Nigerian Journal of Theoretical and Environmental Physics (NJTEP)
ISSN Print: 3026-9601
DOI: https://doi.org/10.62292/njtep.v2i1.2024.19
Volume 2(1), March 2024

Hierarchical and Non-Hierarchical Cluster Classification of Precipitation Time Series Data in Nigeria

***[1]Agada, I. O., [2]Peter, O. and [1]Eweh, E. J.**

[1]Department of Physics, Federal University of Agriculture, Makurdi. Benue State, Nigeria.
[2]Department of Statistics, Federal University of Agriculture, Makurdi. Benue State, Nigeria.

*Corresponding author's email: agadainikpi020@gmail.com

ABSTRACT

Precipitation is highly unpredictable over space and time, and is an indicator for climate change. The present work classified average yearly precipitation over the 36 states and the Federal Capital Territory (FCT) in Nigeria into various groups. The hierarchical and non- hierarchical (k-means) cluster method was used in the classification of 37- years precipitation data (1984-2020). The results showed that six major clusters were formed. The Precipitation across the 36 states and the FCT in Nigeria can be classified as Light precipitation for Sokoto, Borno, Kastina, Yobe and Kano States; Moderate precipitation for Kebbi, Zamfara, Dutse, Yola, Gombe and Bauchi States; Heavy precipitation for Kaduna, Plateau, Kwara, Taraba, Nassarawa, Benue and Kogi States; Very Heavy precipitation for Osun, Ogun, FCT, Ekiti, Niger and Lagos States; Shower for Edo, Abia, Delta, Anambra, Ebonyi, Enugu, Ondo and Oyo States and lastly, Heavy Shower for Akwaibom, Imo, Rivers, Bayelsa and Cross River States.. The rangs of precipitation based on the six classification are 1.5-1.8mm (Light precipitation), 1.9-2.7mm (Moderate precipitation), 2.8-3.6mm (Heavy precipitation), 3.7-4.3mm (Very Heavy precipitation), 4.4-5.4mm (Shower) and greater than 5.5mm (Heavy Shower). The hierarchical cluster method performed better than the non- hierarchical (k-means) cluster method in structuring the precipitation data and shaping it into groups (clusters) such that observations within the same group are more similar than those in different groups. The results of this study can contribute to a better understanding of the climate condition of the36 States and the FCT in Nigeria.

Keywords:
Weather,
Precipitation**,**
Hierarchical cluster analysis,
K-means,
Nigeria,
Rainfall,
Climate.

INTRODUCTION

Precipitation is the meteorological phenomenon that is unpredictable in time and space since its structure and intensity are determined by various atmospheric conditions. Precipitation is the universal term for rainfall, snowfall, and other forms of frozen or liquid water falling from clouds and it varies from year to year, and changes in intensity, frequency, amount and type. Precipitation depends greatly on temperature and the weather situation of a place and it occurs when hydrometeors are large and heavy enough to fall to the Earth's surface. Hydrometeors are liquid and ice particles that form in the atmosphere and it ranges from small cloud droplets and ice crystals to large hailstones. The microphysics of precipitation-particle formation is affected by nucleation, diffusion, super-saturation and collision (Stull, 2017). The formation of new liquid or solid hydrometeors as water vapor attaches to tiny dust particles carried in the air is known as nucleation. Diffusion is the random movement of water-vapor molecules through the air toward existing hydrometeors. Super-saturation specifies the amount of excess water vapor available to form rain and snow. These particles are called cloud condensation nuclei or ice nuclei. Collision between two hydrometeors allows them to combine into larger particles. These processes affect liquid water and ice differently (Stull, 2017).

Precipitation can have wide-ranging effects on human well-being, ecosystems, the amount of surface water and groundwater available for drinking, irrigation, and industry. Changes in precipitation can upset a wide range of natural processes, mostly if these changes occur more quickly than animal and plant species can adapt. As average temperatures at the Earth's surface rise, more evaporation occurs, which in turn increases overall precipitation of a place. As precipitation patterns vary across the regions, so will the precipitation effects of climate change. The potential impacts of heavy precipitation include crop damage, soil erosion, and an increase in flood risk (Duan and Kausha, 2013).

Several methods have been used in investigating rainfall patterns (Şen and Habib, 2001; Jebari *et al.*, 2009; Sabziparvar *et al.*, 2015; Lin *et al.*, 2017; Gupta *et al.*, 2017; Haines and Olley, 2017). Amissah-Arthur and Jagtap (1999) applied both Principal Component Analysis (PCA) and Cluster Analysis to seasonal rainfall in Nigeria in order to cluster 23 stations into six groups. Nnaji *et al.* (2016) grouped 24 rainfall stations into five clusters in Nigeria based on the co-efficient of variation (CV) through hierarchical cluster analysis (HCA). Machiwal *et al.* (2017) classified 32 stations in hot and cold Indian arid regions into entirely and partially arid clusters based on 102 years (1901–2002) rainfall using k-means clustering. Machiwal *et al.* (2019) used hierarchical cluster analysis (HCA) to delineate the spatial patterns of monthly, seasonal and annual rainfall by clustering 62 stations in the western arid region of India based on a 55 year (1957–2011) data set. The results show that Cluster I is characterized by the lowest mean rainfall (western portion), whereas mean rainfall was the highest for cluster IV situated in the eastern portion. Eruola *et al.* (2021) assessed rainfall extremes for agricultural overview in Nigeria using trend analysis and probability of exceedance. The annual rainfall trend indicated variability in the six geopolitical regions with North-East having the lowest range and South-South area with highest. Ologunorisa and Tersoo, 2006; Enete and Ebenebe, 2009; Okonkwo and Mbajiorgu, 2010; Adejuwon, 2011; Azuwike and Enwereuzor, 2011; Ekpoh and Nsa, 2011; Abaje *et al.*, 2012; Babatolu, 2014; Ekwe *et al.*, 2014 studied spatial distribution of rainfall patterns in Nigeria. Researchers are mostly faced with a large number of observations that can be meaningless unless classified into manageable groups.

The comparison of hierarchical and non- hierarchical cluster classification method of precipitation data in the 36 States and the FCT, Nigeria has rarely been studied. Most research works carried out in Nigeria is centered on rainfall trend and not precipitation. The primary objective of this study is to classify 37-years (1984-2020) precipitation data of the 36 states and the FCT in Nigeria by placing the similar observations into groups using Cluster analysis. Cluster analysis is a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess. Cluster analysis has been applied to find patterns in climate, psychiatry, biology and so on. Thus, knowledge about the patterns of precipitation is crucial for sustainable agricultural, water management and to detect climate variability in Nigeria.

