

Redefining Homogeneous Climate Regions in Bangladesh Using Multivariate Clustering Approaches

Sultan Mahmud (✉ smahmud@isrt.ac.bd)

International Centre for Diarrhoeal Disease Research <https://orcid.org/0000-0003-0757-7630>

Ferdausi Mahojabin Sumana

South Asian University

Md. Mohsin

University of Dhaka

Md Hasinur Rahaman Khan

University of Dhaka

Research Article

Keywords: Homogeneous climate regions, Multivariate clustering approaches, Cluster validation, Water resources, Climate, Bangladesh.

Posted Date: June 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-633865/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Natural Hazards on November 28th, 2021. See the published version at <https://doi.org/10.1007/s11069-021-05120-x>.

1 Redefining Homogeneous Climate Regions in Bangladesh Using Multivariate 2 Clustering Approaches

3 Sultan Mahmud; Ferdousi Mahojabin Sumana; Md. Mohsin; Md Hasinur Rahaman Khan

4
5 **Abstract:** The knowledge of the climate pattern for a particular region is important to alleviate the impact of climate change
6 and protect the environment by taking appropriate actions based on geographical knowledge. It is also equally important
7 for water resources planning and management. In this study, the regional disparities and similarities have been revealed
8 among different climate stations or regions in Bangladesh based on different climatological factors such as rainfall,
9 temperatures, relative humidity, sea level pressure, cloud cover, wind speed, the sunshine hour. We have selected one of
10 the best-fitted algorithms for particular climate data from three multivariate clustering approaches named hierarchical
11 clustering, partitioning around medoids (PAM), and K-means clustering by using different validation tests. Four
12 homogeneity tests (Mann-Kendall Test, Pettitt's test, Buishand Range Test, Standard Normal Homogeneity Test) also have
13 been performed for each of the clusters created based on several factors. The results suggest that the climate regions or
14 meteorological stations of Bangladesh can be clustered into two groups based on a combination of climatological variables.
15 According to the findings, there is a huge variation between the two groups in terms of climatological factors. The first
16 group (cluster 1) is located in the northern part of the country that includes drought-prone and vulnerable regions, whereas,
17 the second group (cluster 2) contains rain-prone and hilly regions, which are mostly situated in the southern part. All newly
18 defined clusters show homogeneous behavior with few exceptions such as clusters based on sea level pressure are not
19 homogeneous.

20 **Keywords:** Homogeneous climate regions; Multivariate clustering approaches; Cluster validation; Water resources;
21 Climate; Bangladesh.

22 _____
23 Sultan Mahmud (Corresponding author)

24 International Centre for Diarrhoeal Disease Research, Bangladesh, Mohakhali, Dhaka 1212, Bangladesh

25 Tel.: +8801738761160

26 E-mail: smahmud@isrt.ac.bd

27 Ferdousi Mahojabin Sumana

28 South Asian University, New Delhi

29 Tel: +8801727400272

30 E-mail: ferdausi.sumana@gmail.com

31 Md Mohsin

32 Institute of Statistical Research and Training, University of Dhaka, Dhaka-1000, Bangladesh

33 Tel.: +8801518498289

34 E-mail: mmohsin@isrt.ac.bd

35 Md Hasinur Rahaman Khan

36 Institute of Statistical Research and Training, University of Dhaka, Dhaka-1000, Bangladesh

37 Tel.: +8801725106661

38 E-mail: hasinur@isrt.ac.bd

39

40

1 **1 Introduction**

2 Identifying the homogeneous regions in hydrological and climatological fields is very crucial for
3 policymakers to develop national climate policies to respond to climate behaviors for reducing the
4 impact of climate change (Iyigun et al. 2013). Due to various development activities, climate change
5 is probably the most significant issue affecting the overall social and economic wellbeing of
6 Bangladesh, also forcing to occur hydrological disaster (Shahid and Behrawan 2008). Soon, climate
7 change could be a reason for unexpected changes in the pattern of climatological factors such as
8 precipitation, the temperature that will have a huge impact on biodiversity (Willis and Bhagwat 2009),
9 water resource management (Huq 2001), agriculture (Huq et al. 1999, Shahid 2011), and public health
10 (Mahmood 2012). Bangladesh is one of the most vulnerable countries in the world (Shahid 2010) and
11 the impact of climate change is not identical all over the country. For instance, Haque and Jahan (2015)
12 showed Dhaka, Rajshahi, and Sylhet are facing more natural disasters compared to the coastal regions
13 specifically Barisal, Chittagong, and Khulna in Bangladesh. Rahman et al. (2015), Ahmed and Kim
14 (2003), Poncelet et al. (2010) also observed significant changes among several climatological factors
15 in some regions compared to others. Therefore, it is essential to monitor the parameters governing the
16 climatological and ecological equilibrium for a given region. Towards achieving this aim the
17 climatological factors such as rainfall, temperature, cloud cover, humidity need to be studied in space
18 and time.

19
20 Furthermore, the proper knowledge of the climatological properties of different regions would be
21 useful in water resources planning and management for multiple purposes (Goyal et al. 2019). The
22 identification of a homogeneous climate region also will be helpful to the researchers to understand
23 the atmospheric mechanisms (Abadi et al. 2019). It is also important to know the regional properties
24 to figure out the suitable crops and plants for a particular climate region. In the climatological,
25 ecological, and hydrological discipline, there are several approaches available for identifying
26 homogeneous climate regions known as unsupervised data mining techniques for instance hierarchical

1 clustering, k-means clustering, fuzzy c-clustering, partitioning around medoids (PAM). Those
2 techniques divide data into several subgroups that are called clusters based on similarities or
3 dissimilarities measures. Several studies are available in the literature for identifying the
4 climatological and hydrological disparities among different regions. For example, Siraj-Ud-Doulah
5 and Islam (2019) and Rahman et al. (2013) used hierarchical and k-means clustering respectively to
6 define the homogeneous climatological regions in Bangladesh. In Italy, Di Giuseppe et al. (2013)
7 applied the Partitioning around medoids (PAM) method to delimitate the homogeneous climate zones.
8 In New Zealand, Mosley (1981) used multivariate cluster analysis to define the hydrological regions.
9 Modarres (2006) and Ahmad et al. (2013) respectively identified the homogeneous precipitation
10 regions of Iran and Malaysia using a hierarchical clustering approach. Sahin and Cigizoglu (2010),
11 Goyal et al. 2019, and Basalirwa (1995) respectively in Turkey, India, and Uganda also applied
12 clustering approaches for the same purpose. Furthermore, Abadi et al. (2019) applied a multivariate
13 clustering approach to divide Bolivia into identical and coherent climatological sub-regions. The
14 clustering approaches have been applied to identify the climate zones of the Conterminous United
15 States by Fovell and Fovell (1993). Xiong et al. (2019) also used the same approach to define diverse
16 climate regions. Hargrove and Hoffman (1999) applied different clustering algorithms to define
17 ecoregional borders based on environmental data. The hydro-climatic regions have been defined by
18 applying clustering approaches on climate data in the southern Appalachian Mountains, U.S. Besides,
19 several other studies used clustering approaches for analyzing climate data such as Abadi et al. (2019),
20 Ahmad et al. (2013), Zhang et al. (2016), Sahin and Cigizoglu (2010), Basalirwa (1995) and Kysely
21 et al. (2007). Although a lot of clustering approaches have been used to define homogeneous climate
22 regions all over the world, not all the clustering approaches do perform identically for all problems or
23 data (Rodriguez et al. 2019). For example, Unal et al. (2003) showed the hierarchical clustering
24 algorithm is the best fitted for temperature data in Turkey among the five multivariate clustering
25 approaches. Matulla et al. (2003) also observed huge differences in terms of accuracy of clustering
26 among several multivariate clustering approaches for different seasonal and periodical rainfall data.

1 Moreover, the k-means clustering approach has been found as outperformed for the data with outlines
2 by Chen et al. (2005), and Mingoti and Lima (2006) revealed that the fuzzy c-means provide
3 outstanding performance for simulated overlapping data with outlines but hierarchical and k-means
4 provide identically bad performance.

5
6 The knowledge of the regional variation of climatological factors is very essential to recognize the
7 urgency of further research and improvement in climatology and hydrology. This study can make a
8 positive contribution by providing a clear idea about the regional climate pattern in Bangladesh to the
9 meteorologist, and hydrologist. In the nearest past, for observing the variation in regional
10 climatological factors, many researchers in Bangladesh divided the whole country into several parts.
11 Most studies selected several climate clusters or regions on a subjective basis or based only on one or
12 two climate factors and without any statistical validation. For instance, Mannan and Karmakar (2007)
13 divided Bangladesh into six different parts to observe heavy rainfall variations such as Northeast
14 region (NE region), Northwest region (NW region), East-central region (EC region), West-central
15 region (WC region), Southeast region (SE region), and Southwest region (SW region). Islam et al.
16 (2005), Mahmud, and Islam (2019), Mahmud et al. (2020) also selected some of the climate stations
17 from different geographical parts of Bangladesh by considering each part as homogeneous. However,
18 Siraj-Ud-Doulah and Islam (2019) conducted a study to identify the homogeneous climate region in
19 Bangladesh based only on the rainfall variable. In addition, they have selected the homogeneous region
20 on a subjective basis. To identify homogeneous climate zones, one factor cannot be sufficient.
21 However, our study can be distinguished from others in several ways, and those are: (i) the best
22 clustering approach has been identified and applied for a particular climate feature, (ii) seven validity
23 measures have been calculated to get the optimal number of clusters and best-performing clustering
24 approach, (iii) the clustering approaches applied on eight individuals and combination of more than
25 seven climatological factors including latitude and longitude of climate stations, (iv) the homogeneity

1 test also has been performed to suggest whether the newly determined climate zones are directly useful
2 or not for further climatological or meteorological research.

