International Journal of Basic Science and Technology, Vol. 3 No 1 (2017) p.37-46

Comparison of the Different Hierarchical Clustering Techniques for the Classification of Soils under Oil Palm in Nigeria

[1]Edokpayi, A. A., [2]Agho, C. A. [1]Adeh, S. A. and [3]Okpamen, S. U.

[1]Statistics Division, Nigerian Institute for Oil palm Research (NIFOR), Benin City

[2] Plant Breeding Divisions, Nigerian Institute for Oil palm Research (NIFOR), Benin City, Nigeria

[3] Soils and Land Management Division, Nigerian Institute for Oilpalm Research (NIFOR), Benin City.

Received: 1st June, 2017
Accepted: 3rd Nov.2017
Available online
30th Dec.,2017

Keywords:
Cluster Analysis, Hierarchical clustering methods, Soil data, Oil palm belt

Abstract: Cluster Analysis has been very well used in soil classification, but sustainable guidelines are not readily available for the choice of appropriate clustering technique for soil data. This paper tested the robustness of five common agglomerative hierarchical clustering methods using soil data collected from 25 farmers' field within the oil palm belt of Nigeria using the Cophenetic Correlation Coefficients (CPCC) while the Kappa coefficients of agreements were used to compare grouped formed by the hierarchical method and that of the taxonomic classification. From the dendrogram arising from the clustering analysis, shows some diversity of soils within the study locations. However, the groups formed by the five methods were composed of different numbers of soil individuals indicating that the different methods created different results. The result of the Cophenetic Correlation Coefficients (CPCC) shows the highest values for the Average linkage (0.7337065) while the Complete (0.7255978) ranked second with the Ward method (0.7102611) coming close while Centriod and the Single methods ranked 4th and 5th with a CPCC value of (0.6928591) and (0.5071803) respectively. Kappa coefficients of agreements were generally low for the five methods compared indicating that the classification by the traditional method and the hierarchical methods were not the same. The study therefore shows that the average linkage method ranked best in the classification of the soil data in the study location. From the study, it is suggested that researchers should evaluate carefully the methodology to apply before using any of the hierarchical clustering methods, and that simply applying a particular method to a data set and accepting the solution at face value will not be adequate.

Corresponding author: edokpayiaustin@yahoo.com

Introduction

The oil palm belt of Nigeria covers a wide expanse of land with great diversity of soils (Omereji, 2005). These various soil types accounts for the large differences in the supply of mineral nutrients between locations in the zones such that each location has its own particular soil characteristics. According to Ogeh *et al.* (2012), variations in soil properties have been found to influence soil management and crop production as each soil type has its own unique management practices. Therefore, accurate assessment of the levels and patterns of soils diversity can be invaluable to soil suitability assessments. Also knowledge of the soil diversity can facilitate reliable classification of soils so that site or locations with similar soil properties can be identified and appropriate management technique for each group designed. This can facilitate group specific investigations and management techniques recommendation that will be beneficial to all farmers within a particular cluster. Management technique that will benefit a group is more cost effective than site specific investigation and management technique recommendation.

The traditional soil classification schemes such as the USDA Soil Taxonomy and the FAO UNESCO Legend systems are commonly used to classify soils. These natural system approaches are based on some presumed soil genesis and uses taxonomic criteria involving soil morphology and laboratory tests to inform and refine hierarchical classes. However some of these traditional soil classification techniques have been seriously criticized. Sometimes the soil properties overlap resulting in difficulty in differentiating one soil group from the other, making conclusions imprecise (Kank and Tripath, 2007). As observed by Lamontagne and Camire (1987), a meaningful classification of sandy soils for instance based on field observations is difficult, because these soils have only very subtle differences in texture and structure and profile differentiation is often minimal. So, simply applying the traditional classification method and accepting the solution at face value is generally not adequate. Consequently, evaluation and validation of this traditional method using appropriate classification model is necessary to ensure consistency and plausibility of results.

