Improved Performance of The CHIRPS Monthly Rainfall Estimation Extraction from Google Earth Engine (GEE) platform in South Sulawesi Region

L Bangsawan¹,³, M C Satriagasa² and S Bahri⁴

¹Master Program of Science in Remote Sensing, Faculty of Geography, Gadjah Mada University, Yogyakarta
²Faculty of Forestry, Gadjah Mada University, Yogyakarta
³Sorong Meteorological Station, Indonesia Agency for Meteorology Climatology and Geophysics
⁴Maros Climatological Station, Indonesia Agency for Meteorology Climatology and Geophysics

odebangsawan@gmail.com

Abstract. The integration of the availability and processing of The Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) data by the Google Earth Engine (GEE) platform is used in this study to extract the estimated monthly rainfall in South Sulawesi. Several areas are selected based on the characteristics of the rainy period cycle representing South Sulawesi, namely Makassar, Masamba, Wajo, and Bone. Monthly rainfall estimation data of CHIRPS in the year 2019 were validated by monthly observed rainfall at the same period showing the CHIRPS rainfall estimation has not been maximized with correlation coefficient values are 0.94, 0.63, 0.65, 0.75, and RMSE percentage 54%, 52%, 95%, 64% for each of the study areas. Then the increase in rainfall estimation performance is carried out by applying multiple linear regression method and considering both monthly observed and estimated rainfall during 30 years from 1989 to 2018, latitude and longitude point as well as elevation in every location. The results show an increase of correlation coefficient to 0.95, 0.74, 0.74, and 0.87 and a general decrease of RMSE percentage to 53%, 39%, 80%, and 67%. Thus, monthly rainfall estimation performance improvement is successfully achieved in various rainy period cycles of the study area.

Keywords: Google Earth Engine (GEE), CHIRPS, Multiple linear regression.

1. Introduction
Precipitation is an important element that affects various areas of human life and the environment. By knowing the amount of rain that falls each month, the climate characteristics can be seen in an area to support agricultural activities, irrigation, tourism, mitigating extreme climate, and others [1–3]. Research on rainfall estimation with model and satellite approach has been widely carried out by utilizing data obtained from web providers and then processed with certain applications on each computer. The presence of the GEE platform facilitates the integration between the availability of big
data through open access and processing it via cloud computing to make data processing more efficient and simplify the performance of the computers used [4–6].

The researchers have conducted several studies related to the use of the GEE platform in rain satellite data processing through estimating and evaluating diurnal rainfall variations in the basin area to analyze rainfall patterns using Tropical Rainfall Measuring Mission (TRMM) satellite data. [7, 8]. Spatiotemporal evaluation and rainfall variability have also been carried out using the CHIRPS dataset on the GEE platform [9]. Other researchers use GEE to process long-term satellite rainfall estimation data including decadal, monthly, and seasonality for dozens of years to monitor meteorological drought in their research area, [10]. CHIRPS is a rainfall database that is a combination of three rainfall information, namely global climatology, satellite-based rainfall estimation, and rainfall observed in-situ with 0.05 ° latitude-longitude resolution and quasi-global coverage (50°N–50°S, 180°W–180°E) [11]. This makes CHIRPS very useful for providing rainfall estimation data in areas where there is no rain gauge [12]. CHIRPS data has been evaluated in several countries and has been proven to be one of the best rainfall estimation databases, however, the results still vary depending on the character of the location. [13]. This study aims to determine the performance of the CHIRPS rainfall estimation obtained and processed through the GEE platform and to improve the rainfall estimation performance by considering several geospatial variables such as geolocation and topography as well as considering the rain cycle as a characteristic of rainfall in the study area.

2. Study Area
South Sulawesi is a province in central Indonesia that has a complete cycle of rainy periods, making this area interesting to study.

Figure 1. Rain cycle over South Sulawesi
(Source: Putra, 2014)

Figure 1 shows the rain cycle in South Sulawesi consisting of annual and semi-annual cycles and a combination of both with different rain peaks. The western part of South Sulawesi is dominated by an annual rain cycle with one peak of rain occurring at the beginning of the year. The northern and central parts of South Sulawesi are dominated by a semi-annual rain cycle with two peaks of rain occurring in the fourth and mid-year months.
The combination of annual and semi-annual rain cycles occurs in the southern part, where the peak of rain occurs once in the middle of the year \cite{14}. There is a difference between the rain cycle in the western part of South Sulawesi and the eastern part where the western rain cycle is more influenced by monsoon activity \cite{15,16}. In this study, several regions were selected to represent the rain cycle in South Sulawesi. Makassar, Masamba, Wajo, and Bone were selected to represent areas with annual, semi-annual, and combined annual and semi-annual rain cycles.

