**Response to reviewers’ comments.**

Dear Reviewers,

Thank you very much for your very constructive comments. Please kindly find the response to each comments below. We hope that the revised version can be accepted for publication in ETASR Journal.

Best regards

Reviewer B:

Comment: “I suggest that the authors add to the introduction a clearer presentation on the contextualization and the purpose of the study.  
Answer :” We uodated and revised the last part of the introduction section by clearly mentioning the purpose of the study, as well as the motivation of the study”.

Comment: “I urge the authors to introduce a justification for using soil data by 2016. Why did not they use data from 2017 and 2018?”

Answer: “We added a sentence to justify the use of ground data by 2016. Since it is daily data, the periods of 18 years is sufficient enough to represent the climatological condition. The initial work (data management process) of this study was started in late 2017 and it was taking long time to pre-process it due to the data structure (grid points basis). Therefore, the data in 2017 is used in the clustering process only. We believe that including 1 year data of 2017 during the skill verification would not change the result significantly.”   
  
Comment: “Better to analyze the data in table 1.”

Answer : “ We have added the description after the table in this revised version”  
  
Comment: “To facilitate the reading and understanding of the methodology used, I recommend that the authors add a step-by-step description of how the research was conducted. Can be in flowchart form.

Answer : “We summarizes the steps in Figure 3.”  
  
  
------------------------------------------------------  
Reviewer C:

Comment: “Some formatting issues. Format the paper properly, use ETASR related styles (see template). Also please write references properly (see template).”

Answer: “We updated the paper fllowing the provided templete, including the references”

Clustering of Precipitation Pattern in Indonesia Using TRMM satellite data

Heri Kuswanto

Department of Statisics

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

heri\_k@statistika.its.ac.id

Dedi Setiawan

Department of Statisics

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

[dedi1statistika@gmail.com](mailto:dedi1statistika@gmail.com)

Ardhasena Sopaheluwakan

Agency for Meteorology, Climatology and Geophysics (BMKG)

Jakarta, Indonesia

ardhasena@bmkg.go.id

*Abstract*—This paper identified the climatic regions in Indonesia based on the similarity of rainfall pattern using TRMM data. Indonesia is a tropical climate region and has been clustered into three different clusters i.e. monsoonal, anti-monsoonal and semi-monsoonal types. The cluster was formed by examining rainfall observation dataset recorded at number of stations over Indonesia with coarse spatial resolution. Furthermore, clustering based on higher resolution dataset is needed to characterize the rainfall pattern over areas with no station. TRMM data provides high resolution gridded dataset which has been proven to have a good skill over Indonesia. A statistical test has been applied to evaluate the significance of TRMM bias, and it indicated that 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 pattern. The cluster using Euclidian Distance found that Indonesia can be grouped into three significantly different rainfall patterns. Compared to the existing reference, 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

Introduction

Indonesia is an archipelago located between 100S to 60N, and 950E to 1410E, and hence, part of the regions is on the equatorial line with a high degree of rainfall variability. The Indonesian climate condition is 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 as well as the ocean and the land also influence the climate variability among regions in Indonesia. Aldrian and Susanto [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 rainfall a year called as 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 only have once maximum rainfall a year. Meanwhile, the anti-monsoonal pattern is shown by the unimodal rainfall pattern, opposite to the monsoonal rainfall type [2].

Due to the high degree of rainfall variability, the analysis of precipitation condition in Indonesia requires long historical series with adequate spatial coverage. Rainfall observation in Indonesia is managed by the Agency for Meteorology, Climatology and Geophysics (BMKG) through 116 meteorological stations spreaded out over Indonesia. Moreover, the BMKG also uses satellite product to support the need for fast and continue rainfall data over Indonesia. One of the obvious strengths of satellite data is its spatial coverage which is normally available in a high-resolution grid covering all areas in the regions. One of the satelite products that is commonly used by the BMKG as their 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 and learn precipitation over tropical regions as an effort to study earth as a global system and it was developed firtsly in 1997 [4]. The TRMM satellite measures rainfall intensity in every region various temporal scale. The spatial resolution also varies from 0.25 to 5 degrees. The mission of this satellite is to understand better the precipitation structure and warming in the tropical region in the earth [5]. Among the studies which note the strong performance of TRMM in specific regions compared to other satelite products are [6, 7] for the case of China and Iran respectively.

