African Research Review
An International Multi-Disciplinary Journal, Ethiopia
Vol. 4 (3a) July, 2010
ISSN 1994-9057 (Print) ISSN 2070-0083 (Online)

Quantitative Analysis of Complex Tropical Forest Stands: A Review *(Pp. 367-377)*

Oyebade, B. A. **-** Forest Biometrics & Measurement Unit, Department of Forestry & Wildlife Management, University of Port Harcourt, Nigeria
E-mail:bukkibadef@yahoo.com
TEL: +2348056239312; +2347034464789

Ekeke, B. A. **-** Department of Forestry & Environment, RSUST, Port Harcourt, Nigeria.

Abstract

The importance of data analysis in quantitative assessment of natural resources remains significant in the sustainable management of complex tropical forest resources. Analyses of data from complex tropical forest stands have not been easy or clear due to improper data management. It is pivotal to practical researches and discovery that promote development in forestry and many related disciplines. Many quantitative methods and approaches are strongly dependent on the source, nature, and quality of the data. However, many issues related to data analysis in the tropical complex forests are inimical and may render quantitative methods impossible if not resolved. Data collection in many complex tropical forests is very difficult and oftentimes results in data violating simple assumptions of statistical models. The use of relevant data transformation proffers significant solution to this perennial challenge within the complex tropical forests. This paper therefore reviews statistical issues related to quantitative data collection and analyses in the complex tropical forests and provides pragmatic approaches *for solving data analysis challenges in complex tropical forests' management and planning.*

Keywords: data issues, analysis, complex stands, forestry

Introduction

Statistical methods are used in a wide variety of occupations and help people identify, study, and solve many complex problems. In forestry, these methods enable decision makers and managers to make informed and better decisions about uncertain situations.

Vast amounts of statistical information are available in today's global and economic environment because of continual improvements in computer technology. To provide reliable regional guidance, forest managers and decision makers must be able to understand the information and use it effectively. Statistical data analysis provides hands on experience to promote the use of statistical thinking and techniques to apply in order to make educated decisions in forest management.

Akindele (2008) showed that forestry research is based on scientific method which is popularly known to be the inductive-deductive approach. Consequently, foresters across the globe have become increasingly quantitative in their approaches to research and management; and with rising forest values; there is concomitant increase in the demand for accuracy and precision in management prescriptions and projected outcomes. This invariably depends on data factors and issues with the ability to resolve these associated problems using statistical principles.

Tropical rainforests are notably the complex forests of the world, and they are characterized by profuse growth and regrowth of plant and tree species that gradually occur throughout the year. The tree species are highly diverse but usually have smooth, straight trunks and large, simple leaves. Large vines are common, but the tangled growth of a jungle occurs only where the normal forest area has been abused.

The tropical environment is rich in terms of bio-diversity. The tropical African forest is 18 per cent of the world total and covers over 3.6 million square kilometers of land in West, East and Central Africa. This total area can be subdivided to 2.69 million square kilometers (74%) in Central Africa, 680,000 square kilometers (19%) in West Africa, and 250,000 square kilometers (7%) in East Africa. In West Africa, a chain of rain forests up to

350 km long extends from the eastern border of Sierra Leone all the way to Ghana. The rain forest of West Africa continues from east of Benin through southern Nigeria and officially ends at the border of Cameroon along the Sanaga River (Figure 1).

The rain forests as the richest, oldest, most prolific, and most complex systems on earth, are dying, and in turn are upsetting the delicate ecological balance. A vivid representation of tropical rain forests of the world is given in the figure 1 below

Figure 1: Tropical Rain Forest Distribution

Tropical rain forests are located between the tropics of Cancer and Capricorn i.e. 23°27' north to 23°27' south of the Equator, with the mean annual temperature of about 27° C (equivalent 80° F). Although they cover less than 10 percent of the Earth's surface, tropical rain forests provide habitat for 50 to 90 percent of the world's plant and animal species. Although tropical rain forests now cover just 2 percent of the globe, they are home to more than half the world's living plant and animal species. Rain forests are the predominant natural vegetation throughout the wet tropics. The defining characteristics of tropical rain forest are temperature and rainfall. Wherever temperature is high enough and rainfall heavy and regular enough, there is rain forest.

