Book of Proceedings, 14th Nigeria Association of Hydrological Sciences Conference (Okitipupa 2024) held at Olusegun Agagu University of Science and Technology, Okitipupa, Ondo State, Nigeria, November 5 - 8, 2024

Review of Clustering Algorithms for Regional Drought Characterisation

Chukwu, S. Emeka*,Otache, Y. Martins*, Ikugbiyi, S. Olushola*, Atemoagbo, O. Precious*, Jibril Ibrahim**
*Department of Agricultural & Bioresources Engineering, Federal University of Technology, Minna, Nigeria.
** Department of Agricultural Technology, Niger State College of Agriculture, Mokwa, Niger State, Nigeria
Email: sechukwu6@gmail.com, 08065749076

ABSTRACT

Drought is characterised by a negative water balance originating from a deficiency of precipitation or a lack of available water resources for an extended period of time. However, drought patterns have become more complicated in recent years due to climate change, and thus there is a need to better understand drought patterns and characteristics. Currently, the main limitations of drought analysis is the lack of ability to classify spatial pattern according to its kind and concomitant regional characteristics. This ability is increasingly important because the effects of drought accumulate slowly over a considerable period of time, and move slowly to adjacent positions. Myriads of scholars have considered clustering techniques as the most common approaches.Findings here unveiled the shortcomings and strength of composite clustering algorithms. It is clear here, that the choice of the cluster algorithm is relative subjective, yet should be bore in mind, that is there is need to explore more than one cluster algorithms in fear of losing microscopic precipitation fact and for the ease of analysis of Spatio-temporal phenomena. In recognition of hydrological time series characteristics of trans-boundary interference and spatio-temporal variation, where zones or regions that share common boundary may inherit similar hydro-climatic characteristics that seem different from other part of the region. It is imperative, for hydro-climatic researchers to adopt clustering algorithm that reveals the degree of shared characteristics or membership properties, in this hierarchy that is; FCM, PCA,k-means and SOM, for effective water resource planning and management in the phase of climate change.

Keywords: drought, cluster, water, homogenous and region

1.0 INTRODUCTION

The study of climate regionalisation is essentially the aggregation or delineation of regions with similar hydro-climatic characteristics and spatial continuity. In general, the regional behaviour of droughts has been studied by analysing the point behaviour (data analysis of point rainfall or streamflow) and then mapping the relevant parameters over a region or a country. The regional drought characterisation concept is solely based upon the basic idea of point-drought and drought-affected area; the former being related to a threshold (truncation level or critical level) below which a water shortage exists, whereas the later concept is related to another threshold (critical area) above which the integration of the various areas affected by point-droughts represent a significant portion of the whole region under study (Rodrigues *et al*., 1993). Rodrigues *et al*. (1993) extended the concept of random fields and investigated the regional characteristics of drought. Raziei *et al*. (2015) further developed a statistical drought distribution model on the regional level, which is operational in Portugal and applied to coherent regions in

Europe. The coherent regions were delineated on the basis of Box-Cox parameters in the time series of drought variables with negligible levels of persistence

Currently, the main limitations of drought analysis is the lack of ability to classify spatial pattern according to its kind and concomitant regional characteristics. One important tool which is indispensable in regional analysis is the multiple regression algorithm Wei*et al.* (2020), which involves parameters of drought, geomorphology and climate for the development of regression equations. Another important tool for regionalisation is that of kriging (Loukas, 2004). Among other emerging approaches for drought analysis are pattern recognition (Jimoh *et al,* 2023) and neural networks (Otache, 2008) which are all subsets of machine learning, gainfully engaged in drought analysis especially, the unsupervised component namely; clustering algorithm. It have proven to be the most viable framework in grouping of seemingly homogenous pools into clusters. Therefore, cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some specific sense defined by the analyst) to each other than to those in other groups (clusters). It is a main task of exploratory data analysis, and a common technique for statistical data analysis, used in many fields (Agrawal *et al*., 2005). In light of the foregoing, this research is tilted towards reviews of common clustering algorithms often use in drought analysis and in extension water resources characterisation.

