

RESEARCH ARTICLE

PRINCIPAL COMPONENT ANALYSIS USING MULTIVARIATE METHOD FOR ANALYSING INVENTORY FIELD DATA IN FEDERAL COLLEGE OF FORESTRY, IBADAN, NIGERIA

¹Aghimien, E.V., ²Chenge, I.B., ¹Ekaun, A.A., ³Geplu, O.A

¹Federal College of Forestry, Ibadan, Nigeria

²Federal University of Agriculture, Makurdi, Nigeria

³Forestry Research Institute of Nigeria (FRIN)

ABSTRACT

Multivariate analysis provides statistical methods for study of the joint relationships of variables in data that contain inter-correlations. At present even with the developed techniques, one of the main limiting factors for multivariate analysis operation is the lack of availability of data. The study aimed to develop Principal Component Analysis (PCA) using multivariate method for analysing inventory field data and their applications for sustainable forest management. The research was carried out in the two stands of *Gmelina arborea* and *Tectona grandis* plantation. Total enumeration was carried out and a total of 108 *Gmelina arborea* and 205 *Tectona grandis* species were measured. Parameters measured were diameter at breast height and Total height. DBH was measured at 1.3m above the ground level, using spiegelrelascope and measuring tape. The result revealed that *Tectona grandis* had a standard error for height parameter of 0.354 and diameter at breast height of 0.198 while *Gmelina arborea* had a standard error value for height of 0.248 and diameter at breast height of 0.045, the higher the standard error of the stands, the higher the susceptibility to error but the lower the standard error, the closer to perfection. This implies that diameter at breast height of *Gmelina arborea* had a low standard error of 0.045 which means there is lesser susceptibility to error. Height parameter of *Tectona grandis* and *Gmelina arborea* had a positive low correlation value of 0.027 and diameter at breast height parameter of *Tectona grandis* and *Gmelina arborea* had a low negative correlation value of -0.190, this indicates that there is a weak relationship between the two tree species. Conclusively *Tectona grandis* and *Gmelina arborea* are independent on each other for growth and survival.

KEYWORDS: Principal component analysis, Total height, *Gmelina arborea*, *Tectona grandis*, Multivariate

Received for Publication: 18/02/16

Accepted for Publication: 25/05/16

Corresponding Author: aghimien4@yahoo.com



All rights reserved

This work by [Wilolud Journals](http://www.wiloludjournal.com) is licensed under a [Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/)

INTRODUCTION

General Background

Multivariate analysis provides statistical methods for study of the joint relationships of variables in data that contain inter-correlations. Because several variables can be considered simultaneously, interpretations can be made that are not possible with univariate statistics. Applications are now common in medicine, agriculture, geology, social sciences, and other disciplines. The opportunity for succinct summaries of large data sets, especially in the exploratory stages of an investigation, has contributed to an increasing interest in multivariate methods (Jobson, 1992).

Multivariate statistics is a form of statistics encompassing the simultaneous observation and analysis of more than one outcome variable. The application of multivariate statistics is multivariate analysis methods of bivariate statistics, for example simple linear regression and correlation, are not special cases of multivariate statistics because only one outcome variable is involved (Jobson, 1992)

Multivariate statistics concerns understanding the different aims and background of each of the different forms of multivariate analysis, and how they relate to each other. The practical implementation of multivariate statistics to a particular problem may involve several types of univariate and multivariate analysis in order to understand the relationships between variables and their relevance to the actual problem being studied (Rencher, 2001).

Forest Resources Inventory

All decision-making requires information. In forestry, this information is acquired by means of forest inventories, systems for measuring the extent, quantity and condition of forests (Penman *et al.*, 2003). More specifically, the purpose of forest inventories is to estimate means and totals for measures of forest characteristics over a defined area. Such characteristics include the volume of the growing stock, the area of a certain type of forest and nowadays also measures concerned with forest biodiversity. A forest inventory could in principle be based on a complete census, i.e. on measuring every tree in a given area, but this is usually impossible in forestry because of the large areas involved. Therefore the acquisition of information is typically based on sampling, i.e. only a proportion of the population, a sample, is inspected and inferences regarding the whole population are based on this sample (Annika *et al.*, 2006).