Data analysis of the complex forest ecosystem therefore involves the process of gathering, modeling and transforming the data with the goal of highlighting useful information, suggesting conclusions and supporting decision making. Data analysis in complex tropical forests has multiple facets and approaches, encompassing diverse techniques under a variety of names, and with different methods and application domains for quantitative management of data collected from such ecosystems. Generally, data analysis of the complex forest ecosystem can be statistically divided into descriptive statistics, exploratory data analysis (EDA) and confirmatory data analysis (CDA). The EDA of complex tropical forests data analysis focuses on discovering new features in the data while the CDA centers on confirming or falsifying existing hypotheses about the data collected from complex forest environments.

This paper discusses possible issues associated with quantitative data of complex forest and simple statistical techniques of resolving them before being managed for logical uses and reasonable application.

Data collection, analysis and reporting on complex stand

Data collection design is an important process in complex forest statistical data analysis. Adoptions of the established statistical approaches for data collection in complex populations appear to be the only meaningful data collection methods for complex tropical forest ecosystems. Statistical inference from random samples of the ecosystem should give reliable information about the forest with limited errors.

Fundamental Assumptions of Parametric Models and data

Lack of Normality

One of the basic assumptions for the Analysis of Variance is that data is normally distributed, with mean equal to zero and variance equal to one. Unfortunately, ecological data from tropical forests rarely follow a normal distribution and non-normality appears to be the norm as in biological sciences (Potvin and Roff 1993, White and Bennetts 1996, Hayek and Buzas 1997, Zar 1999). Besides, the normal distribution primarily describes continuous variables whereas in the tropical forests count data are discrete (Krebs 1989). Although parametric statistics are fairly robust to violations of normality, highly skewed distributions can significantly affect the results. Ideally, the sample size should be equal among groups and sufficiently large (e.g., $n > 20$). The significance of non-normality can be tested with several techniques that ensure precision. Graphical examinations of the data are also relevant and appropriate in detecting the extent as well as the type of the problem. Data that are reasonably symmetric about the mean and that do not have a large number of observations in the distribution tails are generally well enough approximated by a normal distribution for most standard analyses for which this is an assumption.

Indexed African Journals Online: www.ajol.info

This affects the test of significance of the parameters, as well as the estimation of confidence intervals for the parameter estimates. Under these circumstances, it is recommended to use data transformations based on the analysis of the residuals.

Independence of Observations

An indispensable condition of a good number of statistical tests is the independence of observations in space and time (usually obtained using random selection). Observations can be counts of individuals or replicates of treatment units in manipulative studies. Krebs (1989) argues that if the assumption of independence is violated, the chosen probability for Type I error (α) cannot be achieved. Analysis of variance (ANOVA) and linear regression techniques are sensitive and relevant to this violation (Sabin and Stafford 1990, Sokal and Rohlf 1995). Mixed-model analysis procedures, which are now obtainable in a number of statistical software packages permit for some relaxation of the assumption of independence. This is pertinent in that the tropical forest ecosystems may not provide data that may not violate the independence assumption even in a single.

Homogeneity of Variances

Another assumption in the Analysis of Variance is that all observations have a common variance. Parametric models frequently assume that sampled populations have similar variances even if their means are different. This assumption becomes critical in studies comparing different groups of species, treatments, or sampling intervals. If the sample sizes are equal, then parametric tests are fairly robust to the departure from homoscedasticity (i.e., equal variance of errors across the data) (Day and Quinn 1989, Sokal and Rohlf 1995). Indeed, equal sample sizes across treatments should be obtained whenever possible since most tests are overly sensitive to violations of assumptions in situations with unequal sample sizes (Day and Quinn 1989). Plotting the residuals of the analysis against predicted values can reveal the nature and severity of the potential problem. Visual inspection of the data can help determine if transformation of the data is needed and can also indicate the type of distribution. Although several formal tests exist to determine the heterogeneity of variances (e.g., Bartlett's test, Levine's test), these techniques assume normal data distribution, which reduces their utility in most ecological and complex forest studies (Sokal and Rohlf 1995). This is consequently of limited application in the analysis of complex tropical forest ecosystem.

Correlated Errors

Correlated error problem is mostly encountered in dynamic surveys where data are repeatedly collected in the same survey units. Under these circumstances, variables tend to present correlation between their residual values, which in turn diminish the precision of estimates, and could even invalidate any test of significance of the estimated parameters. This problem could be overcome by using Time Series Analysis and Generalised Least Squares (GLS) procedures.

