Defining climate zone of Borneo based on cluster analysis

Zulfaqar Sa’adi1 · Shamsuddin Shahid2 · Mohammed Sanusi Shiru3,4

Received: 7 January 2021 / Accepted: 14 June 2021 / Published online: 9 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract
Although Borneo Island is one of the most vulnerable tropical regions to climate change, maps depicting the local climate conditions employing climate classification are still not well defined. The present study attempted regional climate classification to divide the Borneo region into several homogenous groups based on long-term average climate behavior. Daily gridded rainfall and temperature (Tavg, Tmax, and Tmin) data at 0.25° resolution spanning 56 years (1960–2016) was used. The classification was done using non-hierarchical k-mean and several hierarchical methods, namely, Single, Complete, McQuitty, Average, Centroid, Median, and two algorithms of Ward’s method, wardD, and wardD2. The results showed that k-mean, wardD, and wardD2 were able to classify the climate of Borneo into four zones, namely ‘dry and hot’ (DH), ‘wet and hot’ (WH), ‘wet’ (W), and ‘wet and cold’ (WC) with a considerable difference at the boundaries. Spatial relevancy, stability, and variability of the clusters based on correlation and compromise programming showed that the wardD method was the most likely to yield acceptable results with optimum 4-cluster to partition the area into four principal climate zones. The constructed cluster plot, centroid plot, and probability distribution function (PDF) showed a distinct climatic characteristic between the climate zones in terms of rainfall, temperature, and seasonality. The proposed climate zonation for Borneo can help in better understanding climate regionality and climate-related development planning.

1 Introduction
Climate classification has two primary functions: identification, organization, and orderly naming of climatic types for formulating relationships with climatic variables and support policy-makers in decision-making on climate-related socio-economic planning (Belda et al. 2014). Classifying climate is crucial for the Borneo region as its climate is spatiotemporally highly heterogeneous. The mass population of Borneo depends on an agrarian economy. Climate classification can be useful to differentiate the local climate influence on various agriculture types. The crop selection for a certain local climate characteristic significantly impacts the yield and the local economies (Paterson et al. 2015; Ahmed et al. 2016). Besides, understanding the local climate characteristics based on certain classification can help identify flood-prone areas, a common phenomenon in the region (Yusuf and Francisco 2009; Sa’adi et al. 2017). The less rainfall and high-temperature areas susceptible to fire during dry periods can also be identified. Climate classification can also help biological and conservation planning and activities in Borneo, which has one of the highest concentrations of species per unit area in the world (Nakagawa et al. 2007). As small variability in climate can significantly impact biological diversity, the areas that cater to conservation purposes can be better implemented (Sa’adi et al. 2020). Therefore, establishing a specific scheme of climatic divisions will be in the best interest of various policy-makers and stakeholders in Borneo.

Previous climate classification studies in Borneo were mostly based on political boundaries (Aldrian 2001; Hamada et al. 2002; Aldrian and Dwi Susanto 2003), which is not suitable in many circumstances. Therefore, several classification schemes based on various approaches have been
investigated (Köppen 1900; Thornthwaite 1948; Trewartha and Horn 1980; Oliver 2005; Peel et al. 2007; Alvares et al. 2013; Rahimi et al. 2013; Llanos-Herrera 2014; Karki et al. 2016). The most famous examples are Köppen (Köppen 1900), Köppen-Geiger (Peel et al. 2007), Thornthwaite (Thornthwaite 1948), and Köppen–Trewartha (Trewartha and Horn 1980) classifications. The classifications involve the a priori definition of a set of climate types or rules applied in climate classification (Fovell and Fovell 1993).

In those studies, the climatic types were externally specified or indirectly suggested by the data instead of directly issuing them (Fovell and Fovell 1993). These approaches have an essential advantage of directly and quantitatively specifying the type of climate. Nonetheless, there is a disadvantage of the classification rules as they were subjectively formulated. To reduce the subjectivity, Richman and Lamb (1985) propose a regionalization of climate by adopting variable manipulation strategies through principal component analysis (PCA). They found a pattern that emphasizes an anomaly that was relatively strong in different parts of the Central United States, yielding a regionalization of the rainfall-based domain.

There have been several quantitative attempts to classify the climate zones of Borneo. However, there is a constraint in earlier works as those were conducted within a political boundary. Consequently, the studies were concentrated either within the Kalimantan part of Indonesia or the States of Sabah and Sarawak in Malaysia and Brunei (Aldrian 2001; Hamada et al. 2002; Aldrian and Dwi Susanto 2003; Dambul and Jones 2007). Indonesia was climatologically classified into four regions by Hamada et al. (2002). They found climate characteristics were similar in three regions, as previously shown by Aldrian (2001); however, there is an intermediate region with undefined rainy and dry seasons. In another study, Aldrian and Dwi Susanto (2003) classified Indonesia’s climate based on the annual mean rainfall variability using a double correlation method. Based on their result, Borneo has two distinct rainfall regions, classified as A (strong influences of monsoon) and B (associated with inter-tropical convergence zone (ITCZ)) but with high spatial variability. Dambul and Jones (2007) made the latest classification for Borneo, which classified Borneo’s climate zones into six spatial sub-groups and eight weather (temporal) types based on the regional and temporal climatic scheme using k-mean cluster analysis.

Some of the results from these studies are contradictory due to limited climate data and subjectivity in selecting preferable climate indices. There was also a constraint due to the usage of political boundaries as a basis. Some studies rely on one climate parameter (rainfall) only, regardless of the importance of other climate parameters (temperature), which are essential to determine the local climatic pattern. Meanwhile, some studies made climate classification distinctly based on spatial and temporal patterns. Besides, all the previous studies used a limited number of observation data, which often may lead to a vague classification of Borneo’s highly complex geographical region.

Homogeneous climate zones can be analyzed with a clustering method based on meteorological parameters’ (Fovell and Fovell 1993; DeGaetano 1996; Mimmack et al. 2001; Unal et al. 2003). Unlike the commonly used statistical methods, cluster analysis does not have its basis on some theoretical a priori assumed distribution. Both non-hierarchical and hierarchical methods can be used for clustering. The most widely used non-hierarchical clustering method is k-mean and k-medoids, where the clustering algorithm is used for dataset classification into k number of clusters. This approach has the advantages of speed and simplicity, and there is a possibility of the movement of subjects from one cluster to another. The resulting clusters’ dependency on the initial random assignments gives the approach a disadvantage of not yielding the same result with each run due to the initiation position. The hierarchical clustering method is another method where the clustered was mapped into a hierarchy, reflecting inter-cluster similarities or dissimilarities. It can be either top-down or bottom-up. Clusters are built through repetitive joining and merging of the objects with the shortest distance separation. The distance matrix is updated after merging the closest two points, and the process is repeated until the joining of all objects.