MATERIALS AND METHODS

Study Area, Data and Methods

Nigeria lies between 4^0 and 14^0N latitude and longitude 4^0 to 14^0E. It is bounded on the north by the Republic of Niger, east by Cameroon and west by Benin Republic while the southern boundary is Gulf of Guinea which is an arm of the Atlantic Ocean. Nigeria is composed of various ecotypes and climatic zones. The Nigerian climate is characterized mainly by the dry north-easterly and the moist south-westerly winds. The main ecological zones are the tropical rainforest along the coast, savannah in the middle belt and semi-arid zones in the northern fringes (Fig. 1). The daily precipitation data was obtained from National Aeronautic and Space Administration (NASA) (Modern Era Retrospective Analysis Version 2 (MERRA-2)) for the period of thirty-seven (37) years (1984-2020).

Figure 1: Map of Nigeria showing the 36 States and the Federal Capital Territory (FCT)

Cluster Analysis

Several clustering techniques begin not with the raw data but with a matrix of inter-individual measures of distance or similarity calculated from the raw data. Similarity signifies the degree of correspondence among objects across all of the uniqueness used in the analysis. It is a set of rules that serve as criteria for separating or grouping items. The distance measure is often used as a measure of similarity, with higher values representing greater dissimilarity (distance between cases), not similarity instead of the Correlation measure. Several distance measures are available, each with specific characteristics. The most commonly distance function is the Euclidean distance between two vectors $X = (x_1, x_2, \ldots\ldots, x_p)'$ and $Y = (y_1, y_2, \ldots\ldots, y_p)'$ defined as:

$$d(x,y) = \sqrt{(x-y)'(x-y)} = \sqrt{\sum_{j=1}^{p}(x_j - y_j)^2} \quad (1)$$

To adjust for differing variances and covariances among the p variables, the statistical distance can be used, which is given as:

$$d(x,y) = \sqrt{(x-y)'S^{-1}(x-y)} \quad (2)$$

where **S** is the sample covariance matrix (Rencher, 2002).

Other distance measures include: Squared Euclidean, City- block (Manhattan), Chebychev and Mahalanobis distance: The Squared Euclidean is the sum of the squared differences without taking the square root. Mathematically:

$$d^2(x,y) = \sum_{j=1}^{p}(x_j - y_j)^2 \quad (3)$$

The city- block (Manhattan) distance uses the sum of the variables' absolute differences while the Chebychev distance is the maximum of the absolute difference in the clustering variables' values. Frequently used when working with metric (or ordinal) data. Lastly, the Mahalanobis distance is a generalized distance measure that accounts for the correlations among variables in a way that weights each variable equally.

There are three different procedures that can be used to cluster data: hierarchical cluster analysis, *k*-means cluster, and two-step cluster. The choice of a method depends on, among other things, the size of the data file. The two-step procedure can be used If the data file is large (even 1,000 cases is large for clustering) or a mixture of continuous and categorical variables. If the data file is small, the hierarchical clustering method can be used and it enables one to easily examine solutions with increasing numbers of clusters. The *k*-means clustering method produces a partition of the data into a particular number of groups set by the investigator and the data file is moderately sized. A combination approach using a hierarchical approach followed by a non-hierarchical approach (K-means) is often advisable. A hierarchical technique is used to select the number of clusters and profile clusters centers that serve as initial cluster seeds in the nonhierarchical procedure. While a non-hierarchical method (K-means) then clusters all observations using the seed points to provide more accurate cluster memberships. In this way, the advantages of hierarchical methods are complemented by the ability of the nonhierarchical methods to refine the results by allowing the switching of cluster membership.

Hierarchical Clustering

Hierarchical methods and other clustering algorithms represent an attempt to find "good" clusters in the data using a computationally efficient technique. The number of ways of partitioning a set of *n* items into *g* clusters is given by:

$$N(n,g) = \frac{1}{g!}\sum_{k=1}^{g}\binom{g}{k}(-1)^{g-k}k^{n} \qquad (4)$$

The hierarchical clustering method is also called the agglomerative hierarchical methods. Agglomerative hierarchical clustering methods begin by placing each observation into a separate cluster. Clusters are then joined, two at a time, until the number of clusters is reduced to the desired target. The distances between two clusters include: Nearest neighbor (single linkage), Furthest neighbor (complete linkage), Centroid, Median, Group average (average linkage) and Ward's method. The N*earest neighbor (single linkage)* defines the distance between 2 clusters as the minimum distance between any member of one cluster and a member of the other whereas the *Furthest neighbor (complete linkage)* defines the distance between 2 clusters as the maximum distance between any member of one cluster and a member of the other. The *Centroid* is the distance between 2 clusters as the distance between the centroids of each cluster, where the centroid is located at the average values of each variable over all members of the cluster. The *Median* is the distance between 2 clusters as the distance between the medians of each cluster, where the median is located at the median values of each variable over all members of the cluster. The *Group average (average linkage) i*s the distance between 2 clusters as the average distance between all members of one cluster and all members of the other. Lastly, the *Ward's method* defines the distance between 2 clusters in terms of the increase in the sum of squared deviations around the cluster means that would occur if the two clusters were joined.

Non-Hierarchical Method (K-Means)

The method allows the items to be moved from one cluster to another, a reallocation that is not available in the hierarchical methods. The method of *k-Means* works as follows: *k* items are selected to be the initial *seeds* for the *k* desired clusters and the remaining items are assigned to the cluster whose seed is closest to that item. The centroids of each cluster are computed and each item is checked to determine whether it is closer to the centroid of another cluster than to the centroid of the cluster it is currently assigned to. If so, it is assigned to the other cluster and both centroids are recalculated.

The last step is repeated until no further changes take place.

The best way to view the output of a cluster analysis is usually by looking at the *Dendogram.* The dendogram shows the sequence of joins that were made between clusters. Lines are drawn connecting the clustered that are joined at each step, while the vertical axis displays the distance between the clusters when they were joined. In this work, the hierarchical and non-hierarchical (k-means) cluster analysis was requested from SPSS.