The BMKG uses TRMM satellite data to observe the rainfall condition over Indonesia due to the limited number meteorological stations. The TRMM data has also been widely used in weather and climate studies in Indonesia such as the use of TRMM to study extreme weather events carried out by [8]. As-syakur et al. [9] did a study about the advantage of using TRMM based precipitation to investigate the spatiotemporal pattern and rainfall characteristic over Indonesia. Gunawan [10] compared monthly rainfall observed from meteorological stations with TRMM and NOAA surface model using simple mean testing and found that TRMM has a potential to impute the missing spatiotemporal data in some areas which have no data. Moreover, Giarno et al. [11] verified TRMM satellite precipitation data and observation rainfall over Makassar Indonesia using 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 has also been verified in many other countries such as [12] which used TRMM to estimate rainfall in Topajo River in Amazon using bias correction, while [13] verified TRMM 3B42 over China. TRMM applications have also been intensively investigated in other tropical countries close to Indonesia such as in Singapore [14, 15], in Malaysia [16, 17] and many others.

The purpose of this study is to use the TRMM data to update the cluster of climatic regions in Indonesia based on the precipitation pattern. Moreover, the TRMM data will be verified towards observation data in Indonesia prior to clustering process in order to justify that TRMM is a reasonable choice to be used as the basis of clustering. This work differs from the study carried out by [2] in several significant ways. First, we use the verified 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 high resolution dataset used in the clustering process. Moreover, the TRMM data provides rainfall information over regions without stations. Second, this research uses different clustering process in term of the distance measure i.e. standard (Euclidean) distance measure. Meanwhile, [2] used Double Correlation (DC) which treats the data as a cross-sectional series. Another important different is about the period of the examined dataset where this study examines the average monthly rainfall over last 18 years, while [2] used data spanning from 1961 to 1993.

Corresponding author: Heri Kuswanto

# STUDY AREA AND DATA

Indonesia is an archipelago with thousands of islands stretching along the Equator from 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 year round. Moreover, the rainfall quantity and distribution vary due to the location humidity level as well as the **monsoon regime**. The availability of historical rainfall data recorded at meteorological stations is an important aspect to study further the rainfall variability. Figure 1 depicts the location of the 116 meteorological stations over Indonesia.

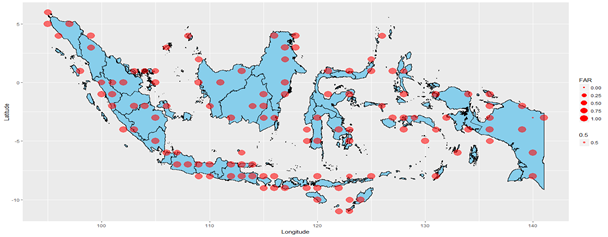


Fig. 1. Location of the meteorological stations in Indonesia

We can see clearly from Figure 1 that there are some areas without meteorological stations which make it difficult to characterize the climatic pattern in those areas. Therefore, using high resolution gridded data from TRMM will make it possible to investigate the climate pattern in a specific region precisely. The climate in Indonesia, according to [2] is classified into three different regions as can be seen in Figure 2.

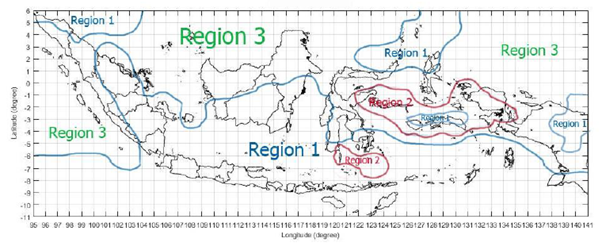


Fig. 2. Climate region in Indonesia (source: Aldrian and Susanto [2])

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 is the semi monsoonal region with two precipitation peaks per year covering most of the northwestern part of Indonesia.