Hierarchical agglomerative methods are used more often than hierarchical divisive methods. Several hierarchical agglomerative methods have been developed over the years, such as Single, Average, MacQuitty, Complete, Median, Centroid, and Ward’s methods. In hierarchical agglomerative methods, each subject starts from its separate cluster. Principally, there is a repeated combination of the two ‘closest’ clusters until all subjects are in a single cluster. The optimum number of clusters is finally chosen from all cluster solutions. The result of hierarchical agglomerative clustering is visualized as a dendrogram.

Ward’s method has been the clustering method commonly employed in climate research (Kalkstein et al. 1987). All variables’ means within each cluster are calculated in the method, then all variables’ means within each cluster are calculated. Then, the distance to the cluster mean of each case is calculated, and in all cases, sums are finally taken. Ward’s method tends to form clusters of relatively small and equal objects’ numbers. This gives the notion that climate zone classification should have approximately the same number of stations (Kalkstein et al. 1987). There has also been a report of the inferior performance of Single, Complete, and Ward’s method under these error perturbations (Milligan 1980). A spatiotemporal homogeneous dataset could therefore be employed to address this drawback.
Understanding the strengths and limitations of various clustering methods is crucial in making an informed decision on which method is most applicable for a particular study. However, no consensus on determining the most appropriate clustering approach to classify climate zone has been reached. Unal et al. (2003) used correlations between the regionalized stations and cluster averages to measure climate parameters’ stability. They found Ward’s method is acceptable to define the climate zone of Turkey. However, their interpretation of the correlation results is vague as the decision is made visually based on correlation. Therefore, group decision-making (GDM) has been applied where the clustering method is given a position based on the frequency of the rank, based correlation, obtained at different grid points. The cluster method’s ranking was made based on the ability of the grid points to simulate the mean climate parameter of the respective climate zone.

This study intends to define Borneo’s spatial and homogeneous climate zone using several forms of non-hierarchical and hierarchical cluster analysis methods. Clustering was done by employing daily gridded-based Princeton (Sheffield et al. 2006) climate datasets of rainfall and temperature (Tavg, Tmax, and Tmin). Though there is flexibility in the methods applied in the previous studies, we have a viewpoint that significant improvement is possible using fine spatiotemporal resolution gridded-based climate data. These datasets have been applied in various climate studies (Khan et al. 2018, 2019; Nashwan and Shahid 2019; Pour et al. 2020). The best clustering method was selected based on its spatial relevancy and correlation ranking by using GDM. The chosen method was then used to define the climate zone of Borneo. The climate zone map developed in this study can help policy-makers and stakeholders in crop selection based on local climate characteristics, identify flood and drought susceptible areas, and biological conservation planning. This study utilizes daily gridded-based Princeton (Sheffield et al. 2006) climate datasets to define Borneo’s spatial and homogeneous climate zone as described in Section 2. The major large-scale climate variability affecting the local climate in Borneo is discussed in Section 3. The methodology involved is described in Section 4. Section 5 gives details on the major findings of this study, discusses possible explanations for the results, and Section 6 summarizes the results and suggests future work.

2 Study area and data

This study covers Borneo island (743,330 km\(^2\) area), the third-largest island in the world. It is located on the south of the South China Sea (SCS) within the geographical domain of 4° 17’ S and 7° 20’ N latitude and 108° 40’ E and 119° 5’ E longitude, as shown in Fig. 1. Borneo is politically divided among the Malaysian States of Sarawak and Sabah, Brunei, and Kalimantan of Indonesia. Borneo is surrounded by several seas, namely, Java Sea, SCS, Celebes Sea, Sulu Sea, Karimata Strait, and Makassar Strait. Land and water interaction characterizes Borneo’s climate causing local diurnal

![Fig. 1 Elevation map, administrative division, and 961 grid point and spatial variation of the mean annual rainfall for the period 1960–2016 based on Princeton dataset over Borneo](image)
variations in winds, temperature, rainfall, and clouds and contributing to inter-annual variations and intra-seasonal oscillations.

Peel et al. (2007) classified most of Borneo as Af (tropical rainforest climate), and the mountainous region in the central part as CFa (humid subtropical climate) and Dwa (monsoon-influenced hot-summer humid continental climate). The temperature in Borneo is uniform all year round, humidity is high, and rainfalls are copiously ranging from 1700 mm near the coastland (at the south and eastern region) to 4100 mm at inland areas (interior region and southwestern coast) (Fig. 1) (Peel et al. 2007; Tangang et al. 2013). The temperature (Tavg, Tmax, Tmin) distribution is also spatially heterogeneous with high temperature along the coastline in the west, south, and east, while lower temperature in the interior’s mountainous ridges and northern region (Fig. 2). Borneo’s climate is influenced by two monsoon regimes: the SW monsoon and the NE monsoon (Chenoli et al. 2018). The spatiotemporal pattern of mean monthly and monsoonal patterns of rainfall and temperature in Borneo are shown in Fig. 3. Geographical location, local littoral circulation and sea currents, orographic influence, and atmospheric circulation characteristics mark the climatological differences in the area.

In defining climatic types and delineating similar climatic zones, many climatological studies have applied various kinds of data. In cluster analysis, the choice of appropriate data is an initial consideration. The most readily available variables playing important roles in local climate modulation are long-term rainfall and temperature data (Fovell and Fovell 1993). Previous studies also defined climate classification based on temperature and rainfall (Rubel and Kottek 2010). Major climate variations in Borneo are influenced by spatiotemporal convective activity and rainfall (Hamada et al. 2002). Besides, with a vast and complex mountainous and hilly topography across the island, the temperature difference is also high. In addition, land and sea interaction brought about the importance of diurnal patterns in regulating the local climate along the shoreline. The slight difference in temperature in a climate-sensitive tropical region of Borneo also emphasizes the need to incorporate all temperature variables, namely, average temperature (Tavg), minimum temperature (Tmin), and maximum temperature (Tmax) as inputs for cluster analysis. These three temperature indices are also important for understanding the diurnal pattern and the extent of certain temperature thresholds in certain climate zones. Therefore, rainfall, Tavg, Tmin, and Tmax are considered to develop Borneo’s homogeneous climate clusters.