RESULTS AND DISCUSSION

Fig. 2 indicates the presence of increasing and decreasing trend in monthly precipitation across the 36 states and the FCT. The trend increased remarkably from January to August (peak precipitation), and a shape decline from September to December across states except Kwara, Kogi, Lagos, Ogun, Ondo, Oyo, Bayelsa, Edo, Delta and Imo. There were two notable peaks observed in Lagos, Ogun, Ondo, Oyo, Bayelsa, Edo, Delta and Imo (Fig. 2). Eruola et al., (2021) stated that the monthly rainfall in the Southern part of Nigeria observed August break within two peaks, whereas the Northern area exhibits a singular peak. The months of January, February, November, and December were moderately dry (Fig. 2). In terms of seasons, Fig. 2 shows that the wet period was from the months of May to October, and the dry period, November to April.

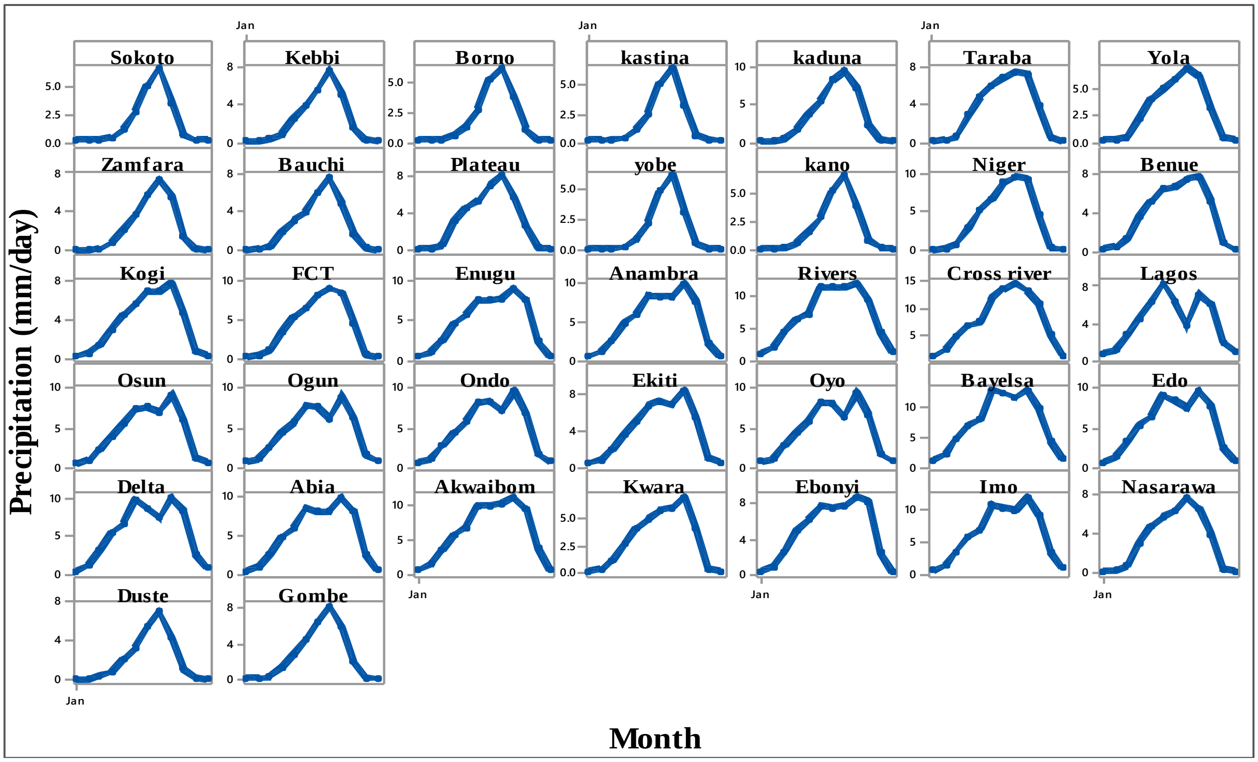

Figure 2: Monthly average Precipitation over 36 states and the FCT, Nigeria

Table 1 is a sample of the proximity distance matrix between 12 states in Nigeria, the 37 states observations was too large to present on a single table. The smallest distance is 0.00 between Plateau and Kaduna, Borno and Sokoto, Zamfara and Kebbi, and the farthest distance is 3.10 between Yobe and Taraba (Table 1).

Table 1: Proximity distance matrix between states in Nigeria

	Kebbi	Sokoto	Borno	kastina	kaduna	Taraba	Yola	Zamfara	Bauchi	Plateau	Yobe	kano
Kebbi	0.00											
Sokoto	0.29	0.00										
Borno	0.24	0.00	0.00									
Kastina	0.38	0.01	0.02	0.00								
Kaduna	0.88	2.17	2.03	2.40	0.00							
Taraba	1.15	2.59	2.43	2.84	0.02	0.00						
Yola	0.35	1.28	1.16	1.45	0.12	0.23	0.00					
Zamfara	0.00	0.32	0.27	0.41	0.82	1.09	0.32	0.00				
Bauchi	0.07	0.65	0.57	0.78	0.45	0.65	0.10	0.06	0.00			
Plateau	0.89	2.20	2.05	2.43	0.00	0.02	0.13	0.84	0.46	0.00		
Yobe	0.48	0.02	0.04	0.01	2.64	3.10	1.64	0.52	0.92	2.67	0.00	
Kano	0.11	0.04	0.02	0.08	1.62	1.99	0.86	0.13	0.37	1.64	0.12	0.00

Firstly, in Table 2, state 5 (Kaduna) is joined with state 10 (Plateau) since the Euclidean distance between these two states is smaller than the distance between any other pair of states except for state 28 (Edo) and state 30 (Abia). The distance is shown in the column labeled "Coefficients". Thirdly, state 31 (Akwaibom) is joined with state 34 (Imo) and so on. The distance between cluster 5 and state 32 (Kwara) is 0.013 millimeters since this is the largest distance between state 32 and any of the members of cluster 5 (distance to state 10). The columns under the heading "Stage Cluster First Appears" show the stage at which a cluster or state being joined first occurred in its current form. Cluster 5 joined at stage 15 was constructed at stage 1 (Table 2).
The "Next Stage" column (Table 2) shows when a cluster constructed at the current stage will be involved in another joining. Thus, cluster 5 as constructed in stage 15 (states 5, 10, and 32) will not be used until stage 20 where it is linked with state 6 (Taraba). It is a lot easier to follow how groups and individual states

join together in this process in a dendrogram (Figure 2), which is simply a tree-like diagram that demonstrates the series of fusions as the clustering proceeds for individual sample members to a single group.