2.0 Overview of Clustering Algorithm

2.1 Hierarchical Clustering

With the help of the HiClim R statistical package in R software as reported in Chukwu (2024), the longitude, latitude, elevation, precipitation concentration degree (PCD), and precipitation concentration period (PCP) of each grid point in the identified region can be used as the attributes in hierarchical clustering to identify the homogeneous climate regions. Among these, latitude, longitude, and elevation are useful for identifying geographically continuous regions and describing atmospheric variables that change over space (Ghosh *et al.,* 2016).The PCD and PCP can reflect the uneven distribution of rainfall within a year. In previous studies, Ghosh *et al.* (2016)demonstrated the reliability of the PCD and PCP for identifying homogeneous drought regions. To eliminate the influence of dimension, the grid characteristics, namely, latitude, longitude, elevation, PCD, and PCP were standardised before hierarchical clustering. However, Chukwu (2024), pointed out few weaknesses common to hierarchical cluster; it rarely provides the best decisions, does not work with a missing data, work poorly with mixed data types and its output dendrogram is commonly misinterpreted. Afterwards, the square of the Euclidean distance was used to calculate the distance between the points. Furthermore, the distance between the classes is often calculated using Ward's method. The appropriate number of clusters can be determined using the silhouette coefficient. The mathematical scheme below illustrate Hierarchical cluster

$$R_i = \sum_{j=1}^{12} r_{ij} \quad (1a)$$

$$R_i = \sum_{j=1}^{12} (r_{ij}\, sin\theta_j) \quad (1b)$$

$$PCD_i = \sqrt{R_{xi}^2 + \frac{R_{yi}^2}{R_i}} \quad (1c)$$

$$PCP_i = arctan\left(\frac{R_{xi}}{R_{yi}}\right) \quad (1d)$$

Where,

i refers to the year (example: i = 1961, 1962, : : : , 2017)

$j$stands for the month (j = 1, 2, : : : , 12) in a year.

r_{ij}represents monthly total precipitation in the j_{th} month of the i_{th}year

$\theta_j$refers to the azimuth of the j_{th}month.

$R_j$expresses the total precipitation amounts of the I_{th} year.

R_{xi}andR_{yi} are the synthetic components in horizontal and vertical directions of the twelve month precipitation vector modules in the i_{th} year, respectively.

PCD_i and PCP_i represent the precipitation concentration degree and precipitation concentration period in the i_{th} year, respectively.

2.2 Empirical orthogonal function (EOF)

EOF decomposition was first established in 1902 as reported by Jolliffe (200).It decomposes elements of a set of space time data into two functions, one dependent only on time and another dependent only on space, to analyse the spatial structure of element fields (Cui *et al.,* 2000). Thus, EOF decomposition is commonly used to analyse spatial and temporal variations in meteorological elements. In contrast, principal component analysis (PCA) is mainly used to find a few independent vectors in a dataset to reduce its dimensionality, thereby simplifying data handling and interpretation. However, the operational steps are basically the same, when the original variables are anomaly data or standardised anomaly data (Cui *et al.,* 2000). EOF decomposition and several variants thereof have been widely applied for analyzing spatial changes in hydrometeorological elements. Notably, EOF decomposition has been used to analyse drought occurrence in both time and space in Korea (Kim *et al.,* 2011). It has also been applied to establish contributions of annual evaporation, annual precipitation, and both warm-and cold-season precipitation, to the annual runoff of rivers in an arid zone to obtain quantitative estimates of the effects of climatic factors on reconstructions of annual runoff series in an ungauged area (Loboda, 2005). A data-interpolating empirical orthogonal function (DINEOF) has been used to analyse distributions of sea surface temperatures in a study demonstrating that DINEOF decomposition can be highly effective, even when very high proportions of data are missing (Sirjacobs *et al*., 2011). In addition, rotated empirical orthogonal function (REOF) decomposition has been applied to analyse space time patterns of groundwater fluctuations in the Choshui River alluvial fan, Taiwan, based on monthly observations of piezometric heads from 66 wells during the period 1997–2002 (Yu *et al.,* 2013). It has mild shortcoming of requiring adequate computation skill and mechanism. Therefore, it is a silent approach when it comes to regionalisation via geographic protocol and gives less insight to drought characteristics of the concern region (Loboda, 2005).

2.3 Principal components analysis (PCA)

Principal components analysis is a statistical procedure that transforms several (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components (PC) (Hair *et al*., 1998). The main objective of the PCA, specifically the principal factor analysis (PFA), are: (1) to reduce the number of variables (2) to detect structures in the relationship between variables (3) to reduce the system's information entropy, that is, the information not directly available about a system due to the uncertainty or randomness of data flow (Deng *et al*.,2014) and (4) to combine correlated variables into factors (Kaiser *et al*., 1960).

The determination of the number of components or factors to retain is often based on the Kaiser's rule, the factors whose eigenvalues are greater than 1 must be retained (Kaiser, *et al.,*1960). The spatial patterns of the eigenvectors (factor loadings) represent the correlation between the original data and the corresponding factor time series. More localised patterns are obtained by applying the Varimax rotation technique to selected factor loadings (Abdi, 2003). The projection of the Standardised Precipitation Index (SPI) fields onto the orthonormal Eigen functions provides the factor score time series (Rencher, 2002). Factor scores are estimates of the actual values of individual cases (observations). For instance, the estimated factor score (regionalised SPI) on factor j for observation, or month, i, F_{ji}can be represented as follows:

$$F_{ji} = Z_{j1}W_{j1} + Z_{j2}W_{j2} + ... + Z_{jk}W_{jk} \quad 2$$

Where W_j is the regression weight, multidimensional value referred to as factor score coefficient; and Z_i is the variable that is the SPI series or desire drought series at a single rain gauge. For any single common factor, an infinite number of sets of scores can be derived that would be consistent with the same factor loadings (Grice, 2001). The factor scores are particularly useful to perform further regional analyses that have been identified in the factor analysis, such as fitting drought characteristics with copulas.