3 **2 Study Area and Data**

4 Our study area, Bangladesh, is situated in South Asia with two different environments such as Bay of
5 Bangle to the South and Himalayas to the North extends from latitudes $20^{\circ}34'$ - $26^{\circ}38'$ north and
6 longitudes $88^{\circ}01'$ - $92^{\circ}41'$ east and it covers an area of more than 140,000 sq km (Ahasan et al. 2011).
7 Bangladesh is a small country with wide climatological diversities which can be observed through
8 different climatological factors. To observe these diversities of climate, we have collected monthly
9 average data for 68 years during the period 1949 to 2017 for a total of 34 rain gauge stations (climate
10 stations so-called climate zone) for some important climate factors from the Bangladesh
11 Meteorological Department (BMD) (<https://www.bmd.gov.bd/>). The different climate factors have
12 been observed in terms of different units such as monthly average rainfall in millimeter (mm), monthly
13 average relative humidity in percentage (%), monthly average cloud cover in Octa, monthly average
14 sea level pressure in Millibar, monthly average sunshine duration in time (hour), monthly average wind
15 speed meter per second (m/s), the monthly average maximum temperature in Celsius, the monthly
16 average minimum temperature in Celsius, latitudes, and longitudes of the climatological stations.
17 Generally, these different units influence the outcome of cluster analysis. To avoid this influence,
18 several studies suggest a data transformation approach which is called data normalization (Goyal and
19 Gupta 2014, Cannarozzo et al. 2009, Dikbas et al. 2012). The normalization approach transforms the
20 values of the variables in a way such that each of the transformed factors will have zero mean and unit
21 variance (De et al. 2008). Furthermore, missing values also very commonly influence the analysis of
22 climatological data. In this study, we have estimated the missing values by using the Multiple
23 Imputations by Chained Equations (MICE) method (White et al. 2011, Mahmud et al. 2020).

1 **3 Methodology**

2 Clustering or grouping of objects is a popular multivariate approach to identify dimensionality,
3 unusual, and unexpected observation, as well as the interesting multivariate relationship among the
4 variables or features (Johnson and Wichern 2002). In climatology and hydrology, clustering is a core
5 technique for disclosing the hidden pattern of the climatological or hydrological features (Ouyang et
6 al. 2010). Most of the techniques create several groups among all the objects from the data using
7 similarity measures or dissimilarity measures (distance measures). There are several multivariate
8 clustering techniques available. In this study, three clustering techniques: hierarchical, k-means, and
9 Partitioning around medoids (PAM), the most popular among several fields including climatology and
10 hydrology, have been chosen which are discussed in the following subsections.

11 **3.1 Hierarchical Clustering**

12 The hierarchical clustering method is one of the best choices for climatological or hydrological pattern
13 discovery (Ouyang et al. 2010). It works out by building a hierarchy of the tree limb using the divisive
14 and agglomerative approaches (Giraldo et al. 2012). The hierarchical clustering methods are usually
15 grouped into two major techniques, namely, agglomerative: a series of successive merges, divisive: a
16 series of successive divisions. Because of the computational and technical feasibility, agglomerative
17 methods are more widely preferred in applications rather than the divisive method. The agglomerative
18 approach starts with initial objectives that mean each of the objects in the data set is initially considered
19 as a cluster. Then they create initial subgroups by combining the most similar or closed objects. After
20 that, these clusters are merged according to their similarity or closeness (Johnson and Wichern, 2002).
21 Moreover, the agglomerative hierarchical clustering approach can be summarized into the following
22 steps (Iyigun et al. 2013):

23 Step1: Start with the initial n objects or clusters, and calculate the proximity matrix for each of the
24 clusters.

1 Step2: In the proximity matrix, search for the minimal distance $dis(cl_i, cl_j) = \min_{1 < k, l < n, l \neq k} dis(cl_s, cl_p)$ where $dis(\dots)$ express
2 the distance measure and combine cluster cl_i and cl_j to construct a new cluster cl_{ij} .
3 Step3: Measure the distances between the clusters cl_{ij} and the other clusters and update the proximity
4 matrix.
5 Step4: Repeat steps 2 to 3 until only one cluster remains.

6
7 However, several proximity measures are available for the agglomerative hierarchical clustering
8 approach namely single linkage, complete linkage, average linkage, and Ward's method which are
9 called linkage matrices. In this study, we have used Ward's method (Ward 1963) that merges two
10 clusters by minimizing the amount of increase in the error sum of squares (ESS) or loss of information.
11 Where, ESS_i is the sum of squared deviations of every object in the i^{th} cluster from the cluster mean
12 (centroid). In the current stage if we have p clusters then ESS is the sum of the ESS_p . At every step, the
13 union of every possible pair of clusters is considered, and the pair of clusters that contribute the
14 smallest increase in the error sum of squares or loss of information are merged.

15 **3.2 K-means Clustering**

16 K-means method is a popular nonhierarchical clustering technique that has been suggested by
17 (MacQueen 1967). It creates a specific number of partitions or groups among items instead of
18 variables. In this technique, each object assigns to a specific partition or cluster in a way such that the
19 cluster's centroid or mean has the lowest distance from that corresponding object compare to other
20 cluster's centroids and these centroids represent the cluster. This clustering, also known as a
21 partitioning approach (Johnson and Wichern 2002), can be described in the following few steps:

22

23 Step1: Specify the number of clusters k and create k partitions among the given items or objects.

- 1 Step2: Allocate each object to the nearest partition or cluster based on distance measure and recalculate
- 2 the centroid for the groups who received the new objects and lost the objects.
- 3 Step3: Repeat step 2 until no new allocation takes place.

4 **3.3 Partitioning Around Medoids (PAM)**

5 Partitioning around medoids (PAM) (Kaufman and Rousseeuw 1987) is an unsupervised clustering
6 algorithm that creates partitions or groups among objects by minimizing the within-cluster sum of
7 squares similar to K-means clustering. PAM, however, is considered to be more robust since it counts
8 other dissimilarities besides the distance measures. PAM clustering approach also selects an object
9 randomly as a medoid instead of the calculating centroid for each partition that represents the cluster
10 and the number of clusters is predefined. Then the remaining objects are assigned to the nearest
11 medoid. This process reiterates until no change takes place in medoids. This algorithm can be
12 explained by the following steps (Bhat 2014):

13 Step 1: Specify the number of clusters k .

14 Step 2: Select k objects randomly as medoids that represent k clusters or groups.

15 Step 3: Calculate swapping cost among each object and select medoid. If the swapping cost is less than
16 zero, then replace that object with the corresponding medoid. Allocate each data object to the cluster
17 with the most identical medoid.

18 Step 4: Reduplicate steps 2 and 3 until there is no update in the medoid.

19 **3.4 Choosing Appropriate Algorithm and Number of Clusters**

20 This is challenging to figure out which method works best for a given set of data. Nevertheless, it is
21 crucial to evaluate the quality of the groups or clusters, particularly how well the clusters are separated
22 from each other (Batistakis and Vazirgiannis 2001). In addition, the significance of the result from
23 most of the clustering approaches still not specified and verification of clustering outcomes is manual,
24 time-consuming, and subjective (Ankerst et al. 1999). Therefore, a suitable cluster validation is an
25 inevitable part of cluster analysis for evaluating the cluster's outcomes (Dubes and Jain 1979). In this

1 study, all of these purposes have been served by computing internal and stability validation measures
2 by using the R package “clvalid” (Brock et al. 2011). Three different measures such as connectivity,
3 Silhouette Width, and Dunn, have been considered for internal validation that reflects the measure of
4 solidarity, association, and isolation of the partitions or clusters. The association or connectivity (Handl
5 et al. 2005) revealed the extent of similarity among the members of a cluster that are placed in the
6 same group or cluster as the nearest neighbor of the other members. By measuring the intra-cluster
7 variation, the solidarity or compactness evaluates the homogeneity of the clusters, notwithstanding,
8 the isolation determines the degree of separation (Brock et al. 2011).

9
10 The tools Silhouette Width and Dunn Index both express the non-linear combination of isolation and
11 solidity, whereas, the combination of all three measures: association, isolation, solidity are expressed
12 by the tools named connectivity. The stability validation also includes four different measures; all of
13 those measures use two different datasets. One is full data and the other is data with a single column
14 removed at a time. For example, the average proportion of non-overlap (APN) is one of the measures
15 that takes into account the average proportion of objects not occupied in a similar cluster from the
16 clustering based on full data and clustering based on data with a single column removed. The average
17 distance (AD) is another important measure that calculates the distance between the objects allocated
18 in a similar cluster from the cluster based on full data and the cluster based on data with a single column
19 removed. The average distance between means (ADM) indicates the distance between the groups’
20 centers for the objects occupied in a similar cluster from clustering based on full data and clustering
21 based on data with a single column removed. The figure of merit (FOM) is a monumental measure of
22 stability validation that assesses the means intra-cluster variance among the removed column, whereas
23 the clustering occurred based on the unremoved column of the data. By considering all these statistical
24 tools or measures, the comparison among hierarchical clustering, K-means, and PAM will be made in
25 this study to find the best clustering approach for a particular climate feature or combination of climate
26 features.

1 **4 Findings and Discussion**

2 In this study, we have used several validation measures to find out the sublime clustering approach for
3 meteorological data of Bangladesh among the hierarchical method, K-means, and Partitioning around
4 medoids (PAM) approach. The validation measure described in Section 3.4 provides a suggestion for
5 the optimal number of clusters for the given data set. The optimal number of clusters or optimal
6 clustering approach indicates which algorithm has the lowest amount of error or cost and maximum
7 separation or accuracy of separation for how many clusters (Sannakki et al. 2011). To obtain an optimal
8 clustering algorithm along with the optimal number of clusters, all three clustering approaches have
9 been performed and the performance of validation measures have been captured over a possible value
10 of K (2 to 10). A comparative view of clustering efficiency or performance in terms of internal validity
11 measures (Silhouette Width, and Dunn, Silhouette Width, and Dunn) and stability validation measures
12 (average proportion of non-overlap (APN), average distance (AD), the average distance between
13 means (ADM), the figure of merit (FOM)) among three clustering approaches have been presented in
14 Table 4. Table 1 shows the summary results from Table 4, which displays the optimal number of
15 clusters and the best clustering approach suggested by each of the selected measures with the
16 corresponding optimal score. The connectivity, for the best algorithm with the optimal number of
17 clusters, should be minimum and the Dunn Index and the Silhouette Width should be maximum. Also,
18 the smallest value of each of the stability measures indicates the best performance of the corresponding
19 clustering approach and the number of clusters.