Table 2: Agglomeration Schedule for precipitation data across states

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	5	10	0.000	0	0	15
2	28	30	0.000	0	0	23
3	31	34	0.001	0	0	30
4	16	25	0.001	0	0	8
5	1	8	0.001	0	0	13
6	17	24	0.001	0	0	19
7	6	35	0.002	0	0	20
8	13	16	0.002	0	4	17
9	2	3	0.003	0	0	18
10	4	11	0.006	0	0	18
11	7	37	0.006	0	0	24
12	22	23	0.009	0	0	22
13	1	36	0.010	5	0	28
14	14	15	0.013	0	0	27
15	5	32	0.013	1	0	20
16	18	33	0.015	0	0	25
17	13	21	0.015	8	0	22
18	2	4	0.021	9	10	21
19	17	26	0.025	6	0	25
20	5	6	0.040	15	7	27
21	2	12	0.066	18	0	31
22	13	22	0.077	17	12	32
23	28	29	0.077	2	0	29
24	7	9	0.081	11	0	28
25	17	18	0.087	19	16	29
26	19	27	0.153	0	0	30
27	5	14	0.208	20	14	32
28	1	7	0.248	13	24	31
29	17	28	0.369	25	23	34
30	19	31	0.558	26	3	33
31	1	2	0.648	28	21	35
32	5	13	0.665	27	22	34
33	19	20	1.345	30	0	36
34	5	17	1.960	32	29	35
35	1	5	5.049	31	34	36
36	1	19	13.750	35	33	0

Lines are drawn connecting the clustered that are joined at each step, while the vertical axis displays the distance between the clusters when they were joined. The dendogram in Fig. 3 shows the result of clustering the 37 states in Nigeria using the Euclidian distance method. At the start, five major clusters are formed; cluster 1 (Kaduna, Plateau, Kwara, Taraba, Nassarawa, Benue and Kogi), cluster 2 (Osun, Ogun, FCT, Ekiti, Niger and Lagos), cluster 3 (Edo, Abia, Delta, Anambra, Ebonyi, Enugu, Ondo and Oyo), cluster 4 (Sokoto, Borno, Kastina, Yobe and Kano) and cluster 5 (Kebbi, Zamfara, Dutse, Yola, Gombe and Bauchi). Akwaibom, Imo, Rivers, Bayelsa and Cross River forms the sixth cluster (cluster 6). A new cluster is formed when the first cluster (1) joined with the second cluster (2) and this process continued until a single cluster was formed. The general shape of the dendogram suggests grouping the 37 states into two groups: Group 1 comprises of states in Cluster 1,2,3,4 and 5, while Akwaibom, Imo, Rivers, Bayelsa and Cross River state (cluster 6) is classified under Group 2.

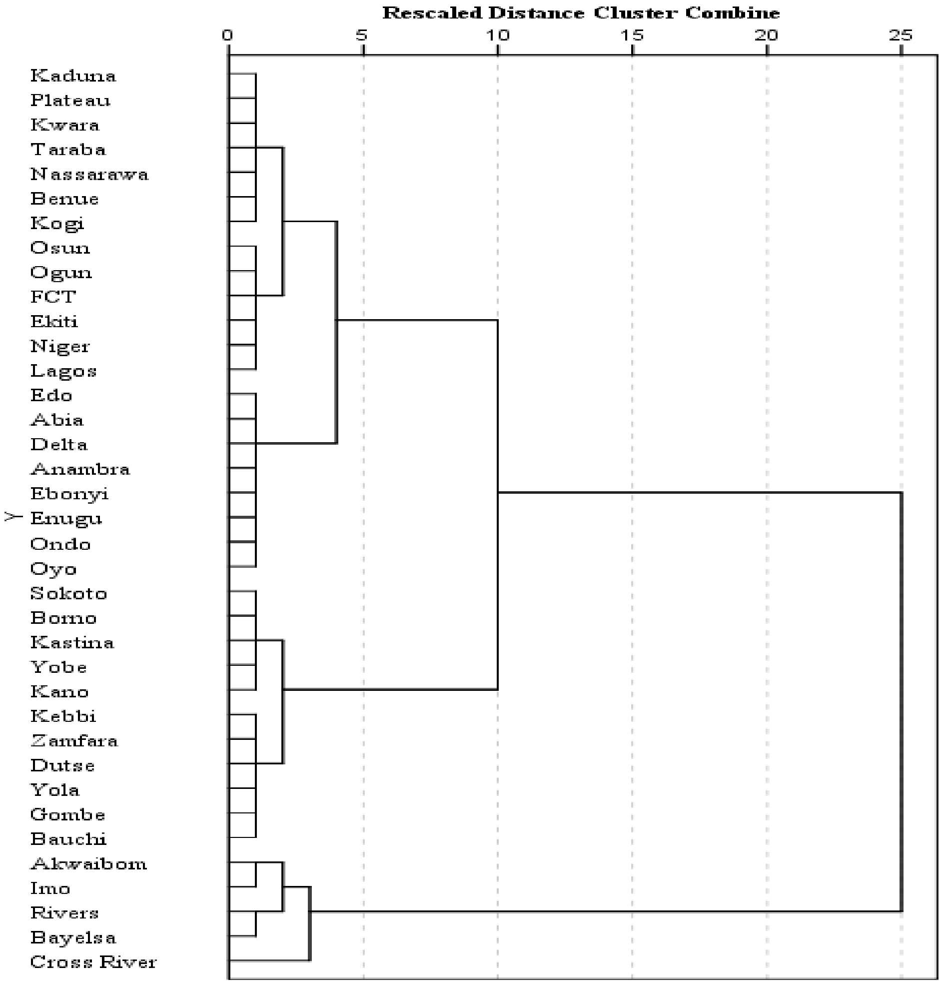

Figure 3: Dendrogram from average linkage clustering of Precipitation data.

This work went further to use the k-means clustering method, since a combination approach using a hierarchical approach followed by a non-hierarchical (K-means) approach is often advisable. The advantages of hierarchical methods are complemented by the ability of the non-hierarchical methods to refine the results by allowing the switching of cluster membership.

