# Improving the accuracy of rainfall data in the Philippines through concurrent use of GPM and ground-based measurements

Archie I. Veloria[1], Gay Jane P. Perez[1], Giovanni A. Tapang[2], and Josefino C. Comiso[1,3]
[1] Institute of Environmental Science and Meteorology, University of the Philippines Diliman, Philippines
[2] National Institute of Physics, University of the Philippines Diliman, Philippines
[3] NASA Goddard Spaceflight Center, Maryland, United States

## Abstract

Reliable rainfall information is necessary in order to mitigate effects of hazards such as flooding and enable research on weather forecasting, agrometeorology, and climatology. Synoptic stations in the Philippines provide the best rainfall data in the country through maintained and well-manned observations. However, these stations are sparse and can represent rainfall only at its immediate vicinity when localized weather systems are considered. Dense network of automatic rain gauges can fill in rainfall information all over the country. While these ARGs can provide near-realtime measurements, these unmanned stations are prone to instrument malfunctions. Satellite products such as GPM provide gridded rainfall that overcomes the limitations of ground-based measurements spatially and temporally. This study focused on making GPM and ground-based rainfall measurements consistent with each other through comparative analyses. Because synoptic stations are point measurements, mismatches in rainfall data dominates the comparison with GPM. Synoptic stations were then used to validate rainfall data from ARGs. While agreements in synoptic and ARG data are distance-dependent, ARGs tend to report less rainfall when relatively close stations in a homogenous location were compared. Generalized reduced gradient algorithm was applied to ARG rainfall to make it consistent with synoptic data. Bias were reduced from 1.93 (1.69) mm/day to 0.19 (0.09) mm/day for dekad (monthly) rainfall. Comparative analysis of ground stations and GPM was done using ample number of ARGs and synoptic stations within the GPM footprint. A correlation of 0.77 and 0.86 were obtained at the dekad and monthly comparison. While a slope of 0.86 (0.89) was obtained for dekad (monthly) rainfall, RMSE (MAE) of 5.13 (3.27) mm/day and 3.09 (2.15) mm/day was observed for dekad and monthly comparison, respectively. Generalized reduced gradient algorithm was applied to further bias correct GPM. Statistics further improved after making the GPM consistent with ground measurements within its footprint. The results of this study may be applied to generate long-term gridded rainfall from bias-corrected GPM at the dekad and monthly scales.

## 1. Introduction

Rainfall is essential to life and is the primary source of fresh water needed by humans, plants and animals. It is responsible in the movement of enormous amount of water and heat through the Earth's atmosphere and is a major part of the Earth's energy budget and climate. In the Philippines, rain is driven mainly by the Monsoons: the Summer Monsoon, called *Habagat*, and the Winter Moonson called *Amihan*. The Summer Monsoon is usually carried by southwest winds and is the predominant weather pattern from April through September. The Winter Monsoon, on the other hand, overlaps with the Summer Monsoon for about a week and becomes part of the weather pattern for the rest of the year through March the following year. The Winter Monsoon usually originates in Siberia or Mongolia and transported through an anti-cyclonic northeast winds. Other sources of rain include the occurrences of tropical cyclones that comes from the mid-Pacific Oceans and reach the Philippines a few times a year.

Reliable information on rainfall in the Philippines at sufficient spatial and temporal resolution is important because of a number of reasons. Firstly, the Philippines is primarily an agricultural country and it is critical to know rainfall patterns and possible occurrences of drought. Secondly, rainfall data provide the means to assess the rate and persistence of precipitation events and how they may affect communities. And third, they are also needed in the development of risk assessment models that are used to provide early warning systems (Dembélé and Zwart 2016). Currently, the key sources of rainfall data are the 51 synoptic stations installed and maintained by the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and about 2,000 automatic rain gauges (ARG) installed by the Department of Science and Technology – Advance Science and Technology Institute (DOST-ASTI). The synoptic data are regarded as the more accurate and more dependable data set because the instruments were built, calibrated, and maintained according to the standards set by the World Meteorological Organization (WMO). The ARGs are unmanned stations that provide real time rainfall measurements every 10/15 minutes and provide much improved spatial coverage over the entire country compared with synoptic data coverage. However, the quality of the data may not be as good as the synoptic measurements because of less attention to proper calibration and maintenance of the sensors.

The synoptic stations were also designed to monitor synoptic scale systems running from 200 km to 10,000 km spatial radius and lifetime of 1 day to 1 month (Lin 2007). Such systems include tropical cyclones, high and low pressure areas, and air masses. However, the insufficient density of these stations leads to errors in representing the areal distribution of rainfall within a synoptic system (Mishra 2013). The low density of synoptic stations also makes it difficult to detect patterns brought by small scale systems such as thunderstorms. The situation is improved significantly with the introduction of ARGs. But even with reliable and quality checked ARGs, there are still considerable gap in the spatial coverage, especially in mountainous and difficult to reach areas.

The situation can be considerably improved through additional use of satellite remote sensing data. Historically, satellite rainfall data have had problems with statistical gaps in coverage because of the dynamic nature of rainfall events and inability of polar orbiting sensors to cover all events especially when they are regional and short term. However, the satellite rainfall products are becoming more and more capable for hydrological and climate studies and now have reasonable spatial and temporal resolution (Dembélé and Zwart 2016). The most popular and widely used product has been that from the Tropical Rainfall Measuring Mission (TRMM) (Fensterseifer et al. 2016). With a global coverage, spatial resolution of 0.25°, and temporal resolution up to 3 hours, TRMM can provide continuous rainfall measurements covering the whole Philippines. However, some studies show that satellite products are prone to over/underestimation depending on the season and location (Aghakouchak et al. 2011; Jamandre and Narisma 2013). The Global Precipitation Measurement (GPM) continued the legacy of TRMM after it has decommissioned in 2014. GPM provides global coverage at higher spatial (0.1°) and temporal (30 minutes) resolution than TRMM (Huffman et. al 2019). Satellite rainfall products have been validated and merged with ground-based measurements on a global scale but it is not known to what extent they are consistent with synoptic and ARG data in the Philippines.

The goal of this study is to develop a rainfall data set that can be used to accurately assess large scale distribution and patterns of rain in the entire country through concurrent use of GPM and synoptic and ARG ground-based measurements. The strategy is to use synoptic measurements as the baseline and use the data to do comparative studies with ARG measurements with emphasis on near simultaneous and co-located ARG measurements to correct for any bias and improve the accuracy of the latter. Since each GPM measurement covers a relatively large area, the average of several ground measurements within the footprint of GPM are then used for comparative studies with GPM. GPM data are then normalized such that they are consistent with ARG data that is also made consistent with synoptic data. The production of the GPM, ARG and Synoptic dataset that are similarly formatted and have consistent values is expected to provide a more powerful tool for rainfall studies.

## 2. Data and Methods

### 2.1 Ground Rainfall Measurements

Rainfall data from ground stations were obtained from synoptic stations and automatic rain gauges (ARGs). Figure 1 shows the location of the synoptic stations and ARGs used in the study.

### 2.1.1. Synoptic stations

Synoptic stations provide rainfall measurements using a tipping bucket rain gauge which are manned and calibrated by weather observers. As these stations follow the standards set by World Meteorological Organization in measuring and recording meteorological parameters, rainfall data provided by PAGASA can be considered the most accurate ground rainfall measurement in the Philippines. Daily rainfall from 55 active stations during the period 2014 to 2017 were considered in the study. The daily rainfall data were then aggregated to dekad (10 day) and monthly totals. Afterwards, accumulated rainfall was expressed in mm/day.