When applied to the SPI or drought series values from a set of rain gauges, it allows their regrouping and consequently, the delimitation of climatic regions in relation to synoptic situations, that is regionalised. SPI series such as the three distinct areas with coherent climatic variability identified by Serinaldi *et al.* (2009) in Sicily from 1926 to 1996, the three homogeneous regions were adopted for drought characterisation in mainland Portugal. Santos *et al.* (2011) employed similar approach in two climatic sub-regions in the western Iran, all the analysis were based on the SPI field and effective delineate component series into homogenous subregion and finding out few remarkable shortcomings that data inputs produce result that are very much off the correct projection of the data. In the case of Raziei *et al.* (2015), the PCA was applied to the unsmoothed and smoothed SPI time series at the timescales of 3 and 6 months and the result was consistence with previous studies carried out by Rodrigues*et al.* (1993) in a related region, in that regard PCA consists of computing the covariance matrix of the SPI series with the corresponding eigenvalues (λ) and eigenvectors (v).

2.4 K-means Clustering Algorithm

K-means cluster analysis is an example of a hard partitioning algorithm (Hartigan *et al*., 1979). A set of *N* data $(x_1, x_2, \ldots x_N)$ in *d* dimensions is partitioned into *K* clusters, where each element in the data set is allocated entirely to a particular cluster. It is an iterative process whereby the data are initially partitioned, the mean position of each group calculated, and then the data partitioned again by allocating each datum to its nearest mean cluster position. The procedure terminates when no datum changes cluster or when the number of iterations reaches a pre-defined maximum. The algorithm is described by the flowchart in Figure 1

