Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 114 (2017) 324–332

Procedia Computer Science

www.elsevier.com/locate/procedia

Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA

Defining Philippine Climate Zones Using Surface and High-Resolution Satellite Data

Irenea L. Corporal-Lodangco[a,]*, Lance M. Leslie[b]

[a]School of Meteorology, University of Oklahoma, Norman, OK 73072, USA
[b]Mathematics and Physical Sciences, University of Technology Sydney, Ultimo NSW 2007, Australia

Abstract

Philippine climate zones traditionally were classified from a rain-gauge network, using the Modified Coronas Classification (MCC). MCC uses average monthly rainfall totals to define four climate zones: Types I-IV. Types I and III have wet and dry seasons, whereas Types II and IV have wet seasons but no dry seasons.

The present study redefines Philippine climate zones by applying cluster analysis to the average monthly rainfall amounts from surface-based rain-gauge observations, and dense, high-resolution satellite data from the Tropical Rainfall Monitoring Mission (TRMM). To determine the optimal number of climate type clusters, both single-linkage hierarchical and K-means cluster analysis algorithms were used, together with known characteristics of Philippine rainfall distributions and attributes.

Employing single linkage hierarchical and K-means methods in tandem identified six different Philippine climate types, which is two climate types more than the currently accepted MCC climate classification. Due to the far greater number of TRMM observations compared with the rain gauge network, the study provides more clearly defined cluster characteristics including the spatial and temporal variability of climate divisions. This study uses known meteorological factors contributing to the identification of six distinct climate types. This paper is intended to assist agricultural stakeholders with planning and decision-making.

Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.

Keywords: cluster analysis; Philippines; climate types; high-resolution rainfall satellite data; K-means; single-linkage

* Corresponding author. Tel.: +1-405-620-5893; fax: +1-405-325-3098.
E-mail address: ireneacl@ou.edu

1877-0509
Peer-review under responsibility of the scientific committee of the Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems.
10.1016/j.procs.2017.09.068

1. Introduction

The Philippines, because it is situated in the tropics, experiences an annual temperature range that is not large enough to produce the four distinct seasons typical of the extra-tropics. The Philippine archipelago has high levels of solar radiation year round as it is located entirely between the Equator and the Tropic of Cancer. Consequently, the single variable best describing the Philippines climate is rainfall, which is greatly influenced by other factors affecting rainfall amounts and patterns, such as prevailing wind systems, tropical cyclone activity, topography and geographical locations, thunderstorms, location of the ITCZ, and surface temperature. The Philippines conventionally is classified into four climates zones, based on the Modified Coronas Classification (MCC) applied to a relatively small network of surface (rain gauge) observations. The primary purpose of this study is to apply cluster analysis to two sets of monthly rainfall data: surface-based rainfall observations, and the much denser Tropical Rainfall Measuring Mission (TRMM) high-resolution satellite rainfall data. TRMM satellite rainfall data are used to identify distinct spatiotemporal rainfall patterns that redefine the MCC-based climate types of the Philippines. The clusters obtained from the rain gauge data are used to assess whether the TRMM rainfall data accurately captures the Philippine general rainfall patterns.

There are techniques available for classification and several studies used cluster analysis to classify rainfall data. Both [1] and [2] employed K-means algorithm, whereas [3] and [4] used agglomerative hierarchical clustering. In this study, the single linkage hierarchical and K-means non-hierarchical cluster methods are used to define the climate zones of the Philippines as a function of rainfall characteristics.

A unique feature of the study is the application of two cluster algorithms on surface-based and satellite rainfall data, and the integration of the known attributes of Philippine rainfall as influenced by the interaction of several factors to define the most appropriate clusters.

2. Data and Methodology

2.1 Rain Gauge Observation Data

Monthly surface-based data, from 52 rain gauge stations, are obtained from the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA). The PAGASA monthly rainfall for each month, January to December, from 1961-2015 was averaged and used in the clustering process.

