Journal of Agriculture, Biology and Applied Statistics
Vol. 5, Nos. 1-2, 2026, pp. 65-75
© ARF India. All Right Reserved
URL: www.arfjournals.com
https://doi.org/10.47509/JABAS.2026.v05i01-02.06

Dissecting Precipitation Dynamics: Factor Vs Cluster Analysis Approaches

S.R. Krishna Priya[1], N. Naranammal[2], K. Gowthami[2] K. Ranjith[3] and R.P. Joshi[4]

[1]*Assistant Professor, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India*
[2]*Research Scholar, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India*
[3]*M. Sc Student, Department of Statistics, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India*
[4]*College of Agriculture, Rewa, JNKVV (M.P.), India*

To Cite this Article

S.R. Krishna Priya, N. Naranammal, K. Gowthami K. Ranjith & R.P. Joshi (2026). Dissecting Precipitation Dynamics: Factor Vs Cluster Analysis Approaches. *Journal of Agriculture, Biology and Applied Statistics,* 5: 1-2, pp. 65-75.

Abstract: Understanding Rainfall variability is crucial for effective climate adaptation, disaster risk management and water resource planning. This study focuses on the comparison of Factor and Cluster Analysis, two prominent multivariate analytical methods, using district-wise rainfall data of Tamil Nadu from 2000 to 2023. The objective is to determine which method more effectively identifies rainfall patterns, as well as flood- and drought-prone zones. The finding highlights the strengths and limitations of both techniques, providing policymakers with valuable insights. The study concludes that integrating these methods can enhance the accuracy of precipitation classification and improve decision-making in environmental and agricultural sectors.

Keywords: Multivariate Techniques, Classification, Factor Analysis, Cluster Analysis, Rainfall Variability

1. INTRODUCTION

Precipitation is considered the most important meteorological variable influencing droughts and floods. Analysing precipitation and drought data provides valuable insights for improving water management, protecting the environment, supporting

Received : 28 July 2026 • **Revised :** 10 September 2026 • **Accepted :** 14 September 2026 • **Published :** 19 October 2026

agriculture and enhancing regional economic development (Gocic & Trajkovic.,2013). Therefore, understanding spatiotemporal variability in precipitation is essential for water resource management, agricultural planning, and mitigation of hydrological extremes such as droughts and floods (Santos *et al.*,2019). Cluster analysis is used to group regions or time periods with similar precipitation characteristics. It is particularly useful in climate zoning, rainfall pattern classification, and hydrological planning. By applying factor analysis, the study summarizes many precipitation-related variables into a few unobserved variables (factors) that explain most of the variability in the data.

Previous studies have demonstrated the extensive utilization of multivariate approaches, including Cluster and Factor Analysis in diverse research domains.

Cluster analysis has been widely employed in various contexts: In the regional agro-food complex development (Stovba *et al.*, 2017), classification of farmers in rural developing areas based on their socio-economic characteristics (Makhura *et al.*, 1998), categorization of Turkey's climatic areas (Unal *et al.*, 2003), and in agricultural communication and media studies using hierarchical cluster analysis (Essary *et al.*, 2022). It has also been utilized to analyze India's agricultural sectors and related industries (Sarojamma *et al.*, 2019).

Factor analysis, on the other hand, has been used primarily for data reduction and latent variable identification. Significant applications include evaluating faculty efficiency within academic institutions (Mamatha *et al.*, 2017), serving in the role of a tool for survey data analysis (Shrestha, 2021), and estimating body weight from biometric measurements (Eyduran *et al.*, 2010).

Several researchers have employed a combination of both cluster and factor analysis to gain deeper insights into complex datasets. These include climatic classification (Ahmed, 1997), interpretation of groundwater chemistry (Ashley and Lloyd, 1978), classification of countries based on development indicators (Berlage and Terweduwe, 1988), and characterization of soybean genotypes (Salimi *et al.*, 2012).

2. MATERIALS AND METHODS

2.1. Data description

The present analysis utilizes district-wise rainfall data from 29 districts of Tamil Nadu over the years 2000-2023. Rainfall data were sourced through the Season and Crop Reports compiled by the Directorate of Economics and Statistics. To classify

regions and identify major rainfall trends, the data were analyzed using Cluster and Factor Analysis.

2.2. Factor Analysis

Factor Analysis (FA) is a Multivariate data reduction method that aims to identify the underlying relationships among observed variables and to reduce data dimensionality by grouping correlated variables into factors (Kim & Mueller, 1978). In this study, FA was employed to extract the major latent constructs influencing [specify the context, e.g., agricultural productivity or rainfall variability].