Table 3: Iteration History (K-means)

	Change in Cluster Centers				
Iteration	**1**	**2**	**3**	**4**	**5**
1	0.000	3.631	1.632	1.633	2.670
2	2.005	0.510	0.416	0.297	0.334
3	0.928	0.250	0.300	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000

Convergence achieved due to no or small change in cluster centers.

The iteration history (Table 3) shows the progress of the clustering process at each step. The maximum absolute coordinate change for any center is 0.000. The current iteration is 4. The minimum distance between initial centers is 6.537. The ANOVA table (Table 4) indicates which variables contribute the most to your cluster solution. Variable with large F (Plateau) values provide the greatest separation between clusters.

Table 4: One way Analysis of Variance (ANOVA)

	Cluster		Error		F	Sig.
States	**Mean Square**	**df**	**Mean Square**	**df**		
Sokoto	13.877	4	0.269	7	51.639	0.000
Kebbi	17.413	4	0.589	7	29.547	0.000
Borno	12.534	4	0.167	7	75.133	0.000
Kastina	13.035	4	0.212	7	61.343	0.000
Kaduna	31.374	4	0.618	7	50.794	0.000
Taraba	23.540	4	0.433	7	54.391	0.000
Yola	18.150	4	0.422	7	43.024	0.000
Zamfara	17.990	4	0.524	7	34.322	0.000
Bauchi	17.353	4	0.296	7	58.683	0.000
Plateau	23.384	4	0.255	7	91.594	0.000
Yobe	11.731	4	0.219	7	53.501	0.000
Kano	13.896	4	0.322	7	43.147	0.000
Niger	36.701	4	0.849	7	43.225	0.000
Benue	23.475	4	0.457	7	51.334	0.000
Kogi	22.019	4	0.571	7	38.540	0.000
FCT	33.120	4	0.684	7	48.397	0.000
Enugu	28.028	4	0.786	7	35.664	0.000
Anambra	31.891	4	0.764	7	41.758	0.000
Rivers	46.345	4	1.651	7	28.075	0.000
Cross River	62.767	4	2.317	7	27.091	0.000
Lagos	18.030	4	1.232	7	14.637	0.002
Osun	26.725	4	0.682	7	39.160	0.000
Ogun	23.745	4	0.707	7	33.606	0.000
Ondo	29.308	4	0.752	7	38.964	0.000
Ekiti	24.593	4	0.569	7	43.238	0.000
Oyo	26.546	4	0.785	7	33.831	0.000
Bayelsa	52.030	4	1.725	7	30.161	0.000
Edo	31.202	4	0.884	7	35.280	0.000
Delta	32.712	4	1.080	7	30.287	0.000
Abia	32.803	4	0.937	7	35.025	0.000
Akwaibom	41.481	4	1.536	7	27.001	0.000
Kwara	16.408	4	0.517	7	31.722	0.000
Ebonyi	28.682	4	0.896	7	32.026	0.000
Imo	44.162	4	1.301	7	33.950	0.000
Nasarawa	22.767	4	0.426	7	53.472	0.000
Dutse	15.850	4	0.393	7	40.312	0.000
Gombe	23.362	4	0.535	7	43.707	0.000

The final cluster centers are computed as the mean for each state within each final cluster. The final cluster centers reflect the characteristics of the typical case for each cluster. States in cluster 1and 3 tend to experience little precipitation. States in cluster 2 and 5 tend to be experience extreme precipitation, while states in cluster 4 tend to experience moderate precipitation (Table 5).

Table 5: Final Cluster Centers

	Cluster				
States	**1**	**2**	**3**	**4**	**5**
Sokoto	0.0096	3.0182	0.5023	0.5993	5.8210
Kebbi	0.0413	4.3161	1.3095	1.4352	6.4106
Borno	0.0107	3.0840	0.9576	0.8099	5.5678
Kastina	0.0237	2.7089	0.4215	0.5632	5.7084
Kaduna	0.0617	6.1197	2.1004	2.4803	8.5477
Taraba	0.1660	6.3853	3.6557	3.6707	6.8874
Yola	0.1092	5.3843	2.9619	2.9523	6.1953

Zamfara	0.0279	4.5626	1.3799	1.3470	6.3885
Bauchi	0.0814	4.2389	1.6035	2.4432	6.5915
Plateau	0.1367	5.3955	2.4917	3.7298	7.4583
Yobe	0.0169	2.6125	0.4622	0.4848	5.3944
Kano	0.0521	3.3216	0.8368	1.0535	5.9008
Niger	0.2387	7.7113	4.4633	3.9312	8.9203
Benue	0.4624	6.8584	4.9717	4.2313	6.8223
Kogi	0.5359	6.6744	4.5897	3.6996	6.8713
FCT	0.2831	7.3658	4.5023	4.1838	8.5135
Enugu	1.1798	8.2222	7.5813	5.1453	7.5919
Anambra	1.2580	9.0388	7.5007	5.1827	8.0684
Rivers	2.4632	11.5442	9.1759	6.6350	11.3130
Cross River	2.8483	12.4543	10.6087	7.0244	13.9781
Lagos	1.3554	7.6917	6.0983	5.3264	5.0107
Osun	0.8864	8.1430	5.8352	4.5855	7.1952
Ogun	1.1818	8.2134	6.0941	4.8273	6.7454
Ondo	1.1888	8.8309	6.6587	5.0237	7.6553
Ekiti	0.7554	7.5902	5.3317	4.2435	7.0151
Oyo	1.2294	8.6568	6.4816	5.0523	7.1087
Bayelsa	2.5624	12.5070	9.4286	7.1603	11.6431
Edo	1.5146	9.3803	7.7665	5.7247	7.9286
Delta	1.7104	9.8955	8.2623	6.0604	7.9786
Abia	1.4553	9.3264	8.2976	5.4624	8.1483
Akwaibom	2.1255	10.6438	9.7004	6.3353	10.1457
Kwara	0.4322	5.8804	3.9868	3.2877	5.7325
Ebonyi	1.2623	8.2453	8.2258	5.5932	7.5767
Imo	2.0284	11.2874	9.1744	6.3182	10.0077
Nassarawa	0.2127	6.0840	3.8377	3.8943	6.9718
Dutse	0.0844	3.8478	1.1012	1.4038	6.2857
Gombe	0.0776	5.1887	1.8933	2.0441	7.4545

Table 6 shows the Euclidean distances between the final cluster centers. Greater distances between clusters correspond to greater dissimilarities (clusters 1 and 5).