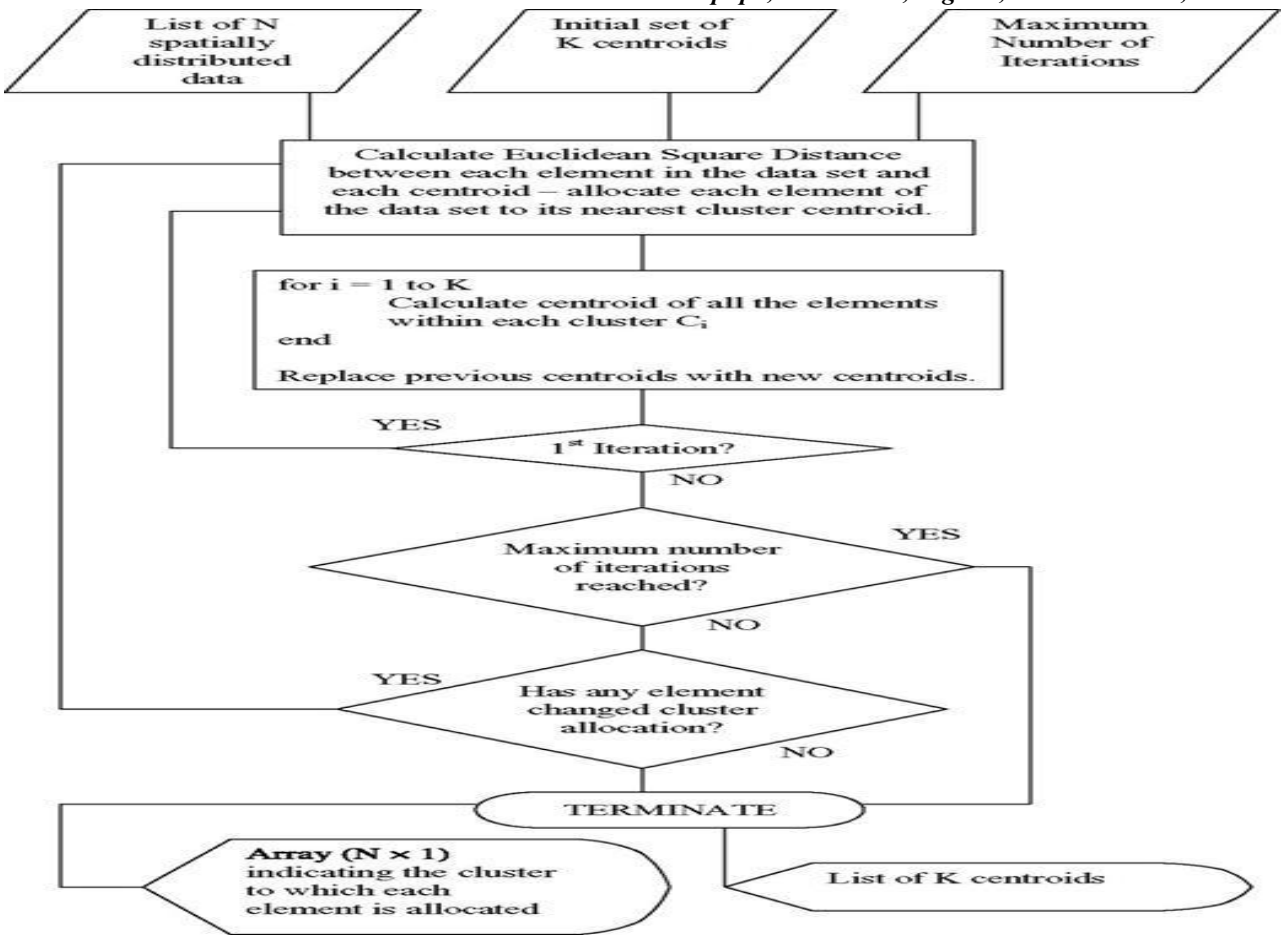

Figure 1: K-means Cluster Algorithm (Rajsekhar*et al*. 2011).

He *et al*. (2015) found that k-means clustering algorithm is able to search the synchronous variability of at-site SPI series and classify the homogeneous sub-regions over the study region as well as indicate that temporal evolutions are quite different over other candidate cluster algorithm explored as verified by clustering validation indices (CVI) and sum squared error (SSE) in Guizhou province in southwest China. In a related study carried out by Pei *et al*. (2019) in Xinjiang province examines the variation of SPI-12 from the meteorological stations. It was reported that k-means cluster was enabled to group the stations with similar temporal trends, thereby identifying three clusters; Cluster 1 is the driest, cluster 2 has a clear alleviating tendency of drought, and cluster 3 shows late occurrence of change point. Rajsekhar *et al*. (2011) used entropy approach for identification of homogeneous drought regions using standardized streamflow index (SSFI) for the Brazos basin in Texas over a time span of 1949–2000. Regions obtained using entropy theory were similar to the regions obtained with k-means clustering method. However, one problem that emerges when k-means method is used is a demand for a priori choose of the number of clusters in a region. To solve this problem, Gap Statistics is often recommended (Pham *et al*., 2005). This Statistics uses the distortion of a cluster and it is determined as described by Carvalho *et al*. (2016).

2.5 Fuzzy c-means clustering

The concept of fuzzy c-means (FCM) clustering was given by Dunn (1973). It is a soft partitioned clustering algorithm, which means a data point belongs to every cluster with different

degrees of belonging (membership), unlike hard partitioned clustering in which a data point belongs completely to one cluster only. The degree of belonging (membership) is inversely proportional to the distance between the cluster centre and the data point. The more the distance of the cluster centre from the data point, lesser the membership value of station for that cluster. Dunn (1973) shows that soft partitioned clustering provides more information and is essential for obtaining homogeneous clusters. The homogeneity of the cluster is evaluated using L-moment-based homogeneity test (H-Test) as reported in Chukwu (2019). The fuzzy c-means (FCM) clustering algorithm optimises the objective function (3a). It consider a cluster c having M objects in which Y_k is the data vector for k_{th} object, k = 1,2,…, M. The fuzzy objective function is

$$J(U, C) = \sum_{K=1}^{M} \sum_{i=1}^{c} u_{ik}^{\alpha} \|Y_k - C_i\|^2 \quad 3a$$

Where u_{ik} denotes the degree of belonging of k_{th} data point in the i_{th} cluster, C_i denotes the centre of the i_{th} cluster, $\|Y_k - C_i\|^2$ is the squared Euclidean distance between k_{th} data point and C_i and a is called fuzzifier or fuzziness index, which can have any value > 1.

Fuzzy c-means algorithm steps is given below:

Cluster centres are assumed randomly.

1. Using cluster centres, membership matrix is calculated using following equation.

$$u_{i<1}^{t} = \left[\sum_{j=1}^{c} \left[\frac{\|yk - ci\|}{\|yk - cj\|} \right]^{\frac{1}{\alpha - 1}} \right]^{-1} \quad 3b$$

Where i = 1; 2, . . ., c and k = 1, 2; . . ., M.

