High-resolution-gridded rainfall dataset derived from surface observation by adjustment of satellite rainfall product

A Rifani[1], **M R Ferdiansyah**[2,*]

[1] Center for Public Weather Services, Indonesian Agency for Meteorology, Climatology and Geophysics, Indonesia, Jl. Angkasa 1 no. 2, Jakarta Pusat, 10610

[2] Center for Applied Climate Information and Services, Indonesian Agency for Meteorology, Climatology and Geophysics, Indonesia, Jl. Angkasa 1 no. 2, Jakarta Pusat, 10610

* Corresponding author's email: muhammad.ferdiansyah@bmkg.go.id

Abstract. A high-resolution-gridded rainfall dataset is essential for many purposes, such as analysis of extreme weather conditions, and natural-disaster mitigation, or to be used as an input to the hydrological model. Satellite-based rainfall products (e.g., Global Satellite Mapping of Precipitation-GSMaP) can solve the spatial and temporal issues despite their rainfall intensity often being under or overestimated. This research aims to provide a high-resolution rainfall dataset by adjusting the 0.1 deg GSMaP rainfall data to the surface rainfall data from several observation points in the greater Jakarta (Jabodetabek area) during January 2020 when several flooding occurred in Jakarta. The adjustment process includes calculating the bias between the satellite estimation in the nearest observation point and interpolating the error back to the 0.01 deg grid by using radial basis function (RBF) to obtain the correction factor in every grid point, GSMaP data then adjusted by the correction factor. We implemented the method during January 2020 when several floods occurred in Jakarta. The result reveals a more realistic rainfall spatial distribution than regularly interpolating the observation data. The validation of adjusted rainfall estimation at the verification points also shows a reduction in domain-wide Root Mean Squared Error (RMSE) by 30 – 80%.

1. Introduction

A high-resolution-gridded rainfall dataset is essential for many purposes, such as for analysis of extreme weather conditions, natural disaster mitigation, or to be used as an input to the hydrological model [1]. For instance, the accumulation of rainfall is one of the meteorological aspects that can lead to hydro-meteorological disasters such as landslides. Such information is demanded particularly for the areas with steep contours and less vegetation. Flood mitigation and analysis of extreme weather especially over urban areas, where high population and emerging socio-economic, also require spatial rainfall information. Moreover, the gridded-rainfall dataset is beneficial as an input for hydrological models to support early warning systems. Generally, such a dataset is provided by interpolating several distributed points of rainfall measurement.

Unfortunately, these measurements are not equally spaced nor always available in areas where rainfall information is required. It then affects the representativeness of the interpolated data or makes the results of the interpolation less reliable. For this reason, satellite-based rainfall products (e.g., Global

Satellite Mapping of Precipitation-GSMaP) are expected to be utilized due to their high spatial coverage and continuous measurement. However, to our best knowledge, recent studies related to validations between satellite products and surface observation over tropical regions are still few. Besides, due to the different measurement methods between instantaneous rainfall from above (by satellite) and the rainfall accumulation (direct measurement by rain gauge) at the surface, the value of satellite products tends to be under/ over-estimated. To obtain more accurate rainfall accumulation and to represent rain events over tropical regions more reliably, the assessment and the adjustment for those satellite products are important.

To address the above-mentioned problems, this study aims to provide high-resolution-gridded spatial data of rainfall that can be used for both the analysis of extreme weather and the mitigation of hydrometeorological disasters. The proposed framework includes the assessment of GSMaP utilization for operational purposes and the adjustment using observational data as the truth value for the GSMaP product.

2. Data and Methods

2.1. Ground Observation Data

The study is conducted over the greater Jakarta area (Jabodetabek), situated between 6.8-5.9 S and 106.3-107.3 E. The Jabodetabek area covers various types of topography from the coastal area at its northern part and mountainous topography at its southern part. About 13 rivers also flow to Jakarta leading to flood events. In several events of heavy rainfall in the south surface flooding often occurred when heavy rainfall in the north. Hence, a dense observation of rainfall is needed to cover these various areas, providing high-resolution and accurate observation data for various purposes. Two climatological stations of Badan Meterologi Klimatologi dan Geofisika (Indonesian Agency for Meteorology, Climatology, and Geophysics: BMKG) provide 99 manned rainfall observation sites in the study area combined with 9 BMKG meteorological and geophysical stations yielding 108 manned observation stations in Jabodetabek.