### 2.1.2. Automatic rain gauges

Multiple projects of the Department of Science and Technology (DOST) have installed approximately 2,000 automatic rain gauges (ARGs) all over the Philippines. These ARGs were incorporated in automatic weather stations, water level monitoring systems, and standalone ARGs. Installation of these ARGs initiated as early as 2011; however, only until 2014 were there almost 1000 ARGs that are fully operational. Currently, the Advanced Science and Technology Institute (DOST-ASTI) receives and stores the data transmitted by these rain gauges. Rainfall data from ARGs during 2014 to 2017 were utilized in the study.

ARGs provide rainfall measurements every 10 or 15 minutes depending on the type of instrument installed. Similar with synoptic stations, this network is composed of tipping bucket rain gauges. Since the ARGs are not supervised, there are always the possibility of instrumental malfunction or other unforeseen sources of error. To ensure quality and optimize accuracy and reliability of the measurements, the following quality checks, as suggested by Combinido et al (2017), were implemented:

- **Geolocation check** to ensure that the reported location of the measurement is consistent with actual location established during installation;
- **Timestamp check** to ensure that the automatically logged data follow the set time intervals per rainfall measurement;
- **Range check** to remove excessively high rainfall measurements do not exceed 20 (30) mm for 10 (15) minute instantaneous sampling; and
- **Internal consistency check** to verify rainfall measurements are within expected values as can be inferred from corresponding temperature and relative humidity measurements.

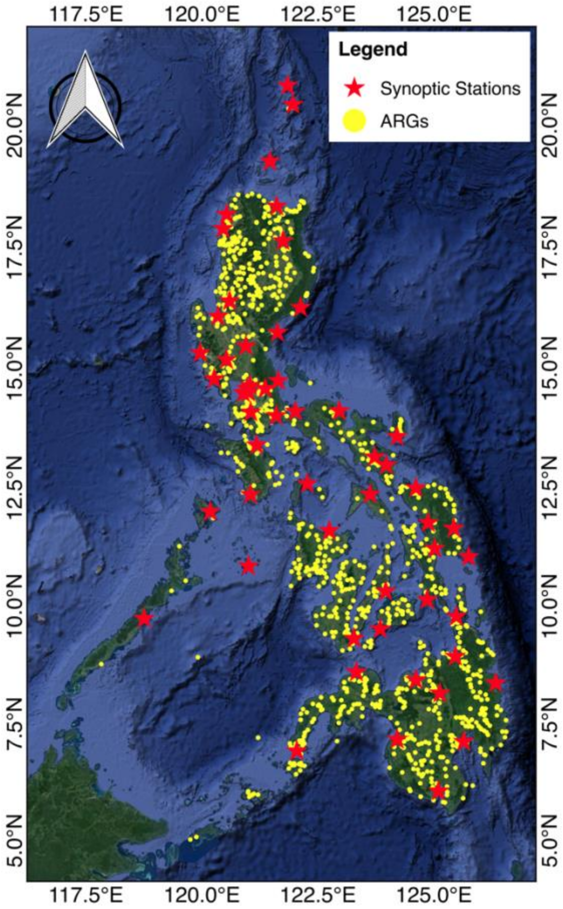

Figure 1. Location of synoptic stations and ARGs scattered all over the Philippines.

Additionally, the following checks were included in the quality assurance procedure to account for further data cleaning:

- **Climatology consistency check** to verify monthly averages are consistent within 1.5 standard deviation with monthly climatology for the same month of the nearest synoptic station; and
- **Zero measurements check** to exclude in this study ARGs that provides 75% or more zero (0 mm) or no rainfall data during the study period.

The quality assured rainfall measurements were then aggregated to daily, dekad, and monthly averages. Location and elevation details of the ARGs were utilized to perform distance-dependent comparative analyses with synoptic stations and satellite pixel.

### 2.2. Satellite Rainfall Measurement – Global Precipitation Measurement

NASA and JAXA launched the Global Precipitation Measurement Core Observatory Satellite (GPM) on February 2014 to continue the legacy of the Tropical Rainfall Measuring Mission (TRMM) that has decommissioned on July 2014. GPM aims to provide the next generation of precipitation products at higher spatial and temporal resolutions.

GPM provides rainfall product known as the Integrated Multisatellite Retrievals for GPM (IMERG). IMERG combines precipitation estimates from both passive microwave (PMW) and infrared (IR) sensors from LEO and geostationary satellites, respectively (Anjum et. al 2018; Huffman et. al 2019). IMERG precipitation are available every 30 minutes at 0.1° spatial resolution. Half-hourly precipitation estimates are composed of instantaneous PMW estimates when available and PMW estimates propagated from previous or future times using Lagrangian time interpolation (Anjum et. al 2018). IR estimates are used to supplement PM estimates. Afterwards, precipitation estimates are calibrated using monthly precipitation data from Global Precipitation Climatology Center (GPCC). IMERG is available in three stages: the early, late, and final runs. Early and late run products includes near real-time precipitation data which are released after four and 12 hours, respectively. Final run undergoes monthly rainfall calibration from GPCC and is available after 3.5 months.

In this study, IMERG late run version 6 was utilized. Half-hourly rainfall estimates from 2014 to 2017 was considered. Half-hourly rainfall were aggregated to daily, dekad, and monthly estimates. The gridded rainfall product of IMERG was then matched with overlapping ground stations based on the boundaries of each pixel.

## 2.3. Comparative Analyses

### 2.3.1. Point-to-pixel comparison of synoptic stations and GPM

Comparative analyses in this study started with the determination of relationship between the ground truth of rainfall measurements in the Philippines, the synoptic station data, and the gridded rainfall product from GPM. The comparison of rainfall measurements was done by utilizing the location of each synoptic station and matching its data with the rainfall amount from the GPM pixel enclosing the station. Only days with recorded rainfall of greater than 0.1 mm from both synoptic stations and GPM were considered in the time series comparison. Dekad and monthly rainfall were obtained from aggregated daily data of synoptic stations and GPM pixels.

Statistical metrics such as correlation coefficient (Siuki et. al 2016; Anjum et. al 2018), root mean square error (Siuki et. al 2016; Anjum et. al 2018; Peralta et. al 2020), mean absolute error (Siuki et. al 2016; Peralta et. al 2020), and bias (Siuki et. al 2016; Anjum et. al 2018) were obtained from comparing the time series data. The slope of the regression line were also used for analyzing the relationship between synoptic and GPM rainfall.

### 2.3.2. Point-to-point comparison of synoptic stations and ARGs

After undergoing quality checks, rainfall measurements from ARGs were validated using synoptic station data. Individual rain gauges were compared to the nearest synoptic station relative to its location. Shortest distance between synoptic stations and ARGs was used as the basis to determine each synoptic and ARG pair. Distance between an ARG and a synoptic station was obtained using the distance formula (Beck et. al 2017) derived from Pythagorean theorem. Statistical metrics mentioned above were calculated to determine the effect of distance to agreement in measurements of ARGs and synoptic stations.

Relatively close ARGs and synoptic stations were also compared to account for systematic errors and biases obtained from unmanned rain gauges. In this case, only ARGs and synoptic stations with a maximum distance of 1 km were considered. Only ARG-synoptic pairs with homogenous locations and an elevation difference of less than 50 m were used in this analysis.