3. Data and Method

The data needed in this research are as follows:

a. CHIRPS monthly rainfall estimation data for the period 1989 to 2019 in the Makassar, Masamba, Wajo, and Bone areas.

b. Monthly rain observation data from rain gauges at related rain gauge stations for the period 1989 to 2019. In months where the monthly rainfall data is not recorded, it is interpolated using the average monthly rainfall data.

c. Geolocation data in the form of longitude and latitude data for the rain gauge of the related area.

d. Topographical data in the form of rain gauge elevation data for the related area.

Geolocation data and rain gauge stations topography are presented in table 1 below:

| Location | Longitude | Latitude | Elevation |
|----------|-----------|----------|-----------|
| Makassar | 119.552   | -5.071   | 14        |
| Masamba  | 120.324   | -2.554   | 50        |
| Wajo     | 120.038   | -4.171   | 20        |
| Bone     | 120.093   | -4.859   | 132       |

The stages carried out in this research are:

a. Extracting 2019 CHIRPS monthly rainfall estimation data at each rain gauge location in Makassar, Masamba, Wajo, and Bone using the GEE platform.

b. Build a multiple linear regression model to improve the performance of the CHIRPS rainfall estimation using CHIRPS monthly rainfall estimation and observed rain gauge data for 30 years from 1989 to 2018 and take into account the geolocation and topographic variables of the related area.

c. Applying the model equation by entering the estimated value of CHIRPS rainfall in 2019, geolocation, and topographic variables then comparing it with the CHIRPS rainfall estimate and observed rain gauge.

d. Conducting model accuracy tests by calculating the correlation coefficient and RMSE percentage of the improved CHIRPS model, CHIRPS, and observed rain gauge in 2019 in the related areas.

The multiple linear regression model is built using the formula below:

\[ y = a + bx_1 + cx_2 + dx_3 + ex_4 \]  \hspace{1cm} (1)

With \( y \) = improved CHIRPS monthly rainfall estimate, \( x_1 \) = CHIRPS monthly rainfall estimate, \( x_2 \) = longitude, \( x_3 \) = latitude, \( x_4 \) = elevation, \( a \), \( b \), \( c \), \( d \) dan \( e \) are constants that would be vary every month due to the difference of observed and CHIRPS monthly rainfall used during 30 years in building the model as showed in table 2. To validate the model results in equation (1) the correlation coefficient formula and RMSE percentage are used as shown in the following formula:

\[ r = \frac{\sum(c-\bar{c})(g-\bar{g})}{\sqrt{\sum(c-\bar{c})^2} \sqrt{\sum(g-\bar{g})^2}} \]  \hspace{1cm} (2)
\[ RMSE = 100 \sqrt{\frac{1}{n} \sum (c-g)^2} \]

where \( c \) = improved CHIRPS monthly rainfall estimate, \( \bar{c} \) = average of improved monthly CHIRPS rainfall estimate, \( g \) = observed monthly rainfall \( \bar{g} \) = average monthly rainfall, and \( n \) = amount of data. Observed monthly rainfall is created from summation of daily rainfall measured by rain gauge at each related month. Correlation coefficient values and RMSE are generally used to validate continuous data in the estimation model [17, 18]. The correlation coefficient shows the closeness of the relationship between two parameters measured on a scale of -1 to 1 where -1 = negative linear relationship, 1 = positive linear relationship and 0 = no relationship. The RMSE percentage shows the large difference between the two parameters so that the smaller RMSE percentage indicates the smaller error between observed and model values.