The suitability of data for FA was assessed using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity. A KMO value above 0.6 and a significant Bartlett's Test ($p < 0.05$) confirmed the appropriateness of the data for factor extraction (Kaiser, 1974).

Principal Component Analysis (PCA) was used as the extraction method, followed by Varimax rotation to ensure a clearer and more interpretable factor structure (Harman, 1976). Factors with eigenvalues greater than 1.0 were retained based on Kaiser's criterion.

Additionally, the Scree Plot has been examined to determine the optimal number of factors to retain. The scree plot displays the eigenvalues in descending order and helps to identify the "elbow point," beyond which the remaining factors contribute minimally to explaining variance (Cattell, 1966).

2.3. Cluster Analysis

Cluster Analysis is a technique used to classify cases (e.g., regions, individuals) into groups based on similarity in their characteristics (Everitt *et al.*, 2011). In this study, Hierarchical Cluster Analysis has been employed to group the cases based on the factor scores obtained from the previous step.

The hierarchical clustering method is used, starting with each object in its own cluster and successively merging the most similar clusters. Ward's linkage method and squared Euclidean distance were selected as the clustering strategy, as they tend to produce clusters with minimum within-group variance (Ward, 1963). The resulting dendrogram was examined to determine the appropriate number of clusters.

3. RESULTS AND DISCUSSION

Summary statistics of rainfall, including Maximum, Minimum, Mean and Standard Deviation, are presented in Table 1.

Table 1: Summary Statistics of Rainfall

Districts	*Minimum (in mm)*	*Maximum (in mm)*	*Mean*	*SD*
Chennai	0	2327	151.91	276.624
Kancheepuram	0	2395	142.95	309.552
Thiruvallur	0	2534	129.51	257.449
Cuddalore	0	4467	146.23	389.125
Villupuram	0	4643	139.48	437.354
Vellore	0	2618	109.33	277.255
Thiruvannamalai	0	453.7	81.52	79.934
Salem	0	344.4	71.11	63.922
Namakkal	0	258.8	56.08	55.626
Dharmapuri	0	913.6	70.79	84.998
Coimbatore	0	541.8	75.26	78.524
Erode	0	343.1	56.47	61.685
Tiruchirappalli	0	376.3	56.85	64.164
Karur	0	339.3	46.61	58.357
Perambalur	0	481.7	65.47	75.870
Pudukkottai	0	474.7	68.09	72.220
Thanjavur	0	723.8	79.09	95.625
Thiruvarur	0	742.8	92.39	117.650
Nagapattinam	0	821	104.86	141.863
Madurai	0	783.9	72.10	87.063
Theni	0	381.8	68.50	70.828
Dindigul	0	405.2	72.38	73.083
Ramanathapuram	0	397.4	66.69	83.651
Virudhunagar	0	326.3	56.55	65.484
Sivaganga	0	318.8	73.00	69.788
Thirunelveli	0	405.4	75.73	80.881
Thoothukudi	0	522	54.23	79.119
Nilgiris	0	781.8	128.11	126.716
Kanniyakumari	0	405.1	91.58	87.205

3.1. Factor Analysis Results

Factor Analysis served as the classification technique in this study, and its results are shown below.

3.1.1. Adequacy Testing

To assess the suitability of the data, the KMO measure and Bartlett's test of sphericity were performed, with results displayed in Table 2.

Table 2: Statistical Tests for Data Adequacy

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.946
Bartlett's Test of Sphericity	Approx. Chi-Square	13504.315
	df	406
	Sig.	<0.001

As shown in Table 2, the KMO measure is 0.946, indicating a high level of sampling adequacy and confirming the suitability of the data for factor analysis. Additionally, Bartlett's test shows a p-value < 0.001, indicating that the correlation matrix significantly differs from an identity matrix. Thus, proceeding with factor analysis is justified.

3.1.2. Total Variance Explained

The results of the variance explained by extracted factors are presented in Table 2.

Table 3: Variance Distribution Across Components

Factor	*Rotation Sums of Squared Loadings*		
	Total	*% of Variance*	*Cumulative %*
1	8.964	30.912	30.912
2	5.792	19.973	50.884
3	5.687	19.612	70.496
4	4.034	13.910	84.406

Table 3 reveals that the four extracted components collectively account for 84.41% of the total variance, confirming the adequacy of the factor solution. Among these, Factor 1 exhibits the greatest influence.

3.1.3. Scree Plot

The Scree Plot shows the classification of districts based on their rainfall patterns into factors using graphical representations, and it is presented in Figure 1.

