

# Clustering of Precipitation Pattern in Indonesia Using TRMM Satellite Data

Heri Kuswanto

Department of Statistics,  
Institut Teknologi Sepuluh Nopember,  
Surabaya, Indonesia  
heri\_k@statistika.its.ac.id

Dedi Setiawan

Department of Statistics,  
Institut Teknologi Sepuluh Nopember,  
Surabaya, Indonesia  
dedi1statistika@gmail.com

Ardhasena Sopaheluwakan

Meteorology, Climatology, and  
Geophysical Agency (BMKG),  
Jakarta, Indonesia  
ardhasena@bmkgo.go.id

**Abstract**—This paper identifies the climatic regions in Indonesia based on the rainfall pattern similarity using TRMM data. Indonesia is a tropical climate region with three main climate clusters, i.e. monsoonal, anti-monsoonal and semi-monsoonal. The clusters were formed by examining rainfall observation datasets recorded at a number of stations over Indonesia with coarse spatial resolution. Clustering based on higher resolution datasets is needed to characterize the rainfall pattern over remote areas with no stations. TRMM provides a high resolution gridded dataset. A statistical test has been applied to evaluate the significance of TRMM bias, and it indicated that the TRMM based satellite precipitation product is a reasonable choice to be used as an input to cluster regions in Indonesia based on the similarity of rainfall patterns. The clustering by Euclidean distance revealed that Indonesia can be grouped into three significantly different rainfall patterns. Compared to the existing references, there have been regions where the rainfall pattern has been shifted. The results in this research thus update the previously defined climate regions in Indonesia.

**Keywords**—cluster; monsoon; TRMM; remote sensing; precipitation

## I. INTRODUCTION

Indonesia is an archipelago located between  $10^{\circ}\text{S}$  to  $6^{\circ}\text{N}$ , and  $95^{\circ}\text{E}$  to  $141^{\circ}\text{E}$ , and hence, part of its regions lies on the equatorial line with a high degree of rainfall variability. The Indonesian climate conditions are influenced by several factors such as Asia-Australia Monsoon, El Nino, La Nina, East-West and North-South circulation and other local influences [1]. Moreover, the latitude and longitude position, topography, the ocean, and the land also influence the climate variability. Authors in [2] found that Indonesia can be clustered into three rainfall patterns (hereafter referred to as climate regions), i.e. monsoon, anti-monsoon and semi-monsoon types. The semi-monsoonal type has two monthly maximum rainfalls in a year and is called bimodal. The monsoon type is influenced by the big scale of ocean wind and this pattern clearly shows the difference between the dry and wet season in a year, and has only one maximum rainfall in a year. The anti-monsoonal pattern is shown by the unimodal rainfall pattern, opposite to the monsoonal rainfall type. Due to the high degree of rainfall variability, the analysis of precipitation conditions in Indonesia requires long historical series with adequate spatial coverage.

Rainfall observation in Indonesia is managed by the Meteorology, Climatology, and Geophysical Agency (BMKG) through 116 meteorological stations spreaded out all over Indonesia. Moreover, the BMKG also uses satellite data to support the need for fast and continuous rainfall data. One of the obvious advantages of satellite data is their spatial coverage, while they are normally available in a high-resolution grid covering all areas of the regions. One of the satellite products that is commonly used by the BMKG as reference is the Tropical Rainfall Measuring Mission (TRMM). TRMM is a collaborative space project between NASA and the Japan Aerospace Exploration Agency (JAXA) to monitor the precipitation over tropical regions as an effort to study earth as a global system and it was developed in 1997 [4]. The TRMM satellite measures rainfall intensity in every region on various temporal scales. The spatial resolution varies from 0.25 to 5 degrees. The mission of this satellite is to better understand the precipitation structure in the tropical region of the Earth [5]. Among the studies which note the strong performance of TRMM in specific regions compared to other satellite products are [6, 7] for the case of China and Iran respectively.