Systematic bias in rainfall measurements from ARGs were then corrected using the generalized reduced gradient algorithm (Fylstra et. al 1998; Gumindoga et. al 2016). In this case, the algorithm was applied using a power transform (Gumindoga et. al 2016)

$$y = ax^b \tag{1}$$

where $y$ represents the corrected rainfall from ARGs, $x$ represents the actual ARG rainfall data, and $a$ and $b$ are coefficients optimized to obtain minimum residuals between synoptic and ARG rainfall measurements. The bias corrected rainfall from ARGs were again compared with the synoptic data using the same statistical metrics.

### 2.3.3. Comparison of GPM and multiple ground measurements within GPM footprint

ARG data made consistent with synoptic measurements were now utilized for comparison with GPM. Both synoptic stations and ARGs were counted as point ground measurements in this analysis.

Because daily measurements are more dynamic and prone to errors as will be shown in the results, only dekad and monthly accumulated rainfall will be considered for the comparative analysis with GPM. A GPM footprint was defined as the extent of an individual GPM pixel with a buffer of 0.05° on all sides. Ground station measurements within the GPM footprint were averaged and then compared with the GPM pixel value. As GPM represents the average rainfall within the grid, only pixels which has a minimum of 10 ground stations within its footprint were considered in this analysis.

After performing quality checks on ARGs, not all dekads and months were comprised of measurements from all present ARGs in the GPM footprint. Some dekadal and monthly rainfall were averaged from a fewer stations less than their actual number within the footprint. Thus, a minimum of five ground stations was set as a threshold for comparison with GPM. Moreover, since the analysis only involves GPM pixels with ample number of ARGs within its footprint, a cumulative distribution function (CDF) plot was obtained in order to determine the contribution of each month in the whole time series. This is to verify that the seasonality of rainfall in the Philippines is well represented in the analysis. A CDF plot showing the number of ARGs averaged within the GPM footprint was also made.

Further bias correction of GPM based on comparative analysis with average ground measurements was done using the generalized reduced gradient algorithm applied on a power transform similar to Eq. 1. Statistical metrics were also calculated to determine changes in the gridded product.

## 3. Results and Discussion

### 3.1. Synoptic stations vs. GPM

Synoptic stations provide the most accurate rainfall measurements in the Philippines following the standards set by WMO. However, these stations are few and sparsely distributed in the country. Synoptic stations are maintained and operated to monitor mainly synoptic scale systems such as tropical cyclones. Thus, these stations are spaced strategically to capture weather and climatic events which can cover the whole country. Rainfall measurements from synoptic stations were compared with gridded rainfall of GPM at the daily, dekad, and monthly scales to validate performance of the satellite product.

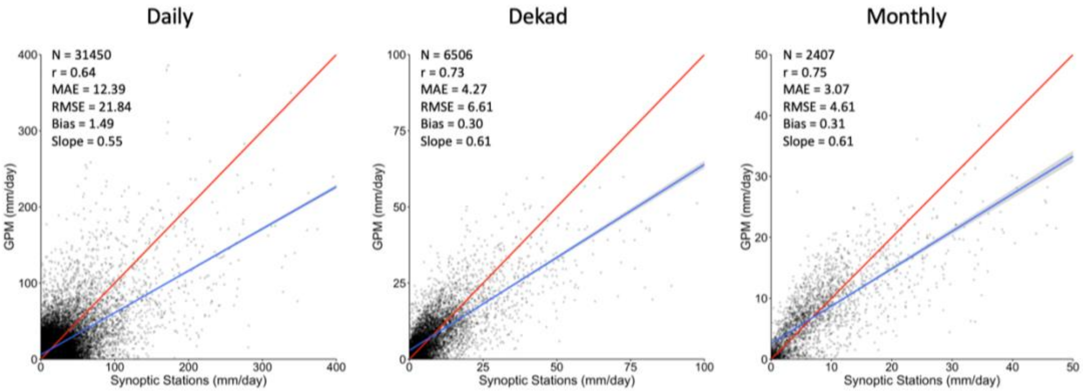

Figure 2. Synoptic rainfall measurements compared with GPM at the daily (left), dekad (middle), and monthly (right) accumulation times. Red line depicts 1:1 rainfall values. Blue line indicates linear regression with uncertainties.

Comparing synoptic measurements with GPM, it was observed that the daily rainfall is very dynamic with widespread measurements and mismatches at higher rainfall amounts as shown in Figure 2. Dekad and monthly measurements have better agreement with linear regressions closer to the 1:1 line. Higher correlation coefficients were also obtained at the dekad and monthly scale but the widespread scatter plots still depict huge errors which are also represented by MAE and RMSE values.

Because synoptic stations best represent rainfall at the synoptic scale, agreements with GPM improved with increasing accumulation period. In this case, a point measurement of rainfall from a synoptic station becomes more representative of a wider area as its rainfall is accumulated longer. On the other hand, GPM depicts the average rainfall estimate over its grid. At lower time scales such as daily and dekad, rainfall from synoptic stations corresponds to the rainfall received in the immediate vicinity of the station while GPM still represents the average rain within a 0.1° x 0.1° grid. With this, ample number of ground stations within the GPM footprint is needed to appropriately compare ground measurements with the gridded product.

## 3.2. ARGs vs. Synoptic stations

The dense network of rain gauges can provide ample number of ground measurements within the GPM footprint for a more appropriate comparison of rainfall. However, these unmanned measurements must be validated even after performing quality checks. The quality assured rain gauge data were then compared with synoptic measurements at the dekad and monthly scale.

Rainfall from each ARGs were compared to the nearest synoptic station data. Analysis was done in consideration of the distance between the ARG and synoptic station. Dekad and monthly rainfall were plotted depending on the distance value as shown in Figure 3. The scatter plots depict a distance-dependent agreement of rainfall measurements between ARGs and synoptic stations. Closer ARG-synoptic station pairs have rainfall lying near the 1:1 line. It can also be observed that even at distances around 5 km, the spread of measurements is still wide showing mismatch in rainfall values. These results suggests just how much synoptic stations can represent rainfall even while considering dekad and monthly time scales. In this manner, synoptic station data can only represent rainfall close to where it’s located. Moreover, the use of ARGs for validating gridded rainfall products were supported as these unmanned stations can provide a more dense data which can be averaged over a certain footprint.

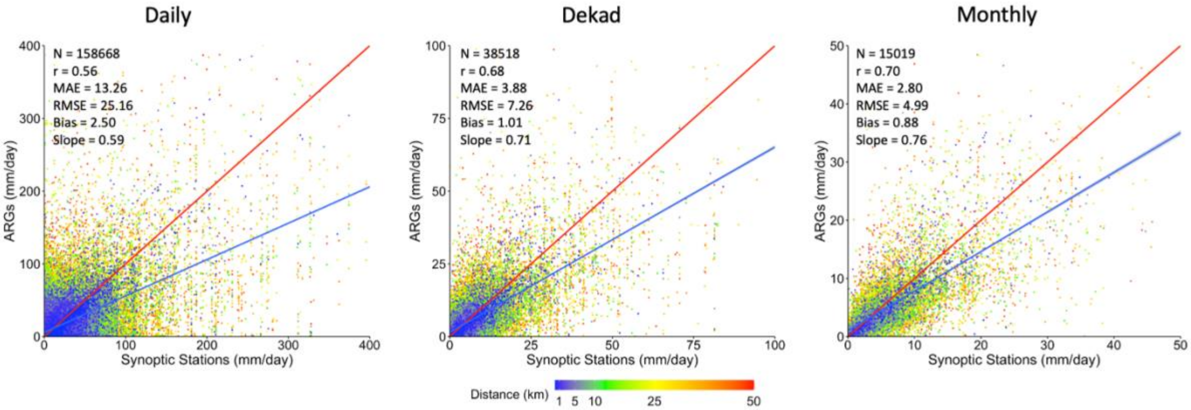

Figure 3. ARG rainfall measurements compared with synoptic stations at the daily (left), dekad (middle) and monthly (right) accumulation times. Colors represent the distance between each ARG and closest synoptic station.