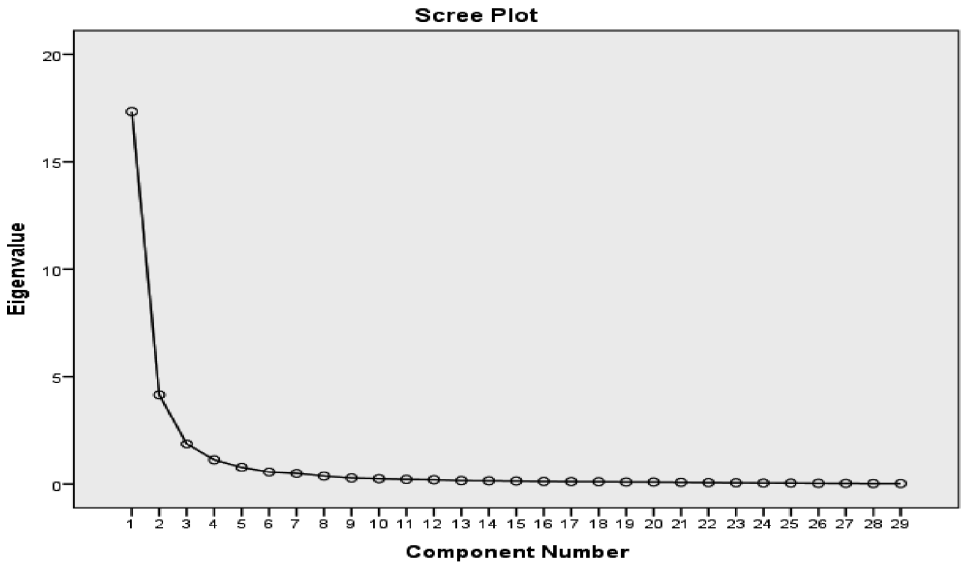


Figure 1: Scree Plot

From Figure 1, the eigenvalues of the first four factors are greater than 1, which indicates that four factors are obtained from 29 districts.

3.1.4. Rotated Component Matrix

Table 4 illustrates the distribution of districts across different factors derived from their rainfall characteristics, with each district aligned to the factor where it shows the strongest loading.

Table 4: Component Loadings Based on Rainfall Patterns

	Factor			
	1	*2*	*3*	*4*
Pudukkottai	0.668	0.284	0.581	0.182
Thanjavur	0.778	0.233	0.500	0.036
Thiruvarur	0.810	0.213	0.412	0.036
Nagapattinam	0.883	0.189	0.295	0.015
Madurai	0.607	0.179	0.394	0.413
Dindigul	0.638	0.135	0.522	0.416
Ramanathapuram	0.893	0.104	0.228	0.243
Virudhunagar	0.745	0.096	0.268	0.462
Sivaganga	0.632	0.155	0.463	0.449
Thirunelveli	0.844	0.122	0.037	0.365

	Factor			
	1	*2*	*3*	*4*
Thoothukudi	0.879	0.104	0.087	0.308
Chennai	0.251	0.883	0.180	0.117
Kancheepuram	0.165	0.948	0.131	0.074
Thiruvallur	0.139	0.931	0.188	0.118
Cuddalore	0.261	0.828	0.049	0.026
Villupuram	0.093	0.894	0.173	0.002
Vellore	-0.066	0.907	0.198	0.095
Thiruvannamalai	0.437	0.412	0.633	0.262
Salem	0.269	0.294	0.751	0.407
Namakkal	0.345	0.153	0.761	0.419
Dharmapuri	0.135	0.150	0.701	0.299
Erode	0.485	0.192	0.587	0.480
Tiruchirappalli	0.569	0.289	0.672	0.205
Karur	0.608	0.181	0.617	0.308
Perambalur	0.598	0.349	0.622	0.139
Coimbatore	0.207	0.027	0.322	0.800
Nilgiris	0.132	0.161	0.258	0.767
Theni	0.521	0.007	0.285	0.697
Kanniyakumari	0.506	0.056	0.250	0.556

As shown in Table 4, the rotated component matrix reveals a four-factor structure categorizing the districts based on their loadings.

- Factor 1 represents Moderate Rainfall Regions – Pudukkottai, Thanjavur, Thiruvarur, Nagapattinam, Madurai, Dindigul, Ramanathapuram, Virudhunagar, Sivaganga, Thirunelveli and Thoothukudi.
- Factor 2 represents High Rainfall Regions – Chennai, Kancheepuram, Thiruvallur, Cuddalore, Villupuram and Vellore.
- Factor 3 represents Low Rainfall Regions – Thiruvannamalai, Salem, Namakkal, Dharmapuri, Erode, Tiruchirappalli, Karur and Perambalur.
- Factor 4 represents Very High Rainfall Regions – Nilgiris, Coimbatore, Theni and Kanniyakumari.