20 **4.1 Climate Zone Based on Single Climatological Factor**

21 In this study, we divided Bangladesh into several clusters by using two different approaches. Firstly,
22 we individually used all the climatological factors and for the later part, we combined all the
23 climatological factors altogether. In this subsection, we will discuss the results of individual factor-
24 wise clustering.

25

1 The hierarchical clustering approach suggested the best algorithm for monthly average rainfall data by
2 most of the validation measures with 2 optimal numbers of clusters. Table 4 shows the lowest value
3 of connectivity (2.93), APN (0), ADM (0), and the highest value of Dunn Index (0.65), Silhouette
4 (0.51) for hierarchical approach with 2 clusters, whereas AD (1.6), FOM (0.53) suggest K-means as
5 the best algorithm with 10 optimal number of clusters. Two clusters or homogeneous groups of climate
6 regions were found that are shown in Fig 1, where, cluster 1 includes 22 climate stations and cluster 2
7 includes 12 climate stations. Likewise, for relative humidity, most of the measures in Table 1 and 4
8 (highest value of Silhouette (0.51) and Dunn Index (0.43) and lowest value of connectivity (6.61) and
9 ADM (0)) suggest hierarchical with 6 optimal number of clusters as the best algorithm. These clusters
10 are depicted in Fig 2.

11
12 The validations measures in Table 1, summarized from Table 4, indicate that the K-means algorithm
13 is the best fitting approach for maximum and minimum temperature along with two and three clusters
14 respectively. Total 23 climate stations belong to cluster 1 while 11 climate stations belong in cluster 2
15 based on the maximum temperature and they are presented in Fig 3. Moreover, Fig 4 shows the clusters
16 of climates stations based on minimum temperature. Cluster 1 contains 11 climate stations and clusters
17 2 and 3 contain 13 and 10 climate stations respectively. Three climate zones also were found based on
18 monthly average cloud cover using the hierarchical algorithm that is presented in Fig 5. Here, cluster
19 1 contains 10 climate stations, 13 climate stations are in cluster 2 and 11 climate stations are in cluster
20 3. Most of the validation measures in Tables 1 and 4 indicate hierarchical as the best clustering
21 algorithm for cloud cover with three clusters. The hierarchical approach is also found as an optimal
22 clustering algorithm for sea level pressure, sunshine hour, and wind speed (see Tables 1 and 4) with
23 the optimal number of clusters two, five, and two respectively. Two climate zones are showed in Fig
24 6 based on sea level pressure where cluster 1 contains one climate station and cluster 2 contains 33
25 climate stations. Furthermore, five clusters were found based on sunshine hour, where cluster 1
26 contains 10 stations, cluster 2 contains 12, cluster 3 contains 2 climate stations, cluster 4 contains 2

1 climate stations and cluster 5 contains 8 climate stations that have been displayed in Fig 7. Similarly,
2 Fig 8 shows two clusters based on wind speed where cluster 1 contains 23 climate stations and 11
3 climate stations belong in cluster 2.

4 **4.2 Climate Zone Based on Combination of Climatological Factors**

5 The homogeneous climate zones are also inquired based on the combination of all the considered
6 climate factors described in Section 2. It is inspected from Tables 1 and 4 that the connectivity for the
7 hierarchical clustering approach is the lowest (2.93) at the number of clusters 2 and it increases with
8 the increase in the number of clusters. Similarly, the highest Dunn index (0.48) and Silhouette Width
9 (0.79) for hierarchical clustering are observed at the same number of clusters and they decrease with
10 the increase in the number of clusters. In addition, the lower amount of overlapping between clusters
11 (APN=0.01) and the lowest amount of average distance within-cluster (ADM=0.08) indicate
12 hierarchical clustering as the best clustering approach at the same number of clusters. Elsewhere,
13 however, the lower amount average distance in the same cluster (AD=1.78) and the lower amount of
14 average intra-cluster variation (FOM=0.88) are observed for the PAM method with the number of
15 clusters 10 and 6 respectively. Therefore, as the best clustering approach for the combination of all the
16 climatological factors considered in this study, we performed hierarchical clustering where each of the
17 two clusters contains 17 homogeneous climate stations (see Fig 9). The locations of the climate regions
18 of each cluster and regional distribution of clustered climate stations based on the combination of all
19 considered factors are presented in Fig 10 with different colors.

20 **4.3 Homogeneity Test of Clusters**

21 This study also checked whether the newly defined clusters are homogeneous or not by performing
22 four statistical tests. Among them, Mann-Kendall (MK) test identifies the monotonic trend among
23 environmental data if there is any. Whereas, Pettitt's test (Pettitt 1979), Buishand Range Test (Buishand
24 R) (Buishand 1982), Standard Normal Homogeneity Test (SNHT) (Alexandersson 1986) identify the
25 change point among environmental data if there is any. The null hypothesis for the MK test is: the

1 annual values of corresponding climate variables are independent and identically distributed and for
2 the rest of the three tests the null hypothesis is: there is no change point among the yearly values of
3 corresponding climate data. The details of theoretical and mathematical developments of these tests
4 are available in (Wijngaard et al. 2003, Sahin and Cigizoglu 2010, Yue and Wang 2004). We have
5 labeled a cluster as doubtful if one or two tests among four indicate that any monotonic trend or change
6 point is detected among the climatological time series data in the corresponding cluster. Again, a
7 cluster is labeled as suspected if more than two tests among four indicate that any monotonic trend or
8 change point is detected among the climatological time series data in the corresponding cluster.
9 Nonetheless, the cluster is labeled as useful if all four tests indicate that there is no change point or
10 monotonic trend among the corresponding environmental time-series data. Table 2 displays the p-
11 value of the corresponding test along with the decision for each of the clusters where the null
12 hypothesis of the corresponding test is rejected if the p-value of that test is less than 0.05 (0.05
13 considered as significant). As we observe from Table 2, all the clusters based on single factor rainfall,
14 relative humidity, minimum temperature, sunshine hour, as well as wind speed and based on the
15 combination of all factors are useful. While the clusters based on cloud cover and minimum
16 temperature are doubtful and on sea-level pressure is suspected.

17 **4.4 Climatological and Geographical Description of Climate Zones**

18 Climate zones based on rainfall: Between two clusters, cluster 1 mostly contains plain and less rain-
19 prone regions and situated in the northern part of Bangladesh while cluster 2 is located in the southern
20 part and contains mostly hilly and most rain-prone regions in Bangladesh. According to the results in
21 Table 3, cluster 1 has a lower amount of average rainfall with lower standard deviation and lower
22 maximum and minimum values compared to other clusters.

23
24 Climate zones based on relative humidity: Here, Bangladesh has been divided into six climate zones.
25 Cluster 1 located in the central part of the country and contains drought-prone climate stations, and the

1 stations with mild climate from the western part build cluster 2. However, all the climate stations
2 located in the southern part of Bangladesh and closed to the sea build cluster 3 and cluster 4 contains
3 climate stations from the eastern part. The climate stations from northeast and southeast formed cluster
4 5 while cluster 6 contains climate stations from the center and southern part of the country. Based on
5 the results from Table 3, cluster 3 has the highest amount of average relative humidity and followed
6 by clusters 4 to 6, whereas, a lower amount of average relative humidity has been found in cluster 1.

7
8 Climate zones based on minimum temperature: Cluster 1 is formed in the northern part of Bangladesh,
9 which contains mostly the plain but climatologically vulnerable regions. On the other hand, most of
10 the climate stations in cluster 2 are from the coastal site near the sea. Cluster 1 has a higher amount of
11 average minimum temperature than the other one (see Table 3).

12
13 Climate zones based on maximum temperature: Three homogeneous climate zones have been found
14 based on maximum temperature. Among them, cluster 1 contains comparatively mild climate regions
15 close to the north, east, and western border of Bangladesh. Mostly vulnerable climate stations, from
16 the central position of the country, formed cluster 2. However, cluster 3 contains climate stations from
17 the coastal regions and moderately vulnerable climate stations. It is apparent in Table 3 that cluster 2
18 has a lower amount of annual maximum temperature and is followed by clusters 3 and 1.

19
20 Climate zones based on cloud cover: The climate stations from northeast and southeast regions build
21 cluster 1 which are the most rain-prone regions of Bangladesh. Cluster 2 mainly contains climate
22 stations from the central area which are mostly climatologically vulnerable while cluster 3 has been
23 constructed by the climate stations from the northwest and southwest part of Bangladesh. Cluster 1 has
24 the highest amount of average, minimum, and maximum cloud cover with less variation and followed
25 by clusters 2 and 3.

26

1 Climate zones based on sea level pressure: Based on the sea level pressure, two clusters have been
2 found. Only a single climate station formed cluster 1, which is located on the coastal area of
3 Bangladesh, whereas, remaining 33 climate stations construct cluster 2 from all over the country. It is
4 remarkable from Table 2 that cluster 2 is not a homogeneous climate zone.

5
6 Climate zones based on the sunshine hour: Here, five homogeneous climate zones have been found.
7 Among them, cluster 1 usually contains climate stations from the northwest part of the country while
8 most of the climate stations in cluster 2 are from southwest and southeast. Moreover, each of clusters
9 3 and 4 contains two climate stations from the southeast part of the country but cluster 3 is more close
10 to the sea in the coastal region. The climate stations from the hilly and coastal regions which are
11 commonly rain-prone regions of Bangladesh, build cluster 5. Cluster 3 has the highest amount of
12 average sunshine hour followed by clusters 1, 2, 5, 4 (see in Table 3).

13
14 Climate zones based on wind speed: Among the two climate zones, cluster 2 contains climate stations
15 mostly from the southern part which are from the coastal region and close to the sea with the highest
16 amount of annual average, minimum and maximum wind speed (see Table 3). On the other hand,
17 climate stations from the rest of the country build cluster 1 with a lower amount of monthly average,
18 minimum, and maximum wind speed (see Table 3).

19
20 Climate zones based on the combination of climate variables: Here, Bangladesh has been divided into
21 two homogeneous climate regions. Among them, the first region (cluster 1) contains cold and drought-
22 prone areas mostly from the northern and central parts of the country. Contrariwise, most of the climate
23 stations from highland or hilly regions, and the coastal part, which are heavy rain-prone regions of
24 Bangladesh formed cluster 2. According to the result from Table 3, the first climate zone (cluster 1)
25 has a lower amount of monthly average rainfall, relative humidity, cloud cover, sea level pressure,
26 sunshine hour, wind speed, minimum, and maximum temperature compared to cluster 2.