While agreements can be observed in Figure 3 for close ARGs and synoptic stations, further analysis was done using ARG-synoptic station pairs with distances less than a kilometer. For this analysis, five pairs were considered having relatively homogenous surfaces and elevation difference of less than 50 m. The ARG-synoptic station pairs were summarized in Table 1. Because of the dynamic nature of daily rainfall, comparison at this scale was not performed.

Table 1. Relatively close ARG-synoptic station pairs.

| **Synoptic Station** | **Synoptic Station Elevation (m)** | **ARG ID** | **ARG Elevation (m)** | **Distance (m)** |
|---|---|---|---|---|
| Catbalogan, Samar | 5 | 84 | 11 | 940 |
| General Santos | 132 | 186 | 140.78 | 30 |
| Iba, Zambales | 5.5 | 160 | 0 | 90 |
| Science Garden | 43 | 190 | 0 | 640 |
| Tacloban | 2.7 | 81 | 5 | 20 |

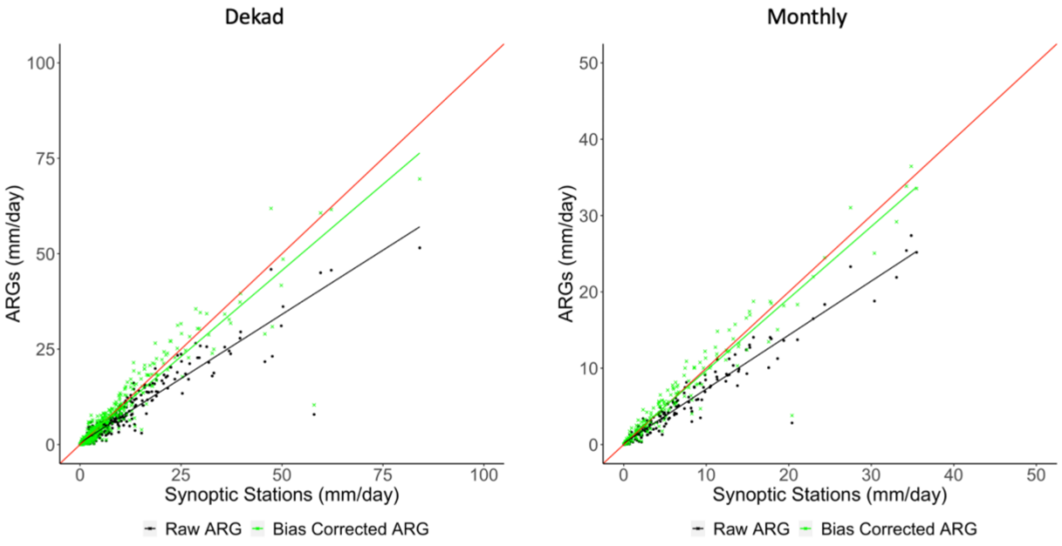

Figure 4. ARG and synoptic station rainfall at distances less than 1 km. Dekad (left) and monthly (right) rainfall represented by black points are raw ARG data while green points are corrected data.

Rainfall from select ARG and synoptic stations at close distances with homogenous locations are expected to have comparable measurements. While high positive correlations were obtained for both dekad (r = 0.94) and monthly (r = 0.97) comparisons, Figure 4 shows that ARGs tend to have less recorded rainfall amount most of time. This is represented by black points in the scatter plot with slope less than one (black regression line). This is also apparent with the rainfall biases of 1.93 mm/day and 1.69 mm/day for both dekad and monthly comparisons, respectively.

Generalized reduce gradient algorithm using a power transform was applied to the ARG measurements in order to reduce residuals of rainfall values when compared to the established synoptic measurements. Optimized values of $a$ and $b$ were obtained, providing minimum residuals in the comparison and were then used as bias correction factors for ARG rainfall. Optimized values of $a = 1.2815$ and $b = 1.0132$ obtained for dekad rainfall while $\mathrm{a} = 1.3494$ and $\mathrm{b} = 0.9959$ for monthly rainfall. Figure 4 shows the scatter plots of rainfall measurements from ARGs before and after bias correction as compared to the synoptic station data. Improvements in the slope of the regression line can be observed with errors and biases in both dekad and monthly rainfall reduced as summarized in Table 2.

Table 2. Statistics before and after bias correction of ARG rainfall data.

| **Parameter** | **Dekad** | | **Monthly** | |
|---|---|---|---|---|
| | **Raw ARG** | **Corrected ARG** | **Raw ARG** | **Corrected ARG** |
| N | 465 | 465 | 177 | 177 |
| r | 0.94 | 0.94 | 0.97 | 0.97 |
| MAE (mm/day) | 2.09 | 1.58 | 1.75 | 1.08 |
| RMSE (mm/day) | 4.54 | 3.39 | 3.03 | 1.88 |
| Bias (mm/day) | 1.93 | 0.19 | 1.69 | 0.09 |
| Slope | 0.67 | 0.90 | 0.70 | 0.94 |