The precipitation data used in this study are secondary data collected from two different sources i.e. TRMM data as the satellite-based data and ground data collected from the meteorological stations over Indonesia.

* TRMM data

Daily precipitation dataset has been collected from the TRMM website. The data is in a grid point basis with resolution of 0.250 x 0.250 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.

* 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 is used only to verify the TRMM skill, where the periods of 1998 to 2016 is already sufficient to describe the climatological condition. The measured precipitation data are freely available from the Agency for Meteorology, Climatology and Geophysics (BMKG) Indonesia. 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 then filled with the precipitation data from the nearest adjacent station.

# 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 possibly carried out a direct comparison, such as spatial interpolation or simple averaging. In this study, the comparison study will be conducted by upscaling the spatial position of the meteorological station with the closest grid. Meanwhile, other grids covering no meteorological stations were excluded from the evaluation.

The analysis is started with evaluation of TRMM precipitation with ground data through visual presentations i.e. time series plot, boxplots and distribution (pdf) plot. Moreover, the skill of TRMM precipitation data is verified by 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 i.e. pseudo-f statistic [18] and Silhouette statistic of [19]. The high pseudo-f represents the optimal number of cluster i.e. members within the cluster are as homogeneous as possible, and members between cluster are as heterogeneous as possible. The pseudo-f statistic is defined as

 (3)

with, , ,



where,

*SST* : sum square of the distance to the overall mean

*SSW* : sum square of the distance of the object t the average of its group

*n* : number of samples

*c* : number of clusters

: number of data in group c

*p* : number of variables

: group-*i* on sample-*j* and *k*-th variable

: the average of all sample on variable-*k*

: 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 , where *a(i)* is the average of distance between members within the cluster, *b(i)* is the minimum value of the average distance of object i with object lies in others cluster, while *S(i)* is Silhouette coefficient of object *i*. The Silhouette coefficient is the average of *S(i)* for every object. The following table provides Silhouette categories following [21].

##### TABLE 1. 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 |

Table 1 provides the Silhoutte coefficient with its corresponding category of cluster result. Based on Table 1, we see that if the coefficient of the cluster is below 0.25, then the cluster is said to be bad. It means that the observations within cluster are inhomogeneous, while observations between cluster are tend to be homogeneous. Meanwhile, 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 3 below summarizes the steps of the research methodology.

