2015). Therefore, high precipitation can occur without significant cloud top and ice-particles, where estimation using infrared sensors and high frequency microwave channels depends on those two variables (Sekaranom and Masunaga 2017). The large moisture supply from Indian Ocean during DJF in the study area possibly favourable to support the intense collision and coalescence in the warm rain development, in addition to condensation due to forced upward motion in the mountainous area. Sekaranom and Masunaga (2017) identify that warm rain mechanism is the most dominant process in the heavy rain development over land surface. However, the effect of topography is not included in their analysis. Therefore, it is still not clear whether warm/cold rain mechanism are more dominant than the another. It is therefore important to identify what factors that are more dominant in producing the biases in mountainous regions in the future research.

Summary

In this study, monthly precipitation data from TRMM 3B42 were compared with monthly weather station measurement in Central Java Region-Indonesia. Data from 58 rain weather stations, covering a 13-year period (1998–2010) were spatially interpolated to evaluate the satellite estimate data. Long-term and seasonal timescales were used to analyse the differences between the ground observation and the satellite precipitation data. Correlation coefficients, MBE, and RMSE analysis were employed to both of the timeframes. Moving average interpolation was utilized to calculate the areal precipitation.

The comparison of the precipitation data from the dense weather station network in Central Java and TRMM 3B42 estimation indicates that the TRMM 3B42 often underestimate precipitation data in rain season. As a result, larger bias and lower correlation occur in the rainy season. The rainy season differences, in this case, should be taken into account, especially since high precipitation amount in the rainy season and hydrometeorological hazards have a close relationship in the study area. The larger error in rainy season, therefore, gives remarks to the satellite precipitation data utilization and application in developing disaster early warning system, mainly due to biases between satellite estimation and ground-based measurement.

Although the TRMM 3B42 algorithm includes an adjustment process with ground observation to improve its accuracy, the result of this research indicates that the satellite precipitation data still need to be improved. It is possible that although the adjustment reduces the global discrepancies, the regional discrepancies might exist. The comparison with the dense weather station network suggests that the errors are higher in specific parts of the study area, particularly the mountainous region. It could be inferred that despite the adjustment of the TRMM 3B42 precipitation rainfall algorithm, gaps between satellite-ground data observation are still exist.

It is also important to note that this research use TRMM 3B42 version 7 only, therefore no comparison for version 6 is conducted in this research. However, it is known that the TMPA version 7 algorithm perform better compared than the version 6 (Prakash et al. 2015). In summary,

the version 7 algorithm use the GPCC data version 4, while the previous version using GPCC version 2. The TMPA algorithm version 7 also use improved AMSU algorithm compared to older one in version 6 (Prakash et al. 2015). The TMPA version 7 also shows improvement in the TRMM combined instrument (TCI) (Liu 2016).

Although there are various improvement in the algorithm, the input datasets itself could contribute to the bias at the estimation, especially due to assumption in ice-particles and cloud top temperature relationship with rainfall, as mentioned in the discussion. There are no significant improvement related to this issue in the input datasets used by TMPA products. For example, the TRMM Microwave Imager (TMI) land algorithm, by design, estimate higher rain-rate with higher ice-scattering, which are not always shows linear relationship (Sekaranom and Masunaga 2017). However, there are no significant improvement in the TMI land algorithm between version 7 and version 6, so that the above linear assumption still exists (Zagrodik and Jiang 2016). Sekaranom and Masunaga (2017) shows that the 3B42 estimation perform better than the TMI estimation over land. This is because the 3B42 algorithm also consider TRMM Precipitation radar (PR) data as input in the TCI, as well as the later adjustment with GPCC data. However, the 3B42 still have lower performance compared to the TRMM PR data (Sekaranom and Masunaga 2017). It is because the TRMM PR data directly measures radar reflectivity as the product of sixth moment of drop size distribution (DSD) (Iguchi et al. 2000), rather depends on ice-particles and cloud top temperatures, which are considered more indirect (Liu and Zipser 2014). As a result, the TRMM PR data are better in identifying warm rain mechanism that occur mostly at humid environment (Sekaranom and Masunaga 2017).