4. Results and Discussion

4.1 Monthly rainfall comparison between CHIRPS, Improved CHIRPS and rain gauge

The CHIRPS monthly rain extortion process in 2019 on the GEE platform uses the JavaScript Application Program Interface (API) programming language which is input into a code editor and then run. The earth engine code editor is a web-based Integrated Development Environment (IDE) that is accessed online to run the JavaScript API. The extraction results are shown in Figure 2 below:

![Figure 2. Display of CHIRPS monthly rainfall estimation using GEE](https://code.earthengine.google.com/)

CHIRPS that are part of the GEE dataset can easily be imported. Likewise, the script that has been available to extract rainfall estimate data is available (available at https://code.earthengine.google.com/). In this study, modifications were made to the available script which was originally used to extract daily rainfall estimates so that it can be used for the extraction of monthly rainfall estimates with the help of API documentation. Such activities in Figure 2 are applied to each of the Makassar, Masamba, Wajo, and Bone research locations where geolocation as a geometry point is an important input in the process. CHIRPS monthly rainfall extraction during the 1989 to 2018 period was also carried out to build a model where the rain value obtained on the console menu was exported to be processed by taking into account the monthly rainfall, geolocation, and topography of all rain gauges in the study area, as mentioned in table 1 so that a multiple linear regression model was created as shown in the table 2 below:
Table 2. Multiple linear regression model

| Month   | Multiple Linear Regression Equation                        |
|---------|------------------------------------------------------------|
| January | $y = 29045.9 + (1.05x_1) - (239.86x_2) + (75.17x_3) - (0.18x_4)$ |
| February| $y = 27693.96 + (1.15x_1) - (229.90x_2) + (48.98x_3) - (0.08x_4)$ |
| March   | $y = 18557.80 + (0.73x_1) - (152.49x_2) + (57.84x_3) - (0.02x_4)$ |
| April   | $y = 12032.96 + (0.52x_1) - (97.31x_2) + (56.06x_3) - (0.03x_4)$ |
| May     | $y = -32072.98 + (0.30x_1) + (268.5x_2) - (9.16x_3) + (0.03x_4)$ |
| June    | $y = -25491.13 + (0.66x_1) + (213.22x_2) - (2.13x_3) - (0.08x_4)$ |
| July    | $y = -11037.39 + (0.82x_1) + (91.95x_2) - (10.52x_3) - (0.09x_4)$ |
| August  | $y = -2992.24 + (0.77x_1) + (25.82x_2) + (19.32x_3) - (0.13x_4)$ |
| September| $y = 3623.98 + (0.76x_1) - (28.14x_2) + (51.17x_3) - (0.19x_4)$ |
| October | $y = 15511.90 + (1.01x_1) - (127.2x_2) + (55.67x_3) - (0.11x_4)$ |
| November| $y = 10016.64 + (0.86x_1) - (82.03x_2) + (39.47x_3) - (0.08x_4)$ |
| December| $y = 40334.90 + (0.92x_1) - (333.24x_2) + (93.50x_3) - (0.23x_4)$ |

Furthermore, the estimated value of CHIRPS monthly rainfall in 2019, geolocation, and topographic variables at each location are substituted to the equation in Table 3 to obtain the improved monthly rainfall estimation of CHIRPS which is then compared with the monthly rainfall estimation of CHIRPS and observed monthly rainfall of rain gauge. The comparison is shown in the curves on figure 3.

Figure 3. Comparison of monthly rainfall estimate curves for CHIRPS, improved CHIRPS and observed rain gauge in (a) Makassar, (b) Masamba, (c) Wajo, and (d) Bone
Figure 3 shows the rainfall estimate cycle patterns of the CHIRPS, improved CHIRPS, and rain gauge in general, which are similar across all locations. Makassar has an annual rain cycle character with one peak visible on the CHIRPS curve still showing the same pattern on the improved CHIRPS curve and rain gauge. The maximum monthly rainfall in Makassar which on the CHIRPS curve occurs in January also occurs in the same month on the improved CHIRPS curve and rain gauge, as well as the minimum monthly rainfall which both occurs in August. The CHIRPS estimation ability performs better during the peak rainy season [19]. The greater availability of water vapor during the peak rainy season makes CHIRPS 'performance more responsive in estimating rainfall [20]. Masamba and Wajo which have semi-annual rain cycle characteristics, there was also a similarity in the form of two rain peaks between the CHIRPS estimates with the improved CHIRPS and the rain gauge. Even so, there is a difference in the time of peak monthly rainfall, wherein Masamba the estimated monthly rainfall peak of CHIRPS and improved CHIRPS occurred in February and April, the rain gauge occurred in April and June. The timing of different peak rainfall events between CHIRPS, improved CHIRPS, and rain gauge also occurred in Wajo where the peak of CHIRPS rains occurred in February and April while the improved CHIRPS was in April and May but the rain gauge was in April and June. For Bone with combined annual and semi-annual rain cycle characteristics, the peak monthly rainfall estimate for CHIRPS and rain gauge both occurred once but for CHIRPS it occurred in April while improved CHIRPS was in May and the rain gauge was in June. This implies difficulty for the CHIRPS and improved CHIRPS to determine the peak monthly rainfall in locations characterized by a semi-annual rainfall cycle and a combination of semi-annual with annual.