2. Using updated membership values calculated in step 2 and equation 3b, new cluster centres

are calculated using following equation

$$C_i = \frac{\sum_{k=1}^{M} u_{ik}^2 \, \text{yk}}{\sum_{k=1}^{M} u_{ik}^2} \quad 3c$$

3. Membership matrix is updated using new cluster centres

$$u_{i<1}^{t+1} = \left[\sum_{j=1}^{c} \left[\frac{\|yk - ci\|}{\|yk - cj\|} \right]^{\frac{1}{\alpha - 1}} \right]^{-1} \quad 3d$$

4. If $\left\| U^{(t+1)} - U^{(t)} \right\| < \varepsilon$, then the algorithm stops. Otherwise, go back to equation (3a).

2.6 The Self-Organizing Map (SOM)

It is also known as Kohonen Map or Self-Organizing Feature Map, is an unsupervised neural network based on competitive learning (Kohonen, 1988 and 2001). It projects high-dimensional input data onto a low dimensional (usually two-dimensional) space. Because it preserves the neighborhood relations of the input data, the SOM is a topology-preserving technique. The machine learning is accomplished by first choosing an output neuron that most closely matches the presented input pattern, then determining a neighborhood of excited neurons around the winner, and finally, updating all of the excited neurons. This process iterates and fine tunes, and it is called self-organising. The outcome weight vectors of the SOM nodes are reshaped back to have characteristic data patterns. This learning procedure leads to a topologically ordered mapping of the input data. Similar patterns are mapped onto neighboring regions on the map, while dissimilar patterns are located further apart. An illustration of the work flow of an SOM

application is portrayed in figure 3. The data time series are rearranged in a 2D array such that the data at each time step are reshaped as a row vector. For each time step, the row vector is used to update the weight of the SOM via an unsupervised learning algorithm. This iterative process is called self-organising. The outcome weight vectors of the SOM nodes are reshaped back into characteristic data patterns.

The SOM was introduced to meteorological and climatic sciences in late 1990s as a clustering and pattern recognition method (Hewitson and Crane, 1994 and 2002). It is found to be a useful tool in meteorological applications of different spatial and temporal scales: synoptic climatology, extreme weather and rainfall pattern analysis, cloud classification, as well as climate change analysis

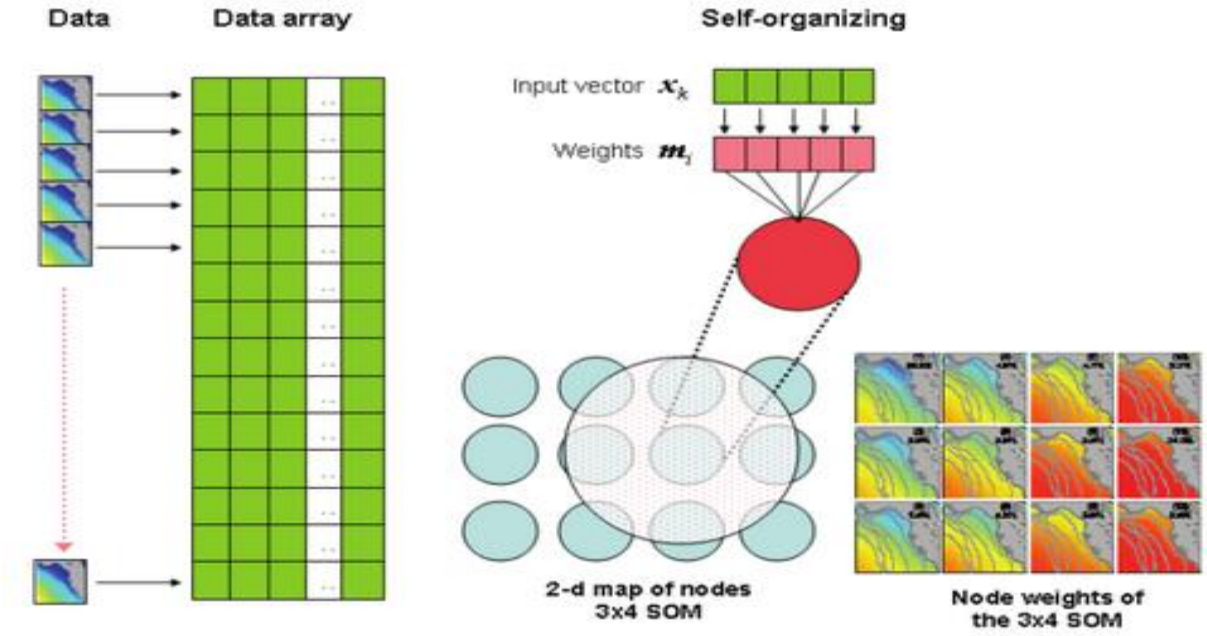

Figure 2: Illustration of SOM mechanisms (Liu *et al.* 2006b)