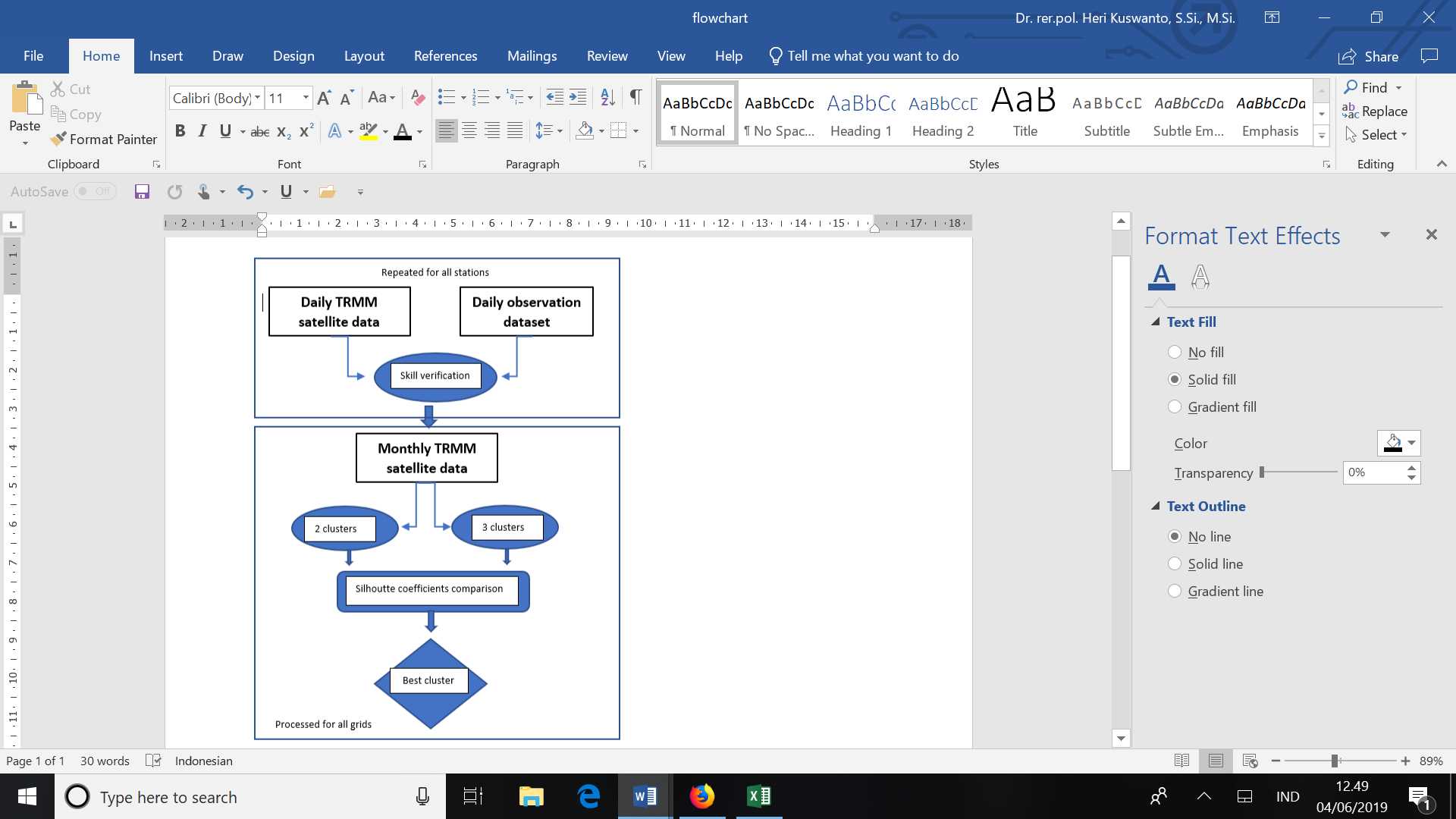


Fig. 3. Steps of the analysis

# RESULTS AND DISSUSSION

## Characteristic of Precipitation in Indonesia Using TRMM Precipitation

In general, Indonesia has been characterized as a tropical country with two seasons i.e. dry season within the periods of October to March and rainy season happening on April to September. However, many researches have shown that there has been a shift on the seasonal periods. Furthermore, as found by [2], the area over Indonesia can be clustered into three climate regions i.e. monsoon, anti-monsoon and semi-monsoonal types. Figure 4 depicts the daily mean precipitation generated from TRMM data over the last 17 years. The picture reveals that the mean precipitation over Indonesia varied in the range of 0 to 18 mm/day. The BMKG defined three categories of mean precipitation in Indonesia i.e. low (0 mm/day to 3.33 mm/day), medium (3.33 mm/day to 10 mm/day) and high (above 13.33 mm/day). We see that Bali and Nusa Tenggara are quite dry, which mean that those regions usually have longer dry seasons compared to other regions in Indonesia. We see also that Papua and West Papua are two regions with very high precipitation rate. Kalimantan and Sumatra have a high precipitation rate, while Java and Sulawesi tend to be medium.