Possible Remedies for violated Parametric Assumptions

Indeed, parametric statistical models are based on a set of assumptions that are essential for models to appropriately fit and describe the data. It is observed that if assumptions are violated, statistical analyses may produce erroneous and extraneous results (Krebs 1989, Sabin and Stafford 1990, Sokal and Rohlf 1995). Thus, it is pertinent for complex forest data analyst to consider whether data will likely fit the assumptions of a selected model. When data contravene the laid down assumptions or when the assumptions are violated, data transformation is suggested for dealing with the problems of such violations.

Data Transformation

If significant violations of parametric assumptions occur, quantitative foresters are advised to implement an appropriate data transformation to resolve the violations. During a transformation, data will be converted and analyzed at a different scale than the original data. Transformations effectively reweigh the data and can result in detecting statistical differences when none could be detected otherwise, so it is important to consider the effects of transforming dependent variables on the eventual output. For tropical forests data analysis, there is the need to back-transform the results after analysis to present parameter values on the original data scale. Table 1 gives examples of common types of transformations that may be recommended for tropical forest analyses. A wisely chosen transformation can often improve homogeneity of variances as well as produce an approximation of a normal distribution. Sabin and Stafford (1990) and Zar (1999) give good overviews of data transformations in ANOVA and regression models that may be applicable to tropical forests.

A primary reason to avoid transformations is that interpreting transformed variables is very difficult (e.g., what is the arcsine square root of a proportion?). As a result, it is recommended that the data be back-transformed after analysis but back-transformations are not always necessarily at the same scale as the original data. Therefore it is important to know the assumptions of the particular statistical model and how transforming will affect their data set. Removing outliers or perhaps using a nonparametric technique may be a better approach than trying to normalize the distribution of data and homogenize variances to meet the assumptions of a parametric model.

Nonparametric Alternatives

When the data collected from the complex forest study violates basic parametric assumptions and transformations fail to remedy the problem, a nonparametric method might be appropriate (Sokal and Rohlf 1995, Conover 1999). Nonparametric techniques have less strict assumptions about the data, are less sensitive to the presence of outliers, and are often more intuitive and easier to compute (Sokal and Rohlf 1995, Hollander and Wolfe 1999). Since nonparametric models are less powerful than their parametric counterparts Day and Quinn (1989); Johnson (1995) and Smith (1995) advocate the use of nonparametric tests if the data meet or appreciate parametric assumptions.

Randomization Tests

These tests are not alternatives to parametric tests, but rather are a means of estimating the statistical significance that relies only on the independence of observations. They are extremely versatile and can be used to estimate significance of test statistics for a wide range of models. Edgington (1995) showed that although randomization tests are computationally difficult even with small sample sizes, there are numerous software packages that have been developed for randomization tests.

Generalised Least Squares (GLS) Apporach

Other parametric techniques such as generalized linear models employ a distribution appropriate for the data instead of trying to normalize them. The works of White and Bennetts (1996) give an example of fitting the negative binomial distribution to point count data for orange-crowned warblers (Vermivora celata) when comparing their relative abundance among forest sites. Zero-inflated Poisson (ZIP) models and negative binomial regression models are recommended for analysis of count data with frequent 0 values (e.g., rare species studies) in which data transformations are not feasible (Heilbron 1994, Welsh et al. 1996, Ridout et al. 1998, Agarwal et al. 2002, Hall and Berenhaut 2002).

Conclusion

In tropical forest analysis traditional knowledge should be extended to meet the current and future challenges in data management for complex tropical forest data. Although, many forestry data collected in this ecosystem is peculiarly borne with attributes that negates the statistical assumptions, it is however statistically possible also to condition these data to normality using some of the methods articulated in the paper. This will without doubt improve the rigor inherent in tropical forest data analysis as well as ascertaining elements of precision and correctness in data analysis and management of the complex tropical forest data.