The BMKG uses TRMM satellite data to observe the rainfall conditions over Indonesia due to the limited number of meteorological stations. The TRMM data has also been widely used in weather and climate studies in Indonesia, e.g. to study extreme weather events [8]. Authors in [9] studied the advantages of using TRMM based precipitation to investigate the spatiotemporal pattern and rainfall characteristics over Indonesia. Author in [10] compared the monthly rainfall observed from meteorological stations with TRMM and NOAA surface models using simple mean testing and found that TRMM has a potential to impute the missing spatiotemporal data in some areas with no data. Moreover, authors in [11] verified the TRMM satellite precipitation data and rainfall observation over Makassar Indonesia using a simple t-test and found that the monthly TRMM precipitation can be used to estimate the rainfall in some areas without meteorological stations. TRMM satellite data have also been verified in many other countries, e.g. authors in [12] used TRMM to estimate the rainfall in Topajo River in Amazon using bias correction, while authors in [13] verified TRMM 3B42 over China. TRMM applications have also been

Corresponding author: Heri Kuswanto

intensively investigated in Singapore [14, 15] and Malaysia [16, 17].

The purpose of this study is to use TRMM data to update the clustering of climatic regions in Indonesia based on precipitation pattern. Moreover, the TRMM data will be verified towards observation data prior to clustering process in order to justify that TRMM is a reasonable choice to be used as a basis of clustering. This work differs from [2] in several significant ways. First, we use the TRMM satellite data as the basis of clustering, instead of using the station observation data. Therefore, the clustering result is more reliable to explain the climate variability among regions due to the high resolution data used in the clustering process. Moreover, the TRMM data provide rainfall information over regions without stations. Also, this research uses a different clustering process in terms of the distance measure, i.e. standard (Euclidean) distance measure, while authors in [2] used double correlation (DC) which treats the data as a cross-sectional series. Another important difference is the period of the examined dataset: this study examines the average monthly rainfall over the last 18 years, while [2] used data spanning from 1961 to 1993.

## II. STUDY AREA AND DATA

Indonesia is an archipelago with thousands of islands stretching along the Equator from the Southeast Asia to Australia. It is a tropical country where the temperature is stable throughout the year, with lows around 22°C to 25°C, and highs around 30°C to 32°C all over the year. Moreover, the rainfall quantity and distribution vary due to the location humidity level and the monsoon regime. The availability of historical rainfall data recorded at meteorological stations is an important aspect to further study the rainfall variability. Figure 1 depicts the location of the 116 meteorological stations over Indonesia.

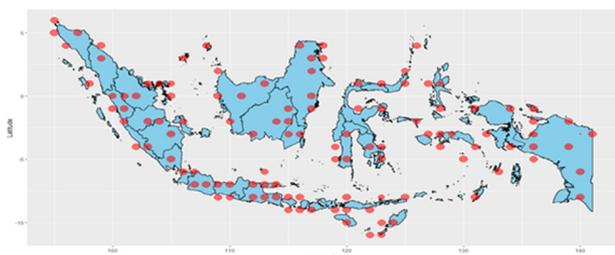


Fig. 1. Locations of the meteorological stations in Indonesia

We can see that there are some areas without meteorological stations making difficult to characterize the climatic pattern. Using high resolution gridded data from the TRMM will allow the investigation of the climate pattern in a specific region precisely. The climate in Indonesia, according to [2] is classified into three different regions. Region 1 is the southern monsoonal region covering the southern and central part of Indonesia. This region is significantly influenced by the Australian monsoon. Region 2 is the anti-monsoonal region, while Region 3, covering most of the northwestern part of Indonesia, is the semi monsoonal region with two precipitation peaks per year. The precipitation data used in this study are

secondary data collected from two different sources, i.e. TRMM satellite-based data and ground data collected from the meteorological stations over Indonesia.