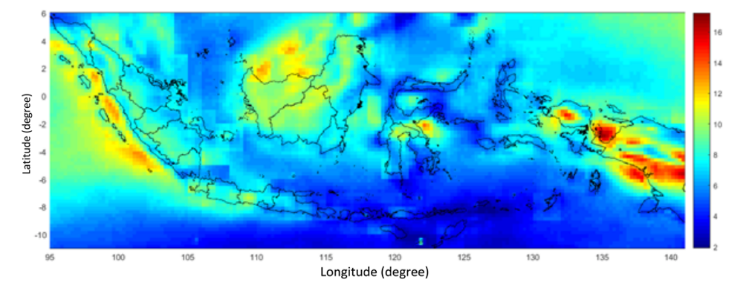
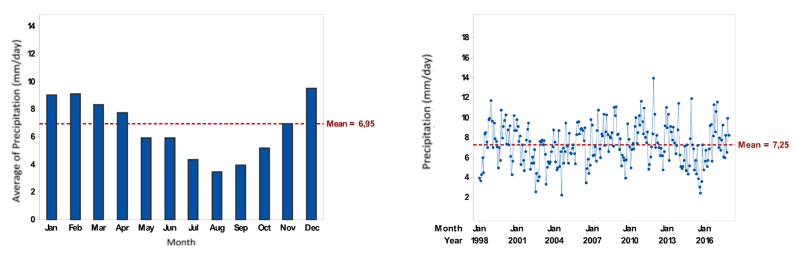
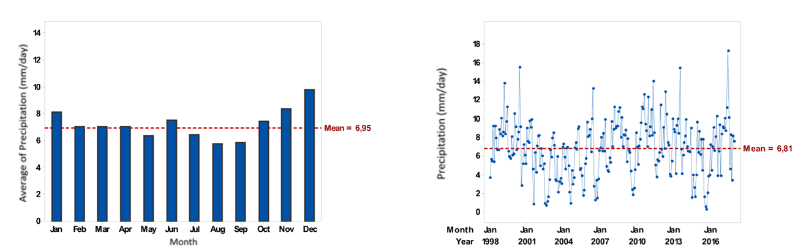


Fig. 4. Average Rainfall in Indonesia (mm/day)

The average daily precipitation per month can be seen in Figure 5. The bar chart shows the mean precipitation over three climate regions as defined by [2]. The region 1 (monsoon type) has a U-shape, consistent with the finding of [2]. In this case, the rainy and dry season 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 the time with the average level of 7.25 mm/day. The anti-monsson regions have a nearly similar pattern as monsoonal type, however, the precipitation rate during the dry season is higher than in monsoonal type. For the semi-monsoonal, the monthly precipitation pattern is very similar to the anti-monsoonal type during the rainy seasons, and the pattern 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.





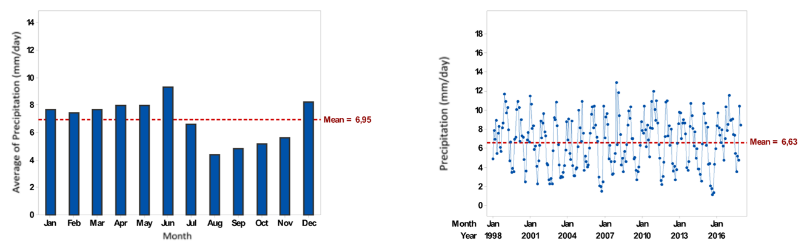


Fig. 5**.** Rainfall patterns (mm/day) over region 1 (upper panel), region 2 (middle panel), region 3 (lower panel)

Figure 6 depicts the distribution of the observed precipitations with the TRMM generated precipitation. The graph provides a nice way to verify the skill of the TRMM precipitation. Furthermore, the mean difference of rainfall from both sources is tested by using the t-test. We see that the TRMM precipitation is under-estimate the observed data over all regions showed by lower mean value than the ground observation. The difference (bias) is around 0.864 mm/day, where the t-test showed that the differences are not statistically significant over those three regions. It means that the TRMM generated precipitation data is 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. 6. Rainfall patterns distribution over regions in Region 1, Region 2 and Region 3 respectively (blue = ground data, red=TRMM data)