Various large-scale climate phenomena influenced the climate in Borneo. It was also inferred that there is high spatial variation over small distances in the region’s climate, which required a high spatial resolution of climate datasets (Chang et al. 2005). Therefore, long-term climate data is needed to improve the resolution of spatial variability over the region. Several studies employed gridded data for climate studies in the areas where reliable long-term climate records are not available (Khan et al. 2018; Nashwan and Shahid 2019; Pour et al. 2020; Ayoub et al. 2020; Iqbal et al. 2021). Concurrently, a better spatial and temporal resolution of climatic data enables a better understanding of the causal links of climate zone in Borneo to larger-scale climate features. In this study, daily gridded-based Princeton (Sheffield et al. 2006) climate datasets were retrieved to tailor the demand for finer resolution of spatial climate information. The Princeton climate datasets provide sufficient spatial resolution of 0.25° (latitude) × 0.25° (longitude), covering 961 grids across Borneo to capture the regional climate. Besides, it also provides the required input climate variables, namely, rainfall and temperature (Tmax, Tmin, Tavg) needed for climate
Defining climate zone of Borneo based on cluster analysis classification in this study. It also caters to a longer historical period, spanning 56 years from 1960 to 2016 (Fig. 1). More details of this dataset can be found at https://rda.ucar.edu/datasets/ds314.0/#description. The climate datasets were initially transformed into annual and monthly series at each grid point before their use.

3 Climate variability over Borneo

The climate in Borneo can be inferred from the published papers on the climate of the Maritime Continent, specifically on the Bornean part of Malaysia and Indonesia (Chang et al. 2005; Tangang et al. 2008, 2012; Juneng and Tangang 2010; Supari et al. 2018; Xavier et al. 2020). The maritime continent monsoon strongly dominates Borneo’s local climate, which can be characterized by a rainy monsoon with predominantly northeasterly winds during boreal winter (DJF) and a relatively drier monsoon with predominantly southwesterly winds during boreal summer (JJA) (Chang et al. 2005). Due to the sheltering effect of the topography associated with a lee or rain-shadow effect, the ridge’s southern site does not experience a sudden increase in rainfall (Chang et al. 2005).

At a regional scale, the climate of Borneo is influenced by two monsoon regimes, namely, the SW monsoon and the NE monsoon (Tangang et al. 2012). The monthly and monsoonal temporal patterns of rainfall and temperature in Borneo are shown in Figs. 2 and 3. The NE monsoon, dominated by northeasterly winds that cross over the SCS, prevails from November to March and brings a more conspicuous sudden increment in rainfall (Tangang et al. 2012; Meteorological Department 2016). The monsoon becomes stronger at the later stage due to the Borneo vortex and cold surge (Tangang et al. 2012). These two major synoptic circulations, namely...
the cold surge (4–5 days) and the Borneo vortex, are linked to the major extreme rainfall event, resulting in floods, particularly in northwestern Borneo (Tangang et al. 2012; Xavier et al. 2020). Without the Borneo vortex’s presence, strong convective activities still largely occur over Borneo, especially in the northwestern region (Chang et al. 2005). In some instances, the interaction between SCS-type cold surge and Borneo vortex could result in the Borneo vortex’s strengthening and bring with it an excessive amount of rainfall. However, in February, a slight decrease of rainfall happens due to the northward equinoct when the subsolar point leaves the Southern Hemisphere and crosses the equator, heading northward. The sun directly overheads the equator at noon during the northward equinoct, causing an increase in temperature and subsequently lower rainfall. The equinoct can also be associated with the ITCZ, a twice-a-year crossing over the equator following the sun. The ITCZ may lead to another peak of monthly rainfall pattern (Mar–May) due to terrain-lifted rainfall at Borneo’s central-mountainous region along the equatorial trough (Chang et al. 2005). The SW monsoon, characterized by low-level southwesterly winds, occurs between May and September and has linkage to a comparatively dry period (Dindang et al. 2013; Diong et al. 2015). During the inter-monsoon months of April and October, there is high rainfall spatial variability due to a 10 km scale or less locally-driven convective activity (Joseph et al. 2008; Tangang et al. 2017). Although the temperature was uniform throughout the year, the monthly temporal pattern of temperature also follows the influence of the monsoon.

Apart from the annual and monsoonal cycle, a considerable inter-annual signature is associated with the El Niño-Southern Oscillation (ENSO). Depending on the occurrences of other climate influences, monsoon, location, and the event’s intensity, the ENSO relationship was generally coherent with Borneo’s rainfall variations (Juneng and Tangang 2010; Tangang et al. 2017). Aldrian and Dwi Susanto (2003) reported that rainfall significantly correlated to ENSO in southern Borneo compare to other regions. Tangang (2012) reported persisting drier-than-normal conditions in Borneo’s southern region, causing surface temperature to soar and often induce large-scale and uncontrollable forest fires. On the contrary, enhancing rainfall during a La Niña event may lead to flood events and landslides (Tangang et al. 2012). However, ENSO influences are not spatially linear due to the influence of cold surge and Borneo vortex at the end of the NE monsoon. This causes a prevalent rainfall even during the El Niño year at the western part of the island, while the northeast experiences drier conditions creating a dipole pattern (Tangang et al. 2017; Supari et al. 2018). Cold surge, Borneo vortex, and other large-scale climate influences might cause ENSO events to affect differently over different regions of Borneo.

On the intraseasonal time scale, Borneo climate variability is also modulated by the Indian Ocean Dipole (IOD), which can last for about 5 to 6 months, usually from July to November (Tangang et al. 2012). During a negative IOD, the maritime continent experiences deficit rainfalls (Vinayachandran et al. 2009; Tangang et al. 2012). On the contrary, high rainfall happens during a positive IOD. However, the extent to which the Borneo climate is affected by the IOD is still not clear. Another phenomenon that occurs on the intra-seasonal time scale and can significantly influence Borneo’s regional climate is the Madden–Julian Oscillation (MJO) (20–90 days) which is most active in the boreal winter (Tangang et al. 2012; Xavier et al. 2020). MJO generally enhanced the large-scale and local-scale rainfall over the west and northwestern Borneo peaked in Nov–Dec and drier conditions over northern Borneo as MJO suppressed February–March (Xavier et al. 2020).

At the sub-daily scales, convective rainfall activities over Borneo usually attain maximum intensity in the evening with considerable seasonal and spatial variation (Tangang et al. 2012). Although there is no sufficient information on the extent, trend, and spatiotemporal distribution of the sub-daily rainfall, the diurnally forced land-sea breeze may interact with larger-scale background wind to produce intensive convective activities leading to extreme rainfall (Tangang et al. 2012). The relatively dry subtropical air also could cool the warm tropical ocean surface in the SCS through enhanced evaporation from enhanced ocean to atmosphere latent heat flux and possibly due to the ocean mixing owing to strong surface winds may contribute to enhanced convection and rainfall (Xavier et al. 2020).