### A. TRMM Data

Daily precipitation dataset has been collected from the TRMM website. The data are in a grid point basis with resolution of  $0.25^0 \times 0.25^0$  over Indonesia regions and hence, we have 12765 grids points. The verification is applied to the daily precipitation data spanning from January 1, 1998, to December 31, 2016, while the time series clustering uses monthly average precipitation series from January 1998 to December 2017.

### B. Ground Data

This research uses the measured daily precipitation from a total of 161 meteorological stations over Indonesia (see Figure 1) from January 1, 1998, to December 31, 2016. Note that the ground data are used only to verify TRMM data, which for the period of 1998 to 2016 are already sufficient to describe the climatological conditions. The measured precipitation data are freely available from BMKG. The daily precipitation was collected from 7 AM to 7 AM Indonesia local time and it has been pre-processed for missing value imputation, where the missing data were filled with the precipitation data from the nearest adjacent station.

## III. METHODOLOGY

In order to verify the precipitation between TRMM satellite products and ground-based observations, the mismatch in the spatial scales between those two data sources needs to be carefully considered. This is mainly because the TRMM precipitation values used in this paper are available at grid scale, while measurements from meteorological stations represent precipitation at point scale. There are several ways to deal with this issue in order to carry out a direct comparison, such as spatial interpolation or simple averaging. In this study, the comparison will be conducted by upscaling the spatial position of the meteorological station with the closest grid. Meanwhile, areas with no meteorological station coverage were excluded from the evaluation. The analysis starts with the evaluation of TRMM precipitation with ground data through visual presentations, i.e. time series plot, boxplots and distribution plot. Moreover, the skill of TRMM precipitation data is verified by a simple t-test. The evaluation is focused on regional basis, instead of a global mean over Indonesia. Furthermore, clustering is done by examining the gridded scale dataset of TRMM precipitation data. The optimum number of clusters is determined by two criterias, pseudo-f statistic [18] and silhouette statistic [19]. High pseudo-f represents the optimal number of clusters, the members within a cluster are as homogeneous as possible, and the members between clusters are as heterogeneous as possible. The pseudo-f statistic is defined as:

$$Pseudo - f = \frac{\left( \frac{R^2}{(c-1)} \right)}{\left( \frac{(1-R^2)}{(n-c)} \right)} \quad (1)$$

$$\text{with } R^2 = \left( \frac{SST - SSW}{SST} \right), \quad SST = \sum_{i=1}^c \sum_{j=1}^{n_c} \sum_{k=1}^p (x_{ijk} - \bar{x}_k)^2,$$

$$SSW = \sum_{i=1}^c \sum_{j=1}^{n_c} \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2$$

where  $n$  is the number of samples,  $c$  the number of clusters,  $n_c$  the number of data in group  $c$ ,  $p$  the number of variables,  $x_{ijk}$  the group- $i$  on sample- $j$  and  $k$ -th variable,  $\bar{x}_k$  the average of all samples on variable- $k$ , and  $\bar{x}_{jk}$  the average of group- $j$  on variable- $k$ .

According to [20], one of the measurements to evaluate the goodness of time series clustering is the silhouette coefficient defined as:  $S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$ , where  $a(i)$  is the average

of the distance between members within a cluster,  $b(i)$  is the minimum value of the average distance of object  $i$  with object lies in other cluster, while  $S(i)$  is the silhouette coefficient of object  $i$ . The silhouette coefficient is the average of  $S(i)$  for every object. Table I provides the silhouette categories [21].

TABLE I. SILHOUETTE COEFFICIENTS CRITERIA

Silhouette coefficients	Category
0.71–1.00	Strong
0.51–0.70	Good
0.26–0.50	Weak
0.00–0.25	Bad

If a cluster is bad, this means that the observations within the cluster are not homogeneous, while the observations between clusters tend to be homogeneous. If the coefficient lies within the interval of 0.71 to 1, the cluster is said to be very good (strong) which means that the observations within clusters are very homogeneous, while observations between different clusters are very heterogeneous. In summary, the higher the silhouette coefficient, the better the clustering results. Figure 2 summarizes the research methodology steps.