1 **5 Conclusion**

2 Determining the regional disparities among the climate regions of a country is an important tool to
3 know about the fundamental features of the geographical environment. This knowledge is crucial to
4 develop water management strategies among other important factors. In this study, some of the
5 important meteorological variables have been considered to study the regional climatological
6 disparities in Bangladesh using a multivariate approach. The results of the statistical validity tests
7 indicate that the hierarchical clustering approach is one of the best-fitted algorithms for most of the
8 climatological features of Bangladesh. The findings of this approach provide two clusters based on
9 rainfall, sea level pressure, and wind speed, six clusters based on humidity and temperature, three
10 clusters based on cloud cover, and five clusters based on sunshine hour. The membership of climate
11 stations for different clusters is not identical for all the climatological variables, which implies the
12 variations are present among the climatological features across the country. The findings of this study
13 provide a concise assessment of the properties of different climate regions of Bangladesh that will help
14 govt. and policymakers to take proper strategies in water management, disaster management, climate
15 protection, etc.

16

17 **Acknowledgment**

18 I would like to express my deep sincere gratitude to QMH. Professor Dr. Ataharul Islam for his
19 valuable and constructive suggestions during the planning and development of this research work. In
20 addition, I would like to thank the Bangladesh Meteorological Department (BMD) for providing the
21 necessary data for this research.

22 **Ethical Approval**

23 The study does not include any human subjects. Hence, no ethical approval was required for this study.

24 **Consent to Participate**

1 The datasets used in this study were collected by the Bangladesh Meteorological Department (BMD)
2 from the designated locations. Moreover, no human subjects were the participants of this study.
3 Therefore, no informed consent to participate was required.

4 **Consent to Publish**

5 It is acknowledged that the persons named at the beginning are the sole authors of the work, that no
6 part of this work has been published elsewhere, and that this work will not be submitted elsewhere
7 until the Journal's editorial process is completed. Also, the owners hereby grant the Journal a global,
8 perpetual, non-exclusive, royalty-free license to publish or distribute the work, to enter into publishing
9 or distribution agreements with others, and to archive the Work.

10 **Authors' Contributions**

11 All authors contributed to the study's conception and design. Specifically, Sultan Mahmud collected
12 the datasets from the BMD and he analyzed the datasets. Ferdausi Mahojabin Sumana conducted a
13 thorough literature review. Sultan Mahmud and Md. Mohsin prepared the first draft with the
14 supervision of Md Hasinur Rahaman Khan. All authors contributed in preparing the final version of
15 the manuscript.

16 **Funding**

17 The authors did not receive financial support from any organization for the submitted work.

18 **Conflict of Interest**

19 The authors do not have any competing interests. The final manuscript was read and approved by all
20 contributors.

21 **Availability of Data and Materials**

22 The datasets used in this study are not publicly available since they are the property of the Bangladesh
23 Meteorological Department (BMD), but they can be available only on reasonable request from the
24 corresponding author.

25 **References**

- 1 Abadi, A. M., C. M. Rowe and M. Andrade (2019) Climate Regionalization in Bolivia: A Combination
2 of Nonhierarchical and Consensus Clustering Analyses Based on Precipitation and Temperature. *Int.*
3 *J. Climatol.*
- 4 Ahasan, M., M. M. Chowdhury and D. Quadir (2011) Prediction of high impact rainfall events of
5 summer monsoon over Bangladesh using high resolution MM5 model. *Sri Lanka Journal of Physics*
6 12: 43-58.
- 7 Ahmad, N., I. Othman and S. Deni (2013) Hierarchical cluster approach for regionalization of
8 Peninsular Malaysia based on the precipitation amount. *Journal of Physics: Conference Series*, IOP
9 Publishing.
- 10 Ahmed, R. and I.-K. Kim (2003) Patterns of daily rainfall in Bangladesh during the summer monsoon
11 season: case studies at three stations. *Physical Geography* 24(4): 295-318.
- 12 Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *Journal of climatology*
13 6(6): 661-675.
- 14 Ankerst, M., M. M. Breunig, H.-P. Kriegel and J. Sander (1999) OPTICS: ordering points to identify
15 the clustering structure. *ACM Sigmod record* 28(2): 49-60.
- 16 Basalirwa, C. (1995). Delineation of Uganda into climatological rainfall zones using the method of
17 principal component analysis. *Int. J. Climatol.*15(10): 1161-1177.
- 18 Batistakis, Y. H. and M. Vazirgiannis (2001) On clustering validation techniques. *Journal of Intelligent*
19 *Information Systems* 17(107): 2-3.
- 20 Bhat, A. (2014) K-medoids clustering using partitioning around medoids for performing face
21 recognition. *International Journal of Soft Computing, Mathematics and Control* 3(3): 1-12.
- 22 Brock, G., V. Pihur, S. Datta and S. Datta (2011) clValid, an R package for cluster validation. *Journal*
23 *of Statistical Software* (Brock et al., March 2008).
- 24 Buishand, T. A. (1982) Some methods for testing the homogeneity of rainfall records. *Journal of*
25 *hydrology* 58(1-2): 11-27.
- 26 Cannarozzo, M., L. Noto, F. Viola and G. La Loggia (2009) Annual runoff regional frequency analysis
27 in Sicily. *Physics and Chemistry of the Earth, Parts A/B/C* 34(10-12): 679-687.
- 28 Chen, B., P. C. Tai, R. Harrison and Y. Pan (2005) Novel hybrid hierarchical-K-means clustering
29 method (HK-means) for microarray analysis. 2005 IEEE Computational Systems Bioinformatics
30 Conference-Workshops (CSBW'05), IEEE.
- 31 De Souto, M.C., De Araujo, D.S., Costa, I.G., Soares, R.G., Ludermir, T.B. and Schliep, A., 2008,
32 June. Comparative study on normalization procedures for cluster analysis of gene expression datasets.
33 In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on
34 Computational Intelligence) (pp. 2792-2798). IEEE.
- 35 Di Giuseppe, E., G. J. Lasinio, S. Esposito and M. Pasqui (2013) Functional clustering for Italian
36 climate zones identification *Theoretical and applied climatology* 114(1-2): 39-54.
- 37 Dikbas, F., M. Firat, A. C. Koc and M. Gungor (2012) Classification of precipitation series using fuzzy
38 cluster method. *Int. J. Climatol.*, 32(10): 1596-1603.
- 39 Dubes, R. and A. K. Jain (1979) Validity studies in clustering methodologies. *Pattern recognition*
40 11(4): 235-254.
- 41 Fovell, R. G. and M.-Y. C. Fovell (1993) Climate Zones of the Conterminous United States Defined
42 Using Cluster Analysis. *Journal of Climate* 6(11): 2103-2135.
- 43 Giraldo, R., P. Delicado and J. Mateu (2012) Hierarchical clustering of spatially correlated functional
44 data. *Statistica Neerlandica* 66(4): 403-421.
- 45 Goyal, M. K. and V. Gupta (2014) Identification of homogeneous rainfall regimes in Northeast Region
46 of India using fuzzy cluster analysis. *Water resources management* 28(13): 4491-4511.
- 47 Goyal, M.K., Shivam, G. & Sarma, A.K. Spatial homogeneity of extreme precipitation indices using
48 fuzzy clustering over northeast India. *Nat Hazards* 98, 559-574 (2019).
49 <https://doi.org/10.1007/s11069-019-03715-z>

- 1 Hassan, B.G. and Ping, F. (2012) Formation of homogenous regions for Luanhe basin-by using L-
2 moments and cluster techniques. *International Journal of Environmental Science and Development*,
3 3(2), 205.
- 4 Handl, J., J. Knowles and D. B. Kell (2005) Computational cluster validation in post-genomic data
5 analysis. *Bioinformatics* 21(15): 3201-3212.
- 6 Haque, A. and S. Jahan (2015) Impact of flood disasters in Bangladesh: A multi-sector regional
7 analysis. *International journal of disaster risk reduction* 13: 266-275.
- 8 Hargrove, W. W. and F. M. Hoffman (1999) Using multivariate clustering to characterize ecoregion
9 borders. *Computing in science & engineering* 1(4): 18-25.
- 10 Huq, S. (2001) *Climate change and Bangladesh*, American Association for the Advancement of
11 Science.
- 12 Huq, S., Z. Karim, M. Asaduzzaman and F. e. Mahtab (1999) *Vulnerability and adaptation to climate
13 change for Bangladesh*, Springer.
- 14 Islam, M. N., T. Terao, H. Uyeda, T. Hayashi and K. Kikuchi (2005) Spatial and temporal variations
15 of precipitation in and around Bangladesh. *Journal of the Meteorological Society of Japan. Ser. II*
16 83(1): 21-39.
- 17 Iyigun, C., M. Türkeş, İ. Batmaz, C. Yozgatligil, V. Purutçuoğlu, E. K. Koç and M. Z. Öztürk (2013)
18 Clustering current climate regions of Turkey by using a multivariate statistical method. *Theoretical
19 and applied climatology* 114(1-2): 95-106.
- 20 Johnson, R. A. and D. W. Wichern (2002) *Applied multivariate statistical analysis*, Prentice hall Upper
21 Saddle River, NJ.
- 22 Kaufman, L. and P. J. Rousseeuw (1987) Clustering by means of medoids. *Statistical Data Analysis
23 based on the L1 Norm*. Y. Dodge, Ed: 405-416.
- 24 Kyselý, J., J. Pícek and R. Huth (2007) Formation of homogeneous regions for regional frequency
25 analysis of extreme precipitation events in the Czech Republic. *Studia Geophysica et Geodaetica*
26 51(2): 327-344.
- 27 MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations.
28 *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA,
29 USA.
- 30 Mahmood, S. A. I. (2012) Impact of climate change in Bangladesh: The role of public administration
31 and government's integrity. *Journal of Ecology and the Natural Environment* 4(8): 223-240.
- 32 Mahmud, S. and M. A. Islam (2019) Predictive ability of covariate-dependent Markov models and
33 classification tree for analyzing rainfall data in Bangladesh. *Theor Appl Climatol* 138(1-2): 335-346.
34 <https://doi.org/10.1007/s00704-019-02812-0>
- 35 Mahmud, S., Islam, M.A. & Hossain, S.S (2020) Analysis of rainfall occurrence in consecutive days
36 using Markov models with covariate dependence in selected regions of Bangladesh. *Theor Appl
37 Climatol* 03159-7. <https://doi.org/10.1007/s00704-020->
- 38 Mannan, M. A. and S. Karmakar (2007). Climatic features of Heavy Rainfall activities in monsoon
39 season and its socio-economic impact in Bangladesh. *Proceedings of SAARC Seminar on Application
40 of Weather and Climate Forecasts in the Socio-economic Development and Disaster Mitigation*.
- 41 Mosley, M.P., (1981) Delimitation of New Zealand hydrological regions. *Journal of Hydrology* 49,
42 173–192.
- 43 Matulla, C., K. P. Edouard, P. Haas, and H. Formayer, (2003) Comparative analysis of spatial and
44 seasonal variability: Austrian precipitation during the 20th century. *Int. J. Climatol.*, 23, 1577–1588
- 45 Mishra, V. and Aadhar, S., 2020. Discussion of “Multivariate Modeling of Projected Drought
46 Frequency and Hazard over India” by Vivek Gupta, Manoj Kumar Jain, and Vijay P. Singh. *Journal
47 of Hydrologic Engineering*, 26(2), p.07020024.
- 48 Mingoti, S. A. and J. O. Lima (2006) Comparing SOM neural network with Fuzzy c-means, K-means
49 and traditional hierarchical clustering algorithms. *European journal of operational research* 174(3):
50 1742-1759.