Table 6: Distances between Final Cluster Centers

Cluster	1	2	3	4	5
1		40.041	28.427	20.253	41.792
2	40.041		14.299	20.303	10.235
3	28.427	14.299		9.681	20.883
4	20.253	20.303	9.681		23.954
5	41.792	10.235	20.883	23.954	

Cluster 1 and 3 contains 3 states each, while cluster 2 and 5 contains 11 states each. Lastly, cluster 4 contains 9 states (Table 8).

Table 7: Number of Cases in each Cluster

Cluster	Number
1	3
2	11
3	3
4	9
5	11
Valid	37
Missing	0

From Fig. 4, the states grouped in cluster 1 are Bayelsa, Rivers and Cross River; cluster 2 are Oyo, Enugu, Ondo, Ebonyi, Ogun, Osun, Edo, Anambra, Abia, Ekiti and Lagos; cluster 3 are Imo, Akwibom and Delta; cluster 4 are Dutse, Kano,Borno, Sokoto, Kastina, Kebbi, Zamfara, Yobe and Bauchi; cluster 5 are Taraba, Nasarawa, Plateau, Kogi, Yola, Benue, Kwara, Kaduna, FCT, Niger and Gombe.

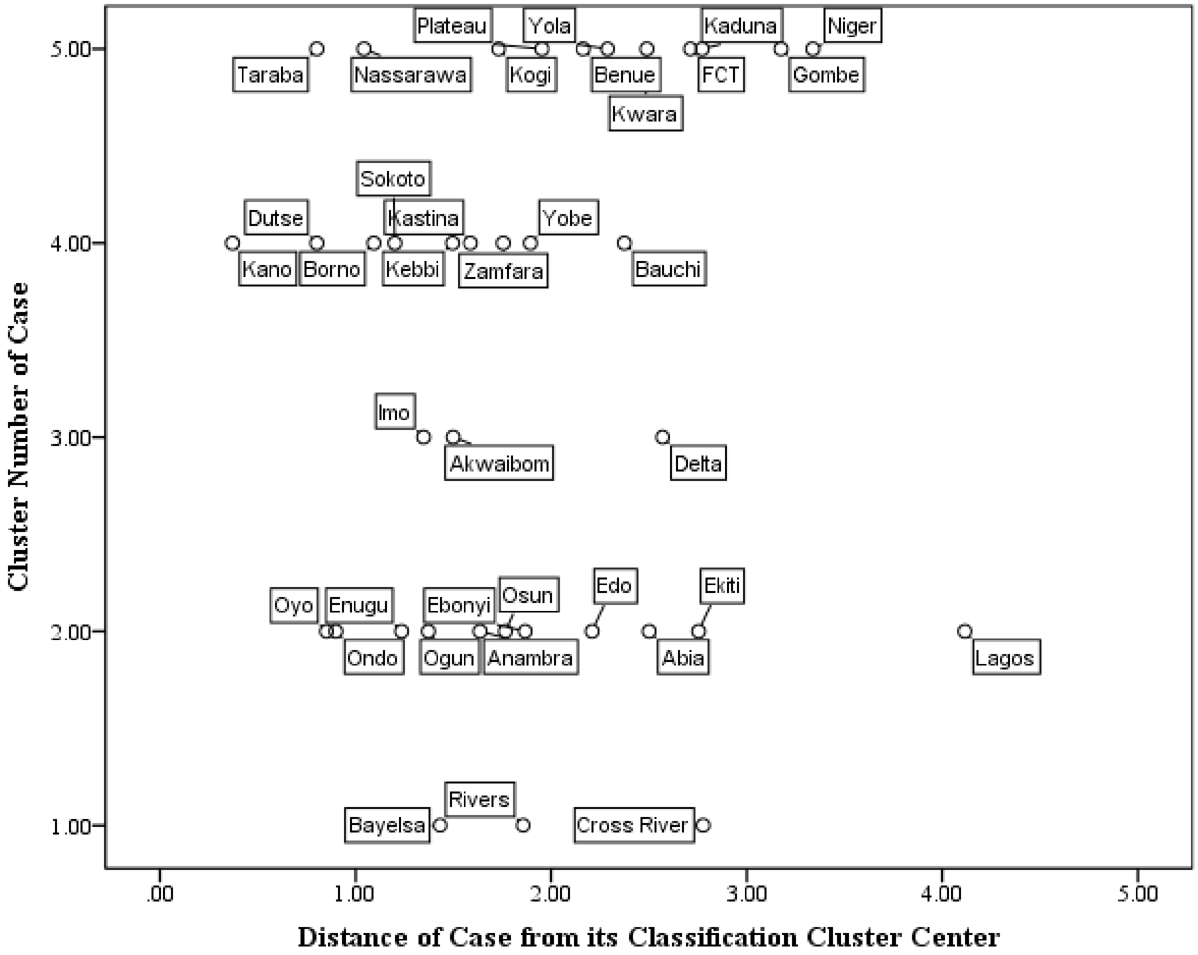

Figure 4: Distance and cluster number of cases from its classification cluster center.

Fig. 5 depicts average yearly precipitation over Nigeria. Comparing Figs. 3, 4 and 5, it is obvious that Fig. 5 validates Fig. 3. The hierarchical approach classified the average precipitation across states in Nigeria better than the non-hierarchical approach (K-means). Hence, precipitation in Nigeria can be classified based on the amount of precipitation as: Light precipitation (Sokoto, Borno, Kastina, Yobe and Kano); Moderate precipitation (Kebbi, Zamfara, Dutse, Yola, Gombe and Bauchi); Heavy precipitation (Kaduna, Plateau, Kwara, Taraba, Nassarawa, Benue and Kogi); Very Heavy precipitation (Osun, Ogun, FCT, Ekiti, Niger and Lagos); Shower (Edo, Abia, Delta, Anambra, Ebonyi, Enugu, Ondo and Oyo); Lastly, Heavy Shower (Akwaibom, Imo, Rivers, Bayelsa and Cross River).