#### IV. RESULTS AND DISCUSSION

##### A. Precipitation Characteristics in Indonesia Using TRMM Data

In general, Indonesia has been characterized as a tropical country with two seasons, the dry season from October to March and the rainy season from April to September. However, many researches have shown that there has been a shift on the seasonal periods. Furthermore, the area can be clustered into three climate regions [2]. Figure 3 depicts the daily mean precipitation from the TRMM data over the last 17 years. The picture shows that the mean precipitation over Indonesia varied in the range of 0 to 18mm/day. The BMKG defined three categories of mean precipitation in Indonesia, i.e. low (0mm/day to 3.33mm/day), medium (3.33mm/day to 10mm/day) and high (above 13.33mm/day). Bali and Nusa Tenggara regions usually have longer dry seasons compared to others. We see also that Papua and West Papua are two regions with very high precipitation rate. Kalimantan and Sumatra have high precipitation rates, while Java's and Sulawesi's tend to be medium.

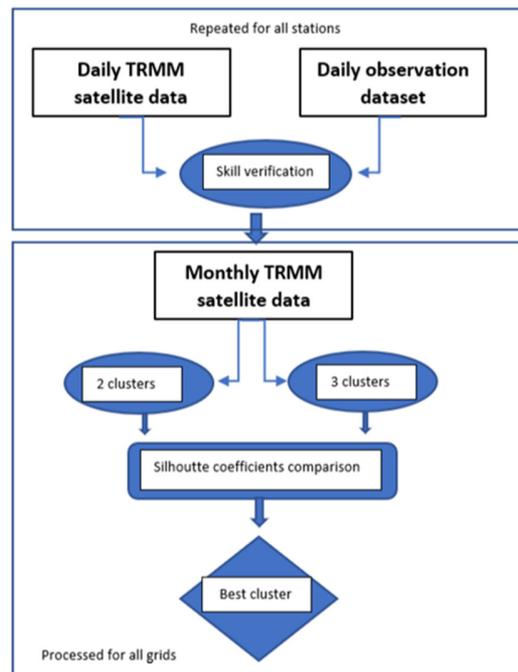


Fig. 2. Steps of the analysis

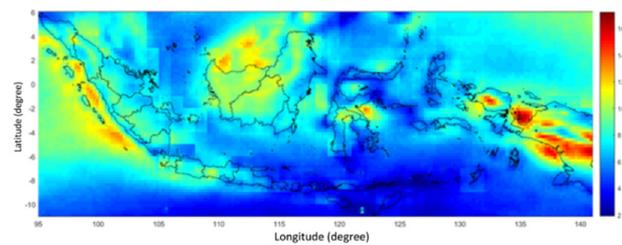


Fig. 3. Average rainfall in Indonesia (mm/day)

The average daily precipitation per month can be seen in Figure 4. The bar chart shows the mean precipitation over three climate regions as defined by [2]. Region 1 (monsoon type) has a U-shape, consistent with the findings of [2]. In this case, the rainy and dry seasons can be differentiated clearly, i.e. November- April for the rainy season and May-October for the dry season. The time series plot shows that the precipitation rate did not change significantly over time with the average level of 7.25mm/day. Anti-monsoon regions have a nearly similar pattern with the monsoonal type, however, the precipitation rate during the dry season is higher than in the monsoonal type. For the semi-monsoonal, the monthly precipitation pattern is very similar to the anti-monsoonal type during rainy seasons, and nearly similar to anti-monsoonal type during dry periods. This paper carries out re-clustering to identify whether the precipitation patterns are significantly different among those three regions. Figure 5 depicts the distribution of the observed precipitations with the TRMM generated precipitation. The graph provides a nice way to verify the efficiency of the TRMM data. Furthermore, the mean rainfall difference from both sources is tested by t-test. We see that the TRMM precipitation is an under-estimate of the observed data over all regions showed by their lower mean

value than the ground observation. The difference (bias) is around 0.864mm/day, while the t-test showed that the differences are not statistically significant over those three regions. This means that the TRMM generated precipitation data are good enough to represent the observation data. As for clustering, the bias can be neglected because the clustering process concerns only on the pattern of TRMM data.