Methods to bridge the gap between ground observation and satellite estimation are still required for the practical purpose, especially for rain-rate to surface discharge calculation. From this research, it is known that the orographic effect could produce larger bias to the 3B42 estimation. It is therefore important to identify more detail the influence of orographic factor towards satellite precipitation, so the bias could be minimized. The possible source of biases due to topographic factor has been discussed in this

paper, including more localized precipitation, cloud vertical structures and ice microphyics that influence 3B42 input datasets, as well as warm rain mechanism. However, it is also possible that the biases are generated from combination of the three above factors, or other factors that are still not considered above. Characterizing of the above factors using a set of satellite combination (i.e. A-Train satellites (Stephens et al. 2002)) might be useful for understanding the process corresponding precipitation over mountainous area, and is suggested for further studies.

Acknowledgements

References

Aerts J., Major D.C., Bowman M.J., Dircke P., Marfai M.A., 2009. *Connecting delta cities: coastal cities, flood risk management and adaptation to climate change*. Netherlands, VU University Press.

Aldrian E., Susanto R.D., 2003. Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology* 23(12): 1435–1452.

Asquith W.H., Famiglietti J.S., 2000. Precipitation areal-reduction factor estimation using an annual-maxima centered approach. *Journal of Hydrology* 230(1–2): 55–69.

As-Syakur A.R., Osawa T., Miura F., Nuarsa I., Ekayanti N.W., Dharma I.B.S., Tanaka T.2016. Maritime Continent rainfall variability during the TRMM era: the role of monsoon, topography and El Niño Modoki. *Dynamics of Atmospheres and Oceans* 75: 58–77

As-Syakur A.R., Tanaka T., Prasetia R., Swardika I.K., Kasa I.W., 2011. Comparison of TRMM multisatellite precipitation analysis (TMPA) products and daily-monthly gauge data over Bali. *International journal of Remote Sensing* 32(24): 8969–8982.

Berg W., L'Ecuyer T., Kummerow C., 2006. Rainfall climate regimes: The relationship of regional TRMM rainfall biases to the environment. *Journal of Applied Meteorology and Climatology* 45(3): 434–454

Chang C.P., Wang Z., McBride J., Liu C.H., 2005. Annual cycle of Southeast Asia-Maritime Continent rainfall and the asymmetric monsoon transition. *Journal of Climate* 18(2): 287–301.

Chen C., Yu Z., Li L., Yang C., 2011. Adaptability Evaluation of TRMM Satellite Rainfall and Its Application in the Dongjian River Basin. *Procedia Environmental Sciences* 10: 396–402.

Collischonn B., Collischonn W., Tucci C., 2008. Daily Hydrological Modelling in the Amazon Basin using TRMM rainfall estimates. *Journal of Hydrology* 360: 207–216.

Feidas H., 2010. Validation of satellite rainfall products over Greece. *Theoretical and Applied Climatology* 99(1–2): 193–216.

Giarno G., Hadi M.P., Suprayogi S., Murti S.H., 2018. Distribution of Accuracy of TRMM Daily Rainfall in Makassar Strait. *Forum geografi* 32(1).

Gopalan K., Wang N.-Y., Ferraro R., Liu C., 2010. Status of the TRMM 2A12 land precipitation algorithm. *Journal of Atmospheric and Oceanic Technology* 27(8): 1343–1354.

Hamada A., Takayabu Y.N., Liu C., Zipser E.J., 2015. Weak linkage between the heaviest rainfall and tallest storms. *Nature Communications* 6: 6213.