An alternative to the traditional classification methods is the numerical techniques. This approach uses rigorous mathematical models that give more homogeneous soil groups and can be more easily applied than the taxonomic method. In this approach, soil individuals are grouped by multivariate method such as cluster analysis. This produces natural groupings without requiring inferences about soil genesis. This method lend themselves not only to summarizing large amounts of data, but also making more rational decisions concerning relationships between soils and between soils and their environments especially when subtle differences exists (Lamontagne and Camire ,1987).

Two broad classes of doing this are the hierarchical and nonhierarchical methods. Nonhierarchical methods identify groups (clusters) of similar samples but do not characterize relationships among clusters (Gauch, 1982). In hierarchical methods, individual and clusters are, most commonly merged (agglomerative) or, less commonly divided (divisive). However, the agglomerative method dominates published uses because they were made popular by early literature and have been object of empirical analysis (Clarke and Warwick, 2001). A variety of agglomerative clustering methods exist depending on which technique or linkage method is used to fuse the objects during the clustering process. Some of the common ones include; Single, Complete and Average linkages, others are the Centroid and Wards methods. The single linkage method is said to work well for elongated cluster with unequal variances and unequal sample sizes, and the most versatile and useful for detecting outliers, but chaining has been identified as a major defect. This property (chaining) may cause the methods to fail to resolve relatively distinct cluster when there are a small number of individuals lying between them (Everitte, 1993). The complete linkage method on other hand, the groups tend to be hyperspherical and, at some stage in the procedure, occupy roughly equal character space. The method is equally strongly affected by outliers. As observed by Lamontagne and Camire (1987), this model of grouping is used less frequently for soil classification, but is frequently used for the classification of vegetation and other environmental data. The average linkage is usually preferred to the single and complete linkage methods for cluster analysis because it uses information about all pairs of distances, not just the nearest or furthest distances. The method and the centriod algorithm tend to produce clusters with rather low within cluster variance and similar sizes. Both methods are said to also be affected by outliers though not as much as the complete linkage. Another major disadvantage of the centroid method outside being affected by outliers is that the distance at which clusters are combined can actually decrease from one step to the next. According to Everitt (1993), this is an undesirable property because clusters merged at later stages are more dissimilar than those at earlier stages. For the Ward's methods, it performs well when the data contained approximately equally sized clusters, but poorly when the clusters are of differently sizes (Everitt, 1987).

As observed by Warsha (2008), the results of hierarchical classifications depend on the choice of the clustering technique (linkage method) and the initial dissimilarity index used to calculate the pair wise dissimilarity between objects. Quinn and Keough (2002) noted that the choice of linkage method is even more critical than the choice of the dissimilarity measure. Vakharia and Wemmerlov (1995), added that the purpose of the analysis, the nature of the data and the standardization of the data, together play a role in determining the optimum clustering technique to use. According to Milligan (1980), no single method could be claimed to be superior for all types of data, as different clustering methodologies may result in different results and interpretation of the result dependent on the methodology used. Though the numerical technique has received application in recent times in the classification of soil data, see Campbell and Jose (1970), Nageswara *et al.*, (2002), Lamontagne and Camire (1987), Mutsuers *et al.*, (1997), in each of these studies different clustering methodologies have been used resulting to different output and interpretation. In most cases, the choice of clustering technique applied is not base on any meaningful guidelines. Therefore for consistency and plausibility of results there is the need to examine all procedures in order to determine the most appropriate method.

The purpose of this study is to apply the hierarchical clustering technique to the classification of soil data. To ensure consistency and plausibility of our results, we apply five hierarchical clustering algorithms, examine their robustness in the classification of the data and determine the most efficient.

Materials and methods

Data used for the study

The data are a subset of data used to classify soils of the oil palm belt in a previous study and were sourced from the Chemistry Division of the Nigerian Institute for Oilpalm Research (NIFOR). The data consisted of 12 soil descriptors (soil properties) collected from twenty five farmer's field across the belt with five each sampled from five different parents' materials. The characterization of the parents' materials and their respective location are presented in Table1, while the site characteristics are presented in Table 2 below.