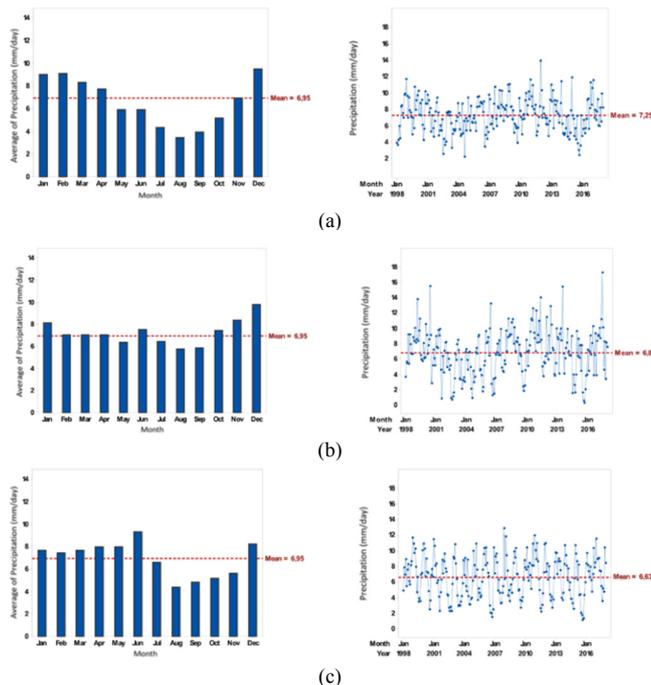


Fig. 4. Rainfall patterns (mm/day) over: (a) region 1, (b) region 2, and (c) region 3

**B. Clustering of Climatic Regions in Indonesia**

Clustering by using the Euclidean distance is done by specifying the number of clusters as 2 and 3. The Euclidean distance measures the distance between two series without taking into account the autocorrelation properties of the series, and it is hereafter referred as standard approach in clustering. The pseudo-f statistic for each number of clusters is 291.691 and 342.51, while the silhouette coefficient is 0.136 and 0.145 respectively. Based on these two criteria, the Euclidean distance suggests that the optimum cluster number is 3. The clusters of the climatic regions can be seen in Figure 6.

The distribution of the climatic regions looks very similar with the findings of [2]. However, we observe that there are shifts over some areas. Most regions belong to Region 1 or Region 3, while Region 2 consists of only a few regions. Region 3 covers mostly the west and northwest part of Indonesia, i.e. most of the areas in Sumatra, West Java, Central Java, Jakarta, Banten, Yogyakarta and the northern part of Kalimantan. Region 1 covers East Java, Bali, Nusa Tenggara, parts of Sulawesi, Riau, South Sumatra, Aceh, the entire Maluku, parts of East Kalimantan and South Kalimantan. Region 2 includes Sulawesi and Papua areas. The rainfall pattern over each region can be seen in Figure 7.