2.2 TRMM Satellite Rainfall Data

The Tropical Rainfall Measuring Mission (TRMM) is a joint U.S.-Japan satellite mission that monitors tropical and subtropical precipitation and estimates its associated latent heating. In this study, the 3B43 product is employed. The dataset is from the Giovanni online data system of the National Aeronautics and Space Administration (NASA) Goddard Earth Sciences - Data and Information Services Center (GES-DISC). The product provides the best precipitation estimate from all global data sources, namely high-quality microwave data, infrared data, and analyses of rain gauges. Spatial resolution is 0.25° by 0.25° and extends from 50°S to 50°N. The temporal resolution is monthly, provided as hourly rain rate (mm/hr) during 1998-2016. Monthly averages are hourly rain rates multiplied by the total hours in each month.

2.3 Methods

Cluster analysis is a tool for classification such that objects in the same cluster have similar properties, and those with dissimilar features belong to another cluster. The single linkage technique falls into a category called agglomerative hierarchical clustering and is a simple, widely used method. The procedure is characterized by the tree-like structure established in the course of the analysis [5]. Initially, this type of procedure starts with each object representing an individual cluster. These clusters are then sequentially merged according to their similarity. First, the two similar clusters (i.e., those with the smallest distance between them) are merged to form a new cluster at the bottom of the hierarchy. Next, another pair of clusters is merged and linked to a higher level of the hierarchy, and so on. This allows a hierarchy of clusters to be established from the bottom up.

Another popular cluster technique is K-means, a simple method that follows a partitioning procedure. It is an entirely different concept from the single linkage method discussed earlier. This algorithm is not based on distance measures from one observation to another observation, but uses the within-cluster variation to form homogenous clusters [5]. Specifically, the procedure segments data such that the within-cluster variation is minimized. The clustering process starts by randomly assigning objects to a number of clusters k, where k is a user-specified parameter. The objects are then successively reassigned to other clusters to minimize the within-cluster variations, which is basically the distance from each observation to the centroid of associated cluster. If the reallocation of an object to another cluster decreases the within-cluster variation, this object is re-assigned to that cluster.

A solution advocated by many experts is to apply a single linkage procedure to determine the number of clusters, followed by K-means [6, 7, 8]. This procedure increases the validity of the solutions [7, 8, 9, 10]. The only cost is the extra time and effort required; a cost outweighed by benefits. The best solutions use single linkage and K-means methods in tandem. The single-linkage method automatically generates the available number of clusters, defined by cluster trees or dendrograms. However, the number of clusters, k, in the K-means method must be pre-specified.

2.4 Optimum Number of Clusters

To determine the most appropriate number of clusters, the dendrogram and cophenetic correlation coefficient (CCC) values from single linkage clustering, and the silhouette coefficient values from K-means clustering were employed. The dendrogram is not a single set of clusters but rather a multi-level hierarchy. This allows users to decide what scale or level of clustering is most appropriate in a particular application requiring a subjective choice. The discrete level of aggregation in the dendrogram (Fig. 1a) shows the merging of clusters of TRMM Philippine rainfall data. A careful analysis of dendrogram provides insights on the optimum number of clusters, as shown in Fig. 1b. The CCC is a measure of how faithfully the dendrogram represents the dissimilarities among observations. It is a measure of the goodness of fit of clustering [11]. The CCC from single linkage clustering has a value of 0.75 that suggests a good linear correlation between the cophenetic distances obtained from the dendrogram, and the original distances (or dissimilarities) used to construct the dendrogram, and also implies that the dendrogram in Fig. 1 is an appropriate summary of the rainfall data.