Table1: Characterization of the Parent Materials and Soil sample collected

S/no	Crystalline metamorphic and igneous rocks	Shale mixed with sandstone and clay	Coastal plain sand	Coastal alluvium	Fresh water swamps
1	Ilesha	Afikpo	Nifor	Bori	Ologbo
2	Ondo	Okigwe,	Agbor	Portharcourt	Kwale
3	Akure,	Owode	Abudu	Calaber	Mosogar
4	Oshogbo	Itori	Okitipupa	Ikotabasi	Degama
5	Ibadan	Umuahia	Evboneka	Abak	Otegbo

Table 2: Soil – site characteristics measured

S/no	soil properties	properties Name	S/no	soil properties	properties Name
1	pH	Soil Ph	7	Mg	Magnesium
2	Orgm	Organic matter	8	S	Sulphur
3	N	Nitrogen	9	Fe	Iron
4	P	Phosphorus	10	Mn	Manganese
5	K	Potassium	11	Cu	Copper
6	Ca	Calcium	12	Zn	Zinc

Hierarchical clustering methods

The basic description according to Quinn and Keough (2002) starts with calculating a matrix of dissimilarity between the objects or variables and two objects which are most similar cluster together to form new objects replacing the merger pair. The dissimilarity between the new set of objects is calculated and again the most similar objects are merged. The process continues until all the objects are linked in a cluster. When items (units or cases) are clustered, proximity is usually indicated by some sort of distance. Distance measures define distances or dissimilarities between observations. More similar observations have shorter distance and more diverse ones have greater distance.

There are in turn a number of distance measures. For continuous variables, dissimilarity measures include; Euclidean distance, Squared Euclidean distance, Manhattan distance, etc. The Euclidean distance is most probably the commonly used type of distance measurement and was used for this study. The advantage of this distance measure is that the distance between any two individual is not affected by the addition of new individuals or objects.

Symbolically, given two distance $\boldsymbol{i}$ and j, the Euclidean distance measure between i^{th} and i^{th} individual is given as;

$$d_{iJ} = \sqrt{\sum_{k=i}^{p} \left(X_{ik} - X_{Jk}\right)^2} \qquad 1$$

where X_{ik} and X_{jk} are the values of the k^{th} variables for observation $\boldsymbol{i}$ and j. With this similarity matrix, a cluster analysis was done.

A clustering criterion usually determines what the clusters look like given the distance measure. Commonly used criteria include single linkage, complete linkage, average linkage, centriod linkage, median and the ward's methods.

Single linkage: This is one of the simplest agglomerative hierarchical clustering methods and is also known as the nearest neighbor. The defining feature of this method is that distance between groups is defined as that of the closest pairs of individual, where only pairs consisting of one individual from each group are considered Everitt(1993). Symbolically, given two groups, R and Q, the single linkage method merges groups based on the minimum distance between two objects in two groups; therefore the distance between cluster R and Q is define by

$$ds(R,Q) = \min d(i,j), \quad i \in R, j \in Q \qquad 2$$

$where\ d(i,j)$, is the distance between the $i^{th}\ and\ j^{th}$ objects and $\boldsymbol{i}$ and j are individual within each group

Complete Linkage Clustering Method: The Complete linkage or further neighbour clustering method is the opposite of the single linkage method in the sense that the distance between groups is now defined as that of the most distant pair of individuals one from each group. Therefore, the distance between clusters R and Q is defined as

$$dc(R,Q) = Max\ d(i,j) \qquad 3$$

$$i \in r, j \in q$$

where $\boldsymbol{i}$ and j are points in cluster R and Q

Group Average Clustering: Here the distance between two clusters is defined as the average of the distances between all pairs of individuals that are made up of one individual from each group. The distance in average linkage is defined a

$$dA(R,Q) = \frac{1}{|R||Q|} \sum_{i \in R, j \in Q} d(i,j) \qquad 4$$

$where\ |R| = the\ number\ of\ object\ in\ cluster R\ and\ |Q| the\ number\ of\ object\ in\ cluster\ Q.$

