Fine-tuning satellite-based rainfall estimates

Hastuadi Harsa\textsuperscript{1}, Agus Buono\textsuperscript{2}, Rahmat Hidayat\textsuperscript{3}, Jaumil Achyar\textsuperscript{4}, Sri Noviati\textsuperscript{1}, Roni Kurniawan\textsuperscript{1} and Alfan S Praja\textsuperscript{1}

\textsuperscript{1} Research and Development Center, Indonesia Meteorology Climatology and Geophysics Agency, Jl. Angkasa I No. 2, Kemayoran, Jakarta 10720, Indonesia
\textsuperscript{2} Computer Science Department, Mathematics and Natural Science Faculty - Bogor Agricultural University, Bogor, Indonesia
\textsuperscript{3} Geophysics and Meteorology Department, Mathematics and Natural Science Faculty - Bogor Agricultural University, Bogor, Indonesia
\textsuperscript{4} Database Center, Indonesia Meteorology Climatology and Geophysics Agency, Jl. Angkasa I No. 2, Kemayoran, Jakarta 10720, Indonesia

E-mail: hastuadi@gmail.com

Abstract. Rainfall datasets are available from various sources, including satellite estimates and ground observation. The locations of ground observation scatter sparsely. Therefore, the use of satellite estimates is advantageous, because satellite estimates can provide data on places where the ground observations do not present. However, in general, the satellite estimates data contain bias, since they are product of algorithms that transform the sensors response into rainfall values. Another cause may come from the number of ground observations used by the algorithms as the reference in determining the rainfall values. This paper describe the application of bias correction method to modify the satellite-based dataset by adding a number of ground observation locations. The bias correction was performed by utilizing Quantile Mapping procedure between ground observation data and satellite estimates data. Since Quantile Mapping required mean and standard deviation of both the reference and the being-corrected data, thus the Inverse Distance Weighting scheme was applied beforehand to the mean and standard deviation of the observation data in order to provide a spatial composition of them, which were originally scattered. Therefore, it was possible to provide a reference data point at the same location with that of the satellite estimates. The results show that the new dataset have statistically better representation of the rainfall values recorded by the ground observation than the previous dataset.

1. Background
Rainfall measurement is carried out on the ground by the observation sites. The measured values are able to represent the rain events. However, the observations are scattered in non-homogeneous distance among others. On the other hand, satellite-based rainfall observations are able to provide data spatially, i.e. in homogeneous distance between their adjacent data locations. Therefore, the satellite-based data can be used as alternative sources where the observation sites do not present.

The data obtained from the satellite estimates are not true rainfall intensity. They are products of transformation algorithms applied to the reflectance received by the satellite sensors into rainfall intensity value. As the products of algorithms, they may contain intrinsic biases. There have been numerous researches regarding the comparison between data produced by satellite estimates and ground
observations. These researches identified bias of the satellite estimates to the ground observations data, so that the bias of satellite products can be suppressed. Toté [1] evaluated the satellite rainfall product for drought and flood monitoring in Mozambique. Vernimmen [2] studied three satellite estimates product in Indonesia and compared them with ground station data. The results were able to reduce the error in the satellites estimates. As-Syakur [3] compared daily, monthly, and seasonal satellite rain rates products with that of rain gauges in Bali.

CHIRPS (Climate Hazard Infra-Red Precipitation with Stations), a relatively new rainfall estimates dataset, is available at Climate Hazard Group (CHG) [4]. CHIRPS is derived from several data sources. It has 0.05° spatial resolution (~5 kilometers) and is delivered in daily, pentadal, and monthly forms. The CHIRPS dataset coverage area spans from 50°S to 50°N for all longitudes. Its temporal availability starts from 1981 to near-present. CHIRPS incorporates data blending procedure between ground measurements and the satellite-only rainfall estimates, another product of CHG: CHIRP (Climate Hazard Infra-Red Precipitation).

2. Data and study area

In this research, the data were located in Java Island. Java has the densest population in Indonesia and also contributes significant agricultural harvests to the national Gross Domestic Product. Rice, corn, and sugar are the examples of common food crops in Java. A better understanding of rainfall characteristics may increase the success of the farmers to supply the demand of these agricultural products.

Figure 1 shows the study area. The area spans from 104.975°E to 115.025°E and from 9.025°S to 4.975°S, as specified by dashed bounding box. The spatial resolution of CHIRPS data is 0.05° or approximately 5 kilometers. By that resolution, there are 202 longitude units and 82 units of latitude, i.e. 16564 data locations (represented in grid points). CHIRPS data are available on the land only, so that the number of grid points having numerical data is 5177. The rest of the grid points do not contain data as they are located on the ocean. The available data locations is indicated by the white area on the map in figure 1, while the ocean is indicated by the grey area.