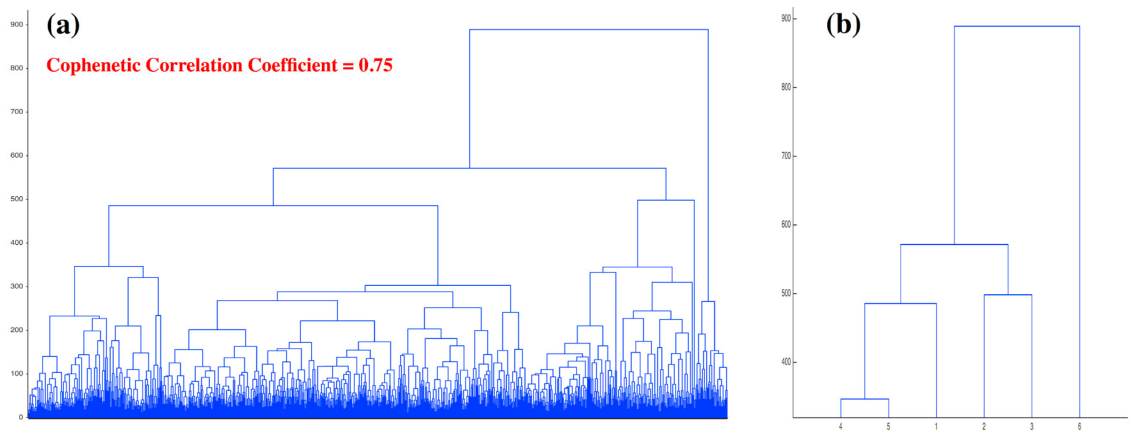

Figure 1: (a) The level of aggregation in the dendrogram from single-linkage cluster analysis of TRMM rainfall data, and the associated cophenetic correlation coefficient is 0.75. (b) A simplified dendrogram that does not display the very lowest levels of the tree but only the chosen number of clusters.

In attempting to assess the possible number of clusters, k, in the K-means method, the pre-determined value for k was sequentially set from 2 to 7. Plots of silhouette coefficient values for various numbers of clusters of TRMM rainfall data are in Fig. 2, and the sums and averages are in Table 1. The silhouette coefficients assist in interpreting and validating consistency within clusters of data objects. Figure 2 provides a clear representation of how well each data object lies within its cluster. It is also a useful tool to assess the overall goodness-of-fit for a given cluster number as the values can be used to compare quantitatively the clusters.

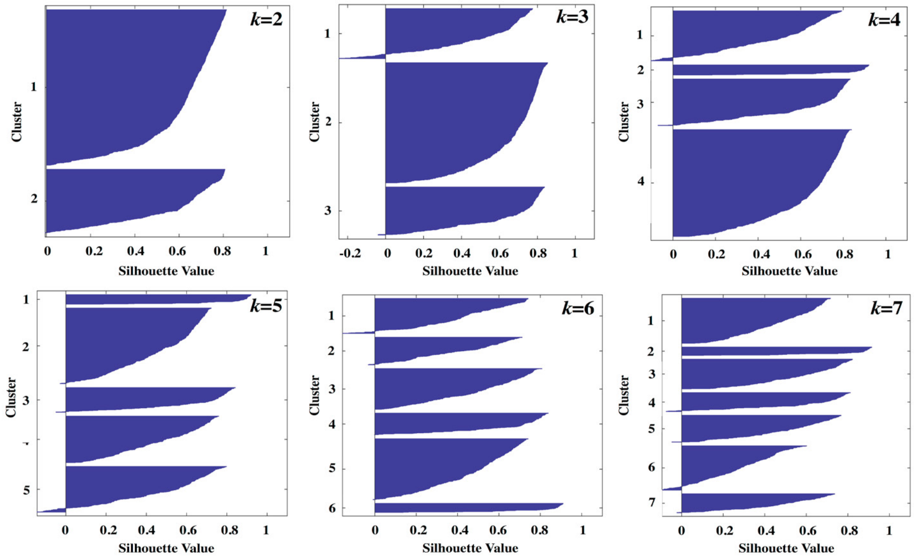

Figure 2: Silhouette plots for the number of clusters k = 2, 3, 4, 5, 6, 7.

	k=2	*k*=3	*k*=4	*k*=5	*k*=6	*k*=7
Sum of positive silhouette values	591.24	601.75	582.22	494.23	501.65	458.58
Average of positive silhouette values	0.6	0.63	0.6	0.52	0.52	0.48

Table 1. Sums and averages of positive silhouette coefficient values obtained from the K-means clustering of TRMM Philippine rainfall data.