The 99 rainfall stations only provide a daily accumulation of rainfall, meanwhile, 9 BMKG stations provide 3 hourly rainfall data. To ensure the data is uniform we consider using daily accumulation rainfall data. The data used are from January 2020 when there were several extreme rainfall events and major flooding occurred in Jakarta. The manned observation site data considered to have high-quality data for the daily accumulation of rainfall rather than automatic observation. The locations of the stations used in this study are shown in figure 1.

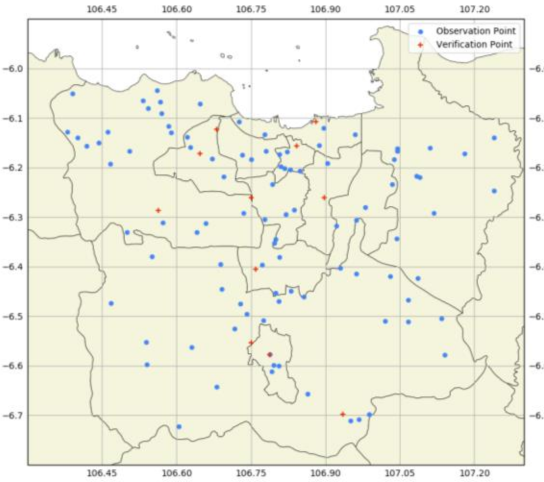

Figure 1. Distribution of the point observations within the study area. Blue circles and red "+" denote the observation points and the selected points for verification (see Table 1 for details), respectively.

2.2. Satellite-derived Rainfall Product

Global Satellite Mapping of Precipitation (GSMaP) is a satellite-based rainfall product that provides rainfall estimation at 0.1 x 0.1 deg spatially and hourly temporal resolution. The GSMaP is one of the Global Precipitation Measurement (GPM) products which also provides satellite-based rainfall estimation. The near-real-time hourly rainfall data (GSMaP NRT) are used in this study, with the lag time of 4 hours behind real-time the data is considered good to be used operationally. Satellite-based precipitation could cover areas with sparse rainfall observation sites, GSMaP could overcome this with its good correlation but relatively underestimate the rainfall compared with the observation [4].

2.3. Spatial Consistency Check

The hourly GSMaP NRT data is accumulated to 24-hour accumulation to match the observation accumulation period. The data was then re-gridded from its original grid (0.1 x 0.1 degree) to 0.01 x 0.01-degree grid which the observation data will be interpolated into this grid domain. A spatial consistency check was applied to the observation data to avoid outliers in space since the interpolation will be performed against the surrounding data. **Spatial Consistency Index (SCI)** following [10] used as in [5] was used to identify outliers as follows:

$$SCI = |R_i - Q_{50}| / (Q_{75} - Q_{50}), \text{ if } Q_{75} \neq Q_{25}$$
$$\text{otherwise, } SCI = |R_i - Q_{50}| / MAD \quad (1)$$

where $\mathbf{R_i}$ represents the rainfall observation (mm); $\mathbf{Q_{25}}$, $\mathbf{Q_{50}}$ and $\mathbf{Q_{75}}$ represent the median, 25th, 50th and 75th percentile of rainfall within a range of the area. The **mean absolute deviation (MAD)** was calculated as follows:

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|R_i - Q_{50}| \quad (2)$$

where **N** is the number of total sites within the range area threshold. Considering the local convective type of rainfall which also frequently occurred during the study period we put 15 kilometres as a search radius range and SCI > 6 to flag the station as outliers. Hence, any station with SCI > 6 is removed from the interpolation. The data is then interpolated to 0.01 x 0.01 grid using **radial basis function (RBF)** inverse multiquadric [2,3]:

$$\varphi(r) = \frac{1}{\sqrt{1+(\varepsilon r)^2}} \quad (3)$$

where **φ(r)** is the radial function, that represents the shape parameter and r is the distance between the unknown point and the observed point. We selected a small value to keep the radial shape localized and not affect points that are quite far from the observation point. RBF method proved to be better than inverse distance weighted as in [11]. The bias adjustment of **rr** then is done by additive bias correction following [6] as follows:

$$rr = rr_{sat} + (rr_{obs} - rr_{sat}) \quad (4)$$

where **rrobs** and **rrsat** are the observed rainfall and the satellite-derived rainfall, respectively.