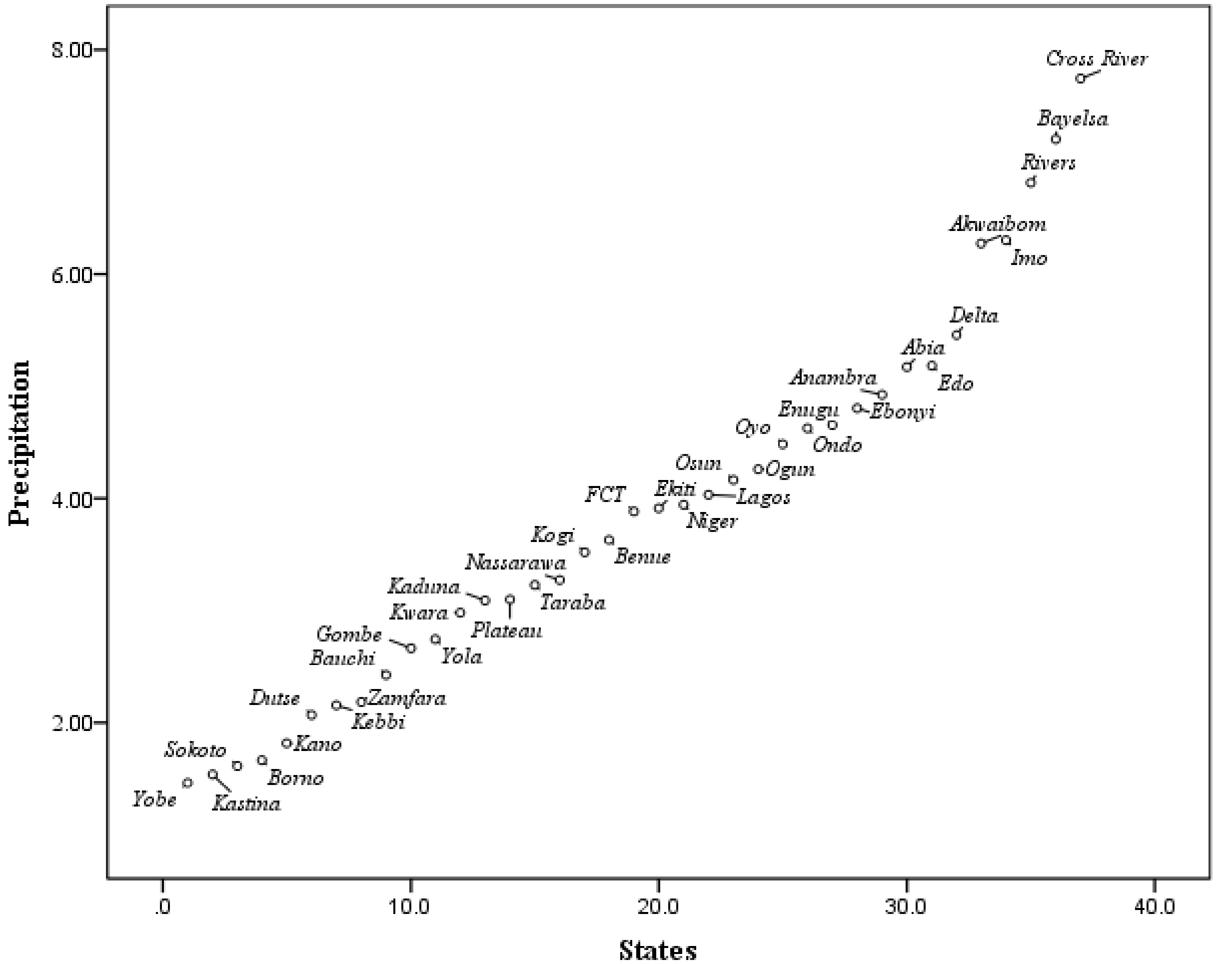

Figure 5: Average yearly precipitation (mm/day) across 36 states and FCT in Nigeria

In this study, Benue was classified under heavy precipitation (cluster 1) in the grouping which agrees with the findings of Aho *et al.* (2019) that Makurdi (Benue) is getting drier in terms of annual total rainfalls, but the occurrence of extreme rainfall (flood causing rainfall) magnitude is increasing. Aho *et al.* (2019) concluded that effective mitigation measures should be put in place for likely floods that may arise owing to the anticipated severe rainfall magnitudes in the coming years in Makurdi. Grouping Akwaibom, Imo, Rivers and Cross River under heavy showers agrees with the work of Egor *et al.* (2015). Egor *et al.* (2015) concluded that the climate of the southern region indicates a tendency towards wet condition rather than dryness and rainfall can be harnessed to create adequate water storage against periods of drought. It has been reported by Eruola *et al.* (2021) that the highest rainfall range is observed in the South- South area while the lowest range is found in the North-East of Nigeria. The South-South states consist of Akwaibom, Imo, Rivers, Edo, Delta, Bayelsa and Cross River, while the North-East states are Yobe, Borno, Taraba, Yola, Gombe and Bauchi. This explains while Akwaibom, Imo, Rivers, Bayelsa and Cross River was classified under heavy showers, but Edo and Delta states was classified under shower from this study. Thus, the amount of precipitation in Akwaibom, Imo, Rivers, Bayelsa and Cross River is more than Edo and Delta state. Looking at the North East state from this study, Yobe and Borno was classified under Light precipitation, while Taraba is classified under heavy precipitation, and Yola, Gombe and Bauchi was classified under moderate precipitation. Adejuwon, (2011) went further to explain that crops produced in the Northern area suffered production loss due to insufficient rainfall. Akwaibom, Imo, Rivers, Bayelsa, Cross River, Yobe and Borno states classification agrees with the work of Eruola *et al.* (2021).

CONCLUSION

The application of hierarchical and non- hierarchical (k-means) cluster method resulted into five major clusters from the 36 states and FCT in Nigeria. The five major clusters are cluster 1 (Kaduna, Plateau, Kwara, Taraba, Nassarawa, Benue and Kogi), cluster 2 (Osun, Ogun, FCT, Ekiti, Niger and Lagos), cluster 3 (Edo, Abia, Delta, Anambra, Ebonyi, Enugu, Ondo and Oyo), cluster 4 (Sokoto, Borno, Kastina, Yobe and Kano) and cluster 5 (Kebbi, Zamfara, Dutse, Yola, Gombe and Bauchi). Akwaibom, Imo, Rivers, Bayelsa and Cross River form a separate cluster (cluster 6). Thus, precipitation across the 36 states and FCT in Nigeria can be classified as Light precipitation (cluster 4); Moderate precipitation (cluster 5); Heavy precipitation (cluster 1); Very Heavy precipitation (cluster 2); Shower (cluster 3): Lastly, Heavy Shower (cluster 6). The general shape of the dendogram suggests grouping the 37 states into two groups: Group 1 comprises of states in Cluster 1,2,3,4 and 5, while cluster 6 is classified under Group 2. The hierarchical cluster method performed better in the classification of the precipitation data over the 36 states and the FCT in Nigeria compared to the non-hierarchical (k-means) cluster method in this work. Overall, this research provides a better understanding of the climate of Nigeria by identifying States with similar weather/climate characteristics.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

Aho, I.M., Akpen, G.D., Ojo, O.O., 2019. Rainfall Variability and Trend Analysis in Makurdi Metropolis, Benue State, Nigeria. Nigerian Journal of Engineering, 26, 17-24.