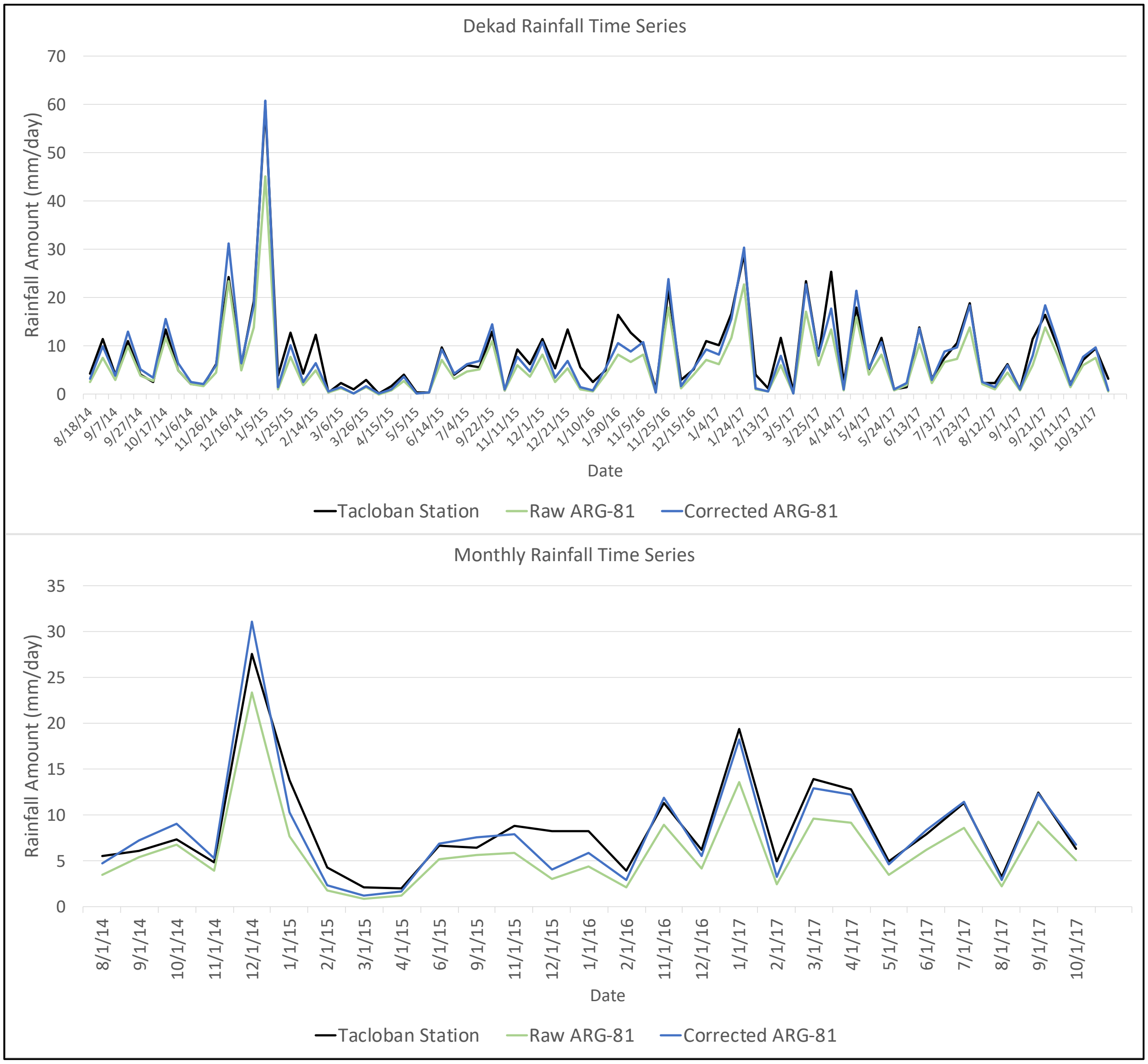

Figure 5. Before and after bias correction of ARG rainfall at Tacloban Synoptic station.

Figure 5 shows the changes made in one of the ARGs close to Tacloban synoptic station. Performing bias correction on both dekad and monthly rainfall from the ARG resulted to time series measurements closer to the synoptic station data depicted by blue lines (bias corrected ARG) closer to black lines (synoptic station).

## 3.3. Average ground measurements vs. GPM

The optimized correction factors were then applied to ARGs for comparison with the GPM gridded rainfall product. Initially, comparative analysis was done for all GPM pixels with ground stations within its footprint. As shown in Figure 6, number of ground stations affects the agreement with rainfall estimates of GPM. As the number of ground stations averaged within the GPM footprint increases, rainfall values tend to lie closer to the 1:1 line. This result is expected because GPM represents average rainfall within its pixel boundaries. Ground stations, on the other hand, only depicts rainfall immediate to its vicinity as with the case of synoptic stations.

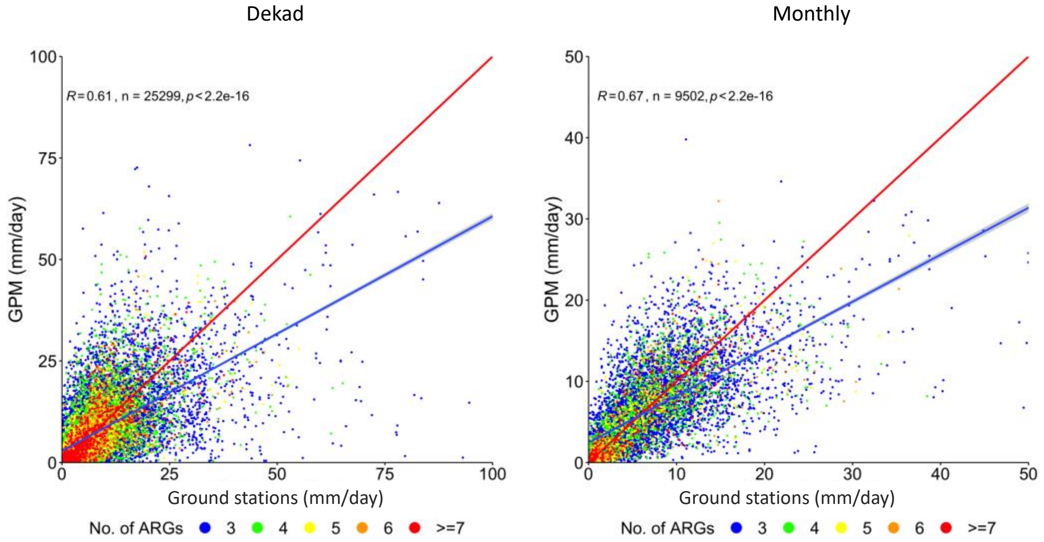

Figure 6. Average ground stations within GPM footprint compared with actual pixel rainfall value. Red line depicts 1:1 rainfall values. Blue line indicates linear regression with uncertainties.

Considering the effect of number of ground stations to the agreement of ground-based rainfall and GPM, analysis proceeded with comparing only GPM pixels with ample number of ARGs and synoptic stations within its footprint. A buffer of 0.5° outside the GPM pixel were applied to nearby ground stations as their measurements still has direct effect to the average rainfall within the pixel. Only GPM pixels with 10 ground stations or more were considered in further analysis. While not all stations can provide rainfall data all the time, a minimum of 5 stations (out of 10 or more stations) must have dekad and monthly measurements to be considered in time series comparison.

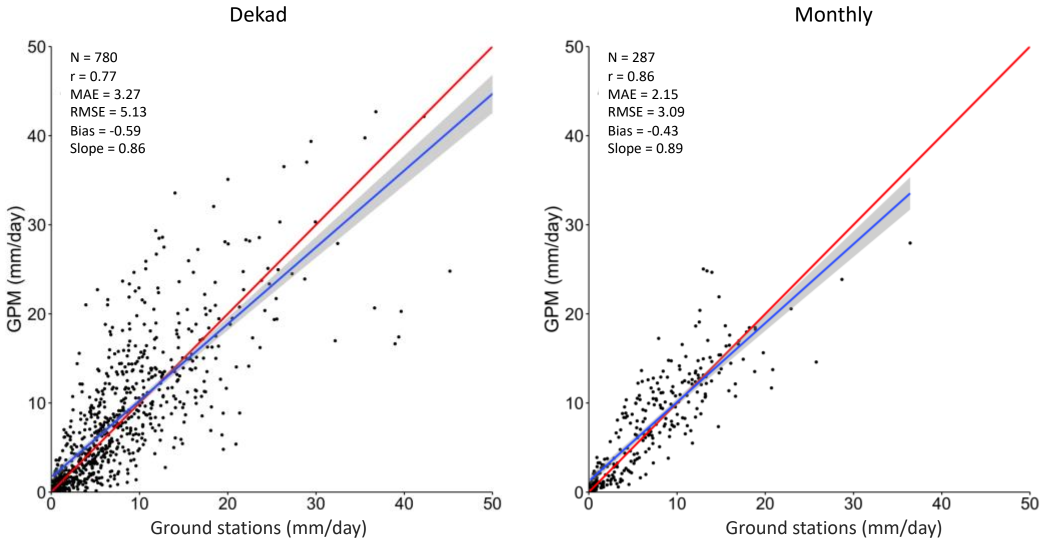

Figure 7. Comparison of GPM with ample number of ground stations within its footprint. Red line depicts 1:1 rainfall values. Blue line indicates linear regression with uncertainties.

Shown in Figure 7, considering only pixels with minimum of 10 ground stations within its footprint provides correlation coefficients of 0.77 and 0.86 for dekad and monthly rainfall comparisons, respectively. Rainfall biases are -0.59 mm/day and -0.43 mm/day for dekad and monthly rainfall. Negative biases depicts that average residuals points to higher GPM rainfall estimates. However, the slope values of less than one indicates a linear trend where ground measurements are greater than GPM rainfall estimates. Looking at the scatter plots, GPM has the tendency to over/underestimate ground rainfall data. To further investigate the over/underestimation of rainfall from GPM, seasonal analysis on these data were done.