The choice of the optimum number of clusters is made possible through the dendrogram, the CCC, the silhouette coefficient values, and coupled with well-known characteristics of Philippine rainfall, as influenced by meteorological factors. Although there are negative silhouette coefficient values and the average silhouette for *k*=6 is 0.52, 0.08 and 0.11 lower than the silhouette coefficients for *k* = (2, 4) and *k*= 3, which are 0.60 and 0.63, respectively. However, after considering Philippine rainfall characteristics in terms of known meteorological aspects, the optimum number of cluster is believed to be 6. These 6 clusters reflect the distinct rainfall patterns and variabilities observed in the Philippines. To acquire a more detailed knowledge of the rainfall patterns, especially the rainfall amount and the clear temporal features of each cluster, the very small number of negative silhouettes in Clusters 1 and 2 of *k*=6 were removed and re-assigned to other clusters. This process is identical with that of [9] and [10].

To explain the choice of *k*=6 instead of *k*=3 (which has the highest silhouette coefficient value), consider the eastern section of the Philippines. All of this region is influenced by the same wind systems and the months with rainfall maximum are predominantly during the boreal winter (northeast) monsoon regime. However, the complex distribution of terrain in the eastern section, the characteristics of air-masses that influence a particular cluster and the tracks of tropical cyclones during the boreal winter monsoon result in significant local rainfall variations of the annual cycle. The result is the distinct climate type of each cluster. Here, the eastern seaboard of the Philippines is categorized into 3 climate types (Clusters 1, 2, and 6; Fig. 5a). What makes each cluster unique is the discrete rainfall patterns, shown in the temporal amplitude of the cluster's annual cycle. The monthly rainfall variation is clearly exhibited in Figs. 3 and 6, and these figures are discussed in Section 3. The high-resolution TRMM rainfall data captures the detailed rainfall pattern reflecting the topographical characteristics of the Philippines. The wind system, terrain, and the types of air-masses that influence Clusters 1, 2, and 6 locations all interplay to produce the clusters distinct climate types. Most of Cluster 1 has relatively low elevation compared with Clusters 2 and 6. Hence, an orographic effect is not present causing its rainfall peak in December to be lower than in Clusters 2 and 6.

The section of Cluster 1 adjacent to Cluster 2 is on the leeward side of the mountains during the boreal summer and winter monsoons, producing a rain shadow effect. Clusters 2 and 6 are mostly mountainous and both are influenced by the orographic effect but they differ in air-mass characteristics that explain the rainfall difference during the boreal winter monsoon months. Cluster 2 during the boreal winter monsoon is greatly influenced by a warm moist air-mass originating from the lower latitudes of the western North Pacific. Cluster 6, however, is influenced during the boreal winter monsoon by the relatively colder air-mass being located in the relatively higher latitude than Cluster 2. Rainfall in the boreal summer months in Cluster 6 is higher than in Clusters 1 and 2, especially in July because of the tropical cyclone tracks that traverse the higher latitudes during that time of the year.

3. Results and Discussion

3.1 Cluster Analysis Results: Satellite-Based Rainfall Estimates vs. Surface-Based Rainfall Measurements

To validate the clustering results using the TRMM rainfall data, cluster analysis also was applied to the Philippine rain gauge data. Figures 3 and 4 show the climate type clusters obtained from TRMM rainfall estimates and surface-based measurements. The TRMM (Figs. 3) and the rain gauge datasets (Figs. 4) yield very similar rainfall patterns for the Philippines. This implies that TRMM satellite provides accurate estimates of the Philippine rainfall. As the TRMM network is so dense (Fig. 5a), TRMM climate type clusters show clearly the distinct rainfall patterns and variabilities important in defining climate types. There are 6 clusters and the corresponding station distributions are color-coded to signify cluster membership. The Philippine network of rain gauges consists of only 52 stations; such a sparse observation network failed to capture the rainfall variability in each climate type cluster (Fig. 5b).