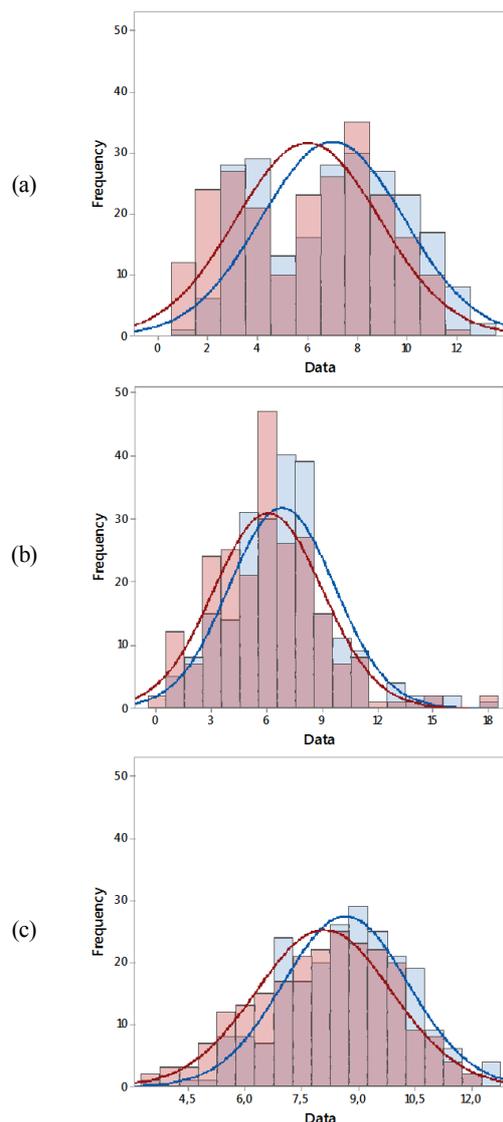


Fig. 5. Rainfall pattern distribution over regions. (a) Region 1, (b) Region 2, and (c) Region 3 (blue = ground data, red=TRMM data)

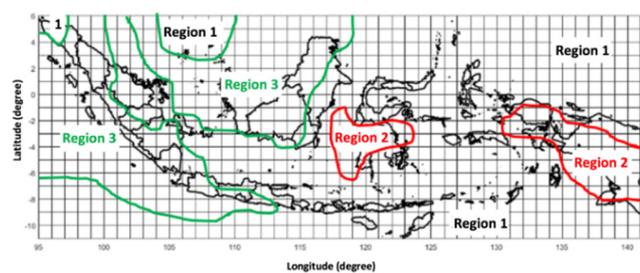


Fig. 6. Clustering of climatic regions (with Euclidean distance)

The rainfall pattern in Region 1 forms a monsoon type or U-shape. The areas in this region had rainy season at the beginning and the end of the year, while the dry season occurred in the middle of the year, from May to October. The peak of the rainy season occurred in January and December while the dry season's peak was in August and September.

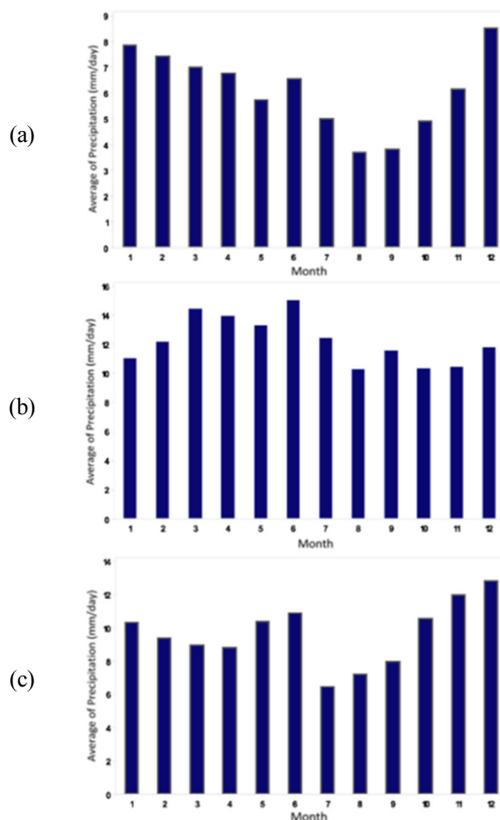


Fig. 7. Rainfall pattern over different regions: (a) Region 1, (b) Region 2, and (c) Region 3

Over Region 2, the rainfall pattern shows an opposite type than that of Region 1, so it corresponds to the term anti-monsoon. Over this region, the rainy season occurred in the middle of the year starting from March to July, while the lowest rainfall happened in the beginning and the end of the year. Region 3 has 2 rainfall peaks in a year, indicating rainy season occurrence twice in a year, in the beginning and in the middle of the year. Based on [3], this pattern is termed as monsoon spring zone or semi-monsoon. The summary of the average precipitation cycle of these three regions is presented in histogram in Figure 8.