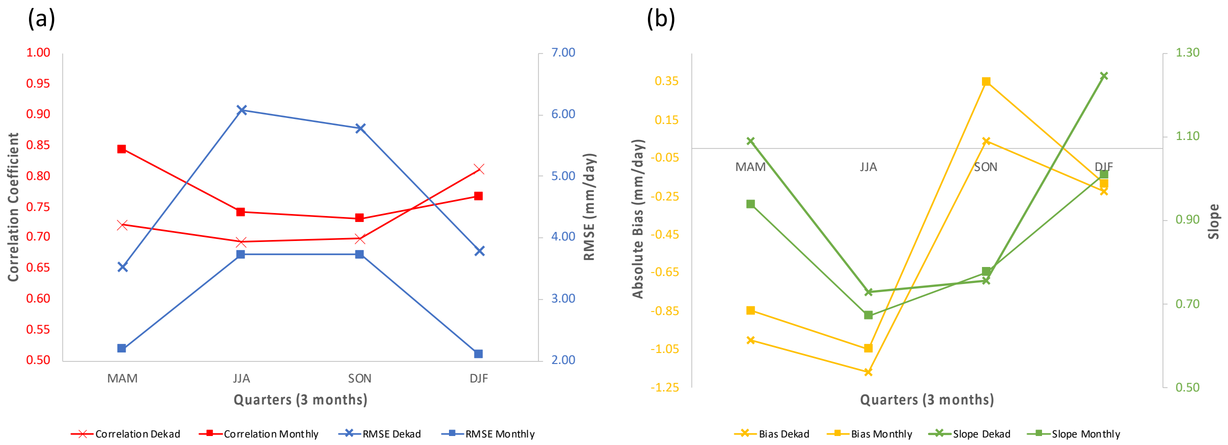

Figure 8. Seasonal statistics of comparing GPM with ground-based rainfall measurements.

Comparison of GPM with ground rainfall measurements were divided into MAM, JJA, SON, and DJF months. Shown in Figure 8, peak RMSE was obtained during JJA months followed by SON for both dekad and monthly rainfall. These months are highly affected by *Habagat* and tropical cyclones which bring a huge amount of rain in the Philippines. Lowest correlations were also obtained during JJA and SON which can be attributed to more scattered points and greater MAE and RMSE values for this months as shown in Figure 9. During JJA months, the minimum rainfall amount on rainy dekads is about 3.0 mm/day. RMSE and biases are lowest during DJF and MAM months which are mainly affected by the dry Winter monsoon (*Amihan*). During these months, rainfall mainly come from isolated micro and mesoscale convective systems rather than synoptic scale processes such as tropical cyclones. MAM and DJF months provide the majority of light to moderate rains while JJA and SON months contribute mainly on the heavy and extreme rains.

Considering the seasonal distribution of rainfall and associated errors and biases in measurements shown in Figure 8 and Figure 9, the slope of the regression lines depicted in Figure 7 might be coincidental. Combining all months in a singular scatter plot neutralizes the slopes and biases of MAM and DJF with that of JJA and SON, creating an overall slope close to one. In this case, further bias correction of GPM cannot be based on the linear regression as it will further increase the errors with respect to ground measurements. With this, bias correction of GPM proceeded with the generalized reduced gradient algorithm using a power transform. This is the same methodology used to bias correct ARGs using relatively close synoptic stations.

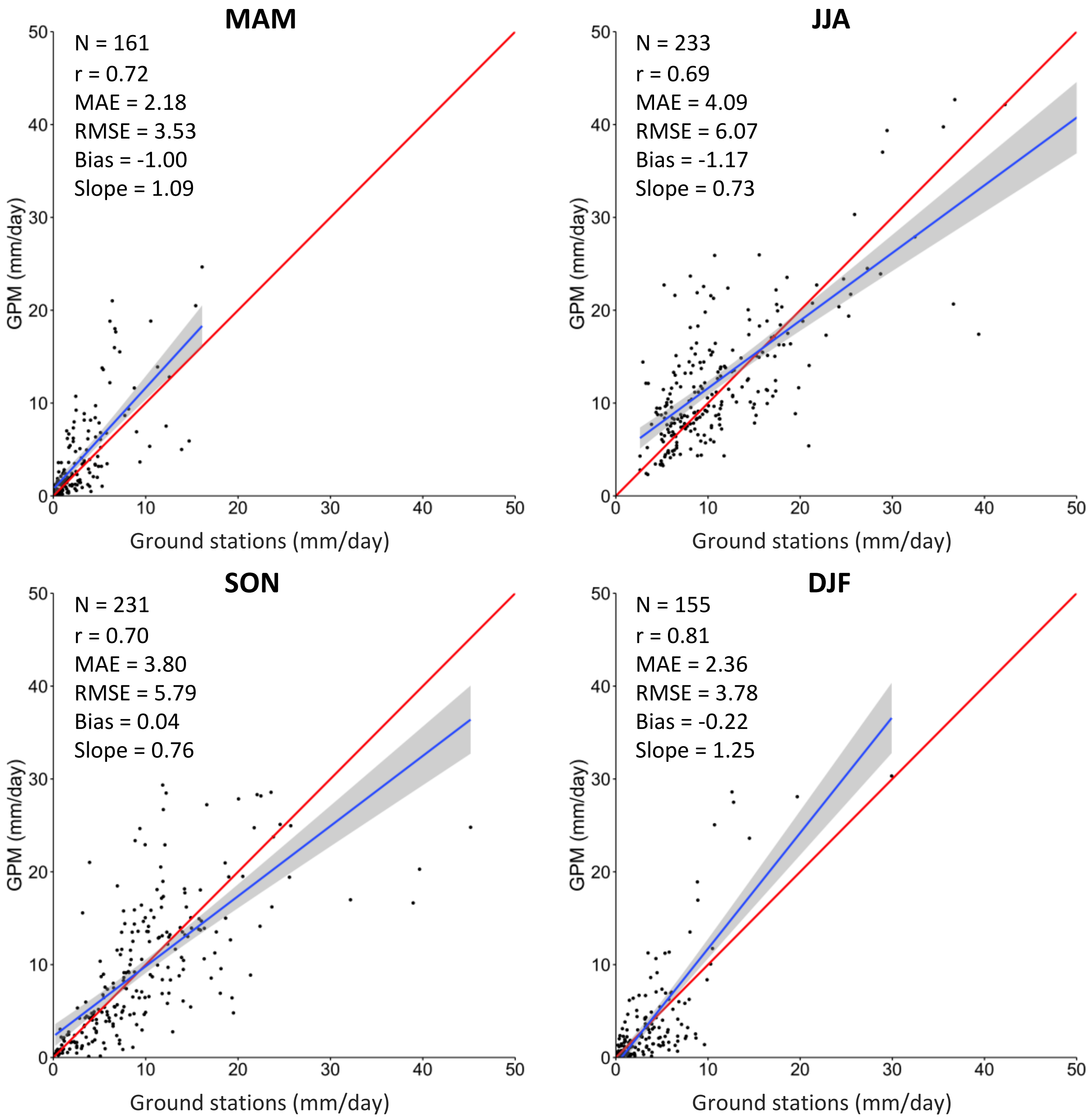

Figure 9. Scatter diagrams of dekad rainfall from ground stations and GPM during MAM, JJA, SON, and DJF months. Red line depicts 1:1 rainfall values. Blue line indicates linear regression with uncertainties.