4.2 Validation

Correlation coefficient values and error for monthly rainfall estimation of CHIRPS, improved CHIRPS, and observed rain gauge are shown in the table 3 below:

**Table 3. Comparison of correlation coefficient values and error for monthly rainfall estimation of CHIRPS, improved CHIRPS, and observed rain gauge**

| Location | CHIRPS vs Rain Gauge | Improved CHIRPS vs Rain Gauge |
|----------|----------------------|-------------------------------|
|          | r  | RMSE (%) | r  | RMSE (%) |
| Makassar | 0.94 | 54 | 0.95 | 53 |
| Masamba  | 0.63 | 52 | 0.74 | 39 |
| Wajo     | 0.65 | 95 | 0.74 | 80 |
| Bone     | 0.75 | 64 | 0.87 | 67 |

Table 3 indicates that there are differences in the correlation value of the CHIRPS rainfall estimate against the rain gauge at various locations. Complex topographic variables also influence the difference in the validation results of the rainfall estimates [17]. On the other hand, in general, there is an increase in the value of the correlation coefficient and a decrease in error against the monthly rain gauge between before and after the improved performance of the CHIRPS monthly rainfall estimation. Makassar has the strongest positive relationship both before and after the improved CHIRPS rainfall estimate with correlation coefficient values of 0.94 and 0.95 which indicate locations with the characteristics of the annual rain cycle as in Figure 3 (a) are more accurate to predict. The annual rain cycle in Makassar resembles the monsoonal rain cycle where areas with a monsoonal rain cycle have high accuracy in CHIRPS rainfall estimation [21]. The topography of Makassar which is a plain area and has a lower elevation than the other areas also supports the high correlation value because CHIRPS performance is proven to be better in relatively flat and open areas [19].

Bone experienced the largest increase in the correlation coefficient value compared to other locations which have as many as 0.12 to became 0.87 but Bone was also the only location that experienced an increase in the RMSE percentage value from 64% to 67%. This is due to the wider
variability of rain values during the peak rainy period as shown in Figure 3 which cannot be perfectly responded by the CHIRPS monthly rainfall estimation even though improvements have been applied to the model used. Masamba and Wajo both experienced an increase in the correlation coefficient up to 0.74. The difference in the increasing value of the correlation coefficient was greater for Masamba than for Wajo that is namely 0.11 compared to 0.09. Yet the difference in decreasing the error value was greater for Wajo than for Masamba that is 15% versus 13%. These indicates that the difference in changes in the value of the correlation coefficient and RMSE is not too significant at locations with a semi-annual rain cycle.

5. Conclusion
Increasing correlation values and decreasing error values generally occur in South Sulawesi after improvements using the multiple linear regression model equation which takes into account monthly observed and CHIRPS rainfall estimation during 30 years from 1989 to 2018, geolocation, and topography in every location. The correlation values which were originally 0.94, 0.63, 0.65, and 0.75 increased to 0.95, 0.74, 0.74, and 0.87, and the RMSE percentage which was originally 54%, 52%, 95%, and 64% were generally reduced to 53%, 39%, 80%, and 67% for the locations of Makassar, Masamba, Wajo, and Bone consecutively. This shows that the improvement in the estimated performance of CHIRPS has been successfully implemented.

Rainfall peak at the location with an annual rain cycle that is affected by monsoonal activity such as Makassar is easier to predict by the resulted model than rainfall peak at the locations with a semi-annual rain cycle and a combination of annual and semi-annual such as Masamba, Wajo, and Bone. It is necessary to apply other models that are more able to respond to the peak rainfall events for various characteristics of the rain cycle in the study area.

Acknowledgments
The author feels grateful to Maros climatological station that is one of the BMKG branch offices in South Sulawesi which has provided rainfall data in this study. High appreciation is also given to GEE developers who have provided registration access for the use and processing of their CHIRPS dataset.