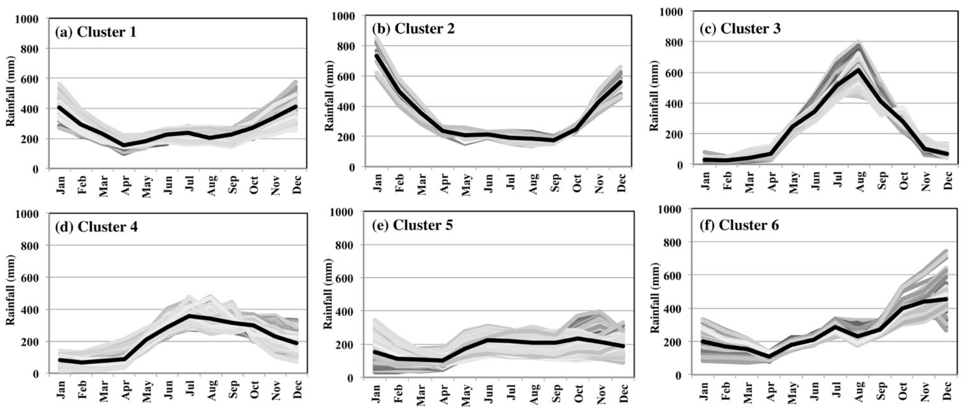

Figure 3: Climate clusters obtained from TRMM rainfall data classification.

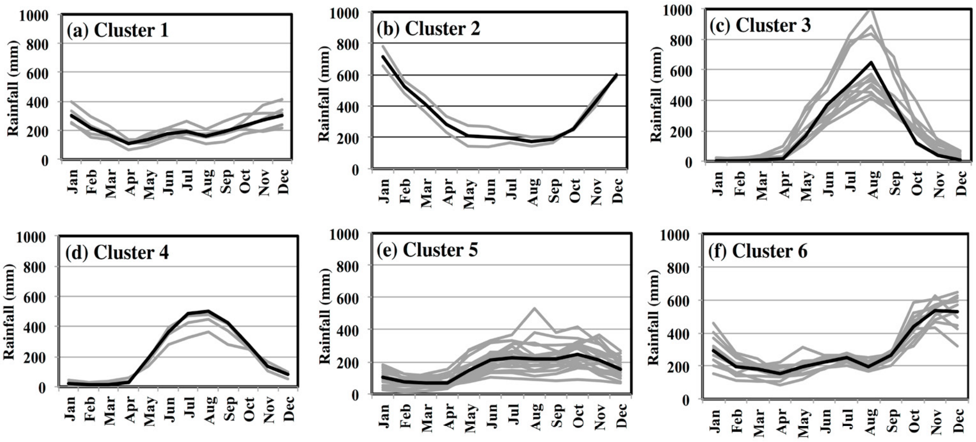

Figure 4: Climate clusters obtained from surface-based rainfall classification.

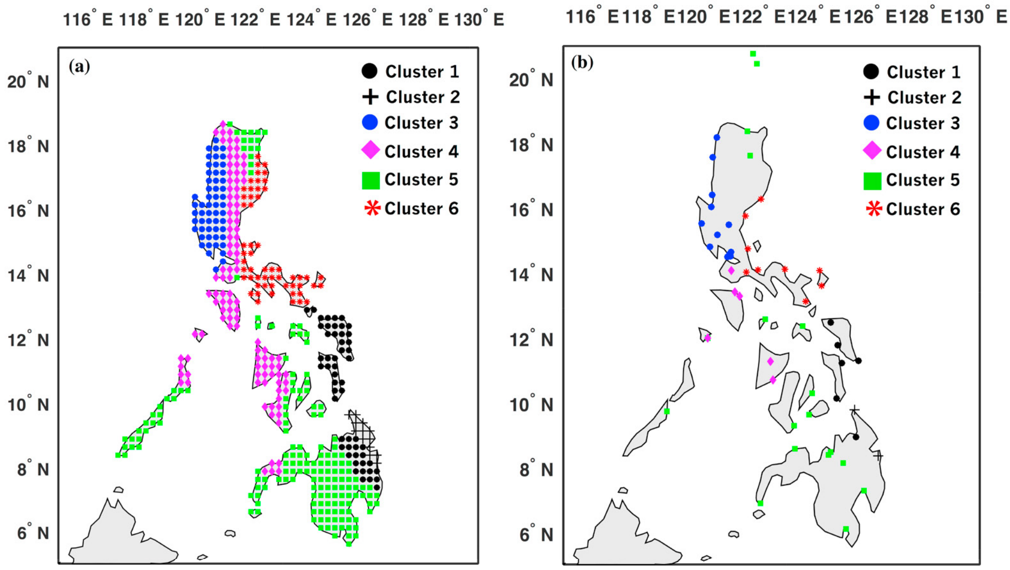

Figure 5: (a) The TRMM rainfall grid points covering the entire Philippine archipelago. (b) Network of Philippine rain gauges. Note that the rain gauge network is not representative of much of the Philippines.