The list of the sites used is shown in Table 1. Root Mean Squared Error (RMSE) was then calculated from each observation point and for the domain during the month. The final output product is a 0.01 x 0.01 deg gridded GSMaP-observation adjusted daily accumulation rainfall. The merging method between satellite product and observation has been applied for some regions, for instance, as introduced [7] and [8] as well as the improvement by geographical weighting implemented by [9]. The SCI in our work can also be considered as the geographically weighting correction to derive a higher resolution of rainfall product.

Table 1. List of stations used for verification.

Station Name	Latitude	Longitude
Cengkareng Meteorological Station	-6.1235	106.68
Kemayoran Meteorological Station	-6.156	106.8423
Tanjung Priok Meterological Station	-6.1077	106.8804
Halim Perdanakusuma Meteorological Station	-6.2616	106.8976
Curug Meteorological Station	-6.287	106.5632

Station Name	Latitude	Longitude
Tangerang Geophysical Station	-6.17195	106.6469
Tangerang Selatan Climatological Station	-6.2615	106.7507
Depok BMKG	-6.40533	106.7589
Darmaga Climatological Station	-6.55358	106.7498
Citeko Meteorological Station	-6.69803	106.935
Cibalagung Rainfall Observation	-6.5783	106.7873

3. Results and Discussion

3.1. Spatial Consistency Check Effect

We evaluate spatial consistency checks towards the observation data to avoid extreme values outside the observation maximum due to the interpolation process. For extreme rainfall cases on 1 and 2 January where almost all of the observation points showed high rainfall value, only around 10% of the site was removed from the dataset leaving 80 - 90 sites to go on interpolation process. Meanwhile, in a localized rainfall event, the site could reach up to 30% of the site to be removed.

The interpolation result after the SCI check (figure 2, upper) in the 1st January case doesn't show significant differences between before and after site removal, meanwhile for the 20th January case (figure 2, lower), a localized high-intensity rainfall at the northwest and localized 0 value at the south are removed maintaining consistency between observation interpolation and GSMaP pattern. The results are consistent with all observation dates remaining. The SCI check result suggests a better consistency pattern with the GSMaP estimation but it is prone to removal of very localized rainfall or zero rainfall observation within its search radius.

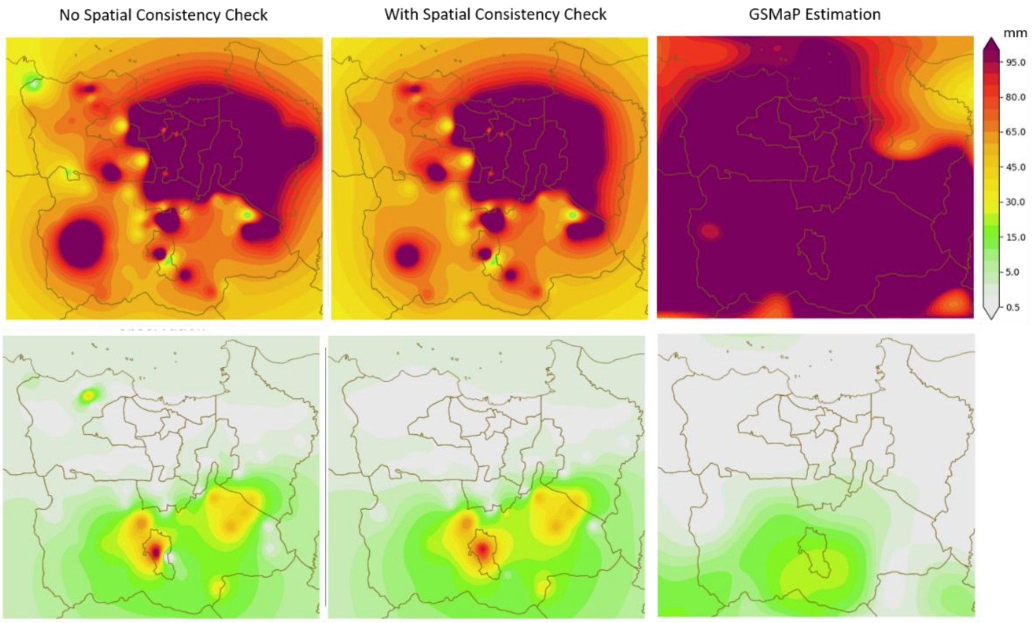

Figure 2. Spatial consistency check result compared with GSMaP estimation for the 1 January case (upper row) and the 20 January case (lower row).