References
[1] Mab P, Ly S, Chompuchan C, Kositsakulchhai E 2019 Evaluation of Satellite Precipitation from Google Earth Engine in Tonle Sap Basin, Cambodia. THA 2019 Int Conf Water Manag Clim Chang Towar Asia’s Water-Energy-Food Nexus SDGs, January 23-25. 2019.
[2] Koutsouris AJ, Seibert J, Lyon SW 2017 Utilization of global precipitation datasets in data limited regions: A case study of Kilombero Valley, Tanzania. Atmosphere (Basel). 8(12).
[3] Awal R, Bayabil HK, Fares A 2016 Analysis of potential future climate and climate extremes in the brazos headwaters Basin, Texas. Water (Switzerland). 8(12).
[4] Amani M, Ghorbanian A, Ahmadi SA, Kakooei M, Moghimi A, Mirmazloumi SM, et al 2020 Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. IEEE J Sel Top Appl Earth Obs Remote Sens.13:5326–50.
[5] Tamiminia H, Salehi B, Mahdianpari M, Quackenbush L, Adeli S, Brisco B 2020 Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS J Photogramm Reman Remote Sens.164 (May):152–70.
[6] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R 2017 Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ. 202:18–27.
[7] Elnashar A, Zeng H, Wu B, Zhang N, Tian F, Zhang M, et al 2020 Downscaling TRMM monthly precipitation using google earth engine and google cloud computing. Remote Sens. 12(23):1–22.
[8] Zhu L, Zhao Y, Rui X, Wei Q 2020 Diurnal Variation of Seasonal Precipitation over the CONUS: A Comparison of Gauge Observations with TRMM Data. Adv Meteorol. 2020.
[9] Banerjee A, Chen R, Meadows ME, Singh RB, Mal S, Sengupta D 2020 An analysis of long-term rainfall trends and variability in the uttarakhand himalaya using google earth engine. Remote
[10] Bayissa Y, Tadesse T, Demisse G, Shiferaw A 2017 Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the Upper Blue Nile Basin, Ethiopia. *Remote Sens.* 9(7).

[11] Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al 2015 The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Sci Data.* 2:1–21.

[12] Belay AS, Fenta AA, Yenehun A, Nigate F, Tilahun SA, Moges MM, et al 2019 Evaluation and application of multi-source satellite rainfall product CHIRPS to assess spatio-temporal rainfall variability on data-sparse western margins of Ethiopian highlands. *Remote Sens.* 11(22):1–22.

[13] Beck HE, Vergopolan N, Pan M, Levizzani V, van Dijk AIJM, Weedon GP, et al 2020 Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Adv Glob Chang Res.* 69:625–53.

[14] Putra MDJ. Annual Rain Cycle Analysis. 2014 *Thesis.* ITB. Bandung. https://digilib.itb.ac.id/index.php/gdl/view/16326

[15] Giarno G, Hadi MP, Suprayogi S, Murti SH 2018 Distribution of Accuracy of TRMM Daily Rainfall in Makassar Strait. *Forum Geogr.* 32(1):38–52.

[16] Setiawan AM, Koesmaryono Y, Faqih A, Gunawan D 2017 Observed and blended gauge-satellite precipitation estimates perspective on meteorological drought intensity over South Sulawesi, Indonesia. *IOP Conf Ser Earth Environ Sci.* Jan; 54(1):012040.

[17] Dinku T, Ceccato P, Grover-Kopec E, Lemma M, Connor SJ, Ropelewski CF 2007 Validation of satellite rainfall products over East Africa’s complex topography. *Int J Remote Sens.* 28(7):1503–26.

[18] Caroletti GN, Coscarelli R, Caloiero T 2019 Validation of satellite, reanalysis and RCM data of monthly rainfall in Calabria (Southern Italy). *Remote Sens.* 11(13).

[19] Paredes Trejo FJ, Barbosa HA, Peñaloza-Murillo MA, Alejandra Moreno M, Farias A 2016 Intercomparison of improved satellite rainfall estimation with CHIRPS gridded product and rain gauge data over Venezuela. *Atmosfera.* 29(4):323–42.

[20] Sacidizand R, Sabetghadam S, Tarnavsky E, Pierleoni A 2018 Evaluation of CHIRPS rainfall estimates over Iran. *Q J R Meteorol Soc.* 144(May):282–91.

[21] Bai L, Shi C, Li L, Yang Y, Wu J 2018 Accuracy of CHIRPS satellite-rainfall products over mainland China. *Remote Sens.* 10(3).