Hashiguchi H., Yamamoto M.K., Yamamoto M., Mori S., Yamanaka M.D., Carbone R.E., Tuttle J.D., 2013. Cloud episode propagation over the Indonesian Maritime Continent from 10years of infrared brightness temperature observations. *Atmospheric Research* 120: 268–286.

Heidinger A.K., Pavolonis M.J., Holz R.E., Baum B.A., Berthier S., 2010. Using CALIPSO to explore the sensitivity to cirrus height in the infrared observations from NPOESS/VIIRS and GOES-R/ABI. *Journal of Geophysical Research: Atmospheres* 115(D4).

Huffman G.J., Bolvin D.T., 2013. *TRMM and other data precipitation data set documentation*. Greenbelt USA, National Aeronautics and Space Administration (NASA).

Huffman G.J., Bolvin D.T., Nelkin E.J., Wolff D.B., Adler R.F., Gu G., Hong Y., Bowman K.P., Stocker E.F., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology* 8(1): 38–55.

Iguchi T., Kozu T., Meneghini R., Awaka J., OkamotoK.I., 2000. Rain-profiling algorithm for the TRMM precipitation radar. *Journal of Applied Meteorology* 39(12): 2038–2052.

JAXA [Japan Aerospace Exploration Agency], 2005. *Tropical Rainfall Measuring Mission TRMM) Precipitation Radar Algorithm: Instruction Manual for Version 6*.

Kuligowski R.J., Yaping L., Yu Z., 2013. Impact of TRMM data on a low-latency, high-resolution precipitation algorithm for flash-flood forecasting. *Journal of Applied Meteorology and Climatology* 52(6): 1379–1393.

Kummerow C., Barnes W., Kozu T., Shiue J., Simpson J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology* 15: 809–817.

Kummerow C., Hong Y., Olson W.S., Yang S., Adler R.F., McCollum J., Ferraro R., Petty G., Shin D.B., Wilheit T.T., 2001. The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *Journal of Applied Meteorology* 40(11): 1801–1820.

Kummerow C., Olson W.S., Giglio L., 1996. A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Transactions on Geoscience and Remote Sensing* 34(5): 1213–1232.

Liu C., Zipser E.J., 2014. Differences between the Surface Precipitation Estimates from the TRMM Precipitation Radar and Passive Microwave Radiometer Version 7 Products. *Journal of Hydrometeorology* 15:2157–2175.

Liu Z., 2016. Comparison of integrated multisatellite retrievals for GPM (IMERG) and TRMM multisatellite precipi-

tation analysis (TMPA) monthly precipitation products: initial results. *Journal of Hydrometeorology* 17(3): 777–790.

Mair A., Fares A., 2010. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering* 16(4): 371–383.

Masunaga H., L'Ecuyer T.S., Kummerow C.D., 2005. Variability in the characteristics of precipitation systems in the tropical Pacific. Part I: Spatial structure. *Journal of Climate* 18: 823–840,

Michaelides S., Levizzani V., Anagnostou E., Bauer P., Kasparis T., Lane J.E., 2009. Precipitation: Measurement, remote sensing, climatology and modeling. *Atmospheric Research* 94(4): 512–533.

Morita J., Takayabu Y.N., Shige S., Kodama Y., 2006. Analysis of rainfall characteristics of the Madden–Julian oscillation using TRMM satellite data. *Dynamics of Atmospheres and Oceans* 42(1): 107–126.

Naylor R.L, 2007. Assessing risks of climate variability and climate change for Indonesian rice agriculture. *Proceedings of the National Academy of Sciences* 104(19): 7752–7757.

Naylor R.L., Falcon W.P., Rochberg D., Wada N., 2001. Using El Nino/Southern Oscillation climate data to predict rice production in Indonesia. *Climatic Change 50*(3): 255–265.

Peña-Arancibia J.L., van-Dijk A.I., Renzullo L.J., Mulligan M., 2013. Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and South and East Asia. *Journal of Hydrometeorology* 14(4): 1323–1333.