At present even with the developed techniques, one of the main limiting factors for multivariate analysis operation is the lack of availability of data. Therefore the level at which forest stands can be analysed is often dictated by the data available. If, for instance, individual trees are not numbered and identified, individual tree-based approaches are not possible. Without sound data, especially measured from permanent sample plots, it is very difficult to develop “perfect” analysis. Further to that, it is essential to collect the data covering all geographical regions and site types to run such analysis. It is very important to define the data characteristics. Often the result of the



analysis are good, as are the decision tools, but the decisions may still not be as good as they should be due to lack of specific quality in the data.

Principal Component Analysis (PCA) inference is especially useful in curbing the researcher's natural tendency to read too much into the data. Total control is provided for experiment wise error rate; that is, no matter how many variables are tested simultaneously, the value of α (the significance level) remains at the level set by the researcher (Rencher, 2001).

The research work is aimed at developing PCA using multivariate method for analysing inventory field data and their applications for sustainable forest management.

METHODOLOGY

The Study Area

This study was conducted in Federal College of Forestry, Ibadan, Oyo State. The area lies between latitude $7^{\circ}26'N$ and longitudinal $3^{\circ}15'E$ the climate pattern of the area is tropically dominated by annual rainfall ranges from 1,300-1,500mm and average relative humidity of about 65% the average temperature is about $26^{\circ}C$. The eco-climatic of the dry season usually commence from November to March while the raining season is usually between April-October (FRIN 2014).

SAMPLING PROCEDURE

The experiment was carried out in two stands of Federal College of Forestry plantation and a complete inventory was carried out on the stands.

Highlights of Tree Parameters Measured

Measurements were carried out on all the tree stands in the study area. Trees in each stand were measured to obtain some variable such as diameter at breast height (DBH) and total height. The tree variables measured are:

a. Total height: - This is the vertical distance between the ground level and the tip of a tree. It is obtained by taking the reading at the top (RT) and reading at the base (RB) which is usually negative (when on an elevated ground) and positive (when in a depressed ground or valley). It was measured using the Spiegel Relaskop. The formula used to obtain the total height (THT) using metric scale type is shown below,

$$H = RT - RB$$

Where H = Height

RT = Reading at the top

RB = Reading at the base



b. Diameter at Breast height (DBH): This is the diameter measurement taking for a standing tree at height 1.3m above the ground level.

STATISTICAL ANALYSES

PCA and descriptive statistics were used to analyse data with the use SPSS software.

RESULTS AND DISCUSSIONS

The research work was carried out on two stands of *Gmelina arborea* and *Tectona grandis* plantation. The parameters assessed were height and the diameter at breast height (DBH).

Table 1: Descriptive statistics of *Gmelina arborea* and *Tectona grandis*

	N	Min.	Max.	Mean	Std. Error	Std. Deviation
HT (<i>T.grandis</i>)	205	8	36	22.03	.354	5.066
DBH (<i>T.grandis</i>)	205	13	26	17.16	.198	2.840
DBH(<i>G. arborea</i>)	108	2	3	1.35	.045	.469
HT (<i>G. arborea</i>)	108	11	24	17.05	.248	2.580

Table 1 shows that *T. grandis* had standard error value for height of 0.354 and DBH of 0.198 while *G. arborea* had a standard error value for height and DBH of 0.248 and 0.045, the higher the standard error of the stands, the higher the susceptibility to error but the lower the standard error, the closer to perfection. This implies that the DBH of *G. arborea* has a low standard error of 0.045 which means there is lesser susceptibility to error.

Table 2: Eigenvalues of the Correlation Matrix of Principle Component Analysis of Stands

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
<i>T. grandis</i>	1.092	54.600	54.600
<i>G. arborea</i>	0.908	45.400	100.000

Table 2 shows eigenvalues of the correlation matrix of Principal Component Analysis. It also shows that *Tectona grandis* had cumulative percentage of 54.60 with same variance while *Gmelina arborea* had cumulative percentage of 100.00 with variance of 45.40 which means that *Tectona grandis* had the highest percentage variance.