Centroid Clustering: With this method, groups once formed are represented by their mean values for each variable, that is their mean vector, and inter- group distance is now defined in terms of distance between two such mean vectors. Let define the distance between cluster centroids $R_i and\ Q_j$

$$d(R_i,Q) = d(R_i,Q) \qquad 5$$

$where \quad C_i = \frac{1}{|ri|} \sum_{xRi} x \quad C_j = \frac{1}{|qj|} \sum_{xQj} x$

Ward Method: Ward (1963) proposed a clustering procedure seeking to form the partitioning P_n, P_{n-1}, - - -, P_1 in a manner that minimizes the loss associated with each groping. And to quantify that loss in a form that is readily interpretable. At each step I the analysis, union of every possible pair of clusters is considered and the two clusters whose fusion results in the minimum increase in information loss are combined. Information loss is defined by Ward's in terms of an error sum – of squares criterion, ESS. The error sum of squares (SSE) is defined as

$$SSE = \sum_{i-1}^{k} \sum_{i=1}^{ni} (Yij - \overline{Y\imath})^2 \qquad 6$$

Where yij is the j^{th}object in the i^{th} cluster

Dendrogram

The output of the clustering procedures is a tree-like structure called dendrogram with the x – axis showing the objects and the y – axis indicating the level of similarity or dissimilarity of the groupings. Similarity between the clusters diminishes moving from lower to upper levels. In the dendrogram, object may be compared at any level of similarity. The level at which the tree is cut determines the number of clusters formed (Romesburg 1984). The decision of where to cut the tree to form groups is subjective In that no criteria are used and a practical rule is to cut the tree at a level that will produce clusters that is maximally related to variables of interest

Cophenetic Correlation Coefficients

The method most commonly used criterion for assessing the robustness of the various agglomerative methods is the Cophenetic Correlation Coefficients (CPCC). The correlation measures how well the clustering was able to match the original dissimilarity in the data. It is simple a correlation between the original dissimilarity matrix and the Cophenetic matrix which is the total dissimilarity produce after clustering. I.e. distance at which two objects become members of the same cluster.

Kappa Statistic:

To get the result which would be comparable with the original group membership, we extracted the partition, a part of the dendrogram with the same number of clusters or groups as the original number of groups. Since the original number of groups was five, we extracted five groups from the dendrogram to form an R X C contingency table. The levels of agreements between the cluster groups and the original soil group were calculate using Kappa statistics.

$$K = \frac{Observed\ agreement - chance\ agreement}{1 - chance\ agreement} \quad 7$$

In terms of symbol, this is

$$K = \frac{P_o - P_c}{1 - P_c}$$

Where P_o is the proportion of observed agreements and P_c is the proportion of agreements expected by chance. Kappa gives us a minimum rating of the degree to which this occurs.

The calculation is based on the differences between how much agreement is actually present (observed agreement) compared to how much agreement would be expected to be present by chance alone.

Interpretation of Kappa

Kappa	Agreement
< 0	Less than chance agreement
0.01- 0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61- 0.80	Substantial agreement
0.81-0.99	Almost perfect

Following, the consensus already reached by several authors (Fleiss, 1981, Gardner, 1995), Kappa value less than 0.4 indicates low agreement and values above 0.7 indicates relative high agreement. In the context of this study, high agreement corresponds to the clustering solution matching the actual grouping of the soils.

Tukey's test for Detecting Outliers

This approach uses the interquartile range defined as follows;

[Q1–K(Q3–Q1),Q3+K(Q3-Q1)] 8

Where, Q1 and Q3 are the lower and upper quartile respectively and K = 1.5 indicating an outlier and K = 3 indicating data far out.

Results and Discussion

The results for outlier detection, Kappa statistic, CPCC and the hierarchical clustering procedures performed using R Statistical package is shown below. Our data set shows the presence of outliers in some of the variables. 7.70 for pH was higher than the range of 2.24- 7.65 while 21.96 was above the range for Orgm. 0.48, 55, 36 and 1.33 were outside the range for N, P and K respectively while 6, and 3.67, 2.96 for Ca and Mg . 48.30, 68.80 and 73.60 were also outside the range for Mn.