Prakash S., Mitra A.K., Momin I.M., Pai D.S., Rajagopal E.N., Basu S., 2015. Comparison of TMPA-3B42 versions 6 and 7 precipitation products with gauge-based data over India for the southwest monsoon period. *Journal of Hydrometeorology* 16(1): 346–362.

Prasetia R., As-Syakur A.R., Osawa T., 2013. Validation of TRMM Precipitation Radar satellite data over Indonesian region. *Theoretical and Applied Climatology* 112(3–4): 575–587.

Qian J.-H., 2008. Why precipitation is mostly concentrated over islands in the Maritime Continent. *Journal of Atmospheric Sciences* 65: 1428–1441.

Ropelewski C.F., Halpert M.S., 1987. Global and regional scale precipitation patterns associated with the El Niño/ Southern Oscillation. *Monthly Weather Review* 115(8): 1606–1626.

Saji N.H., Goswami B.N., Vinayachandran P.N., Yamagata T., 1999. A dipole mode in the tropical Indian ocean. *Nature* 401(6751):360.

Salahuddin A., Curtis S., 2011. Climate extremes in Malaysia and the equatorial South China Sea. *Global and Planetary Change* 78(3): 83–91.

Schneider U., Fuchs T., Meyer-Christoffer A., Rudolf B., 2008. *Global precipitation analysis products of the GPCC*. Germany, Global Precipitation Climatology Centre (GPCC).

Schott F.A., McCreary J.P., 2001. The monsoon circulation of the Indian Ocean. *Progress in Oceanography* 51(1): 1–123.

Sekaranom A.B., Masunaga H., 2017. Comparison of TRMM-Derived Rainfall Products for General and Extreme Rains over the Maritime Continent. *Journal of Applied Meteorology and Climatology* 56(7): 1867–1881.

Shige S., Kida S., Ashiwake H., Kubota T., Aonashi K., 2013. Improvement of TMI rain retrievals in mountainous areas. *Journal of Applied Meteorology and Climatology* 52(1): 242–254.

Sohn B.J., Choi M.J., Ryu J., 2015. Explaining darker deep convective clouds over the western Pacific than over tropical continental convective regions. *Atmospheric Measurement Techniques* 8: 4573–4585.

Song H.-J., Sohn B.-J., Hong S.-Y., Hashino T., 2017. Idealized numerical experiments on the microphysical evolution of warm-type heavy rainfall. *Journal of Geophysical Research: Atmospheres* 122(3): 1685–1699.

Stephens G.L., Vane D.G., Boain R.J., Mace G.G., Sassen K., Wang Z., Illingworth A.J., O'connor E.J., Rossow W.B., Durden S.L., Miller S.D., 2002. The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bulletin of the American Meteorological Society* 83(12): 1771–1790.

Vernimmen R.R.E., Hooijer A., Aldrian E., van-Dijk, A.I.J.M., 2012. Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. *Hydrology and Earth System Sciences* 16(1): 133.

WMO [World Meteorological Organization], 2002. *Statement on Status of the Global Climate in 2001, WMO-No. 940*. Genewa Switzerland, WMO.

WMO [World Meteorological Organization], 2004. *Statement on Status of the Global Climate in 2003, WMO-No. 966*. Genewa Switzerland, WMO.

WMO [World Meteorological Organization], 2010. *Statement on Status of the Global Climate in 2009, WMO-NO. 1055*. Genewa Switzerland, WMO.

Yatagai A., Kamiguchi K., Arakawa O., Hamada A., Yasutomi N., Kitoh A., 2012. APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society* 93(9): 1401–1415.

Zagrodnik J.P., Jiang H., 2013. Investigation of PR and TMI Version 6 and Version 7 Rainfall Algorithms in Landfalling Tropical Cyclones Relative to the NEXRAD Stage-IV Multisensor Precipitation Estimate Dataset. *Journal of Applied Meteorology and Climatology* 52: 2809–2827.