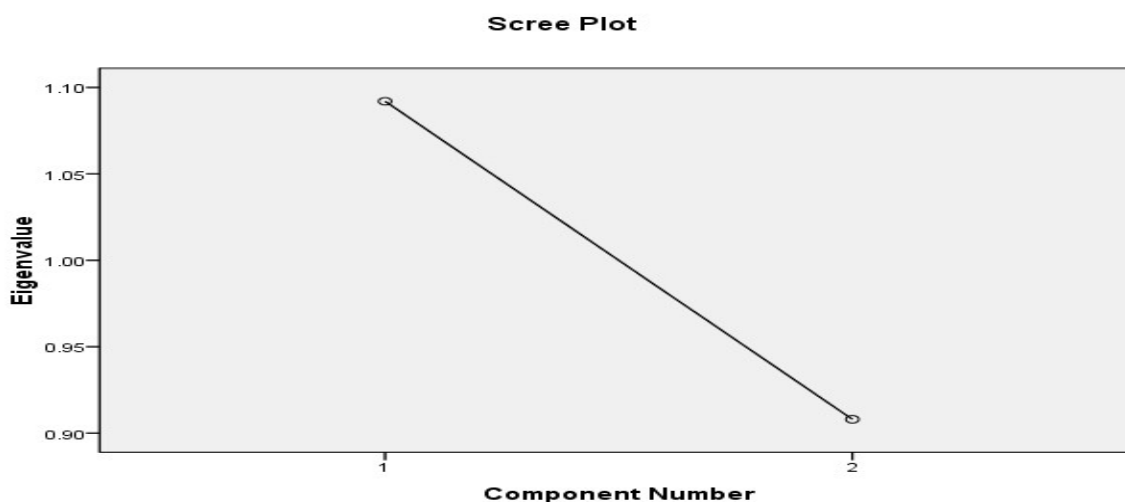


Figure 1: Scree plot for the component number analysis

Table 3: Pearson correlation of *Tectona grandis* and *Gmelina arborea*

	(H) <i>T. grandis</i>	(DBH) <i>T. grandis</i>	(DBH) <i>G. arborea</i>	(H) <i>G. arborea</i>
(H) <i>T. grandis</i> .	1			
(DBH) <i>T. grandis</i> .	.099	1		
(DBH) <i>G. arborea</i>	-.109	-.190*	1	
(H) <i>G. arborea</i>	.027	-.113	.273**	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed)

Table 3 shows the correlation between the two stands and it is therefore stated that the height of *Tectona grandis* and *Gmelina arborea* had a positive low correlation value of 0.027, this indicates that there is a weak relationship between the two stands. While DBH of *Tectona grandis* and *Gmelina arborea* has a negative low correlation of -0.190, this also indicate that there is a weak relationship between the two stands as a result of variation in the number of tree species population.

CONCLUSION

Principal component analysis techniques are now accessible using computer, it reduces the data collected from the field when analysing and only the most dedicated number cruncher would consider doing real-life-sized problems in multivariate statistics without a computer. Fortunately, excellent multivariate programs are available in a number of computer packages.



All rights reserved

This work by [Wilolud Journals](http://www.wiloludjournals.com) is licensed under a [Creative Commons Attribution 3.0 Unported License](http://creativecommons.org/licenses/by/3.0/)

REFERENCES

- Annika, K. and Matti, M. (2006). *Forest Inventory; methodology and applications*. Springer Netherland. 362p
- FRIN (2014): Forestry Research Institute of Nigeria (FRIN) Metrological Report 2014.
- Jobson, J. D. (1992), *Applied Multivariate Data Analysis, Volume II: Cateogorical and Multivariate Methods*, New York: Springer-Verlag.
- Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K. and Wagner, F. (Eds.). 2003. *Good practice guidance for land use, land-use change and forestry*. Intergovernmental Panel on Climate Change: IPCC National Greenhouse Gas Inventories Programme. Kanagawa Japan: Institute for Global Environmental Strategies (IGES) for the IPCC.
http://www.ipccnggip.iges.or.jp/public/gpglulucf/gpglulucf_contents.htm
- Rencher, A.C. (2001). *Methods of multivariate analysis*. 2nd Ed. John Wiley and Sons Inc. Canada.