4 Methods
4.1 Non-hierarchical and hierarchical cluster methods

In this study, non-hierarchical $k$-mean cluster method and several hierarchical agglomerative cluster methods, namely Single, Complete, McQuitty, Average, Centroid, Median, and two algorithms of Ward’s (wardD and wardD2) were used to assign climate data (rainfall, $T_{avg}$, $T_{max}$, $T_{min}$) into different clusters according to their similarity in the output representation. Annual series of each climate parameter for each of the 961 grid points over Borneo was used for cluster analysis. Based on the earlier assessment of the optimum number of clusters required and visualization of the cluster formation, the best cluster method was selected and ranked using GDM to determine the best methodologies going forward. After the climate zones were defined, the temporal analysis of the monthly
and monsoonal patterns for each zone was done for further assessment and classification process.

**K-mean** method is a commonly used non-hierarchical clustering method. In this method, the number of desired clusters is specified in advance, and the ‘best’ solution is chosen. Initially, cluster centers are selected. Then, each subject is assigned to its ‘nearest’ cluster, defined based on the distance to the centroid. After that, the centroids are determined based on the formed clusters. The distance is re-calculated from each subject to each centroid, and observations that are not the closest are moved until the centroids remain relatively stable. **K-mean** is an agglomerative method which means once the cluster is formed, there is no possibility of splitting. Therefore, there is a need for randomization and repetition of the initial point of clusters, allowing for the derivation of stable final clusters. The method is also sensitive to the initial choice of cluster centers. The non-hierarchical cluster method is suitable when large climate datasets are involved.

The different hierarchical agglomerative methods differently determine how the clusters to be joined at each stage. The simplest hierarchical agglomerative clustering method is the **Single** method (Massaro 2014). In this method, the distance between two clusters is defined as the distance between the two closest members or neighbors. There is a comparison of the distance between all the entries in each cluster. Even though there are simplicity and straightforwardness, the cluster structure is not considered, leading to a chaining problem in which clusters become long and straggly. However, it is superior to other methods when the natural clusters are not elliptical or spherical, and it is insensitive to ties. The **Complete** method (Sørensen 1948; Lance and Williams 1967) was developed to overcome the **Single** method chaining problem by producing compact clusters of similar size. The distance is defined as the maximum distance between members. Even though this method’s chaining problem is averted, it is not robust for a priori known cluster. Cluster structure is also not being taken into account, and the method is also sensitive to outliers.

The **McQuitty** method (Mcquitty 1968) builds a dendrogram reflecting the structure present in a pairwise distance matrix. The nearest two clusters are combined into a higher-level cluster at each step. The arithmetic mean of the distances between the cluster and the higher-level cluster is simply its distance to another cluster. In the **McQuitty** method, the average of the cluster’s distances is taken, not considering the number of cluster points. This method has received much fewer applications due to the disregarding of the cluster sizes in calculating the average distances. This implies smaller clusters receiving larger weights in the clustering process. However, according to Corliss et al. (1974), the **McQuitty** method is suitable for clusters in similar sizes, as it can make a decision based on a known *a priori* to eliminate size differences between the resulting groups.

Most of the other hierarchical methods’ shortcomings are overcome by the **Average** method (Lance and Williams 1967). In this method, the cluster’s distances’ average is taken between all subjects’ pairs in the two clusters whilst compensating for the points numbers in that cluster. Unlike the **Single** and **Centroid** methods, there is a slight tendency of the **Average** method to form chains. Also, unlike Ward’s technique, in which the sum of squared distances is minimized within-cluster, there is a minimization of within-group variance and maximization of between-group variance. In climatological research, it gives the most realistic results (Kalkstein et al. 1987).

Among the widely used hierarchical clustering methods, the **Centroid** method is highly robust, but it also experiences the chaining problem (Milligan 1980). It has been shown when applied in climatic research; the **Centroid** method can produce one larger cluster and many small clusters (Kalkstein et al. 1987). The **median** method (Gower 1967) has the same relationship with the **Centroid** method. In this method, the new centroid’s computation is made after a fusion, disregarding the number of objects present in the two clusters in question. Then, the new centroid is obtainable as the simple average of the two former centroids. Small clusters are being attributed to greater weight in this method. The **median** method is most suitable for situations in which the restriction that the operation of averaging data can be meaningful. The inter-cluster mid-point and median point are measured for the **Centroid** and **Median** method, respectively. Meanwhile, the increase in the error sum of squares (ESS) after two clusters fusing is calculated in Ward’s method. To minimize ESS increase, successive clustering steps are chosen. Literature showed two different algorithms for the Ward clustering. The first, which uses option ‘ward.D’ does not implement Ward’s clustering criterion, whereas the second, which uses option ‘ward.D2’ implements the criterion (Murtagh and Legendre 2014). The dissimilarities are squared before cluster updating in the latter.

### 4.2 Euclidean distance

Euclidean distance (the square root of the sum of the squared distances over all variables) is the distance metric commonly applied in atmospheric sciences, including cluster analysis to determine the distance between two observations. It measures the shortest distance between two points. The dissimilarity measure between the clusters applied in this study is the squared Euclidean distance between cluster means. Standardization of each variable is done before distance calculation since observations such as rainfall and temperature with different scales may contribute unequally to the distance calculated. Euclidean distance is defined as:
\[ d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \]  

where \( x_{ij} \) is the Euclidean distance. The Euclidean distance between two objects \( i \) and \( j \) in the \( n \times p \) data matrix \( X \) is simply the squared difference between them for each of the \( p \) variables, summed over the variables (Fovell and Fovell 1993).

### 4.3 Number of clusters

Defining the optimum cluster numbers through the cluster’s termination at a certain step is one of the main issues of cluster analysis. In this study, D index (Charrad et al. 2014) based on clustering gain on intra-cluster inertia was used. Intra-cluster inertia measures the degree of homogeneity between the data associated with a cluster. It calculates their distances compared to the reference point representing the cluster centroid. The optimal cluster configuration can be identified by the sharp knee that corresponds to a significant decrease in the first differences of clustering gain versus the number of clusters. This knee or great jump of gain values can be identified by a significant peak in the second differences of clustering gain. The mapping of the output in terms of climate zoning from each clustering method was used to determine the spatial relevancy of the cluster numbers. Cluster plots, centroid plots, and PDF were also constructed to confirm the different groups of the climate zone.

### 4.4 Ranking of clustering methods

The ranking of the clustering method in estimating and reflecting its similarity with the respective climate zone by several grid points is challenging. This was due to the possibility of clustering methods showing various accuracy degrees at different locations. In overcoming this, information aggregation methods, including a majority of ranks, mean ranking, and frequency of occurrence, were useful (Ahmed et al. 2019; Muhammad et al. 2019). Information from various sources is integrated into the methodology to assist in the process of decision-making. In this study, GDM was used for empirical models’ ranking. The ranking procedure is outlined as follows.