Table 3: Testing for Outlier

Tukey Method Variables	Q1(lower quartile)	Q3(upper quartile	Range of Minor Outliers	Range of Major Outliers
pH	4.65	5.40	3.525-6.525	2.24-7.65
Orgm	1.8	3.8	1.2-6.8	4.2-9.8
N	0.10	0.18	0.02-0.3	0.14-0.42
P	4.5	9.5	3-17	10.5-24.5
K	0.06	0.20	0.15-0.41	0.36-0.61
Ca	0.30	1.30	0.2-2.8	2.7-4.3
Mg	0.18	0.75	0.67-1.61	1.53-2.46
S	9.5	21.3	8.2-39.0	26.2-56.7

Fe	38.0	135.50	108.25-281.75	254.5-428.0
Mn	2.10	10.35	10.25-22.75	22.65-35.1
Cu	0.60	3.50	4.0-7.85	8.1-12.2
Zn	3.35	7.45	2.8-13.6	8.95-19.95

The dendrogram resulting from the analysis was used to form the groups. An examination of the dendrogram (Appedix1) shows that the population can be divided into five candidate groups. The groupings by the respective linkage methods, their CPCC and Kappa statistics are presented below.

Table 3: Single Linkage method

Cluster No	Cluster Membership
1	Afikpo, Okigwe, Ologbo, Kwele, Nifor, Mosogar, Otegbo , Agbor , Abudu, Bori, Portharcourt, Calaber, Itori, Akure
2	Ondo, Ikotabasi, Okitipupa, Umuahia, Abak, Ilesha, Degema, Evboneka
3	Oshogbo
4	Ibadan
5	Owode
CPCC	0.5071803
Kappa	0.15

Table 3 shows the number of group and group membership using the single linkage. Two major groups were identified with the first group having fourteen memberships while the second group had eight members. Three other locations were separately outside the two groups. The table shows that soil individuals within each cluster are composed of one or two dominant soil series. Group 1 completely lump together 1 member from Crystalline metamorphic and Igneous rocks, 3 from Shale mixed with sand stone and clay, 2 from Coastal plain sand,3 and 4 each from Coastal alluvium and Fresh water swamps respectively. The second group also comprises of 2 members each from Crystalline metamorphic and igneous rocks, Coastal plain sand, and Coastal alluvium while Shale mixed with sand stone and clay and Fresh water swamps have I member each. The CPCC value of 0.5071803 was very low compared to other clustering methods. This result is similar to Blashfield, (1976) when four

hierarchical methods were compared. The poor performance could be attributed to chaining and the equal sample size of the data used for the study as the approach is said to work well with unequal sample size. The Kappa value of 0.15 was low indicating a poor agreement between numerical method and taxonomic method.

Table 4: Complete Linkage method

Cluster No	Cluster Membership
1	Afikpo, Okigwe, Ologbo, Evboneka
2	Otegbo, Agbor, Mosogar, Abudu, Kwele, Nifor, Calaber
3	Bori, Portharcourt, Itori, Akure
4	Ilesha, Degama, Oshogbo, Ibadan
5	Ondo, Ikotabasi, Umuahia, Abak, Okitipupa, Owode
CPCC	0.7255978
Kappa	0.25

Grouping formed by the complete linkage method is presented in Table4. Group 1 was dominated by Shale mixed with sand stone and clay with 2 soil individuals and one each from Coastal plain sand and Fresh water swamps. Group 2 combined 2 members from Coastal plain sand, three from Fresh water swamps and one from Coastal alluvium with non Crystalline metamorphic and igneous rocks and Shale mixed with sand stone and clay. Of the four members in group three, one each from crystalline metamorphic and igneous rocks and Shale mixed with sand stone and clay while the remaining two was from Coastal alluvium. Group four comprises of three members from crystalline metamorphic and igneous rocks while the remaining one was from Fresh water swamps. Group five combined one each from crystalline metamorphic and igneous rocks and Coastal plain sand and two each from Shale mixed with sand stone and Coastal alluvium respectively. The CPCC value was high (0.7255978) indicating that the method performed equally well in soil classification.