3.2. Cluster Analysis Results

To examine rainfall similarities, Cluster Analysis has been performed on the districts of Tamil Nadu, with the results detailed in the following section.

3.2.1. Dendrogram of Various Districts Classification

Figure 3 shows the classification of the various districts of Tamil Nadu based on their rainfalls using Cluster Analysis in dendrogram.

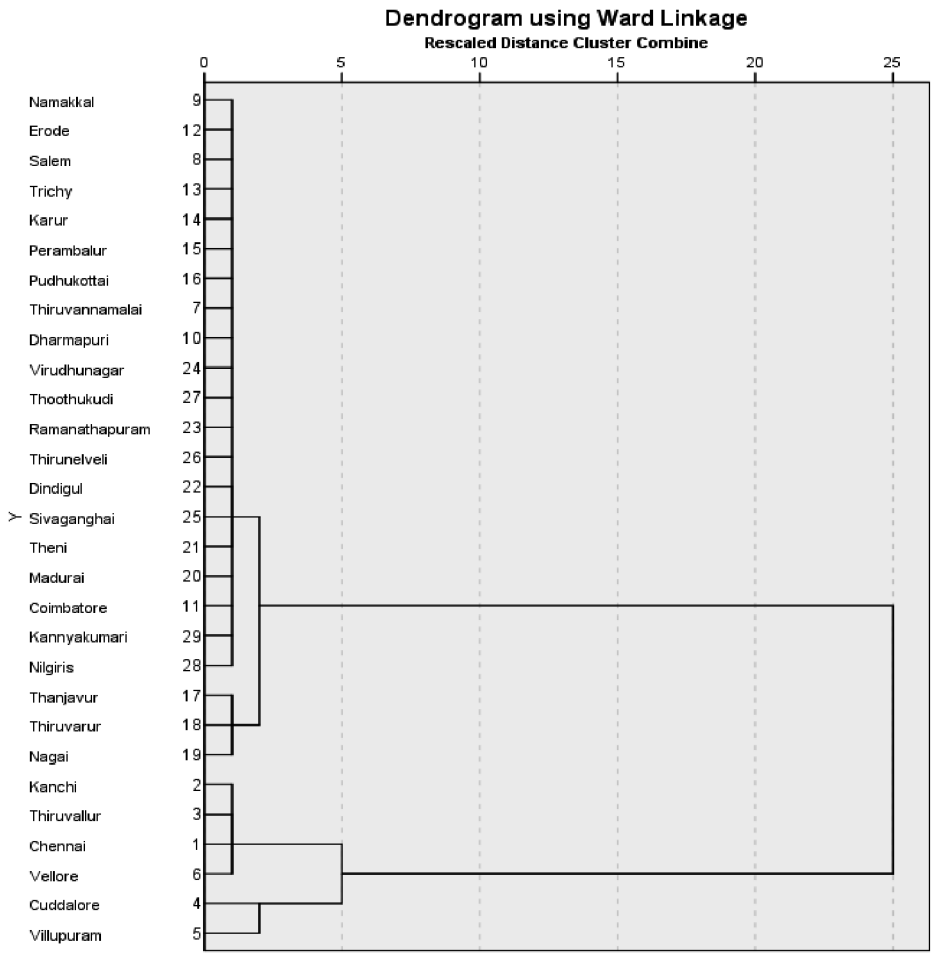


Figure 2: Dendrogram for various districts of Tamil Nadu

From the figure 2, twenty-nine districts are clustered into five clusters based on their rainfall patterns.

- Cluster 1 – Cuddalore (Coastal Region).
- Cluster 2 – Villupuram (Inland Region).
- Cluster 3 – Thanjavur, Thiruvarur, Nagapattinam (Delta Regions).

- Cluster 4 – Chennai, Kancheepuram, Thiruvallur and Vellore (Urban and Coastal Regions).
- Cluster 5 – Thiruvannamalai, Salem, Namakkal, Dharmapuri, Coimbatore, Tiruchirappalli, Karur, Perambalur, Pudukkottai, Madurai, Theni, Dindigul, Ramanathapuram, Virudhunagar, Sivaganga, Thirunelveli, Thoothukudi, Erode, Nilgiris and Kanniyakumari (Central and Interior Tamil Nadu).