1. Initially, the cluster methods were rank using each climate parameter’s correlation at each grid point with their mean of the respective climate zone. The highest correlation was ranked 1st.
2. The occurrence frequency of each clustering method of getting a certain rank at all grid points was estimated through a \( 4 \times 3 \) matrix.
3. The rank positions were given weight as the inverse of the rank \( (w_r = \text{rank}^{-1}) \).
4. The occurrence frequency of a model at a certain rank, obtained in step 2, was multiplied by the weight of the rank obtained in step 3.
5. The overall score of each clustering method \( (W_m) \) was estimated by adding the output of step 4 using Eq. 2. The empirical models were ranked according to the calculated overall weight, where the highest weighted model was ranked top (1st position).

\[ W_m = F_1(w_{r1}) + F_2(w_{r2}) + F_3(w_{r3}) + \cdots + F_n(w_{rn}) \]  

### 4.5 Seasonality index

Each zone’s climate is classified by its relative seasonality index \((SI)\) based on rainfall amounts. The climate zones relative seasonality index is using the following equation (Abaje et al. 2010; Guhathakurta and Saji 2013; M. K. Patil 2015):

\[ SI = \frac{1}{R} \sum_{n=1}^{12} \left| \frac{x_n}{R} - \frac{\bar{R}}{12} \right| \]  

where \( x_n \) is the average rainfall for the month \( n \) and \( \bar{R} \) is the average annual rainfall. The index value varies from 0 (equal rainfall for almost all months) to 1.83 (most rain occurring in a single month).

Rainfall regime was divided into seven classes based on seasonality index defined as very equable \((SI \leq 0.19)\), equable but having a definite wetter season \((SI \text{ varying between 0.20 and 0.39})\), rather seasonal having a short drier season \((SI \text{ varying between 0.40 and 0.59})\), seasonal \((SI \text{ between 0.60 and 0.79})\), markedly seasonal having a long drier season \((SI \text{ between 0.80 and 0.99})\), most rain in 3 months or less \((SI \text{ between 1.00 and 1.19})\), and extreme, almost all rains in 1–2 months \((SI \geq 1.20)\) (Kanellopoulou 2002).

### 5 Results

#### 5.1 Optimum number of clusters

The graphical method of D index was used to determine the number of clusters based on the sharp knee that corresponds to a significant decrease of the first differences of clustering gain versus the number of clusters (Charrad et al. 2014) (Fig. 4). The significant knee (the significant peak in D index second differences plot) corresponds to a significant increase of the measured value. The significant knee was noticed in the graph up to four clusters for \( k\)-mean, \( \text{wardD} \), and \( \text{wardD2} \). The cluster solution showed that clustered group formation is optimal for the ‘4 clusters solution’ with
almost half of the group members forming at this solution. The remaining in similar groups is up to the ‘10 clusters solution’, which is the highest expected solution. Besides, the maps prepared based on optimal cluster number (3, 4, 5, 6, 8, and 10), presented in Fig. 5, indicates 4-cluster is the optimal solution as it visually provided a meaningful result. The lowest variability of each climate parameter in each of the cluster was obtained for 4 clusters. The 3- and 5-cluster solutions were relatively better, showing uniform climate zones across Borneo. However, the variability of each climate parameter was high, particularly in the high elevation area. The 5-cluster solutions gave a large difference between the size of the biggest and the smallest clusters. Besides, the two smallest clusters located in Borneo’s interior region showed similar characteristics. Meanwhile, the 6-, 8-, 10-cluster solutions were heterogeneous, which showed that the similar groups were spatially separated, thus not suitable for climate zone classification. Therefore, the 4-cluster solution was used in the subsequent analysis.

5.2 Selection of cluster methods

The visualization of the output in terms of climate zoning from the non-hierarchical and hierarchical agglomerative clustering methods was used to initially select the methods that give the most relevant results (Fig. 6). The Single, Average, Complete, McQuitty, Median, and Centroid methods performed poorly in classifying Borneo’s climate zone within the hierarchical agglomerative clustering methods. Here, most of the grid points were assigned into one single cluster, showing the chaining problem. On the other hand, the k-mean, wardD, and wardD2 distributed the grid points more evenly across the clusters. They appear to be more robust, producing visually acceptable spatial classification into four zones. There was a difference in the clusters’ size, the smaller cluster overlapping on top of the bigger cluster. This spatial pattern followed Borneo’s elevation, local climate, and geographical characteristics. The largest zone covered the southern and eastern coastal regions of Borneo, characterized by its geographical influence of the Indonesian throughflow (ITF) from the Sulu Sea, Makassar Strait, Celebes Sea, Java Sea, and Karimata Strait, which encompasses some of the warmest ocean temperatures around the globe known as the ‘boiler box’ of the Tropics (Ramage 1968; Gordon and Fine 1996). The ITF brings sea surface currents from the warm pool area, located northeast of Irian Jaya Island (New Guinea). Hence, the SST over the eastern and southern parts of Borneo is mainly determined by the warm pool condition. Due to the sun’s position in the Southern Hemisphere during the NE monsoon, the ITF brings cooler water from the warm pool to the Makassar Strait, Sulu Sea, Sulawesi Sea, and the Java Sea. This cooler SST prevents convective zone formation for the climate zone of DH. Aldrian and Dwi Susanto (2003) also showed SST-related function in this area.

The second-largest zone facing the SCS influences the cold surges, NE monsoon, and Borneo vortex associated with strong northeasterly and easterly winds. Albeit the surge from the north is relatively cooler and drier, as it travels over the warmer parts of the southern SCS, it
becomes moister, enhancing deep convection episode which causes heavy rain that falls along the northwestern coastline of Borneo and deep into the interior region (Tangang et al. 2008). The two smallest zones were located in the central region, influence by the elevated orography. However, there was a difference in the size and boundaries of climate zones defined by each method, although giving a similar pattern.

The cluster plot against 1st two principal components and the centroid plot against 1st two discriminant functions of k-mean, wardD, and wardD2 (not shown) revealed the grid points’ distribution partitioned into four climate zones. There was a similarity between the clusters’ partition between these three methods. Overlapping between clusters was also observed, which may be due to the location of grid points closed at each other’s boundaries. However, the dissimilarity between the clusters remains profound. The validation of the cluster solutions showed that the wardD (1.865–1.920) method gives the lowest average distance, followed by wardD2 (2.263–2.331), and k-mean (2.880–2.927) methods. This suggests that the wardD method performs better in grouping the climate zone of Borneo.