Table 5: Average Linkage method

Cluster No	Cluster Membership
1	Afikpo, Okigwe, Ologbo, Evboneka
2	Ondo, Ikotabasi, Okitipupa, Umuahia, Abak,Ilesha, Degama
3	Oshogbo, Ibadan, Owode
4	Kwele, Nifor, Mosogar, Calaber
5	Bori, Portharcourt, Itori, Akure, Otegbo, Agbor, Abudu
CPCC	0.7337065
Kappa	0.05

Grouped formed using the Average linkage method is presented in Table 5. Group one has four members that combines two members from Shale mixed with sand stone and one each from Coastal plain sand and Fresh water swamps. Group two comprises of seven members with two each from crystalline metamorphic and igneous rocks and Coastal alluvium, one each from Shale mixed with sand stone, Coastal plain sand, and Fresh water swamps. Two members from crystalline metamorphic and one igneous rocks and Shale mixed with sand stone constituted the membership of group three while two members from Fresh water swamps and one each from Coastal plain sand and Coastal alluvium make up the membership of group four. Group five has membership spread across the five main soil types. Of the seven members in this groups Coastal plain sand and Coastal alluvium has two members while the remaining soil types has one member each. This approach recorded the Highest CPCC value of (0.7337065) suggesting that this grouping is more robust than other methods. This could be attributed to the fact that the approach uses all information about all pairs of distance and is mildly affected by outliers.

Table 6: Centroid method

Cluster No	Cluster Membership
1	Akure
2	Itori, Bori, Portharcourt, Calabar, Kwele, NIFOR, Mosogar, Abudu, Otegbo, Agbor
3	Owode
4	Okitipupa, Ondo, Ikotabasi, Umuahia, Abak, Evboneka, Ologbo,Afikpo, Okigwe
5	Ibadan, Oshogbo, Ilesha, Degema
CPCC	0.6928591
Kappa	-4.23

Table 6 shows the numbers of grouped formed and grouped membership when the centroid clustering linkage was applied to the data set. Groups one and three have one member each while groups two; four and five has ten, nine and four memberships respectively. Group two comprises of three members each from Coastal plain sand, Coastal alluvium and Fresh water swamps while the remaining one came from Shale mixed with sand stone. Group four was dominated by members from Shale mixed with sand stone and Coastal plain sand with three and four members each while Coastal alluvium and Fresh water swamps has one member each. The membership in group five was shared between crystalline metamorphic and one igneous rocks and Fresh water swamps. The coefficients of agreement was very low (-4.23) and the CPCC value of (0.6928591) was equally low indicating that this approach will not be a choice in soil classification.

Table 7: The Ward's method

Cluster No	Cluster Membership
1	Afikpo, Okigwe, Ologbo, Evboneka
2	Bori, Portharcourt, Itori, Akure,
3	Otegbo, Agbor, Abudu, Kwele, Nifor, Calaber,Mosogar
4	Oshogbo, Ibadan,Owode
5	Ilesha, Degama,Ondo, Ikotabasi, Umuahai, Abak, Okitipupa
CPCC	0.7102611
Kappa	0.2

Five groups were identified using the Ward's hierarchical method with 4, 4, 7, 3, and 7 group memberships respectively (Table 7). Of the four members in group one, two came from Shale mixed with sand stone while one each came from Coastal plain sand and fresh water swamps. Group two comprises of two members from Coastal alluvium while one each from crystalline metamorphic and igneous rocks and Shale mixed with sand stone. Group three comprises of seven members with three each coming from Coastal plain and fresh water swamps, while the remaining one was from Coastal alluvium. The three members in group four were shared between crystalline metamorphic and igneous rocks and Shale mixed with sand stone with two and one membership respectively. Group five comprises of members from the five soil types with two each from crystalline metamorphic and igneous rocks and Coastal plain while one each came from the remaining soil types. The CPCC value was very high (0.7102611) and this could be due the fact that the clusters are near equally sixed.