3.2. GSMaP Adjustment Result

We present two adjustment results where flooding events were observed in Jakarta as a representative of the result when extreme wide-spread rainfall events were observed. As a comparison, we also present two adjustment results where only localized rainfall was observed by the stations but not shown in GSMaP. The major flooding rainfall in Jakarta occurred on 1st January where the highest rainfall was observed at Halim Perdanakusuma Meteorological Station, accumulating 377.2 mm of rainfall in 24 hours. Several rainfall observation sites in Jakarta also report an accumulation of rainfall exceeding 100 mm. Interpolation of observation data showed the widespread heavy rainfall is concentrated in the east and northeast areas, meanwhile localized high and low-intensity rainfall located at the west and south area (figure 3a). On the contrary, the GSMaP showed a widespread extreme rainfall intensity covering almost all of the domain and concentrated to the south. This results in the difference map which showed the GSMaP overestimated to the south and underestimated the rainfall in the north and northeast.

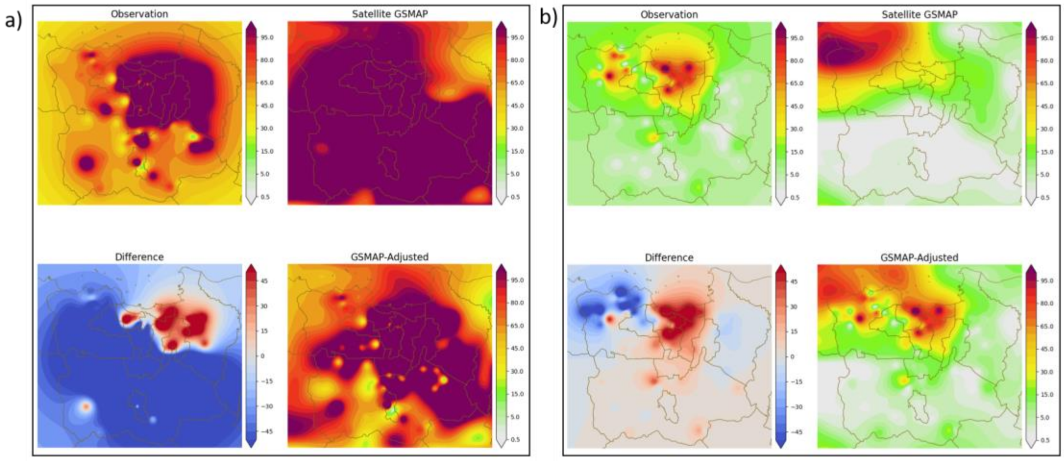

Figure 3. GSMaP adjustment for cases 1 January 2020 (a) and 25 January 2020 (b). A positive value of differences (red) indicates that GSMaP is underestimation and a negative value (blue) for overestimation.

The adjustment results in a more northeast-shifted area of precipitation, with several points reporting lower rainfall accumulation separating the northeast and the southwest rain area. For the second flooding case of 25 January (figure 3b), the GSMaP precipitation shows high-intensity rainfall in the northwest close to the coastal area, while only moderate-intensity rainfall was detected in the Jakarta area. Contrary to the observation, the observation shows high to extreme rainfall intensity in North Jakarta with several observation sites reporting more than 100 mm of rainfall in 24 hours. Bias between observation and GSMaP also showed high positive bias in Jakarta means the GSMaP underestimated the rainfall. Finally, the adjustment results in more shift to east rainfall while preserving the northwest coastal rainfall area.

In several cases, GSMaP cannot detect a localized convective system both in spatial and intensity. For the example case of 3 January (figure 4a) where a widespread light to moderate rainfall was detected in the west but not captured by GSMaP. For this kind of case, the adjustment result tends to shift towards the observation. More localized rainfall shown in the southeast in the 12 January case (figure 4b) where GSMaP also did not detect the rain. Also, the adjustment result will shift towards the observation.