Based on the climate characteristic of each climate zone, it can be divided into four zones, namely ‘dry and hot’ (DH), ‘wet and hot’ (WH), ‘wet’ (W), and ‘wet and cold’ (WC) as shown in Table 1. It showed that the amount of rainfall and temperature for each climate zones were different from each other. For example, under the k-mean method, the DH climate zone receives 25–26% less rainfall compared to other climate zones with the highest temperature for Tavg (27 °C), Tmin (23 °C), and Tmax (31 °C). Meanwhile, the WH climate zone receives the highest amount of rainfall (3411 mm), although having a similar high temperature (Tavg, Tmin, and Tmax) with DH. W climate zone also receives a comparatively high amount of rainfall (3376 mm) with WH and WC, but with a lower temperature for Tavg (25 °C), Tmin (21 °C), and Tmax (30 °C) compared to DH and WH. WC climate zone also has a comparatively high amount of rainfall (3387 mm) with WH and W, but the lowest temperature for Tavg (23 °C), Tmin (18 °C), and Tmax (27 °C). DH climate zone covered the highest grid points (361), mainly in the coastal area of the south and the eastern regions of Borneo. The second highest grid point (318) was WH, mainly covering the central, northwest, and western

Fig. 5 The maps for different cluster solutions using k-mean method
regions of Borneo. The third-highest was W, with 213 grid points covering the mid-elevation and hilly areas below the mountain range in Borneo’s central region. The lowest grid point (69) was WC, characterized by Borneo’s highest elevation area and the mountain range. A similar pattern was found for each climate zones for wardD and wardD2, although some differences in boundaries were observed. The standard deviation for rainfall showed a distinct difference and incomparability between the clustering method due to the different sizes of the cluster being determined by each method. The inconsistencies also can be deduced from the synoptic climate influence such as cold surge and Borneo vortex that may affect the rainfall pattern at different intensities for certain months at the end of NE monsoon. The temperature (Tavg, Tmin, Tmax) for WC showed the highest standard deviation than other climate zones due to the different elevations of the mountainous region.

The four different climate zones derived from this study showed that the large-scale climate phenomena and maritime continent climate govern the local characteristic in
different parts of Borneo. Therefore, it is hypothesized that the specific influence of large-scale climate phenomena like El Niño–southern oscillation (ENSO), cold surge, Madden–Julian oscillation (MJO), Indian Ocean Dipole (IOD), and monsoon can be better understood based on the grouping of the area with similar climate characteristic. Regions having a similar annual cycle would possibly respond specifically and homogeneously to certain climate phenomena compared to others.

Among the final clustering methods assessed in this study, k-mean, wardD, and wardD2 methods showed comparable performance. A further comparative assessment was made by using an annual correlation boxplot and GDM. Boxplot of annual rainfall correlations between cluster average and grid points within the climate zone of ‘DH’, ‘WC’, ‘WH’, and ‘W’ under k-mean, wardD, and wardD2 are shown in Fig. 7. Here, the boxplot was constructed from the correlation coefficients estimated between rainfall (temperature) of the grid points of a climate zone and the zonal average rainfall (temperature) of that zone. Compared to Borneo’s mean annual climate time series, the correlation plot showed higher correlation and low variability in the time series when Borneo was partition into several climate zones with similar climate characteristics. The GDM results in Fig. 8 show that the wardD method gives the best correlation rank under all climate parameters. Therefore, the wardD method was selected as the best method to classify Borneo’s climate zone.

5.3 Climate zone of Borneo

The PDFs of rainfall, Tavg, Tmax, and Tmin were constructed to show each climate zone’s characteristics, namely, DH, W, WH, and WC based on the wardD method, as shown in Fig. 9. Rainfall of DH was the lowest, and WH was the highest. Meanwhile, rain for W and WC was comparable with each other. The differences in the amount and rainfall rate between climate zones were influence by the topography and geographical location, which govern the local climate. The DH showed the lowest rainfall due to the mountainous range that blocks the incoming wind from the SCS, coupled with the warm seas surrounding the southern and eastern parts of Borneo cause by the ITF pathway. The highest rainfall was WH, which faces the SCS, being the first area to receive the NE monsoon’s impact, cold surge, and Borneo vortex. Meanwhile, W and WC, located further inland in the hilly and mountainous region, receive a comparable amount of rainfall.

Although there was a similarity (with differences in kurtosis) between W and WC in rainfall distribution, temperature difference among the zones was notable. The climate classification in zones W and WC were due to the altitudinal difference in the mountainous regions. The W climate zone showed higher rainfall kurtosis, indicating more uniform rainfall compare to WC climate zone. Borneo is a climate-sensitive tropical region. Therefore, the temperature difference between W and WC indicated each zone as a different environment for vegetation and species. WC has the lowest temperature in terms of all temperature indices as this zone is a mountainous and high elevation area. This was followed by W, WH, and DH. There was a distinct difference between each climate zone with temperature, except WH and DH, which were almost similar. This is due to the similarity in WH and DH in terms of low-lying topography. However, DH is slightly warmer than WH due to drier conditions, less rainfall, and warm seas in the surrounding.

Figure 10 shows the mean monthly rainfall and temperature (Tavg, Tmax, and Tmin) pattern of each climate zone in Borneo. The mean monthly rainfall of Borneo was also
Defining climate zone of Borneo based on cluster analysis

inserted in the figure for comparison. A similar bimodal pattern was noticed for all climate zones where it showed a slight rainfall decrease in Feb, suggesting a strong influence of equinox phenomena and ITCZ across Borneo. However, a different decreasing rainfall pattern was observed during the SW monsoon between the four climate zones. A relatively decreasing amount of rainfall for 6 months (May–Oct) was observed for the DH climate zone, 5 months (May–Sept) for the WH climate zone, and 4 months (Jun–Sep) for W and WC climate zone. This indicates a different length and different arrival/withdrawal month of the SW monsoon in each climate zone. The arrival of SW monsoon in DH is in May and the withdrawal in October. The SW monsoon in W and WC climate zones arrives in June and withdraws in September. For WH, the SW monsoon arrives in May and ends in September. The higher peak of rainfall during NE monsoon for all climate zones was due to the Borneo vortex’s influence and cold surges coming from the north out of the SCS. The results showed that the local influences of the monsoon can be different across Borneo (Wang et al. 2004; Chenoli et al. 2018). The differences can be observed in the monthly

Fig. 7 The boxplot shows correlations between the annual cluster average and grid points in the same cluster for the period 1960 to 2016

Fig. 8 The level plot on the ranking of the clustering method using GDM
rainfall distribution in different climate zones. Therefore, in this study, the local influence of NE and SW monsoon was redefined for each climate zone according to the monthly distribution of climate variables (Fig. 10).