Azuwike, D.O, Enwereuzor, A.I., 2011. Effect of Rainfall Variability on Water Supply in Ikeduru L.G.A. of Imo State, Nigeria. International Multidisciplinary Journal, Ethiopia, 5, 22. DOI: 10.4314/afrrev.v5i5.18

Adejuwon, J.O., 2011. Rainfall seasonality in the Niger Delta Belt, Nigeria. Journal of Geography and Regional Planning, 5, 51-60. DOI: 10.5897/JGRP11.096

Abaje, I.B., Ati, O.F., Iguisi, E.O., 2012. Recent Trends and Fluctuations of Annual Rainfall in the Sudano-Sahelian Ecological Zone of Nigeria: Risks and Opportunities. Journal of Sustainable Society, 1, 44-51. DOI: 10.11634/21682585140399

Amissah-Arthur, A., Jagtap, S.S., 1999. Geographic variation in growing season rainfall during three decades in Nigeria using principal component and cluster analyses. Theoretical and Applied Climatology, 63, 107–116.

Babatolu, J.S., Akinnubi, R.T., 2014. Influence of Climate Change in Niger River Basin Development Authority Area on Niger Runoff, Nigeria. Journal Earth Science Climate Change, 5, 230. DOI: 10.4172/2157-7616.1000230

Duan, S.W., Kaushal, S.S., 2013. Warming increases carbon and nutrient fluxes from sediments in streams across land use. Biogeosciences, 10, 1193–1207. DOI: 10.5194/bg-12-7331-2015

Egor, A.O., Osang, J.E., Uquetan, U.I., Emeruwa, C., Agbor, M.E., 2015. Inter-Annual Variability of Rainfall in Some States of Southern Nigeria. International Journal of Scientific and Technology Research, 4, 134-140.

Ekwe, M.C., Joshua, J K., Igwe, J.E., Osinowo, A.A., 2014. Mathematical Study of Monthly and Annual Rainfall Trends in Nasarawa State, Nigeria. IOSR Journal of Mathematics, 10, 56-62. DOI: 10.9790/5728-10135662.

Enete, I.C., Ebenebe, I.N., 2009. Analysis of rainfall distribution over Enugu during the little dry season (1990-2005), Journal Geography and Regional Planning, 2,182-189. DOI: 10.4314/1wati.V6i2.46603

Eruola, A.O., Makinde, A.A., Eruola, G.A., Ayoola, K.O., 2021. Assessment of the Rainfall Exceedance in Nigeria. Nigerian Journal of Technology (NIJOTECH), 40, 751 –761. DOI: 10.4314/njt.v40i4.22

Ekpoh, I.J, Nsa, E., 2011. Extreme Climatic Variability in North-western Nigeria: An Analysis of Rainfall Trends and Patterns. Journal of Geography and Geology, 3. DOI: 10.5539/jgg.V3n1p51.

Gupta, A., Kamble, T., Machiwal, D., 2017. Comparison of ordinary and Bayesian kriging techniques in depicting rainfall variability in arid and semi-arid regions of north-west India. Environmental Earth Sciences, 76, 512. DOI: 10.1007/S12665-017-6814-3.

Haines, H.A., Olley, J.M., 2017. The implications of regional variations in rainfall for reconstructing rainfall patterns using tree rings. Hydrological Processes, 31, 2951–2960. DOI : 10.1002/hyp.11238

Jebari, S., Berndtsson, R., Uvo, C., Bahri, A., 2009. Regionalizing fine time-scale rainfall affected by topography in semi-arid Tunisia. Hydrological Sciences Journal, 52, 1199–1215. DOI:10.1623/hysj.52.6.1199

Lin, G.F., Chang, M.J., Wu, J.T., 2017. A hybrid statistical downscaling method based on the classification of rainfall patterns. Water Resources

Management, 31, 377–401. DOI : 10.1007-S1269-016-1532-2.

Machiwal, D., Sanjay, K., Hari, M.M., Priyabrata, S., Ranjay, K. S., Dharam, V.S., 2019. Clustering of rainfall stations and distinguishing influential factors using PCA and HCA techniques over the western dry region of India. Meteorology Application, 26, 300–311. DOI : 10.1002/met.1763.

Machiwal, D., Dayal, D., Kumar, S., 2017. Long-term rainfall trends and change points in hot and cold arid regions of India. Hydrological Sciences Journal, 62, 1050–1066. DOI : 10.1002/joc.4875

Nnaji, C.C., Mama, C.N., Ukpabi, O., 2016. Hierarchical analysis of rainfall variability across Nigeria. Theoretical and Applied Climatology, 123, 171–184.

Ologunorisa, T.E, Tersoo, T., 2006. The Changing Rainfall Pattern and Its Implication for Flood Frequency in Makurdi, Northern. Nigeria Journal Application Science Environment Management, 10, 97 – 102. DOI : 10.4314/jasem.v10i3.17327.

Okonkwo, G.I., Mbajiorgu C.C., 2010. Rainfall intensity-duration-frequency analysis for Southeastern Nigeria. Agriculture Eng Int: CIGR Journal, 12, 22—30.

Rencher, A.C., 2002. Methods of Multivariate Statistical. New York: Wiley.

Sabziparvar, A.A., Movahedi, S., Asakereh, H., Maryanaji, Z., Masoodian, S.A. 2015. Geographical factors affecting variability of precipitation regime in Iran. Theoretical and Applied Climatology, 120, 367–376. DOI: 10.1007/S00704-014-1174-3

Şen, Z., Habib, Z., 2001. Monthly spatial rainfall correlation functions and interpretations for Turkey. Hydrological Sciences Journal, 46, 525–535. DOI: 10.1080/02626660109492848

Stull, R.B., 1988. “An Introduction to Boundary Layer Meteorology”, Kluwer Academic Publisher, Dordrecht.