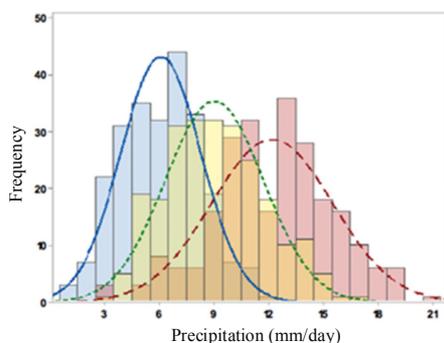


Fig. 8. Histogram of the average precipitation over the 3 different regions. Blue: Region 1, red: Region 2, green: Region 3

It can be seen that Region 1 is drier than the two other regions with lower deviation. This indicates that the rainfall intensity in Region 1 tends to be more stable. Region 3 seems to be the wettest region with the largest deviation indicating higher variability in the rainfall intensity, while Region 2 has a normal rainfall intensity.

## V. CONCLUSION

This paper identified the climatic regions in Indonesia using the satellite TRMM data as input. It has been proven that the TRMM has no significant bias towards the ground dataset, which means that the TRMM data are a good way to overcome the weaknesses of ground data, especially in the case of data scarcity due to the limited number of meteorological stations. Clustering by Euclidean distance suggested three clusters in line with the findings in [2]. We concluded that the current Indonesia climate can still be clustered into 3 zones, i.e. monsoonal type, anti-monsoon type and semi-monsoonal type. We observed that the average monthly precipitation cycle in some areas shifted from one to another type within last the 24 years. This fact is supported by the findings of [22-25], among others, which proved that climate change shifted the rainfall pattern. The climate change impact on the rainfall pattern in Indonesia has been well documented in [26, 27].

## REFERENCES

- [1] M. C. Wheeler, J. L. McBride, Australian-Indonesian Monsoon, Springer, 2015
- [2] E. Aldrian, R. D. Susanto, "Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature", International Journal of Climatology, Vol. 23, No. 12, pp. 1435-1452, 2003
- [3] NASDA, TRMM Data Users Handbook, NASDA, 2001
- [4] <https://trmm.gsfc.nasa.gov/>
- [5] J. Simpson, C. Kummerow, W. K. Tao, R. F. Adler, "On the tropical rainfall measuring mission (TRMM)", Meteorology and Atmospheric Physics, Vol. 60, No. 1-3, pp. 19-36, 1996
- [6] J. Liu, Z. Duan, J. Jiang, A. X. Zhu, "Evaluation of three satellite precipitation products TRMM 3B42, CMORPH, and PERSIANN over a subtropical watershed in China", Advances in Meteorology, Vol. 2015, Article ID 151239, 2015
- [7] M. Darand, J. Amanollahi, S. Zandkarimi, "Evaluation of the performance of TRMM multi-satellite precipitation analysis (TMPA) estimation over Iran", Atmospheric Research, Vol. 190, pp. 121-127, 2017
- [8] R. Prasetya, A. R. A. Syakur, T. Osawa, "Validation of TRMM precipitation radar satellite data over Indonesian region", Theoretical and Applied Climatology, Vol. 112, No. 3-4, pp. 575-587, 2012
- [9] A. R. As-Syakur, T. Tanaka, T. Osawa, M. S. Mahendra, "Indonesian rainfall variability observation using TRMM multi-satellite data", International Journal of Remote Sensing, Vol. 34, No. 21, pp. 7723-7738, 2013
- [10] D. Gunawan, "Perbandingan curah hujan bulanan dari data pengamatan permukaan, satelit TRMM and model permukaan NOAH", Jurnal Meteorologi and Geofisika, Vol. 9, No. 1, pp. 1-10, 2008
- [11] M. P. H. Giarno, S. Suprayogi, S. H. Murti, "Distribution of accuracy of TRMM daily rainfall in Makassar strait", Forum Geografi, Vol. 32, No. 1, pp. 38-52, 2018
- [12] B. Collischonn, W. Collischonn, C. E. M. Tucci, "Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates", Journal of Hydrology, Vol. 360, No. 1-4, pp. 207-216, 2008
- [13] T. Zhao, A. Yatagai, "Evaluation of TRMM 3B42 product using a new gauge-based analysis of daily precipitation over China", International Journal Climatology, Vol. 34, pp. 2749-2762, 2013