3.2 Seasonal Rainfall Distribution

The more detailed rainfall characteristics from TRMM data clustering depicts both the distinct temporal features and more precise rainfall amounts revealed by the ensemble plots in Fig. 3. It is clear that the seasonal evolution of Philippine rainfall differs markedly between clusters, especially the variability in the rainfall totals. The disparities in rainfall seasonality between clusters are attributed to the variations in factors such as the prevailing wind systems, tropical cyclone activity, topography, thunderstorms, location of the ITCZ, and surface temperature that affect the rainfall amounts and patterns. Box plots of TRMM climate type clusters were generated to further emphasize the contrast in seasonal cycles of each cluster. The distribution characteristics of rainfall data in each cluster are depicted in Fig. 6, conveying the locations and variations in the monthly rainfall values. The lower and upper ends of the boxplots are the lowest and highest rainfall values, respectively, that signify the values outside the middle 50%. The central rectangle spans from the first quartile (25th percentile) to the third quartile (75th percentile) that also defines the interquartile range that represents the middle 50% of rainfall data. A segment inside the rectangle shows the median (red bar) marking rainfall data midpoints. Blue asterisks denote the mean, and the red plusses indicate monthly outliers.

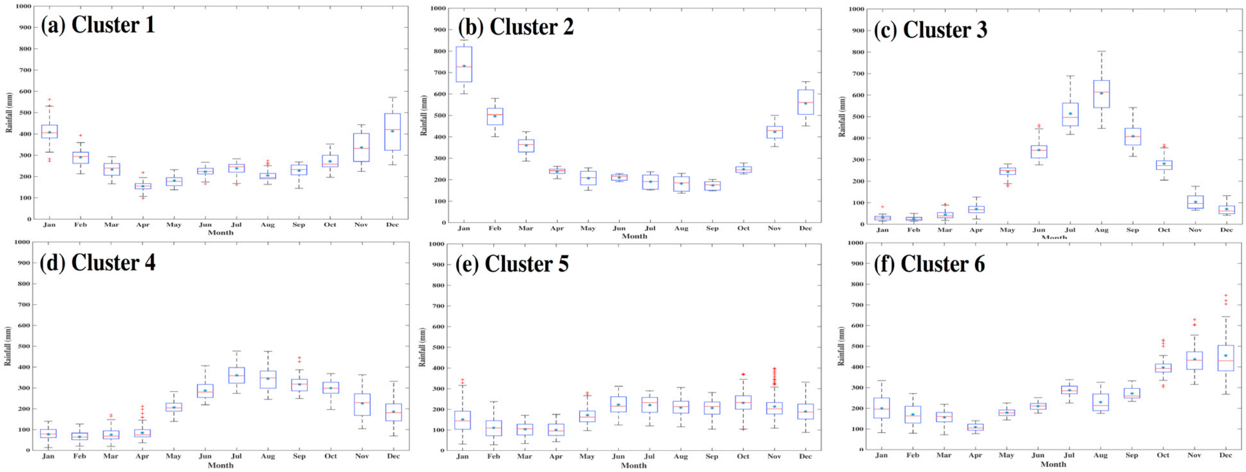

Figure 6: Monthly boxplots of Clusters 1-6 clearly showing the distinct rainfall amounts and temporal distributions of each climate.

There are 53 TRMM data points in cluster 1 (black dots in Fig. 5a), accounting for 12% of all TRMM rainfall grid points. Monthly boxplots in Fig. 6a show the amount and temporal rainfall variabilities in Cluster 1. Rainfall gradually increases from May, with about 180 mm as the median and average values, and continue to increase until July then slightly dips in August. Rainfall again increases in September and peaks in December, with over 400mm median and mean rainfall values. December has the heaviest rainfall with the lowest 25% ranging from 250-320mm, whereas the middle 50% ranges from 320-500mm, and the highest 25% can reach about 580mm. Rainfall starts to decline in January until it reaches its minimum in April with both median and mean rainfall values of 150mm; less than half the December mean and median. The lowest rainfall in April is around 100mm whereas the highest value can be 200mm. The rainfall outliers represented by red positive signs are present in January-February, April, and June-August.