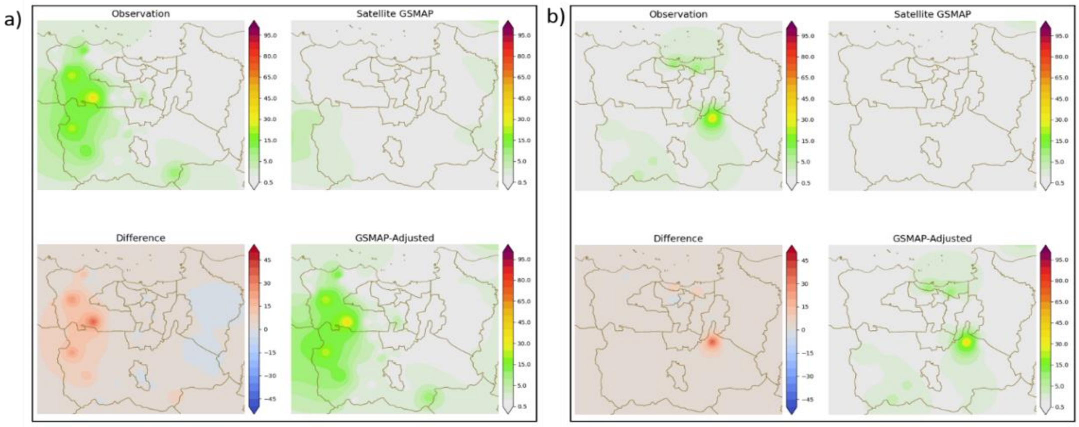

Figure 4. Same as figure 3 except for 3 January 2020 (a) and 12 January 2020 (b).

3.3. Validation
Root-Mean-Square-Error (RMSE) was calculated over the domain by using 11 stations as listed in the table during the month. The result of the RMSE from before adjustment shows a relatively high bias specifically when extreme rainfall events are observed (figure 5). This is due to the underestimation of the GSMaP rainfall and also the removal of observation stations from the interpolation process. Blue bars represent the domain RMSE before the adjustment process, and orange bars represent the domain RMSE after the adjustment process. In general, the RMSE after adjustment shows a significant reduction for almost all study periods with the reduction of RMSE ranging from 20% to 80%. The highest RMSE reduction (81%) occurred in the 19 January case, where only a small amount of rainfall inland was detected by the GSMaP meanwhile the observation shows a light-moderate localized rainfall far inland. The flooding case on 1 January and 24 January, shows a reduction of domain RMSE 33% and 32% respectively.

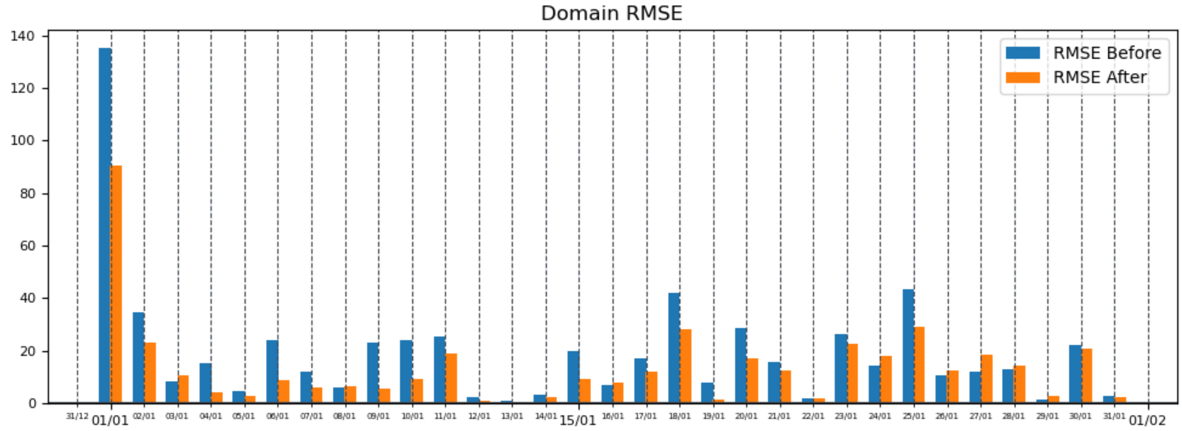

Figure 5. Domain-wide RMSE during the study period (1-31 January 2020) was calculated from 11 verification points as a sample. RMSE before (after) the adjustment is represented in blue (orange) bars.