Each of the classified climate zones also showed a different amount of monthly rainfall than the mean monthly rainfall of Borneo. The DH climate zone showed a lower amount of rainfall each month, ranging between 7.0 and 22.8%. On the contrary, the WH climate zone showed a higher amount of rainfall for each month ranging from 8.7 to 27%. The W climate zone showed higher rainfall during April to December, ranging from 0.3 to 29.1% and less
rainfall during January to March (ranging from 2.2 to 6.8%). Meanwhile, WC climate zone showed a higher amount of rainfall during March to December (2.8–36.4%) and less rainfall during January to February (3.6–4.8%). Although there is a similar monthly rainfall distribution, the temperature was different in W and WC climate zones. The WC climate zone showed lower Tavg (23 °C), Tmin (18 °C), and Tmax (27 °C) compared to W climate zone which showed higher Tavg (25 °C), Tmin (21 °C), and Tmax (30 °C).

The SI (Abaje et al. 2010; Guhathakurta and Saji 2013; M. K. Patil 2015) indicated Borneo’s climate zones as ‘seasonal’. The highest to the lowest degree of seasonality was for different zones were as follows: DH (0.729), W (0.691), WC (0.687), and WH (0.672). The SI indicates the type of climate in relation to water availability. Higher SI means higher variability of water resources, and thus, less reliability. The results presented in this study indicate less water availability or the possibility that seasonal scarcity of water resources is high in the DH zone. Therefore, the Borneo forest fire risk is higher in DH climate zone compare to others. This is consistent with previous findings, where increased forest fire risk in El Niño years is found in Kalimantan due to reliability drier condition compared to other years (Langmer and Siegert 2009). Borneo’s local ecosystem is highly sensitive to a slight difference in the climate. Therefore, although the climate of all the zones is ‘seasonal’, variations of SI within a single class are important for the region due to its high sensitivity to local ecology. Each of the classified climate zones showed a distinct behavior and distinguishable climatic characteristic based on temporal pattern, rainfall amount, temperatures (Tavg, Tmax, and Tmin), seasonality, and the arrival/withdrawal of SW monsoon, as shown in Table 2.

Compared to the climate classification by Peel et al. (2007), our results gave more detailed information on the spatiotemporal pattern of the rainfall and temperature between the determined climate zones. Our study found a different characterization of climate (DH, W, WH, and WC) that was previously being defined broadly as Af by Peel et al. (2007). Peel et al. (2007) classify only a small patch of the mountainous region in the central part of the island as CFa (Humid subtropical climate) and Dw (Monsoon-influenced hot-summer humid continental climate). Our study showed that Borneo’s monsoon influence is widespread and temporally varied depending on the climate zone. The NE and SW monsoon’s length and intensity were found to be different between the climate zones (Fig. 10). Although there is no apparent seasonal pattern for temperature in Borneo, the present study found that the different climate zones can have a different uniform temperature distinct from each other, particularly for Tmax and Tmin due to the elevation difference and rain-shadow effect. In concurrent with the Borneo climate classification by Dambul and Jones (2007), a similar wet characteristic (type 3, western wet) was found at the western part of Borneo and wetter condition (type 4, central wet) along the coast of the northwestern area. However, we found a relatively lower rainfall at the east coast of northern Borneo than the wetter condition (type 5, eastern wet) found by Dambul and Jones (2007). The difference might be due to the threshold-based separation of climate parameters (rainfall and temperature) and the limited number of the station used by Dambul and Jones (2007). Indeed, for type 8 (mixed weather), a hotter condition in the northeast coast and colder condition in the western part of Borneo was similar to our findings. However, our study gives a better spatiotemporal analysis of the climate zone in Borneo compare to the previous study by Dambul and Jones (2007) which employed a limited number of station datasets.

### Table 2 Proposed characteristic of climate zone using the wardD method

| Climate zone | WH | W | WC | DH |
|--------------|----|---|----|----|
| Rainfall (mm) | 2550–4300 | 2450–4100 | 2450–4100 | 1700–3350 |
| Tavg (°C) | 27 | 25–26 | 22–23 | 27–28 |
| Tmin (°C) | 22–23 | 21 | 18 | 23–24 |
| Tmax (°C) | 31–32 | 29–30 | 27–28 | 31–32 |
| Seasonality | 0.672 | 0.691 | 0.687 | 0.729 |
| NE monsoon months | Oct–Apr | Oct–May | Oct–May | Nov–Apr |
| SW monsoon months | May–Sep | Jun–Sep | Jun–Sep | May–Oct |

6 **Conclusion**

The results of this study showed the capability of several non-hierarchical and hierarchical agglomerative clustering methods in classifying the local climate zone of Borneo. The application of fine spatiotemporal resolution gridded data allowed more comprehensive classification of Borneo’s climate than that obtained in previous studies. The k-mean, wardD, and wardD2 methods classified the climate of Borneo into four zones, namely ‘dry and hot’ (DH), ‘wet and hot’ (WH), ‘wet’ (W), and ‘wet and cold’ (WC). The spatial characteristic of each climate zone follows the influence of large-scale climate phenomena and land-sea influence surrounding the island. The wardD method showed the best performance based on cluster plot, centroid plot, and climate parameters ranking using boxplot correlation and GDM approach. The distinct behavior of climate parameters for each climate zone can be differentiated and visualized using PDF. Based on the coverage, temporal pattern, amount of rainfall, temperatures (Tavg, Tmax, and Tmin), seasonality, and the arrival/withdrawal month of SW monsoon (Figs. 9
and 10), each climate zone are distinguishable from the general climate characteristic of Borneo. The local influences of NE and SW monsoon were redefined to reflect local climate conditions of each climate zone (Fig. 10). The present study determined a proper classification of climate in Borneo with higher certainty for the first time, which can improve understanding of the physics that governs the local climate. The presented results can also be used by various stakeholders, especially in contributing to climate change research and adaptation, conservation of water resources, infrastructure, agricultural planning, and biological conservation. In the future, other climate variables like humidity and wind can be considered for climate zonation. Besides, other climate datasets like TRMM, CHIRPS, CPC, and PERSIANN can be used to assess uncertainty in climate zonation due to grid-ded datasets. It would also be interesting to understand how Borneo’s climate zone has changed over time and would change in the future due to climate change. Besides, other clustering methods, distance measurement algorithms, and ranking techniques can be used to improve the classification and certainty of the results. The changes in the climate of the identified zones due to population growth and land-use change can also be assessed.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Zulfaqar Sa’adi and Shamsuddin Shahid. Zulfaqar Sa’adi wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conceptualization: Zulfaqar Sa’adi, Shamsuddin Shahid; Data curation: Zulfaqar Sa’adi, Shamsuddin Shahid, Mohammed Sanusi Shiru; formal analysis: Zulfaqar Sa’adi, Shamsuddin Shahid; funding acquisition: Zulfaqar Sa’adi; Investigation: Zulfaqar Sa’adi, Shamsuddin Shahid; methodology: Zulfaqar Sa’adi, Shamsuddin Shahid; project administration: Zulfaqar Sa’adi; resources: Zulfaqar Sa’adi; software: Zulfaqar Sa’adi, Shamsuddin Shahid; supervision: Zulfaqar Sa’adi; validation: Zulfaqar Sa’adi, Shamsuddin Shahid, Mohammed Sanusi Shiru; visualization: Zulfaqar Sa’adi, Shamsuddin Shahid; writing—original draft: Zulfaqar Sa’adi; writing—review and editing: Zulfaqar Sa’adi, Shamsuddin Shahid, Mohammed Sanusi Shiru.