- 1 Modarres, R. (2006) Regional precipitation climates of Iran. *Journal of Hydrology (New Zealand)*: 13-
2 27.
- 3 Ouyang, R., L. Ren, W. Cheng and C. Zhou (2010) Similarity search and pattern discovery in
4 hydrological time series data mining. *Hydrological Processes: An International Journal* 24(9): 1198-
5 1210.
- 6 Pettitt, A. (1979) A non-parametric approach to the change-point problem. *Journal of the Royal*
7 *Statistical Society: Series C (Applied Statistics)* 28(2): 126-135.
- 8 Poncelet, A., F. Gemenne, M. Martiniello and H. Bousetta (2010) A country made for disasters:
9 environmental vulnerability and forced migration in Bangladesh. *Environment, forced migration and*
10 *social vulnerability*, Springer: 211-222.
- 11 Rahman, A., M. J. H. Jiban and S. A. Munna (2015) Regional variation of temperature and Rainfall in
12 Bangladesh: Estimation of Trend. *Open Journal of Statistics* 5(07): 652.
- 13 Rahman, M. M., S. Sarkar, M. R. Najafi and R. Rai (2013) Regional extreme rainfall mapping for
14 Bangladesh using L-moment technique. *Journal of Hydrologic Engineering* 18(5): 603-615.
- 15 Rodriguez, M. Z., C. H. Comin, D. Casanova, O. M. Bruno, D. R. Amancio, L. d. F. Costa and F. A.
16 Rodrigues (2019). "Clustering algorithms: A comparative approach." *PloS one* 14(1).
- 17 Sahin, S. and H. K. Cigizoglu (2010) Homogeneity analysis of Turkish meteorological data set.
18 *Hydrological Processes: An International Journal* 24(8): 981-992.
- 19 Sannakki, S. S., V. S. Rajpurohit, V. Nargund, A. Kumar and P. S. Yallur (2011) Leaf disease grading
20 by machine vision and fuzzy logic. *Int J* 2(5): 1709-1716.
- 21 Shahid, S. (2010) Recent trends in the climate of Bangladesh. *Climate Research* 42(3): 185-193.
- 22 Shahid, S. (2011) Impact of climate change on irrigation water demand of dry season Boro rice in
23 northwest Bangladesh. *Climatic change* 105(3-4): 433-453.
- 24 Shahid, S. and H. Behrawan (2008) Drought risk assessment in the western part of Bangladesh. *Natural*
25 *hazards* 46(3): 391-413.
- 26 Siraj-Ud-Doulah, M. and M. N. Islam (2019) Defining Homogenous Climate zones of Bangladesh
27 using Cluster Analysis. *International Journal of Statistics and Mathematics* 6(1): 119-129.
- 28 chcUnal, Y., T. Kindap and M. Karaca (2003) Redefining the climate zones of Turkey using cluster
29 analysis. *Int. J. Climatol.*, 23(9): 1045-1055.
- 30 Ward J (1963) Hierarchical groupings to optimize an objective function. *J Am Stat Assoc* 58:236–244
- 31 White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: issues and
32 guidance for practice. *Stat Med* 30(4):377–399
- 33 Wijngaard, J., A. Klein Tank and G. Können (2003) Homogeneity of 20th century European daily
34 temperature and precipitation series. *Int. J. Climatol.* 23(6): 679-692.
- 35 Willis, K. J. and S. A. Bhagwat (2009) Biodiversity and climate change. *Science* 326(5954): 806-807.
- 36 Xiong, J., R. Yao, S. Grimmond, Q. Zhang and B. Li (2019) A hierarchical climatic zoning method
37 for energy efficient building design applied in the region with diverse climate characteristics. *Energy*
38 *and Buildings* 186: 355-367.
- 39 Yue, S. and C. Wang (2004) The Mann-Kendall test modified by effective sample size to detect trend
40 in serially correlated hydrological series. *Water resources management* 18(3): 201-218.
- 41 Zhang, Y., S. Moges and P. Block (2016) Optimal cluster analysis for objective regionalization of
42 seasonal precipitation in regions of high spatial–temporal variability: application to Western Ethiopia.
43 *Journal of Climate* 29(10): 3697-3717.

Figures

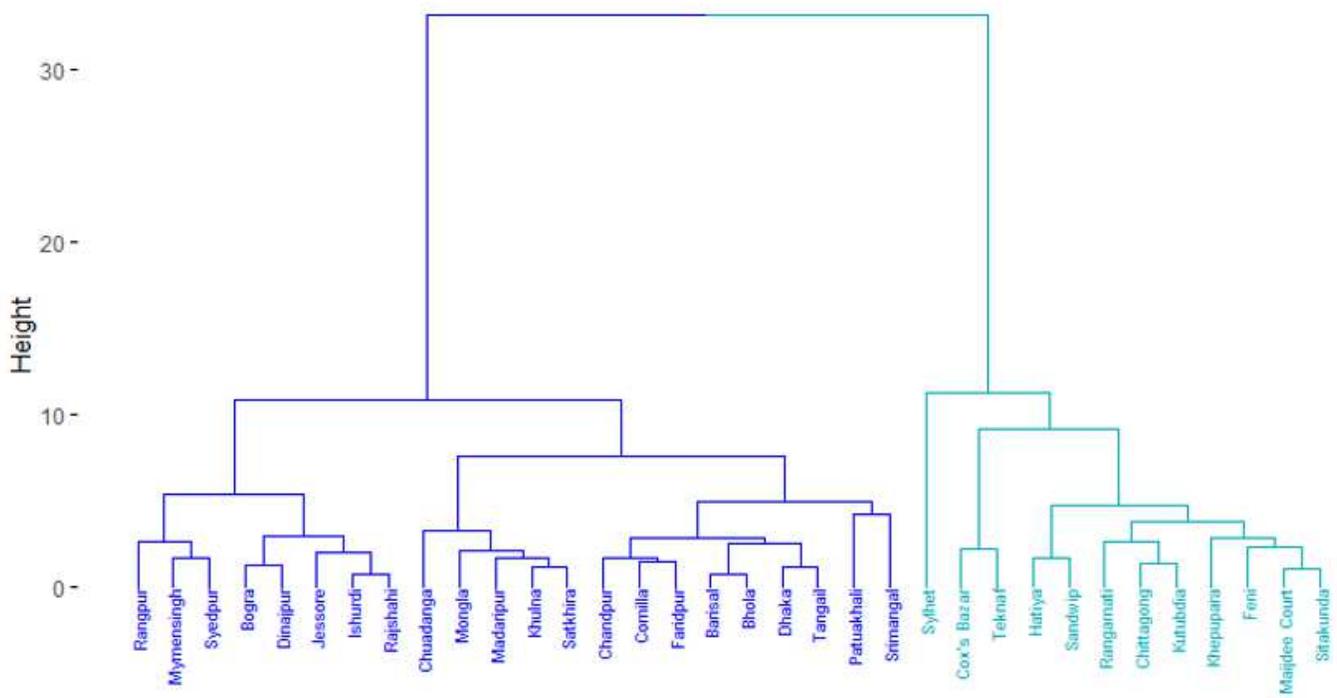


Figure 1

Cluster of climate stations based on monthly average rainfall. Two groups or clusters were found. Cluster 1 include 22 climate stations while 12 climate stations are in cluster or group 2