## Clustering of Climatic Regions in Indonesia

Clustering the area by using the Euclidean distance is done by specifying the number of clusters as 2 and 3 clusters. The Euclidean distance measures the distance between two series without taking into account the autocorrelation properties of the series, and it is hereafter refereed as standard approach in clustering. The pseudo-f statistics for each number of cluster are 291.691 and 342.51, while the Silhoutte coefficient for each cluster are 0.136 and 0.145 respectively. Based on these two criterias, the Euclidean distance suggests that the optimum cluster is 3. The clusters of the climatic regions can be seen in Figure 7.

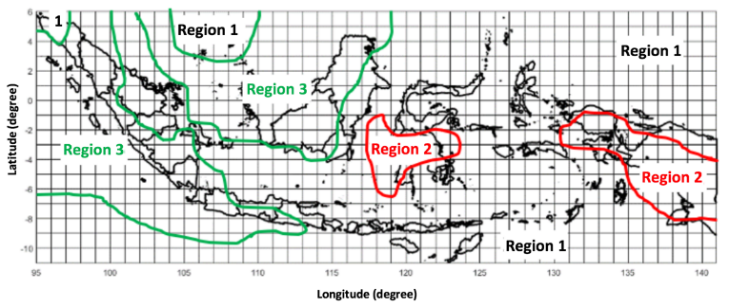


Fig. 7. Clustering of climatic regions with Euclidean distance

The distribution of the climatic regions look very similar with the finding of [2]. However, we observe that there are shifts over some areas. Most of the regions belong to Region 1 and Region 3, while Region 2 consists only few regions. Region 3 covers mostly west and nortwest part of Indonesia. 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 include Sulawesi and Papua areas. Region 3 covers most of the areas in Sumatra, West Java, Central Java, Jakarta, Banten, Yogyakarta and the northern part of Kalimantan. The rainfall pattern over each region can be seen in Figure 8.

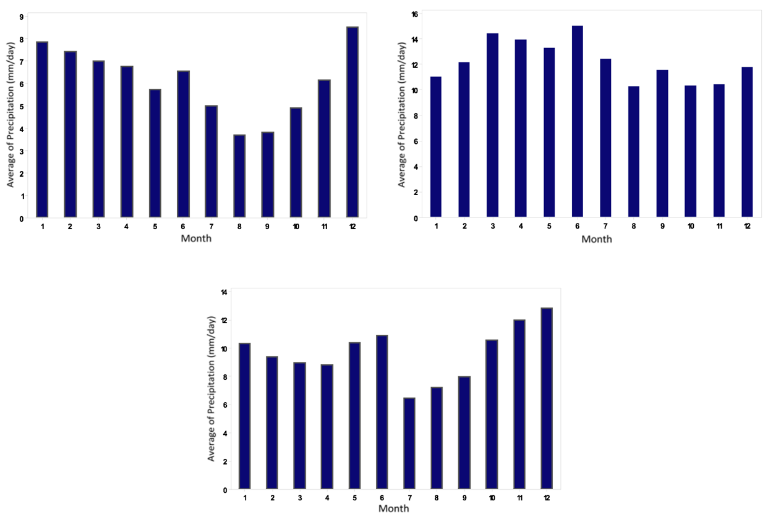
****

Fig. 8**.** Rainfall pattern over different region : Region 1 (upper left), Region 2 (upper right), Region 3 (lower) panel

Rainfall pattern in Region 1 forms a monsoon type or U-shape. The areas in this region had rainy season at the beginning and end of the year, while the dry season occurred in the middle of the year which is around May to October. The peak of the rainy season occurred in January and December while the dry season was in August and September. Over Region 2, the rainfall pattern shows an opposite type of Region 1 or it corresponds to as 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. It indicates that the areas within this region have rainy season twice a year i.e. in the beginning and in the middle of the year. Based on the study carried out by [3], this pattern is called as monsoon spring zone or semi-monsoon. Furthermore, the summary of the average precipitation cycle within those three different regions is presented in histogram as can be seen in Figure 9.