Conclusion and Recommendation

Though, there are significant variations in the groupings of the soils by the five hierarchical methods, but when sample sizes are equal and there is presence of outlier in the data set, the Average linkage method performed better. That, classification of the soils based on hierarchical method differs markedly from the taxonomic approach, and therefore numerical technique should serve as a complementary tool for the classification soil, since it uses rigorous mathematical models that give more homogeneous soil groups. From the study, different clustering methodologies results in different outcome which may make interpretation and correct decision about the pattern of soil distribution within oil palm belt imprecise. Therefore researchers who are attempting to classify soil data set into segments need to evaluate and validate all clustering methods so that the appropriate clustering technique can be determined for the data sets.

References

Blashfield, R. K. (1976). Mixture model tests of cluster analysis. Accuracy of four agglomerative hierarchical methods. *Psychol*. Bulletin, 83,377-385

Everitt, B. S. (1987).Introduction to optimization methods and their application in Statistics, Chapman and Hill, London

Everitt, B. S. (1993). Cluster Analysis, Wiley and Sons Inc, New York – Toronto

Fleiss, J. L. and Zubin, J. (1981). On the method and theory of clustering. *Multivariate Behaviour* Res., 4, 235 -250

Gardner, W. (1995). On the reliability of sequential data: measurements meaning and correction, In Analysis of Change (Edited by John M.Gottman) Erlbaum

Gauch, H .G. (1982). Multivariate analysis in Community ecology, Cambridge University Press.

Kang, B. T. and Tripath, B. (1981). Soil Classification and Characterization. A technical paper

Lamontagne,L and Camire,C (1987). Soil Analysis and Numerical Classification of the Lanoraie Delta , QUEBEC; *Canadian Journal of Soil Science*, vol.67, No 3.

Milligen, G.W.(1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45, 325* – 342.

Mutsaers, H. J.W., Weber, G.K., Walker, P., Fisher, N.M. (1997). *A Field Guide for On* –Farm

Experimentation. A publication by the International Institute of Tropical Agriculture (IITA)

Nageswara, R.D.V. , Mulcahy, M. J.and Mcarthur, W.M. (2002). Numerical Classification of some soils under rubber in Kerala. *Journal of the Indian Society of Soil Science*, Vol.50,4

Ogeh, J.S. and Osiomwan (2012). Evaluation of the effect of oil palm on some physical and chemical properties of Rhodic paleudults. *Nigerian Journal of Basic and Applied Sciences*, 20(1): 78-82.

Omereji, G.O. (2005). The Oil Palm Industry in Nigeria: Cultivation. Processing and Trade. Mindex Publishing

Quinn, G.P. and Keough, M.J. (2002). Experimental Design and Data Analysis for Biologists. Cambridge University Press.

Romesburg, H. C. (1994). Cluster analysis for researchers. Lifetime learning publications, Belmont, Calif.

Sneath, P. H. A. (1957). The application of computer to taxonomy, *J. Gen. Microbiology* 17,201-226

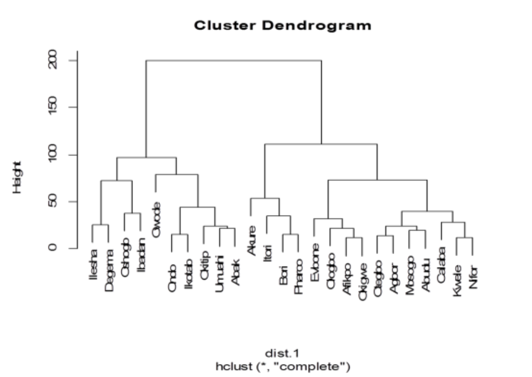
Vakharia, A.J and Wemmerlov, U (1995). A Comparative Investigation of hierarchical clustering techniques and dissimilarity measures applied to the cell formation problem. *Journal of Operations Management*, 13 (2): 117 - 138

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function, *J. Amer. Statist. Assoc*. 58, 236 -244.