Table 1: Some common data transformations relevant to complex forest studies

Transformation type	When appropriate to consider using	Transformation	Back transformation
Square root Logarithmic Inverse	Use with count data following a Poisson distribution; more generally, when variances are proportional to means. In some instances, addition of 3/8 will improve normality. Use with count data when means are proportional to standard deviations. A rule of thumb suggests its use when the largest value of the dependent variable is at least 10 times the smallest value. Use when data residuals exhibit a severe funnel shaped pattern of distribution curve, which is often the case in data sets with many near-zero values.	$y' = \sqrt{y}$ $y' = \sqrt{y + 3/8}$ $y' = \log_e(y+c)$ where c = 0 if all y > 1 and c = 1 otherwise $y' = 1/y$ Note: Inverse transformations will cause very large values to be very small and very small values to be very large. Thus, one must reverse the distribution before transforming by multiplying a variable by -1, and then adding a constant to the	$y = y'2$ $y = \exp(y')-c$ $y = 1/y'$ $y = (\sin y')^2$

		distribution to bring the minimum value above 1.0. Once the inverse transformation is complete, the ordering of values will be identical to the original data.	$y=\begin{cases}(\lambda y'+1)^{1/\lambda} & \text{if } \lambda\neq 0\\ \exp(y') & \text{if } \lambda=0\end{cases}$
Arcsine square root	Appropriate for proportional or binomial data. This transformation is beneficial if it improves normality for nonbinomial proportions. Most efficient when most proportions occur at ends of the scale (0.0–0.25 and/or 0.75–1.0), and least effective when proportions are distributed in the middle (0.25–0.75).	y′ = arcsin(square root[y]), where y is a proportion.	
Box-Cox objective approach	If it is difficult to decide on what transformation to use, this procedure finds an optimal model for the data. Box-Cox approaches may address skewed residual distributions and heterogeneous variance.	$y'=\begin{cases}(y^{\lambda}-1)/\lambda & \text{if } \lambda\neq 0\\ \log_e y & \text{if } \lambda=0\end{cases}$ where, λ is an estimated parameter	

Source: adapted from Sabin and Stafford (1990).

References:

Akindele, S.O. (2008). The place of biometrics in Forestry Research. In Research for Development in Forestry, Forest Products and Natural Resources Management (Eds). Onyekwelu, J.C., Adekunle, V.A.J. and Oke, D.O.) Proceedings of the 1st National Conference of the Forest and Forest Products Society, Federal University of Technology, Akure. Nigeria. 16th - 18th April, 2008. Pp 253-256.

Agarwal, D.; Gelfand, A.E. and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics*. 9: 341-355.

Conover, W. J. (1999). *Practical nonparametric statistics*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons. 592 p.

Day, R.W. and Quinn, G.P. (1989). Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs.* 59: 433-463.

Edgington, E.S. (1995). *Randomization tests*. 3rd ed. New York: Marcel Dekker. 409 p.

Hall, D.B. and Berenhaut, K.S. (2002). Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. *Canadian Journal of Statistics*. 30: 415-430.

Hargis, C.D. and Woodbridge, B. (2006). A design for monitoring northern goshawks (Accipiter genitilis) at the bioregional scale. *Studies in Avian Biology. 31: 275-288.*

Hayek, L.C. and Buzas, M.A. 1997. *Surveying natural populations*. New York: Columbia University Press. 563 p.

Heilbron, D.C. 1994. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*. 36: 531-547.

Hollander, M. and Wolfe, D.A. 1999. *Nonparametric statistical methods*. 2nd ed. New York: John Wiley & Sons. 816 p.

Krebs, C.J. 1989. *Ecological methodology*. New York: Harper Collins. 654 p.

Potvin, C. and Roff, D.A. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics. *Ecology*. 74: 1617-1628.

Ridout, M.; Demétrio, C.G.B. and Hinde, J. 1998. Models for count data with many zeros. Proceedings of the International Biometric Conference. 19: 1-13.

Sabin, T.E. and Stafford, S.G. 1990. *Assessing the need for transformation of response variables. Corvallis, OR*: Oregon State University Forest Research Laboratory Special Publication 20. 31 p.

Sokal, R.R. and Rohlf, F.J. 1995. Biometry: the principles and practice of statistics in biological research. 3rd ed. New York: W. H. Freeman. 887 p.

Southwood, T.R.E. 1992. *Ecological methods: With particular reference to the study of insect populations*. 2nd ed. London: Chapman and Hall. 524 p.

Smith, S.M. 1995. Distribution-free and robust statistical methods: Viable alternatives to parametric statistics? *Ecology*. 76: 1997-1998.

Welsh, A.H.; Cunningham, R.B.; Donnelly, C.F. and Lindenmayer, D.B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling. 88: 297-308.

White, G.C. and Bennetts, R.E. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology*. 77: 2549-2557.

Zar, J.H. 1999. *Biostatistical analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.660 p.