The 1 January case shows an example of a combination of underestimation and removal of extreme points. The removal of Halim Perdanakusuma station which is a station with the highest rainfall record (377.2 mm) made bias for that particular point reach over 100 mm and even after the adjustment process the bias does not reduce significantly. The same manner is also shown by almost all of the verification points. For the case where there is almost no rainfall or only a slight amount of rainfall observed by both the observation and GSMaP, the RMSE is reduced to near zero. Specifically, biases from all observation points do not always show a reduction from their bias state after the adjustment process. Negative bias from a station could be flipped to positive bias after adjustment, and vice versa. In some cases, the adjustment process could increase the bias at that particular point. This is only for a small number of cases within the study period. The detailed bias results from all observation stations are shown in figure 6.

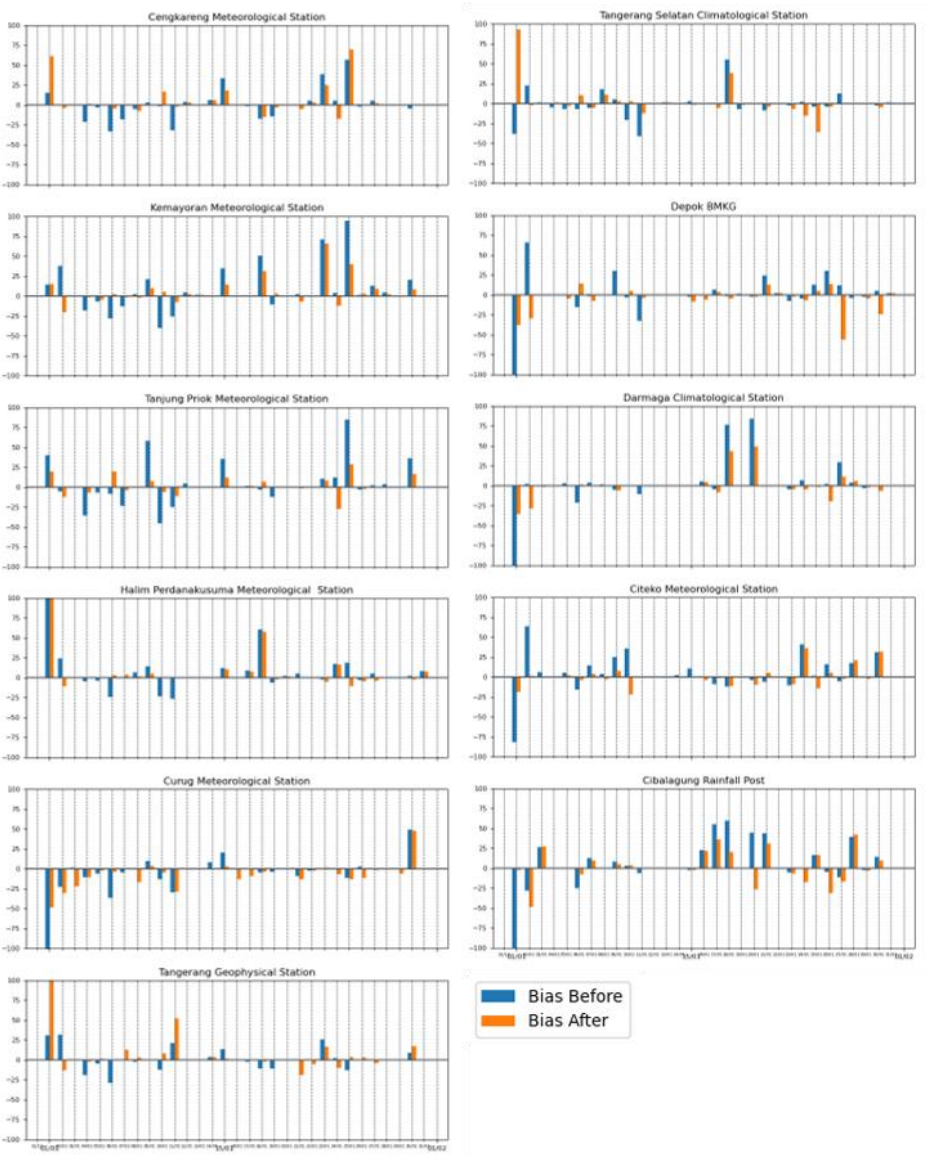

Figure 6. Bias before (blue bars) and after (orange bars) adjustment for each verification point.