- [14] M. L. Tan, Z. Duan, "Assessment of GPM and TRMM precipitation products over Singapore", *Remote Sensing*, Vol. 9, No. 7, Article ID 720, 2017
- [15] J. Hur, S. V. Raghavan, N. S. Nguyen, S. Y. Liong, "Are satellite products good proxies for gauge precipitation over Singapore?", *Theoretical and Applied Climatology*, Vol. 132, No. 3-4, pp. 921-932, 2018
- [16] S. N. M. Zad, Z. Zulkafli, F. M. Muharram, "Satellite rainfall (TRMM 3B42-V7) performance assessment and adjustment over Pahang River Basin, Malaysia", *Remote Sensing*, Vol. 10, No. 3, Article ID 388, 2018
- [17] T. Omotosho, J. S. Mandeep, M. Abdullah, A. Adediji, "Distribution of one-minute rain rate in Malaysia derived from TRMM satellite data", *Annales Geophysicae*, Vol. 31, No. 11, pp. 2013-2022, 2013
- [18] A. R. Orpin, V. E. Kostylev, "Towards a statistically valid method of textural sea floor characterization of benthic habitats", *Marine Geology*, Vol. 225, No. 1-4, pp. 209-222, 2006
- [19] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53-65, 1987
- [20] J. F. Hair, R. E. Anderson, R. L. Tatham, W. C. Black, *Multivariate Data Analysis with Reading*, Prentice Hall, 1995
- [21] L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, 1990
- [22] T. H. Udayashankara, "Impact of climate change on rainfall pattern and reservoir level", *Journal of Water Resource Engineering and Management*, Vol. 3, No. 1, pp. 10-14, 2016
- [23] A. G. Pendergrass, D. L. Hartmann, "Changes in the distribution of rainfall frequency and intensity in response to global warming", *Journal of Climate*, Vol. 27, No. 22, pp. 8372-8383, 2014
- [24] J. Crossman, M. N. Futter, P. G. Whitehead, "The Significance of Shifts in Precipitation Patterns: Modelling the Impacts of Climate Change and Glacier Retreat on Extreme Flood Events in Denali National Park, Alaska", *PLOS ONE*, Vol. 8, No. 9, Article ID e74054, 2013
- [25] J. D. Miranda, C. Armas, F. M. Padila, F. I. Pugnaire, "Climatic change and rainfall patterns: Effects on semi-arid plant communities of the Iberian Southeast", *Journal of Arid Environment*, Vol. 75, No. 12, pp. 1302-1309, 2011
- [26] M. Case, F. Ardiasyah, E. Spector, *Climate Change in Indonesia-Implications for Humans and Nature*, WWF, 2007
- [27] M. Measay, "Indonesia: A vulnerable country in the face of climate change", *Global Majority E-Journal*, Vol. 1, No. 1, pp. 31-45, 2010