Generalized reduced gradient algorithm was performed on GPM rainfall with respect to ground station data. The correction factors were optimized as $a = 1.8057$ and $b = 0.7257$ for dekad rainfall while $\mathrm{a} = 1.2743$ and $\mathrm{b} = 0.8669$ for monthly rainfall. Bias corrected GPM was obtained using the equation $y = ax^b$. Results of bias correction is shown in Figure 10. Based from the figure and statistics when compared to ground measurements, bias corrected GPM improved on the correlation coefficient, MAE, RMSE, and bias values for both dekad and monthly rainfall. The slope of the line decreased; however, since the slope of raw GPM is only close to one coincidentally, improving the errors and biases in the comparative analysis is more desirable.

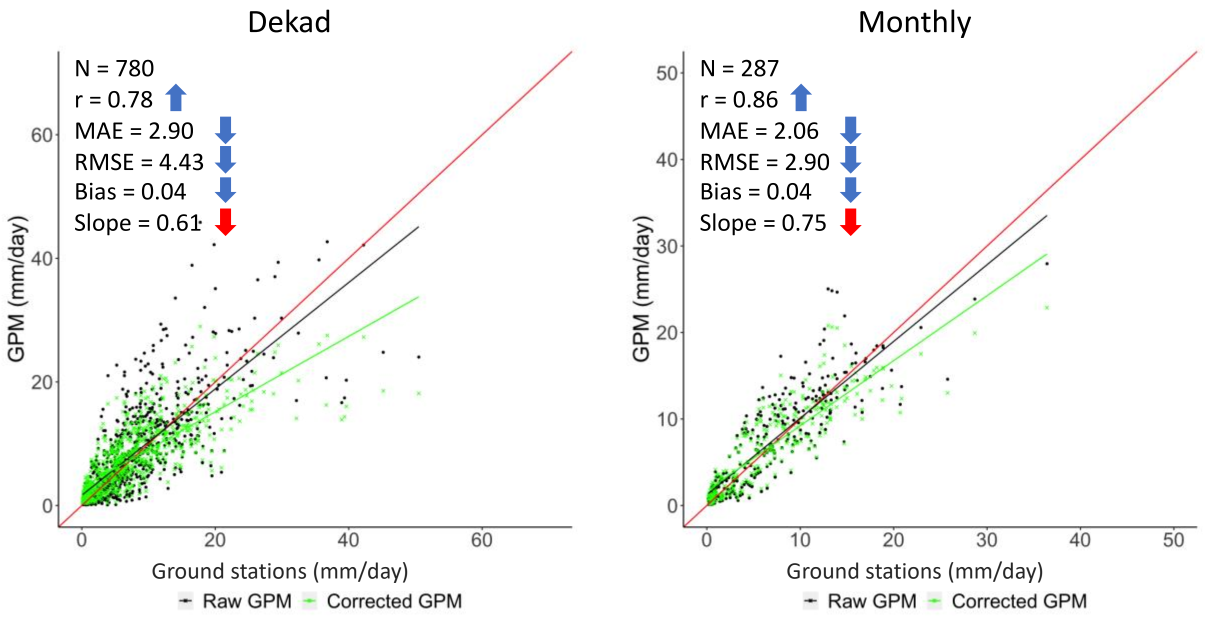

Figure 10. Bias corrected GPM performance compared to ground station rainfall.

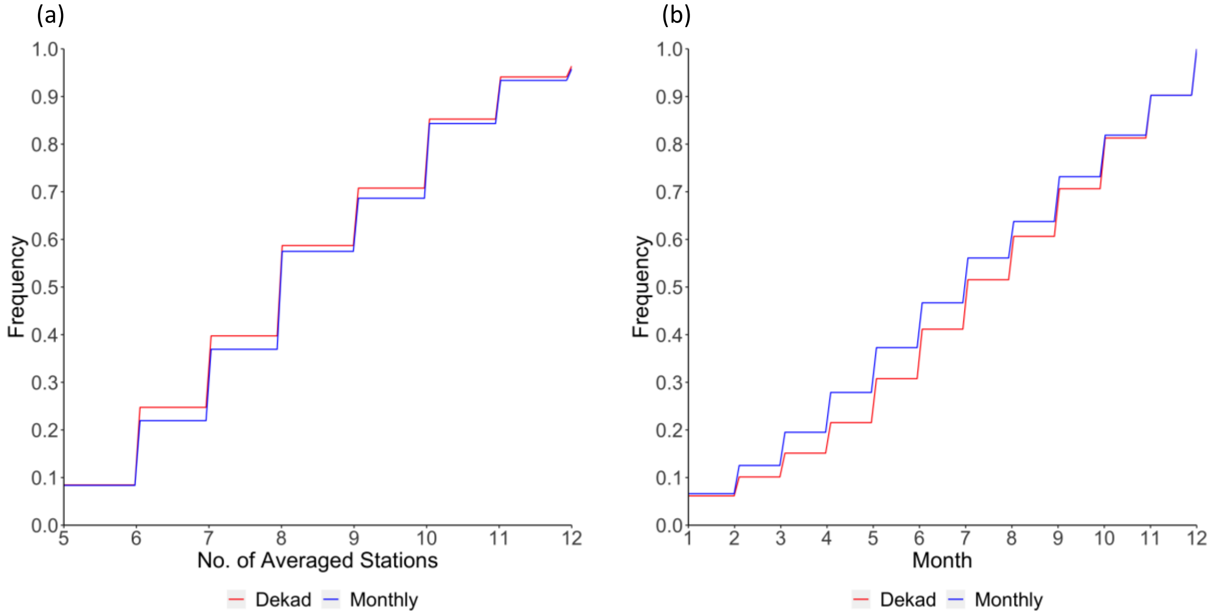

Figure 11. Distribution of data points utilized for analyzing relationship of averaged ground stations with GPM.

To verify representativeness of comparative analysis made, CDF of data points utilized were obtained as shown in Figure 11. The CDF shows that the results were generated using a majority of 8 ground stations averaged within the GPM footprint. Lastly, the results also represent an almost homogenous distribution of rainfall data among different months of the year.

## 4. Summary and Conclusion

As an archipelago in the tropics, Philippines receives huge amount of rain that varies depending on the location and season (Coronas 1920). Majority of agriculture and forests in the country rely heavily on rainfall; hence reliable rainfall information is crucial in the Philippines. Synoptic stations provide the best data of rainfall in the Philippines because these are measurements from manned and well-maintained stations. While serving its purpose of monitoring synoptic systems, representation of rainfall at a more localized scale is needed for studies such as weather and agrometeorological research. A dense network of automatic rain gauges can provide the rainfall information at a more localized perspective all over the country. However, reliability of these data is subject to quality checks. Satellite products such as GPM can provide more robust rainfall data with consistent spatial and temporal information. Satellite estimates, however, should also be validated in order to maximize the benefits of having a gridded rainfall data. Thus, this study was conducted to make these datasets consistent with each other and come up with a product that can better represent rainfall in the Philippines.

Comparative analysis of synoptic station rainfall with overlapping GPM pixel value was done. Correlation coefficient of 0.64, 0.73, and 0.75 were obtained for daily, dekad, and monthly rainfall comparisons. While these correlations are good, the errors and widespread mismatches between the measurements suggest that point data from synoptic stations may not represent average rainfall within a 0.1° x 0.1° grid especially at the daily scale which is affected mainly by localized systems.