From April to October, rainfall in Cluster 2 has a monotonic rainfall distribution and the minimum rainfall amount is experienced during this period, when the summer (southwest) monsoon is affecting the Philippines. Minimum rainfall with 150mm is observed in August and the lowest median and mean (both 180mm) are in September, the months coinciding with the peak of summer monsoon. Rainfall increases in November and reaches its highest in January, concurrent with the peak of winter (northeast) monsoon then the rainfall decreases steadily. The months with increased rainfall also coincide with the TCs following a low-latitude track that brings considerable amount of rainfall in the lower eastern seaboard of the Philippines. January is the wettest month with median and mean of 730 mm of rains. The middle 50% of January rainfall represents the 25-75 percentiles, ranging from 650mm to 820mm. Black plus signs in Figure 5a represent the smallest cluster with 13 data points, accounting for only 2.8% of all data points**.** Rainfall in Cluster 3 (55 blue dots in Fig. 5a, 12.2%) has a steady then a sharp increase commencing in May coinciding with the early onset of summer monsoon. Rainfall is highest in August, corresponding to the peak of summer monsoon and TC activity in the Philippines [12]. August has the highest rainfall variation ranging from 440mm to 800mm. The middle 50% ranges from 550mm to 680mm. November to December receive lower rainfalls, with medians of 100mm and 70mm, respectively. January to May are dry months with averages and medians below 50 mm. Cluster 4 (magenta in Fig. 5a) has relatively dry months from January to April with 70-90mm as rainfall medians and means, with a significant increase starting in May and continuing until July, which is also the wettest month, with 270mm minimum and 480mm maximum; the mean and median both are 350mm. Rainfall gradually and monotonically decrease from August until October, then drop further in November and December. Cluster 4 has 105 data points that comprise 23.4% of the total data points.

Clusters 5 is the biggest climate group in the Philippines (168 green data points in Fig. 5a, 37%) and has a flatter seasonal cycle as compared with other clusters and has resemblance with Cluster 4 but individually there are evident differences. The highest median is in July and October with 230mm, although June to October have comparatively almost the same 25 to 75 percentiles from 180mm to around 270mm. Rainfall in November starts to decrease very gradually until reaching its minimum in April with 100mm as the average and median rainfall. Most of the data points in cluster 5 are found in the inner section of the country. Locations under cluster 5 are neither influenced by the winter monsoon nor by the summer monsoon. Local meteorological systems contribute to its fairly wet rainfall months. Cluster 6 (red asterisks in Fig. 5a) is consists of 55 TRMM data points and accounts for 12.2% of the total TRMM data points. It has 2 maxima, in July and in December, although December is the wettest month of Cluster 6 and has the biggest rainfall variation from 280mm to 650mm. A sharp decrease in rainfall occurs in January with a median and mean of about 200mm. April is driest, with mean and median of only 110mm.

The rainfall distributions in Clusters 1 to 6, as shown in Fig. 3, all differ from each other, which supports the conclusion that there are 6 distinct climate zones in the Philippines, not 4 as has long been accepted. Figure 6 shows the monthly boxplots of each cluster that further emphasize the divergent characteristics of each of the clusters.

4. Conclusions

It has long been accepted that the Philippines has four climate zones based on the Modified Coronas Classification (MCC) in the 1920s that employed rainfall amounts and distributions, as near surface temperatures are relatively uniform over the Philippines. This study also adopts a rainfall based classification, but differs from the MCC approach because it examines more closely the annual evolution of the rainfall totals and patterns, by employing high resolution TRMM satellite rainfall data and using a combination of cluster analysis techniques, together with the known meteorological aspects of the region. Philippine rainfall is dependent on shifts in major wind patterns, most notably the summer and winter monsoons, and on the complex topography of the Philippines. The TRMM high-resolution rainfall data has enabled a much clearer definition of the Philippines rainfall variability, which is vital in identifying distinct climate types and accurately capturing the Philippines rainfall patterns. The small number of rain gauges prevents the capturing of the Philippines rainfall characteristics, especially those affected by varying topographical features.