Many applications of SOM also abound in the field of hydrology for instance, Malmgren *et al*. (1999) used the SOM in climate zonation on the island of Puerto Rico in the Caribbean. Hsu *et al*. (2002) applied the SOM in a rainfall-runoff linear forecast model, called Self-Organising Linear Output map (SOLO). Similarly, Tadross *et al*. (2005) extracted characteristic rainfall patterns over South Africa and Zimbabwe from rainfall data products, and studied the rain-fed maize for the region. Gutierrez *et al.* (2005) applied the SOM to analyse atmospheric patterns over Peru and local precipitation observations at two nearby stations for the purpose of downscaling multimodel seasonal forecasts. Despite all this, the approach has both strength and weakness. This was perfectly illustrated in few instances, traditionally, empirical orthogonal function (EOF) or principal component analysis (PCA) method are known for extraction of patterns of variability in meteorological and oceanographic data. Liu (2007) used both EOF and SOM to extract ocean current patterns from the same data set (a long time series of velocity from a moored ADCP array), and found that the SOM patterns were more accurate and intuitive than the leading mode EOF patterns. Reusch *et al*. (2005b) also tested the SOM against the PCA method using synthetic datasets composed of positive and negative modes of four idealised North Atlantic sea

level pressure fields, with and without noise components. Findings shows that the SOM was more robust than the PCA in extracting the predefined patterns of variability. While K-means is another popular artificial neural network widely used for clustering, after comparing the SOM and k-means methods, Baçãoet *al.* (2005) proposed the use of SOMs as possible substitutes for the k-means clustering algorithms**.** However, wide applications as a tool for feature extraction and clustering, the SOM remains a black box to most meteorologists and oceanographers. SOM new users may be perplexed by the choice of SOM parameters, because different parameter choices may result in different SOM patterns. This challenge may prevent some potential new users from pursuing further SOM applications.

2.7 Discriminant Analysis

The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936 (Cohen, 2003). It is different from an ANOVA or MANOVA, which is used to predict one (ANOVA) or multiple (MANOVA) continuous dependent variables by one or more independent categorical variables. Discriminant function analysis is useful in determining whether a set of variables is effective in predicting category membership (Green *et al*., 2008). Linear discriminant analysis (LDA), normal discriminant analysis (NDA), or discriminant function analysis is a generalisation of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterises or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. Discriminant analysis is used when groups are known *a priori* (unlike in cluster analysis). Each case must have a score on one or more quantitative predictor measures, and a score on a group measure (Bökeoğlu *et al.,* 2017). In simple terms, discriminant function analysis is the act of distributing things into groups, classes or categories of the same type.

Discriminant analysis works by creating one or more linear combinations of predictors, creating a new latent variable for each function. These functions are called discriminant functions. The number of functions possible is either $N_g - 1$ where $N_g =$ number of groups, or P (the number of predictors), whichever is smaller. The first function created maximises the differences between groups on that function. The second function maximises differences on that function, but also must not be correlated with the previous function. This continues with subsequent functions with the requirement that the new function should not be correlated with any of the previous functions. Given group j, with $\mathbb{R}_j$ sets of sample space, there is a discriminant rule such that if $x \in \mathbb{R}_j$, then $x \in j$. Discriminant analysis then, finds "good" regions of $\mathbb{R}_j$ to minimise classification error, therefore leading to a high "percent correct classified" in the classification table (Hardle *et al*., 2007). Each function is given a discriminant score to determine how well it predicts group placement.

Compared with the other methods, Fisher discriminant analysis has advantages. First, it can take multiple factors into consideration **(**Xu *et al.* 2004**).** Therefore, it was applied to detect drought occurrences, it comprehensively takes the soil water and atmospheric moisture conditions into consideration. This is meaningful because atmospheric aridity or associated drought component

has been reported to be more impactful for water and carbon fluxes under climate change (Xu *et al.* 2021). Second, the discriminant formula is created based on local observation; hence, it can be applied directly to the given research site without parameter localisation. Moreover, it is easy to apply because the established discriminant function is not complex and only employs the soil and atmospheric water conditions as the discriminating variables. This would help scientists better conduct field experiments. For example, if an *in situ* water condition control experiment is conducted, drought could be easily and rapidly determined by employing the easily measured soil water content (SWC) and vapor pressure deficit (VPD).

3.0 Conclusion

Findings here unveiled the shortcomings and strength of composite clustering algorithms. It is clear here, that the choice of the cluster algorithm is relative subjective, yet should be bore in mind, that is there is need to explore more than one cluster algorithms in fear of losing microscopic precipitation fact and for the ease of analysis of spatio-temporal phenomena. In recognition of hydrological time series characteristics of trans-boundary interference and spatio-temporal variation, where zones or region that share common boundary may inherit similar hydro-climatic characteristics that seem different from other part of the region. It is imperative, for hydro-climatic researchers to adopt clustering algorithm that reveals the degree of shared characteristics or membership properties, in this hierarchy that is; FCM, PCA, k-means and SOM, for effective water resource planning and management in the phase of climate change.