3.3. Comparative Interpretation of Factor and Cluster Analysis Results

Criteria	*Factor Analysis*	*Cluster Analysis*
Objective	Identifies underlying latent variables (factors) influencing rainfall patterns	Groups districts based on surface-level similarities in rainfall data
Statistical Strength	Explains 84.4% of total variance using four distinct factors	Lacks clear variance explanation; grouping not statistically validated
Classification Accuracy	Produces meaningful categories: low, moderate, high, and very high rainfall regions	Some misclassifications observed; less reflective of true climatic patterns
Interpretability	Factors represent deeper structure and are easier to interpret meaningfully	Cluster groupings are less precise and may not reflect actual variations
Usefulness in Applications	Suitable for disaster management and water resource planning	Less useful due to limited classification accuracy
Data Reduction Capability	Effectively reduces dimensionality while preserving key information	Does not reduce dimensions; works with raw similarities
Overall Insight	Provides deeper, statistically validated insights into rainfall variation	Offers basic grouping with limited analytical depth

4. CONCLUSION

This study highlights the significance of applying multivariate statistical techniques, specifically Factor Analysis and Cluster Analysis, for understanding the structure and classification of precipitation data. Factor Analysis proved to be an effective method for reducing data dimensionality by identifying latent variables that explain the variability in rainfall patterns. It provided deeper insights into the interrelationships among meteorological parameters.

In contrast, Cluster Analysis was useful in grouping regions with similar precipitation characteristics, supporting spatial classification. However, Factor Analysis demonstrated superior performance by revealing the underlying structure of the data and identifying key influencing factors more effectively than Cluster Analysis.

These findings underscore the value of Factor Analysis in future studies for building predictive models and understanding climatic variability. The application

of these methods can support improved planning in agriculture, water resource management, and climate-based decision-making.

Conflict of Interests: The authors declare that there is no conflict of interest related to this article.

REFERENCES

Cattell, R.B. 1966. The scree test for the number of factors. *Multivariate Behavioral Research.*, **1**(2): 245–276.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). Wiley.

Gocic, M., & Trajkovic, S. (2013). Analysis of precipitation and drought data in Serbia over the period 1980–2010. *Journal of Hydrology.*, *494*: 32-42.

Harman, H. H. (1976). *Modern Factor Analysis* (3rd ed.). University of Chicago Press.

Hou-Liu, J., & Browne, R. P. (2022). Factor and hybrid components for model-based clustering. *Advances in Data Analysis and Classification*, 1-26.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika.*, 39(1): 31–36.

Kim, J. O., & Mueller, C. W. (1978). *Introduction to Factor Analysis: What It Is and How to Do It*., Sage Publications.

Mafata, M., Brand, J., & Buica, A. (2022). Data fusion using Multiple Factor Analysis coupled with non-linear pattern recognition (fuzzy k-means): application to Chenin blanc. *OENO One*, *56*(3), 413-425.

Mamatha, H.K., Chaithra, N. and Balasubramanian, S. (2017). Factor analysis scores in multiple linear regression model for prediction of faculty performance in higher education institutions. *International Journal of Computer Engineering and Applications.,* 11(5): 2321-3469.

Mansouri Daneshvar, M. R., Bagherzadeh, A., & Alijani, B. (2013). Application of multivariate approach in agrometeorological suitability zonation at northeast semiarid plains of Iran. *Theoretical and applied climatology*, *114*, 139-152.

Salimi, S., Laniji, H.S., Abadi, G.M., Salimi, S. and Moradi, S. (2012). Genetic diversity in soybean genotypes under drought stress condition using factor analysis and cluster analysis. *World Applied Sciences Journal.,* 16(4): 474-478.

Santos, C. A. G., Brasil Neto, R. M., da Silva, R. M., & Costa, S. G. F. (2019). Cluster Analysis Applied to Spatiotemporal Variability of Monthly Precipitation over Paraíba State Using Tropical Rainfall Measuring Mission (TRMM) Data. *Remote Sensing.*, *11*(6): 637

Sarojamma, B., Geetha, K. and Reddy, B.H.M. (2019). Cluster Analysis for Agriculture and Other Allied Sectors. *Think India Journal.,* 22(10): 4186-4189.

Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics.,* 9(1): 4-11.

Stovba, E., Stovba, A., Abdrashitova, A. and Baygildina, A. (2017). Use of Methods of Cluster Analysis in Designing Strategy of Region's Agro-Food Complex. *Advances in Economics, Business and Management Research.,* 38: 648-652.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association.*, 58(301): 236–244.