4. Summary

Our results show that satellite-based precipitation could solve spatial issues such as lack of observation points over an area even though in some cases the data experienced underestimation and failed to capture localized rainfall detected by the observation point. The observation data could provide a more reliable measurement of rainfall but it could not cover all areas or be placed at a regular grid. Combining the advantages of both platforms could provide a more realistic spatial distribution of rainfall and intensity. This is proven by RMSE calculated before and after the adjustment process could reduce the area bias from 30% to 80%. For extreme rainfall cases, this adjustment process could give a good representation of rainfall concentration areas which could provide more accurate and reliable information with 30% reduction of domain RMSE. Meanwhile, for slight rain cases or localized rain cases that cannot be detected by GSMaP, the adjustment process could give 80% of area bias reduction.

Applying RBF interpolation for gridding the observation data shows better results by keeping the interpolation result localized so that it does not affect the data at a far point. There are many works been done to compare better interpolation methods for sparsely distributed rain gauge observation [12,13,14]. Comparing different interpolation method effect on this bias adjustment scheme should also be considered for future work. Consistency checks performed are proven could maintain consistency between GSMaP and Observation data pattern and avoiding extreme value exceeding maximum observation value arises after interpolation. Determining the optimum threshold for SCI index and search radius for consistency check so that a very localized rainfall in a remote point could be kept as reliable data could be one of the future works to improve this bias adjustment correction framework.

Acknowledgment
We would like to extend our acknowledgment to Tangerang Selatan Climatological Station and Bogor Climatological Station for providing the rainfall observation data from their rainfall observation network.

References
[1] Dinku T, Hailemariam K, Maidment R, Tarnavsky E and Connor S 2014 *Int. J. of Climatology* 34 pp 2489–2504
[2] Hardy, R. L. 1971 Multiquadric equations of topography and other irregular surfaces *J. of geophysical research*, *76*(8) pp 1905-1915
[3] Hardy, R.L., 1990. Theory and applications of the multiquadric-biharmonic method 20 years of discovery 1968–1988. Computers & Mathematics with Applications, 19(8-9), pp.163-208.
[4] Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., ... & Okamoto, K. I. 2007 *IEEE Transactions on Geoscience and Remote Sensing 45*(7) pp 2259-2275
[5] Ryu S, Song J J, Kim Y, Jung S H, Do Y and Lee G 2020 *Asia-Pacific J. of Atmospheric Sciences* pp 1–15
[6] Vila, D. A., De Goncalves, L. G. G., Toll, D. L., & Rozante, J. R. 2009 *J. of Hydrometeorology 10*(2) pp 533-543
[7] Dinku, T., Hailemariam, K., Maidment, R., Tarnavsky, E., & Connor, S. 2014 *Int. J. of Climatology* 34(7) pp 2489-2504
[8] Yang, Z., Hsu, K., Sorooshian, S., Xu, X., Braithwaite, D., Zhang, Y., & Verbist, K. M. 2017 *Journal of Geophysical Research: Atmospheres* 122(10) pp 5267-5284
[9] Chao, L., Zhang, K., Li, Z., Zhu, Y., Wang, J., & Yu, Z. 2018 J. of Hydrology 558 pp 275-289
[10] Kondragunta, C.R., 2001. In *AGU Spring Meeting Abstracts* (Vol. 2001, pp. H22A-07).
[11] Palaseanu, M. and Pearlstine, L., 2008. *Computers & Geosciences*, 34(7), pp.815-826.
[12] Goovaerts, P., 2000. *Journal of hydrology*, 228(1-2), pp.113-129.
[13] Lloyd, C.D., 2005. *Journal of Hydrology*, 308(1-4), pp.128-150.
[14] Kurtzman, D., Navon, S. and Morin, E., 2009. *Hydrological Processes: An International Journal*, 23(23), pp.3281-3291.