References

Abdi, H. (2003). Factor rotations in factor analyses. Encyclopedia for Research Methods for the Social Sciences; Sage: *Thousand Oaks, CA, USA*, 792–795.

Agrawal, R., Gehrke, J., Gunopulos, D. &Raghavan, P. (2005). Automatic Subspace Clustering of High Dimensional Data. *Data Mining and Knowledge Discovery,* 11(1),5–33. Doi:10.1007/s10618-005-1396-1. S2CID 9289572.

Bação, F., Lobo, V. &Painho, M. (2005). Self-organising maps as substitutes for k-means clustering, Lecture Notes in Computer Science, Sunderam, V. S., Albada, G. V., Sloot, P. &Dongarra, J. J., (Eds). 3516, 476-483, ISSN 0302-9743, *Berlin Heidelberg, Springer-Verlag.*

Bökeoğlu, Ç. Ö., and Büyüköztürk, Ş.(2017). Discriminant function analysis: Concept and Application. Eğitimaraştırmalarıdergisi, (33), 73-92.

Carvalho, M. J., Melo-Gonçalves, P., Teixeira, J. C., &Rocha, A. (2016). Regionalisation of Europe Based on a K-Means Cluster Analysis of the climate change of temperatures and precipitation. *Physics and Chemistry of the Earth*, 94, 22–28. https://doi.org/10.1016/j.pce.2016.05.001.

Chen, C. H. & Liu, H. C.(2008), A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. *Hydrology Processes*, 22, 3697–3711.

Chukwu, S. E. (2019). *Analysis of Hydrological Drought Duration-Severity for Sokoto-Rima River Basin, Nigeria*. [Unpublished M.Eng Thesis]. Federal University of Technology, Minna, Nigeria.

Chukwu, S. E. (2024). *Establishment of Appropriate Time Scale Resolution for Regional Drought Characterisation.* [Unpublished PhD Dissertation]. Federal University of Technology, Minna, Nigeria.

Cohen, (2003). Applied Multiple Regression Correlation Analysis for the Behavioural Sciences 3rd ed. Taylor & Francis Group components analysis and an artificial neural network, *Journal of Climate*, 12, 977–985, ISSN 0894-8755.

Cui, D. X. & Liu, W. Y. (2000). Application of empirical orthogonal function resolution to analysis of crustal vertical deformation field. *Earthquake*, 20, 82–86.

Deng, X. &Tian, X. (2014). Entropy principal component analysis and its application to nonlinear chemical process fault diagnosis. Asia-Pac. *Journal of Chemical Engineering, 9*, 696–706.

Dunn, J.C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3(3), 32–57**.**

Ghosh, S. & Srinivasan, K.(2016).Analysis of Spatio-temporal Characteristics and Regional Frequency of Droughts in the Southern Peninsula of India. *Water Resource Management.*30, 3879–3898.

Green, S.B., Salkind, N. J. &Akey, T. M. (2008). Using SPSS for Windows and Macintosh: Analyzing and understanding data. New Jersey: Prentice Hall.

Grice, J.W. (2001). Computing and evaluating factor scores. *Psychology Methods*, 6, 430.

Gutierrez, J.M., Cano, R., Cofino, A.S. &Sordo, C.(2005). Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps, *Tellus A*.57:3, 435-447, ISSN 0280-6495.

Hair,J., Black,W., Babin, B., Anderson, R.,&Tatham,R.(1998).Multivariate Data Analysis; PrenticeHall: *Upper Saddle River*, 5, 207–219.

Hardle, W. &Simar, L. (2007). Applied Multivariate Statistical Analysis. *Springer Berlin Heidelberg*, 289–303.

Hartigan, J.A. & Wong, M., (1979). Algorithm AS136: a K-means clustering algorithm, *Application Statistics,* 28, 100–108.

He, Y., Ye, J.Y. & Yang, X.Y. (2015). Analysis of the spatio-temporal patterns of dry and wet conditions in the Huai River Basin using the standardized precipitation index. *Atmosphere Resource.* 166, 120–128.

Hsu, K.L. Gupta, H. V., Gao, X., Sorooshian,, S. & Imam, B. (2002). Self organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, *Water Resources Research*, 38 (12) 1302. doi:10.1029/2001WR000795, ISSN 0043-1397.

Jimoh, O, D., Otache, M.Y., Adesiji, A.R., Olaleye,R.S., &Agajo, J.(2023). Characterisation of Meteorological Drought in Northern Nigeria Using Comparative Rainfall-Based Drought Metrics. *Journal of Water Resource and Protection*, 15(2), 51-70.

Jolliffe, I. T. (2002). Principal Component Analysis. *Springer-Verlag*, 2nd Edition, New York.