The observation data were available from Indonesia Meteorology Climatology and Geophysics Agency (BMKG). There are two types of BMKG ground observation data: the open-access data and the reserved data. The BMKG open-access data locations are shown by the yellow rectangle while the red circle show the BMKG reserved data locations. There were 29 locations represented the open-access data and 563 locations represented the reserved data. The BMKG open-access data locations are World Meteorology Organization (WMO) members. CHG also collects the observation data from those open-access locations to perform the data blending process between the ground-based observation and the
satellite-only data. Both the observation data and the CHIRPS data collected for this research are in the length of 12 years: from January 1st, 2005 until December 31st, 2016. The data are in daily temporal resolution. These data denote the rainfall intensity values and have the units of millimeter per day.

3. Methods
This research aims to modify the CHIRPS data using additional ground observation data. It is expected that this procedure is able to make CHIRPS data characterize the ground observation data. In order to accomplish this objective, there are two methods utilized: Quantile Mapping (QM) and Inverse Distance Weighting (IDW).

3.1. Quantile mapping
QM is a non-parametric bias correction (BC) method [5]. This method transform all value in a dataset onto new values. The transformed values construct a new dataset. The values in the new dataset can eventually have the occurrence probability as desired. This process requires a determination of new probability function to map a value having a particular probability occurrence from previous dataset.

Piani [6] demonstrated the use of QM to correct the bias of global simulated daily precipitation and temperature. The illustration of QM is given by figure 2. Suppose that there are two datasets as plotted at the top row, the data in red will be adjusted to be closer to the data in black. By calculating the mean and standard deviation of each dataset, then the Probability Density Function (PDF) of the values occurred in each dataset can be obtained as shown by the left chart at middle row. Note that in generating the PDFs here, the values are presumed to be in Gamma distributed data, as the rainfall values are. Nevertheless, another distribution function may be used regarding to the nature of the data. After the PDFs are available, the Cumulative Distribution Function (CDF) of each PDF can be calculated, as shown by the right chart at middle row. The transformation of a value (4.9 for example as shown in the

![Figure 2. Quantile mapping illustration.](image)
chart) is performed by looking up its location at the CDF of the data being corrected (around 0.7 at the red CDF curve) and looking up the values from the chosen CDF (black curve) with the same location of CDF (also 0.7). The location 0.7 belongs to the value 18 in black CDF. Therefore, every time value 4.9 occurred in the red dataset, it is replaced by the value 18. These steps are completed for every value found in the data being corrected and yield a new dataset as plotted by the blue lines at the bottom figure.

The blue data, derived from the previous data in red, have more similar mean and standard deviation with the reference data in black. The blue dataset has similar oscillation with that of the red dataset. The blue data set is still a representation of the red dataset, but it is now shifted up and stretched.

To accommodate the non-stationary characteristic of rainfall data, the time series of every CHIRPS grid and the ground observation locations were grouped in month (January to December). The mean and standard deviation were calculated for each month. Hence, there were 24 features calculated for every CHIRPS grid and ground observation locations. The monthly mean and standard deviation of every CHIRPS grid were then used to construct the monthly PDF and CDF of their corresponding grid.

3.2. Inverse Distance Weighting

After the monthly CDFs of every CHIRPS data were available, the next process was determining the reference CDFs. Each CHIRPS grid must have its own reference CDF. The reference CDF was constructed from the interpolation of mean and standard deviation of the ground observation locations. The interpolation procedure was necessary because the position of CHIRPS data locations did not exactly match the location of the ground observation. The number of ground observation locations included in interpolation process for one CHIRPS grid was determined to be three locations. The reference CDFs was provided for each CHIRPS grid.

The interpolation method used in this research was Inverse Distance weighting (IDW), as highlighted by Beek [7] and Sluiter [8] regarding its usage in meteorological data. Chen [9] also used IDW to estimates the spatial rainfall distribution. In this paper, IDW was applied to interpolate the mean and standard deviation of three nearest ground observation location to one CHIRPS grid. The power parameter for distance was set to 2. An illustration of IDW is shown by figure 3. The overall procedures in this research were as follows:

- Extract monthly mean and monthly standard deviation for each observation and for each grid point of estimates (yields 12 mean and 12 standard deviation values) from daily data.
- Generate monthly PDF of all CHIRPS grid.
- Generate monthly CDF from the monthly PDF of all CHIRPS grid.