Distance-based comparison of synoptic station data and ARG showed that synoptic rainfall, even while accumulated in dekads and months, is consistent only at its immediate vicinity. Good agreement between these ground measurements can only be observed at close distances less than 1 km. Looking at the distribution of data through scatter diagrams, ARG reports less rainfall amount compared to synoptic stations. Correction on these data was done provided that the compared measurements are from stations with homogenous locations. Generalized reduced gradient algorithm was performed on the ARG data to reduce bias in measurements, also improving the errors and slope of the regression line as well. Consistency between synoptic rainfall and ARG was observed on dekad and monthly time series after the bias correction.

With ample number of ground stations within the GPM footprint, comparison between average ground measurements and GPM rainfall estimates show promising correlation and bias both at the dekad and monthly times. The slope of the regression line is close to one; however, the error statistics and obvious mismatches in the scatter diagrams show that the slope might be coincidental. Seasonal comparison of rainfall provides a better picture of the distribution of measurements. Peak errors were obtained during JJA and SON months affected mainly by the Summer monsoon and tropical cyclones which provide moderate to heavy rain. DJF and MAM provide lowest errors because light to moderate rain predominates during these months. Combining the highs and lows of errors in one scatter diagram neutralizes the regression line. In this case, further bias correction of GPM cannot be based on linear regression. Generalized reduced gradient algorithm using a power transform was applied to GPM to minimize biases down to 0.04 mm/day for both dekad and monthly rainfall. Correlation coefficients rose to 0.78 and 0.86 for dekad and monthly comparison, respectively. Lastly, MAE and RMSE were also reduced through the bias correction.

This study has provided a way to make synoptic stations, ARGs, and GPM rainfall measurements consistent with each other. A gridded dataset at the dekad and monthly scale can provide rainfall information to various research applied to agrometeorology and mesoscale weather systems. Further, bias correction may be applied to historical gridded product of GPM for long-term satellite product consistent with ground station measurements.

**References**

Aghakouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite - retrieved extreme precipitation rates across the central United States. 116(September 2010), 1–11. https://doi.org/10.1029/2010JD014741

Anjum, M.N., Ding, Y., Shangguan, D., Ahmad, I., Ijaz, M.W., Farid, H.U., Yagoub, Y.E., Zaman, M., & Adnan, M. (2018). Performance evaluation of latest integrated multi-satellite retrievals for T Global Precipitation Measurement (IMERG) over the northern highlands of Pakistan. Atmospheric Research, 205, 134-146. https://doi.org/10.1016/j.atmosres.2018.02.010

Arulraj, M., & Barros, A.P. (2017). Shallow precipitation detection and classification using multifrequency radar observations and model simulations. Journal of Atmospheric and Oceanic Technology, 34, 1963-1983. https://doi.org/10.1175/JTECH-D-17-0060.1

Beck, H., van Dijk, A., Levizzani, V., Schellekens, J., Miralles, D., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. Hydrology and Earth System Sciences, 21, 589-615. https://doi.org/10.5194/hess-21-589-2017

Chen, F. & Liu, C. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy and Water Environment, 10, 209–222. https://doi.org/10.1007/s10333-012-0319-1

Combinido, J. S., De Paz, M., & Carlos, J. (2018). Automated quality control of ASTI automatic weather station (AWS) meteorological measurements: Quality control algorithm Version 1.0.

Coronas, J. (1920). The climate and weather of the Philippines, 1903 to 1918. *Bureau of printing*, Manila, p 195.

Dembélé, M., & Zwart, S. J. (2016). Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. International Journal of Remote Sensing, 37(17), 3995–4014. https://doi.org/10.1080/01431161.2016.1207258

Fensterseifer, C., Allasia, D. G., & Paz, A. R. (2016). ASSESSMENT OF THE TRMM 3B42 PRECIPITATION PRODUCT IN SOUTHERN BRAZIL 1. 52(2), 367–375. https://doi.org/10.1111/1752-1688.12398

Fylstra, D., Lasdon, L., Watson, J., & Waren, A. (1998). Design and Use of the Microsoft Excel Solver. Interfaces, 28, 29-55. https://doi.org/10.1287/inte.28.5.29

Gumindoga, W., Rientjes, T.H.M., Haile, A.T., Makurira, H., & Regianni, P. (2016). Bias correction schemes for CMORPH satellite rainfall estimates in the Zambezi River Basin. Hydrology and Earth System Sciences Discussions [preprint]. https://doi.org/10.5194/hess-2016-33

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., … Wolff, D. B. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. Journal of Hydrometeorology, 8(1), 38–55. https://doi.org/10.1175/JHM560.1

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J., & Xie, P. (2019). Algorithm Theoretical Basis Document (ATBD) Version 06: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG); NASA/GSFC: Greenbelt, MD, USA.

Jamandre, C. A., & Narisma, G. T. (2013). Spatio-temporal validation of satellite-based rainfall estimates in the Philippines. Atmospheric Research, 122, 599–608. https://doi.org/10.1016/j.atmosres.2012.06.024

Khodadoust Siuki, S., Saghafian, B., & Moazami, S. (2017). Comprehensive evaluation of 3-hourly TRMM and half-hourly GPM-IMERG satellite precipitation products. International Journal of Remote Sensing, 38(2), 558–571. https://doi.org/10.1080/01431161.2016.1268735

Lin, Y. (2007). Mesoscale Dynamics. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511619649

Mishra, A. K. (2013). Effect of rain gauge density over the accuracy of rainfall: A case study over Bangalore, India. SpringerPlus, 2(1), 1–7. https://doi.org/10.1186/2193-1801-2-311

Moron, V., Lucero, A., Hilario, F., Lyon, B., Robertson, A. W., & DeWitt, D. (2009). Spatio-temporal variability and predictability of summer monsoon onset over the Philippines. *Climate Dynamics*, *33*(7–8), 1159–1177. https://doi.org/10.1007/s00382-008-0520-5

Peralta, J. C. A., Narisma, G. T., & Cruz, F. A. (2020). Validation of High-Resolution Gridded Rainfall Datasets for Climate Applications in the Philippines. Journal of Hydrometeorology, 21, 1571–1587. https://doi.org/10.1175/JHM-D-19-0276.1

Siuki, S., Saghafian, B., & Moazami, S. (2016). Comprehensive evaluation of 3-hourly TRMM and half-hourly GPM-IMERG satellite precipitation products. International Journal of Remote Sensing, 38:2, 558-571. https://doi.org/10.1080/01431161.2016.1268735

Tan, M. L., & Duan, Z. (2017). Assessment of GPM and TRMM precipitation products over Singapore. Remote Sensing, 9(7), 1–16. https://doi.org/10.3390/rs9070720

Verdin, J., & Klaver, R. (2002). Grid-cell-based crop water accounting for the famine early warning system. Hydrological Processes, 16, 1617-1630. https://doi.org/10.1002/hyp.1025

Vila, D., De Goncalves, L. G., Toll, D., & Rozante, J. R. (2009). Statistical Evaluation of Combined Daily Gauge Observations and Rainfall Satellite Estimates over Continental South America. Journal of Hydrometeorology, 10, 533–543. https://doi.org/10.1175/2008JHM1048.1

World Meteorological Organization. (2018). Guidelines on the definition and monitoring of extreme weather and climate events. WMO Tech Rep., 43. https://www.wmo.int/pages/prog/wcp/ccl/documents/GUIDELINESONTHEDEFINTIONANDMONITORINGOFEXTREMEWEATHERANDCLIM ATEEVENTS_09032018.pdf