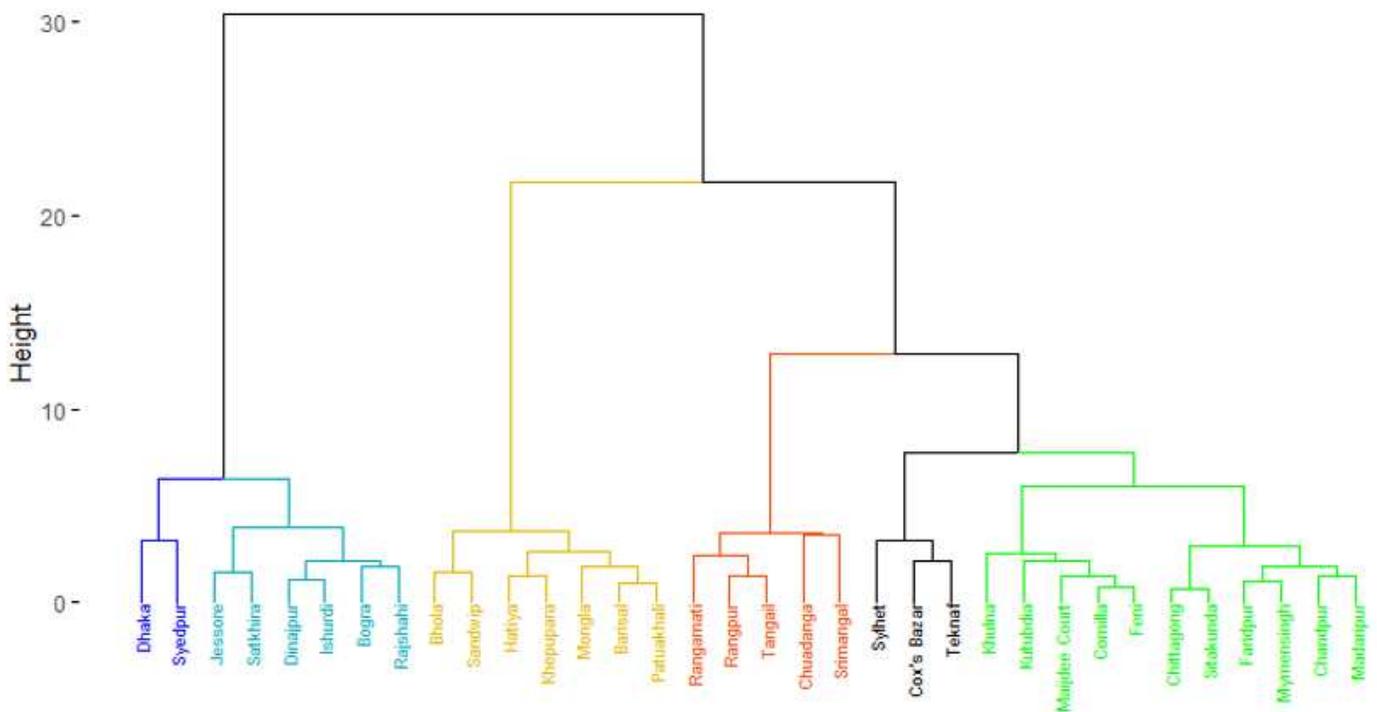


Figure 2

Cluster of climate stations based on relative humidity. Six cluster were found. Cluster 1 contain 2 stations, cluster 2 contain 6 stations, cluster 3 contain 7 stations, cluster 4 contain 5, cluster 4 contain 3, and cluster 6 contain 11 stations.

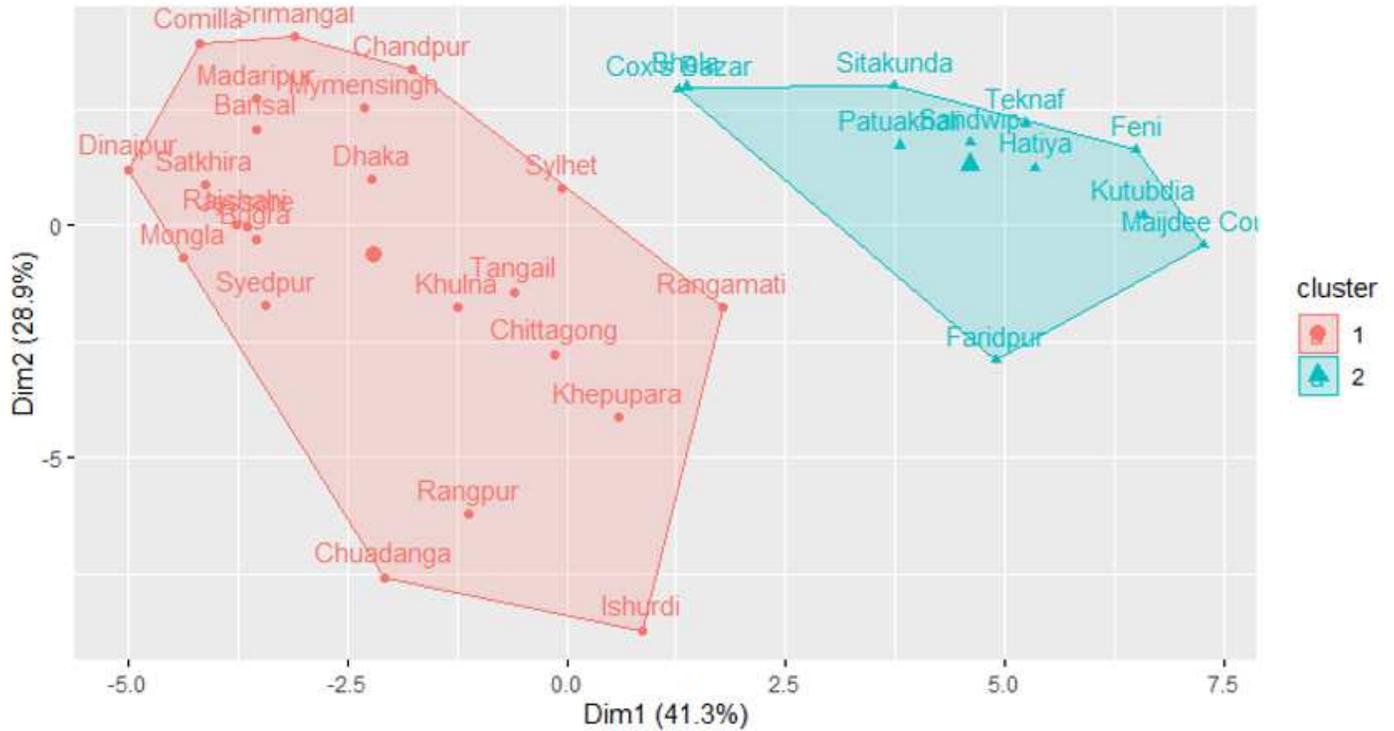


Figure 3

Cluster of climate stations based on combination of monthly maximum and average of maximum temperature. Two clusters were found. Cluster 1 includes 23 climate stations and cluster 2 includes 11 climate stations.

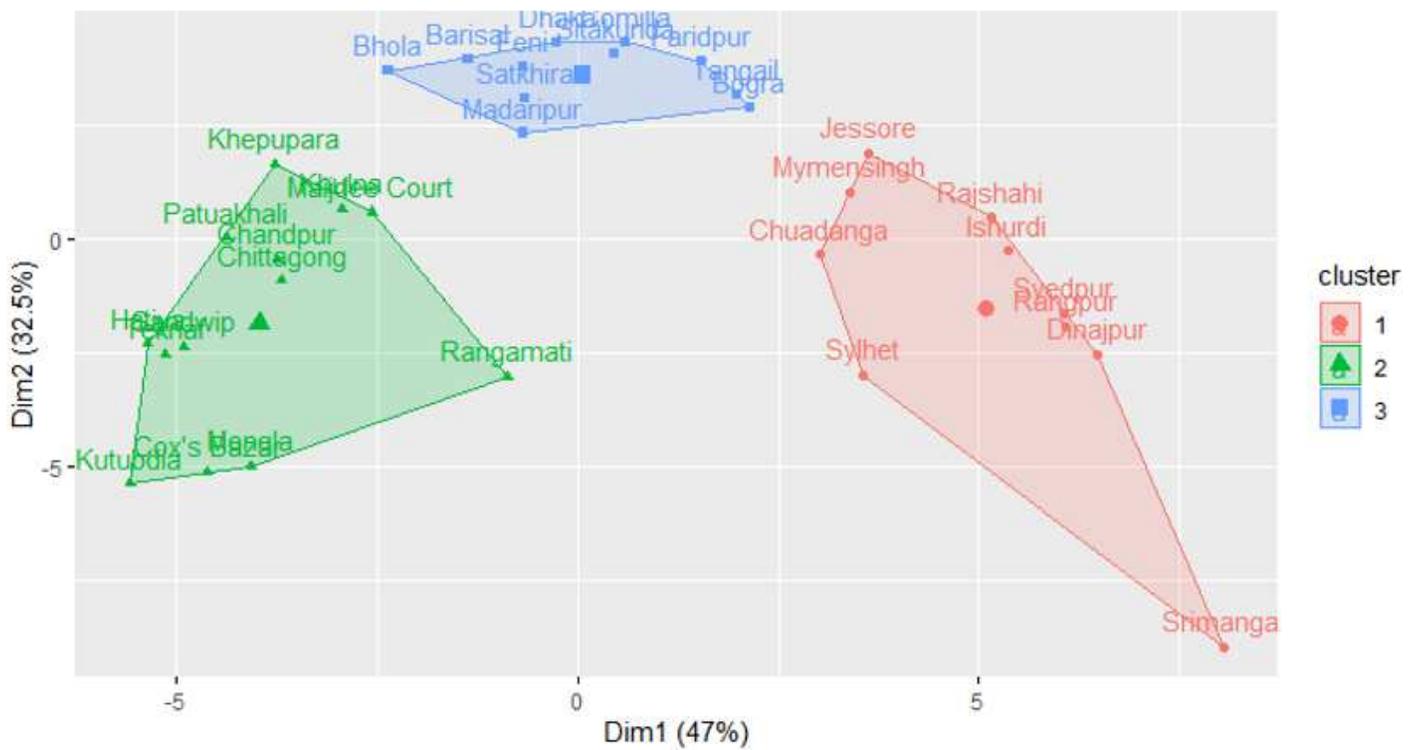


Figure 4

Cluster of climate stations based on combination of monthly minimum and average of minimum temperature. Three clusters were found. Cluster 1 includes 11 stations while 13 climate stations belong in cluster 2 and 10 climate stations in cluster 3.