Kaiser, H .F.(1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement.*20.pp141–151.

Kim, D. H., Yoo C. & Kim, T. W., (2011). Application of spatial EOF and multivariate time series model for evaluating agricultural drought vulnerability in Korea. Advances in Water Resources.5: pp1-3.

Kohonen, T. (2001). Self-Organizing Maps. Springer-Verlag, ISBN 3-540-67921-9, *New York, Berlin, Heidelberg.*

Kohonen, T.(1988). Self-Organization and Associative Memory, *Springer-Verlag*, ISBN 0-387-18314-0, New York, Berlin, Heidelberg.

Loboda, N. S. Glushkov, A. V. &Khokhlov, V. N. (2005). Using Meteorological Data for Reconstruction of Annual Runoff Series over an ungauged area: empirical orthogonal function approach to Moldova-Southwest Ukraine region, *Atmospheric Research*. 77, 100–113.

Loukas, A. &Vasiliades, L. (2004). Probabilistic Analysis of Drought Spatiotemporal Characteristics in Thessaly Region, Greece. *National Hazards Earth System Science*, 4, 719–731.

Malmgren, B. A. &Winter, A. (1999). Climate zonation in Puerto Rico based on principal.

Otache, M.Y. (2008). *Contemporary Analysis and Modelling of Benue River Flow Process.*[Unpublished doctoral dissertation]. Hohai University, Nanjing, P. R. China.

Pei, X., Xiaohui L., Yuhu, Z., Mingna, W., Ihnsup H., &Qiuhua, C. (2018). Cluster analysis of drought variation and its mutation characteristics in Xinjiang province, during 1961–2015. *Hydrology Research*, 49(4), 1-15.

Pham, D.T., Dimov, S. S. & Nguyen, C. (2005). Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: *Journal of Mechanical Engineering Science*. 219(1), 103–119, doi.org/ 10.1243/095440605X8298.

Rajsekhar, D., Mishra, A. K. & Singh, V.P. (2011) Drought regionalisation of brazos river basin using an entropy approach. In: 2011 Symposium on data-driven approaches to droughts. Paper 40

Raziei, T., Martins, D., Bordi, I., Santos, J., Portela, M.M. & Pereira, L. (2015). A SPI Modes of Drought Spatial and Temporal Variability in Portugal: Comparing Observations, PT02 and GPCC Gridded Datasets. *Water Resource. Management,* 29, 487–504.

Rencher, A.C. (2002). Methods of Multivariate Analysis, 2nd ed.; *John Wiley & Sons Hoboken*, *NJ, USA*.492: pp 380–404.

Reusch, D. B., Alley, R. B. &Hewitson, B. C. (2005b). Towards ice-core-based synoptic reconstructions of west antarctic climate with artificial neural networks. International Journal of Climatology, Vol. 25, 581-610, ISSN 0899-8418.

Rodrigues, R., Santos, M.A. &Correia, F.N. (1993). Appropriate Time Resolution for Stochastic Drought Analysis. *Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization*, 247-265.

Serinaldi, F., Bonaccorso, B., Cancelliere, A. and Grimaldi, S. (2009). Probabilistic characterization of drought properties through copulas. *Physics and Chemistry of the Earth.* Parts A/B/C34(10-12), 596-605.

Sirjacobs, D. A. and Alvera-Azcarate, A. (2011). Cloud filling of ocean colour and sea surface temperature remote sensing products over the Southern North Sea by the Data Interpolating Empirical Orthogonal Functions methodology, *Journal of Sea Research*, 65(1), 114–130, 2011.

Tadross, M. A., Hewitson, B. C. and Usman, M. T. (2005). The Interannual variability of the onset of the maize growing season over South Africa and Zimbabwe. *Journal of Climate*.18 pp3356-3372, ISSN 0894-8755.

Wei, F., Shengzhi, H. Qiang, H., Guohe, H., Hao, W., Guoyong, L. & Lu, W.(2020). Identifying Drought Propagation by Simultaneously Considering Linear and Nonlinear Dependence in The Wei River Basin of The Loess Plateau, China. *Journal of Hydrology,* 591(2), 98-116.

Xu, M.J, Zhang T. & Zhang, Y.J. (2021) Drought Limits Alpine Meadow Productivity in Northern Tibet. *Agricultural for Meteorology*, 303, 108371.

Xu, Y, Yang, J.Y, &Jin, Z. (2004) A Novel Method for Fisher Discriminant Analysis. *Pattern Recognition*, 37, 381–384.

Yu, M., Li, G., Hayes, M.J., Svoboda, M. & Heim, R.R. (2013). Are droughts becoming more frequent or severe in China based on the Standardised Precipitation Evapotranspiration Index: 1951–2010? *International Journal of Climatology*, Doi 10.1002/joc.3701.