![Figure 3. Inverse distance weighting illustration.](image)
• Provide interpolated mean and standard deviation using IDW from three nearest ground observation for a CHIRPS grid.
• Generate monthly reference PDF of all CHIRPS grid.
• Generate monthly CDF from all monthly reference PDF.
• Transform the value of CHIRPS using QM based on its monthly CDF and its monthly reference CDF.

4. Results and Discussion
The default condition of the data from the ground observation and CHIRPS is shown by the boxplots in figure 4. Both datasets have the same monthly rainfall pattern, i.e. monsoon, which constitute the ‘U’ letter in their mean values. The estimates (red) have higher rainfall mean than the observation (white), but lower at the standard deviation. IDW adjusted the mean and standard deviation of estimates so that they are closer to the observation as shown by the blue boxplot in figure 3 as well. Note that the boxplot in figure 4 is derived from the condition of all ground observation locations and their nearest CHIRPS grid. The example of spatial monthly mean condition is shown by figure 5. Only the mean condition of
the wettest and driest month (January and August) are displayed. The dots represent the ground observation locations. They are drawn in front of the CHIRPS grids. Therefore, if they still appear, there are still differences between the CHIRPS grids and the ground observations. It is found that CHIRPS still has many different values with the ground observations. The result of bias correction has modified the mean of CHIRPS to be closer to the ground observation as indicated by less visible dots appearing. The modification also affected the standard deviation.

Figure 6 shows the Quantile-Quantile (QQ) plot of CHIRPS versus the ground observation it is found that the data are distributed in Gamma as indicated by the denser distribution near zero value and getting sparse as the values getting higher. The bias corrected of CHIRPS lies nearer the perfect slope than the CHIRPS original data. The time series example of the new dataset which is derived from CHIRPS is shown by figure 7. The time series charts are taken from the same locations of figure 6. The time series
indicate that the new dataset (blue) has bigger range than CHIRPS (red). Some statistical calculations of the CHIRPS and the new dataset compared to the ground observation dataset is shown by figure 8. The monthly mean of statistical parameters calculation was performed on every locations of the ground observations, CHIRPS, and the new dataset. To compare the results, the monthly mean of statistical parameters of CHIRPS and the new dataset were each subtracted from the monthly mean of statistical parameters of ground observations. Note that the locations included in figure 8 are all ground observation locations and their nearest grid of CHIRPS and the new dataset.

5. Conclusion
This research has produced an alternative dataset of daily rainfall intensity derived from the satellite based combined with ground station (CHIRPS) products. The new data were formed by modifying the probability of the values obtained from CHIRPS using Quantile Mapping (QM). To establish the QM process, a parameter estimation procedure was first conducted to construct new spatial Probability Density Function (PDF) and Cumulative Density Functions (CDF) subsequently over grids of satellite estimates using Inverse Distance Weighting (IDW). IDW provided the spatial interpolation of the CDF parameters obtained from the observation as the reference, which was located sparsely. The results show that IDW-QM succeeds in pulling the satellite estimates closer to the observation.

References
[1] Toté C, Patricio D, Boogaard H, Wijngaart RVD, Tarnavsky E and Funk C 2015 Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique Remote Sensing 7 1758-1776
[2] Vernimmen RRE., Hooijer A, Mamunun and Aldrian E 2012 Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia Hydrology and Earth System Sciences 16 133–146
[3] As-Syakur A R, Tanaka T, Prasetya R, Swardika I K and 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
[4] Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A and Michaelsen J 2015 The Climate Hazards Infrared Precipitation with Stations—A New Environmental Record for Monitoring Extremes *Scientific Data* 2

[5] Valdés-Pineda R, Demaría E M C, Valdés J B, Wi S and Serrat-Capdevil A 2016 Bias correction of daily satellite-based rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa *Hydrology and Earth System Sciences*

[6] Piani C, Haerter JO and Coppola E 2010 Statistical bias correction for daily precipitation in regional climate models over Europe *Theoretical and Applied Climatology* 99 187–192

[7] Beek EG 1991 Spatial interpolation of daily meteorological data; theoretical evaluation of available techniques. *DLO The Winand Staring Centre* Report 53.1

[8] Sluiter R 2009 Interpolation methods for climate data: literature review *Roy. Netherlands Meteorological Inst. Tech. Rep.*

[9] Chen F W and Liu C W 2012 Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan *Paddy Water Environ* 10 209