Warsha, S. (2008). Robustness of three hierarchical agglomerative clustering techniques for ecological data. A thesis submitted to the Department of Mathematics, University of Iceland.

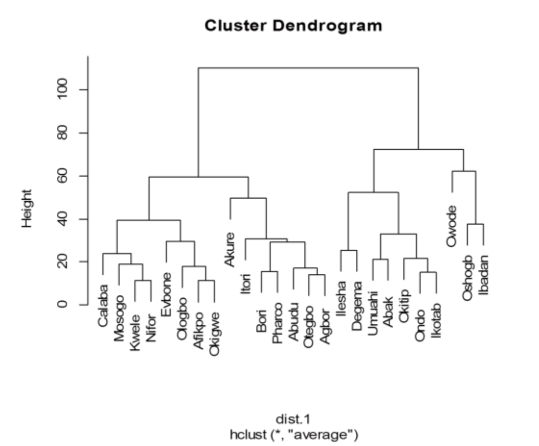
Appendix

COMPLETE LINKAGE



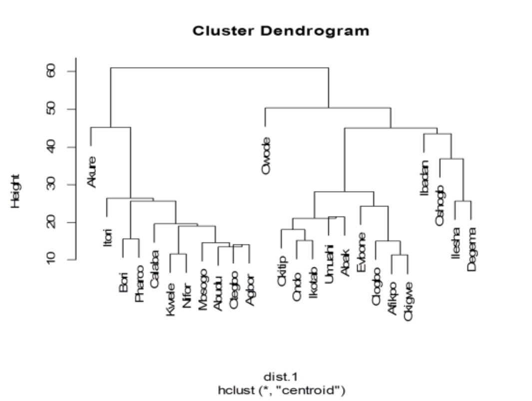
COPHENETIC CORRELATION (COMPLETE LINKAGE): 0.7255978

AVERAGE LINKAGE

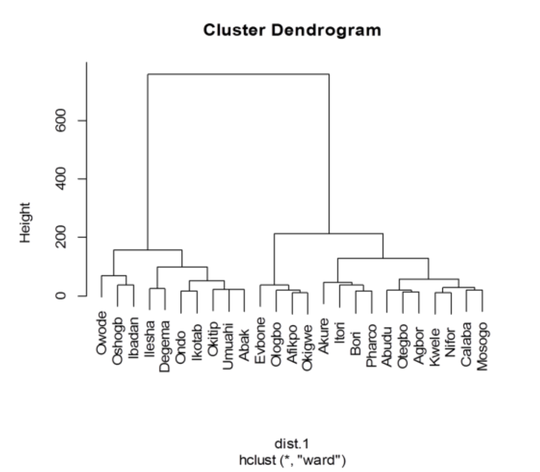


COPHENETIC CORRELATION (AVERAGE LINKAGE): 0.7337065

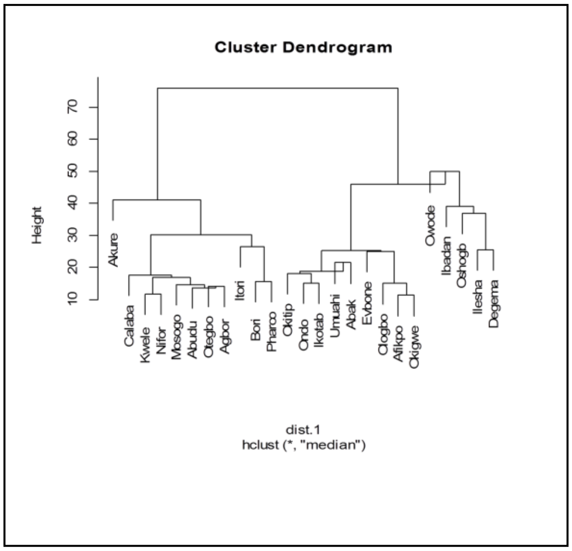
CENTROID



WARD METHOD



COPHENETIC CORRELATION (CENTROID LINKAGE): 0.6928591



COPHENETIC CORRELATION (WARD LINKAGE): 0.7102611