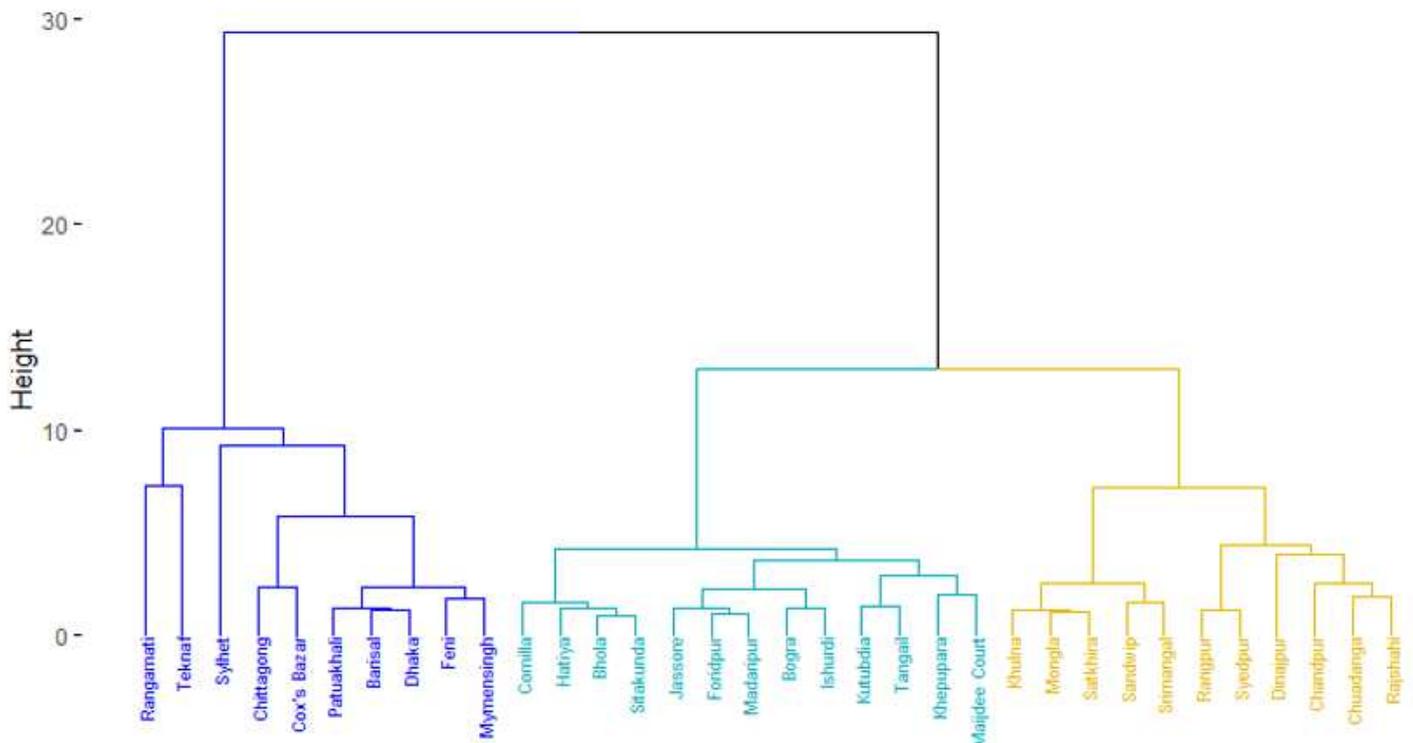


Figure 5

Cluster of climate region based on monthly average cloud cover. Three clusters or groups are found. Cluster 1 contain 10 climate stations while 13 climate stations are in cluster 2 and 11 climate stations are in cluster 3

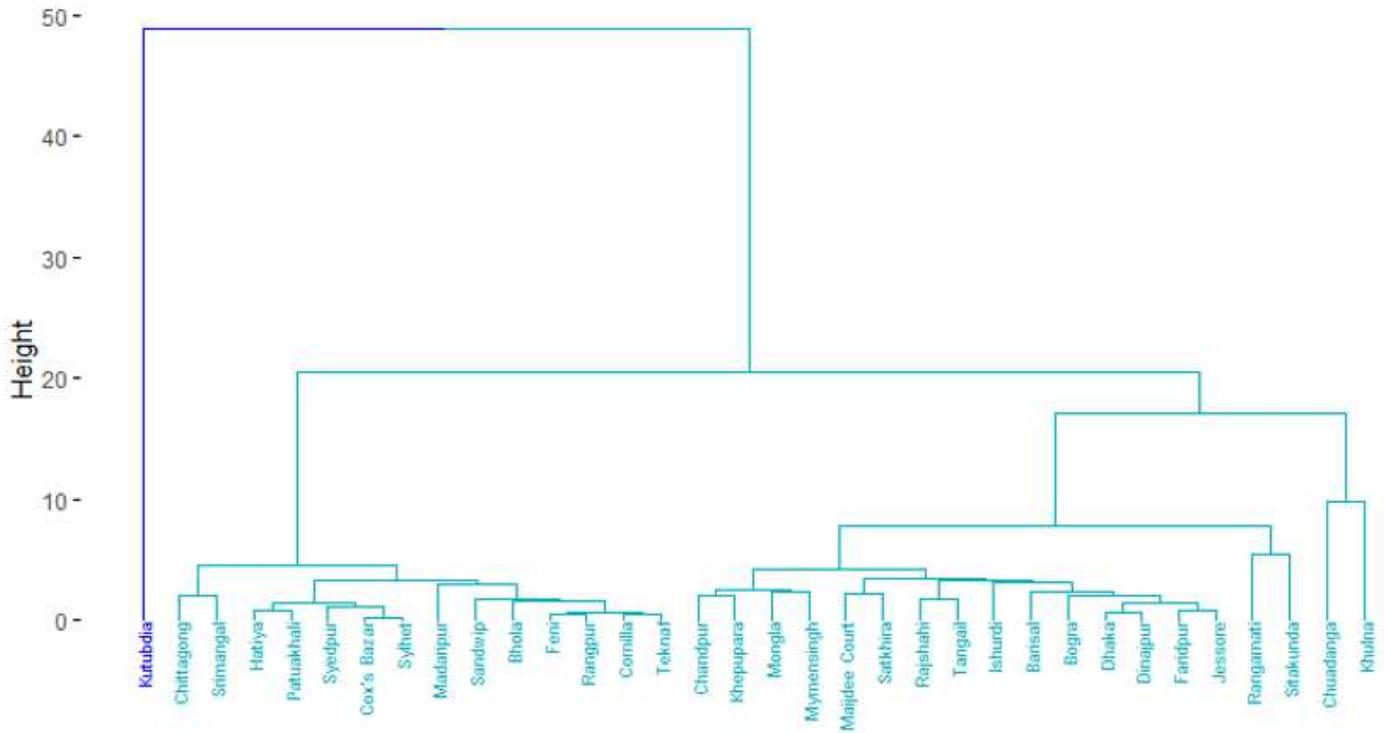


Figure 6

Cluster of climate regions based on monthly average sea level pressure. Two clusters or groups were found. Cluster 1 contains only one station while rest of the climate stations (33) belong to cluster 2

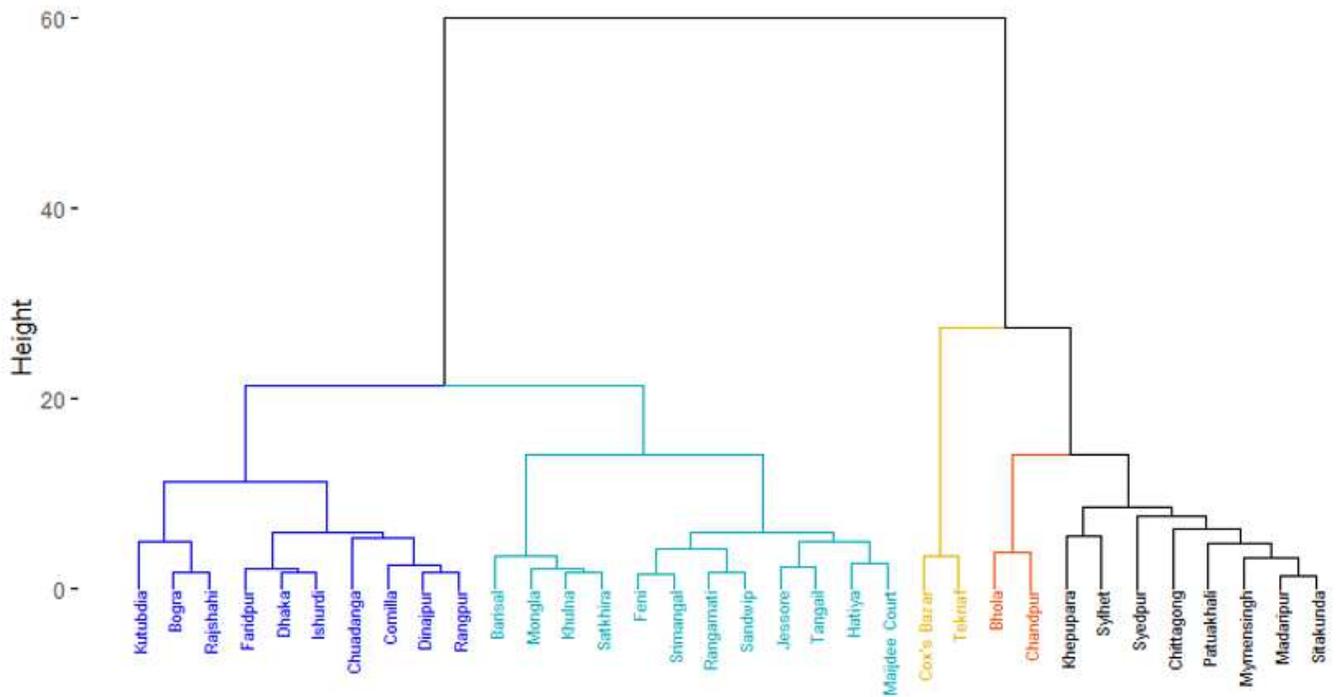


Figure 7

Cluster of climate regions based on monthly average sunshine hour. Five clusters were found. Cluster 1 contains 10 stations while 12 climate stations belong in cluster 2, 2 climate stations in cluster 3, 2 climate stations in cluster 4, and 8 climate stations in cluster 5.