Funding This work was supported by the Water Security and Sustainable Development Hub funded by the UK Research and Innovation’s Global Challenges Research Fund (GCRF) (grant number: ES/S008179/1) Universiti Teknologi Malaysia under High Impact Research Grant (grant number: Vot No 04G46).

Data availability Not applicable.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication We (Zulfaqar Sa’adi, Shamsuddin Shahid and Mohammed Sanusi Shiru) hereby declare that we participated in the study in the development of the manuscript titled (Defining Climate Zone of Borneo based on Cluster Analysis). We have read the final version and give our consent for the article to be published in TAAC.

Conflict of interest The authors declare no competing interests.

References

Abaje IB, Ishaya S, Usman S (2010) An analysis of rainfall trends in Kafanchan, Kaduna State, Nigeria. Res J Environ Earth Sci 2:89–96

Ahmed F, Al-Amin AQ, Mohamad ZF, Chenayah S (2016) Agriculture and food security challenge of climate change: a dynamic analysis for policy selection. Sci Agric. https://doi.org/10.1590/0046-012320150141

Ahmed K, Shahid S, Sachindra DA et al (2019) Fidelity assessment of general circulation model simulated precipitation and temperature over Pakistan using a feature selection method. J Hydrol. https://doi.org/10.1016/j.jhydrol.2019.03.092

Aldrian E (2001) Climate categories in Indonesia based on rain pattern using “double correlation” method. J Sci Technol Weather Modif. https://doi.org/10.29122/jstwm.v2i1.2142

Aldrian E, Dwi Susanto R (2003) Identification of three dominant rain-fall regions within Indonesia and their relationship to sea surface temperature. Int J Climatol. https://doi.org/10.1002/joc.950

Alvares CA, Stape JL, Sentelhas PC et al (2013) Köppen’s climate classification map for Brazil. Meteorol Zeitschrift. https://doi.org/10.1127/0941-2948/2013/0507

Ayoub AB, Tangang F, Juneng L et al (2020) Evaluation of gridded precipitation datasets in Malaysia. Remote Sens 12:613. https://doi.org/10.3390/rs12040613

Belda M, Holtanová E, Halenka T, Kalvová J (2014) Climate classification revisited: from Köppen to Trewartha. Clim Res. https://doi.org/10.3354/cr01204

Chang CP, Wang Z, McBride J, Liu CH (2005) Annual cycle of Southeast Asia - maritime continent rainfall and the asymmetric monsoon transition. J Clim. https://doi.org/10.1175/JCLI-3257.1

Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) An examination of indices for determining the number of clusters: NbClust Package. J Stat Softw. https://doi.org/10.18637/jss.v061.i06

Chenoli SN, Jayakrishnan PR, Samah AA et al (2018) Southwest monsoon onset dates over Malaysia and associated climatological characteristics. J Atmos Solar-Terrestrial Phys 179:81–93. https://doi.org/10.1016/j.jastp.2018.06.017

Corllis JO, Sneath PHA, Sokal RR (1974) Numerical taxonomy: the principles and practice of numerical classification. Trans Am Microsoc Soc. https://doi.org/10.2307/3225339

Dambul R, Jones P (2007) Regional and temporal climatic classification for Borneo. Geogr Malaysian J Soc Sp

DeGaetano AT (1996) Delineation of mesoscale climate zones in the northeastern United States using a novel approach to cluster analysis. J Clim. https://doi.org/10.1175/1520-0442(1996)009<0009:DOMZFl>2.0.CO;2

Dindang A, Taat A, Beng PE et al (2013) Statistical and trend analysis of rainfall data in Kuching, Sarawak from 1968–2010. JMM Res Publ 6:17

Diong JY, Yip WS, Mat Adam MK et al (2015) The definitions of the southwest monsoon climatological onset and withdrawal over Malaysian region. JMM Res Publ 3:1–30

Springer
Tangang F, Farzanmanesh R, Mirzaei A et al (2017) Characteristics of precipitation extremes in Malaysia associated with El Niño and La Niña events. Int J Climatol. https://doi.org/10.1002/joc.5032

Tangang FT, Juneng L, Salimun E et al (2008) On the roles of the northeast cold surge, the Borneo vortex, the Madden-Julian Oscillation, and the Indian Ocean Dipole during the extreme 2006/2007 flood in southern Peninsular Malaysia. Geophys Res Lett. https://doi.org/10.1029/2008GL033429

Tangang FT, Juneng L, Salimun E, et al (2012) Climate change and variability over Malaysia: gaps in science and research information. Sains Malaysiana

Tangang FT, Juneng L, Salimun E, et al (2013) Projected precipitation changes over Malaysia by the End of the 21st Century Using PRECIS Regional Climate Model. Clim Chang Isl Coast Vulnerability 3–20

Thornthwaite CW (1948) An approach toward a rational classification of climate. Geogr Rev. https://doi.org/10.2307/210739

Trewartha G, Horn L (1980) An introduction to climate

Unal Y, Kindap T, Karaca M (2003) Redefining the climate zones of Turkey using cluster analysis. Int J Climatol. https://doi.org/10.1002/joc.910

Vinayachandran PN, Francis PA, Rao SA (2009) Indian ocean dipole: processes and impacts. Curr Trends Sci. https://eprints.iisc.ac.in/id/eprint/35256

Wang B, LinHo ZY, Lu MM (2004) Definition of South China Sea monsoon onset and commencement of the East Asian summer monsoon. J Clim 17:699–710. https://doi.org/10.1175/2932.1

Xavier P, Lim SY, Bin Abdullah MFA et al (2020) Seasonal dependence of cold surges and their interaction with the madden–julian oscillation over Southeast Asia. J Clim. https://doi.org/10.1175/JCLI-D-19-0048.1

Yusuf AA, Francisco H (2009) Climate change vulnerability mapping for Southeast Asia vulnerability mapping for Southeast Asia East. https://doi.org/10.1158/1541-7786.MCR-07-0267

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.