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



Fig. 9.Histogram of the average precipitation over three different regions

# CONCLUSION

This paper identified the climatic regions in Indonesia using the satelite TRMM output as the input. It has been proven that the TRMM has no significant bias towards ground dataset, which means that the TRMM data is a good proxy to overcome the weaknesses of groud data, especially in the case of data scarcity due to limited number of meteorological stations. The cluster by using Euclidiean distance suggested three clusters, in line with the work of [2]. We concluded that the Indonesia climate is currently still can 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 have shifted from one to another type within last 24 year. This fact is supported by the finding of many researches which proved that climate change has shifted the rainfall pattern (see [22, 23, 24] and [25] among others). The climate change impact on rainfall pattern in Indonesia has been well documented in [26, 27] and many others.

REFERENCES

1. M.C. Wheeler, J. McBride, “Australian-Indonesian monsoon”, in Lau WKM and Waliser DE (eds), Intraseasonal variability in the atmosphere-ocean climate system”, Springer, pp. 125-173, 2015.
2. E. Aldrian, D. Susanto, “Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature”, International Journal of Climatology, Vol. 23, No. 2, pp. 1435-1452, 2003.
3. NASDA. “TRMM data users handbook”, National Space Development Agency of Japan. Earth Observation Center, Japan, 2001.
4. S. A. Braun, “Tropical Rainfall Measuring Mission”, Senior Review Proposal, pp. 1-55, 2001.
5. J. Simpson, C. Kummerow, W. K. Tao, R. F. Adler, “On the Tropical Rainfall Measuring Mission (TRMM)”, Meteorology and Atmospheric Physics, Vol. 60, 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, Article ID 151239, pp. 1-13, 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. 63, No. 190, pp. 121-127, 2017.
8. R. Prasetia, A. R. As-Syakur, T. Osawa, “Validation of TRMM precipitation radar satellite data over Indonesian region”, Theoretical and Applied Climatology, Vol. 112, Issue 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 Remote Sensing, Vol. 34, 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. Giarno, S. Suprayogi, S. H. Murti, “Distribution of accuracy of TRMM daily rainfall in Makassar Strait”, Forum Geografi, Vol. 32, No. 1, 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**”,** Int. J. Climatol. Vol. 34, No. 8, 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, pp. 721, 2017.
15. J. Hur, S. V. Raghavan, N. S. Nguyen, S. Y. Liong, “Are satellite products good proxies for gauge precipitation over Singapore?”, Theor. Appl. Climatol., Vol. 132, Issue 3-4, pp. 921-932, 2018.
16. S. N. 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, pp. 388, 2018.
17. T. V. Omotosho, J. S. Mandeep, M. Abdullah, A. T. Adediji, “Distribution of one-minute rain rate in Malaysia derived from TRMM satellite data”, Ann. Geophys. Vol. 31, 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, pp. 209-222, 2006.
19. P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, Computational and Applied Mathematics, Vol. 20, pp. 53-65, 1987.
20. J. Hair, R. Anderson, R. L. Tatham, W. C. Black, “Multivariate data analysis with reading*”,* Englewood Cliffs: Prentice-Hall., 1995.
21. L. Kaufman, P. J. Rousseeuw, “Finding Groups in Data”, John Wiley & Sons, New York, 1990.
22. T. H. Udayashankara, M. B. M. Sadashiva, M. Madhukar, “Impact of climate change on rainfall pattern and reservior level”, Journal of Water Resource Engineering and Management, Vol. 3, No. 1, pp. 10-14, 2016.
23. A. G. Pendergrass, D. L. Hartmann, D.L. “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, 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”, Gland, Switzerland: World Wide Fund for 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.