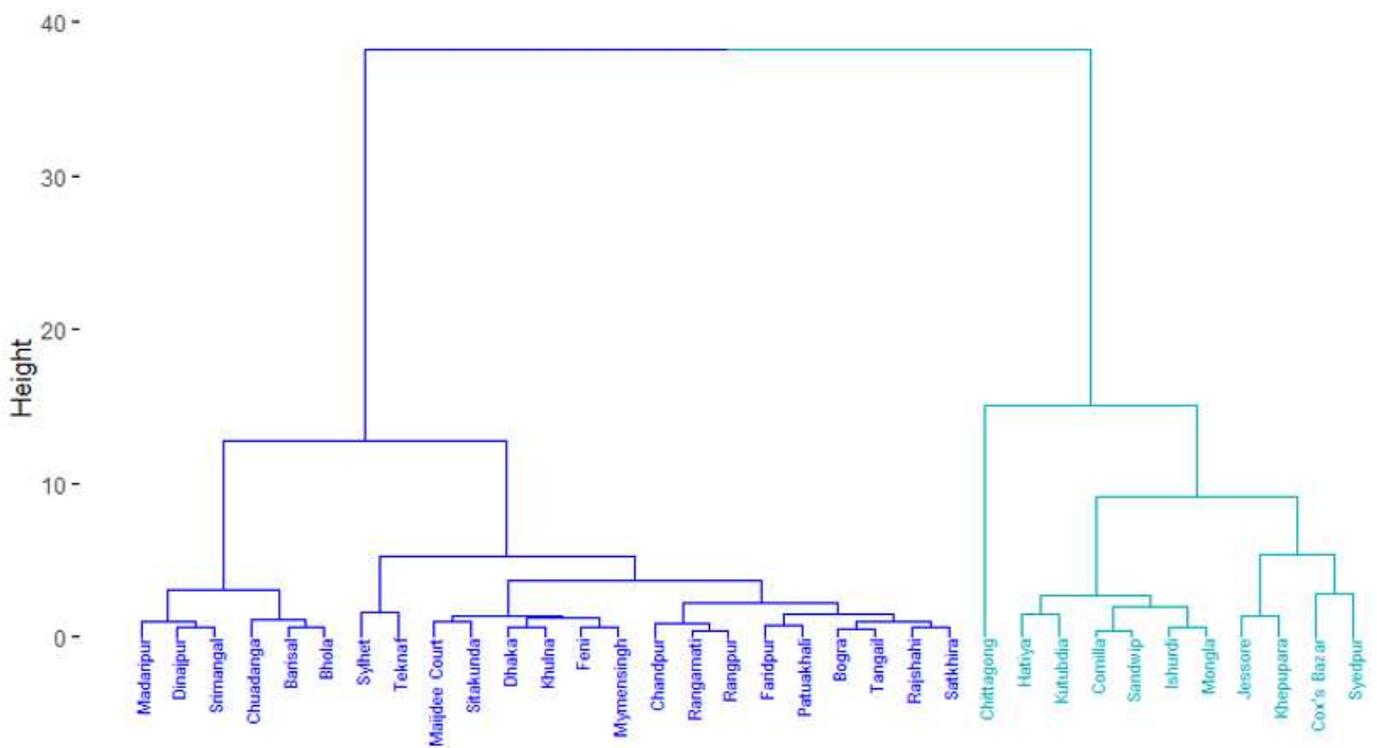


Figure 8

Cluster of climate regions based on monthly average wind speed (m/s). Two clusters or groups were found. Cluster 1 contains 23 climate stations while rest 11 belong to cluster 2

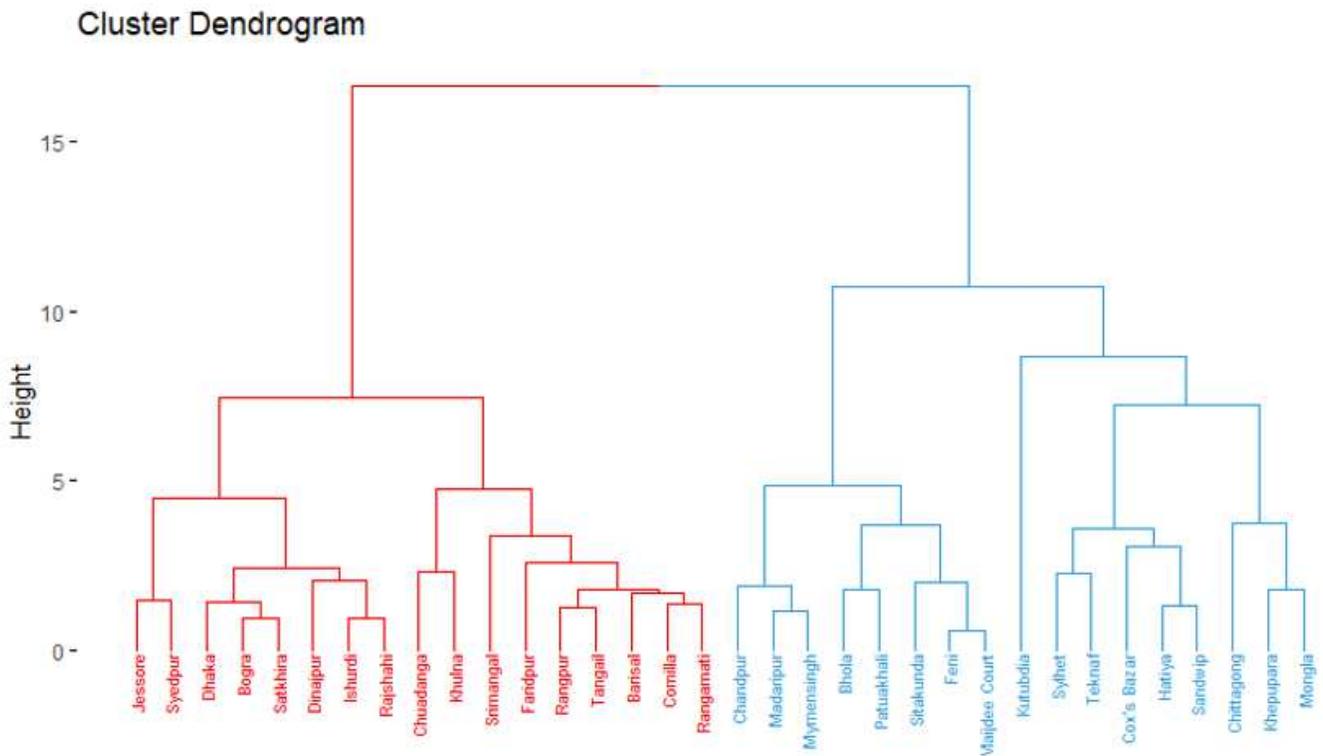


Figure 9

Cluster of climate stations based on all climatological factors combined. Two cluster were found. Each cluster contains 17 climate stations

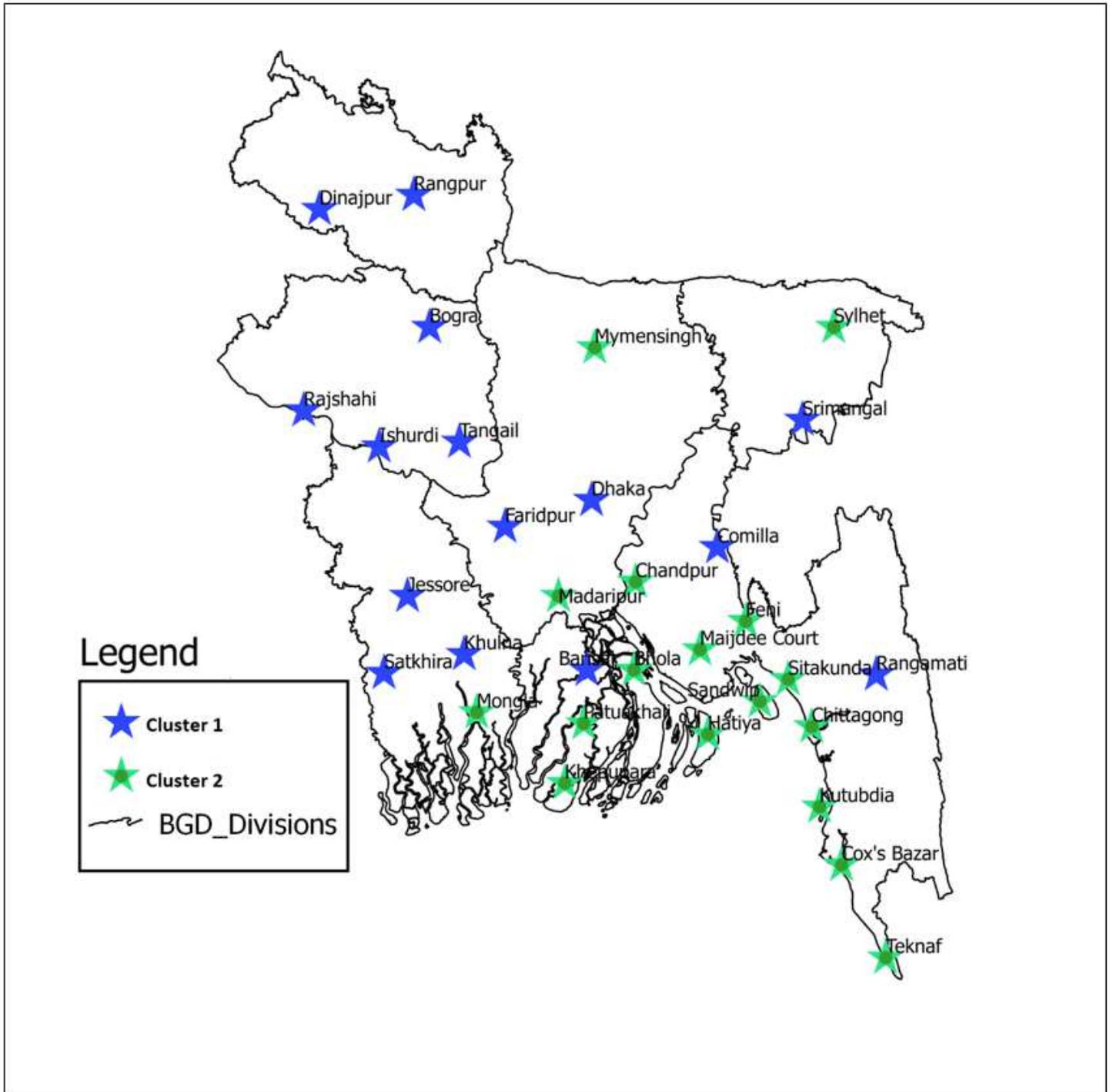


Figure 10

Distribution of clustered climate stations of Bangladesh based on all combined climatological factors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.pdf](#)