The cluster analysis techniques are not new, together with the use of dendrograms, cophenetic correlation coefficients, the well-known K-means approach, and supported by silhouette coefficient values. However, the critical meteorological knowledge of the rainfall totals and distributions rely on recent studies of Philippines climatology by [12]. Recent studies by the authors [13] and others have revealed subtleties resulting from the interplay between the monsoonal patterns and the phases of ENSO. These interactions often are seasonal (quarter-yearly), or longer, influences on the major rain-bearing systems, particularly the tracks of tropical cyclones (TCs). For example, in El Niño years, TCs frequently recurve or decay before reaching the Philippine region, producing *below normal* numbers and landfalls in less active season (LAS) and more active season (MAS) of TC activity. In La Niña years, TC numbers and landfalls are *below normal* in January-March and July-September, but *above normal* in April-June and October-December.

Consequently, the Philippine climate zone classification was found to be more complex than suggested by the MCC, with a total of 6 distinct climate zones identified using the combination of cluster analysis and the known, observed, features of the Philippine climatology. The increase to 6 climate zones, compared with 4 climate zones identified by the MCC, is strongly supported by the findings of this study.

Acknowledgements

We acknowledge the mission scientists and principal investigators of NASA GES DISC who provided the 3B43 data used in this research effort.

References

[1] Nasseri, M., and B. Zahraie. (2011) Application of simple clustering on space-time mapping of mean monthly rainfall pattern. *Int. J. Climatol*. **31**: 732–741, DOI: 10.1002/joc.2109.

[2] Sonmez, I., and A. Komuscu. (2011) Reclassification of rainfall regions of Turkey by K-means methodology and their temporal variability in relation to North Atlantic Oscillation (NAO). *Theor. Appl Climatol*. **106**, doi:10.1007/s00704-011-0449-1.

[3] Marzban, C., and S. Sandgathe (2008) Cluster analysis for object-oriented verification of fields: A variation. *Mon. Wea. Rev*., **136**, 1013–1025, doi:10.1175/2007MWR1994.1.

[4] Lyra G.B., J. F. Oliveira-Júnior, M. Zeri (2014) Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil. *Int J Climatol* **34**:3546–3558. doi:10.1002/joc.3926E.

[5] Mooi and M. Sarstedt (2011) A concise guide to market research: The process, data, and methods using IBM SPSS statistics, Springer.

[6] Hair J.F., R. Tatham, W. C. Black (1992) Multivariate Data Analysis With Readings, Macmillan, New York.

[7] Milligan, G. W. (1980) An examination of the effect of six types of error perturbation on fifteen clustering algorithms, *Psychometrika*, **45**, 325-342.

[8] Punji, G., and D. W. Stewart (1983) Cluster analysis in marketing research: Review and suggestions for application. *J. Marketing Res*., **20**, 134-148.

[9] Corporal-Lodangco, I. L., L. M. Leslie (2014) Cluster Analysis of North Atlantic Tropical Cyclone. *Procedia Computer Science* **36**: 293-300. doi:10.1016/j.procs.2014.09.096.

[10] Corporal-Lodangco, I.L., L. M. Leslie (2016) Cluster Analysis of Philippine Tropical Cyclone Climatology: Applications to Forecasting. *J. Climatol. & Wea. Forecasting* **4**: 152. doi:10.4172/2332-2594.1000152.

[11] Sokal, R. R. and F. J. Rohlf (1962) The comparison of dendrograms by objective methods. *Taxon*, **11**:33-40. 10.2307/1217208.

[12] Corporal-Lodangco, I., L., L. M. Leslie (2016) Climatology of Philippine tropical cyclone activity: 1945-2011. *Int. J. of Climatol*. doi:10.1002/joc.4931.

[13] Corporal-Lodangco, I.L., L. M. Leslie, P. J. Lamb (2016) Impacts of ENSO on Philippine tropical cyclone activity. *J. Climate* **29**: 1877-1897.