ENVIRONMENTAL ANALAR

A Journal of Environmental Analysis

NO. 4, Dec. 1999

ISSN 1595 - 0174

ENVIRONMENTAL SYSTEMS CLUB

ESC

ENVIRONMENTAL SYSTEMS CLUB INC.

Env. Anal. 4 (1999) pp 268-283

ENVIRONMENTAL ANALAR

PRINCIPAL COMPONENTS ANALYSIS OF DATA FROM THE STUBBS CREEK FOREST RESERVE 1: THE VEGETATION SUB-SYSTEM

I. E. Ukpong and J. N. Essien

Department of Geography and Regional Planning
University of Uyo, Akwa Ibom State, Nigeria.

Accepted 26th Oct., 1999

Abstract

Thirty 10 x 15m vegetation quadrats in the Stubbs Creek Forest Reserve were sampled by the point-quarter method for the A and B strata and by transect survey for the C stratum. The data was ordinated using principal components analysis. The aim of the analysis was to classify the vegetation samples according to their structural similarities, with a view to discerning the influence of environmental gradients on the vegetation. While seven clusters were apparent in the A stratum, the B and C strata could only be classified into four groups each, showing that in theoretical hyperspace, the A stratum was the least homogeneous among the strata. The implications of these observations to the management of the Stubbs Creeks Forest Reserve are attempted.

1.0 Introduction

The aim of this paper is to analyse the gradients in a forest vegetation system using principal components analysis (PCA) as a major tool. Consequently, vegetation data from the Stubbs Creek Forest Reserve in Akwa Ibom State, Nigeria, is used to evaluate the effectiveness of PCA as a multivariate technique. The basic premise is that the vegetation is a multivariate rather than a collection of independently varying attributes.

Principal components analysis an eigenvalue technique, may be used in the interpretation of the structure within the correlation (or variance-covariance) matrix of a multivariate data. The principal factors that are extracted from the data and analysed are eigenvectors of a real matrix which has been factored into its characteristic roots and corresponding characteristic vectors (Barkham and Norris, 1970, Cooley and Lhones, 1971). The factor pattern of the PCA is therefore the same as its factor structure. Hence the factor pattern is an orthogonal factor solution i.e. the factor variables from the PCA may be independent in both directions in any attribute/factor matrix. The pitfall with PCA arises mainly due to orthogonality although with suitable standardization and adjustment of data, the method has been found to be appropriate in ecological research in that it requires a minimum of assumptions and is computationally unambiguous (Aweto 1978, Swaine and Greig-Smoth 1980, Ter Braak and Prentice 1988, Ukpong 1994).

In the present investigation, PCA was used in two main contexts: First, to derive stand (vegetation) loadings on the principal axes of variation. Second, to derive species (taxon) loadings on the principal axes of variation (Essien 1998). In the species analysis, the coverage values of 23 species of the Stubbs Creek Forest Reserve, which had frequencies of occurrence 5% and above were used. In the stand analysis, attention was centred on the total attribute value for all species in each stand to group the sample into abstract vegetation communities. Hobbs and Grace (1981) have observed that the solutions of species and stands may be identical but the information referring to species is usually more interesting since stand are often unknown to the other reader.

The R-type matrix (Goodall 1954) was used for some analyses. The R-technique is based on a components analysis of a correlation matrix which standardizes the variables to zero mean and unit variance. This differs from the Q-technique (Orloci 1966) that uses a variance-coveriance matrix. The use of the R-technique eliminated a situation in which the attribute that had the largest variance necessarily determined the principal axes (Austin 1968).

Although there was no intention of using an alternative tool to PCA in this study, it was thought necessary to find out if another variant of factor analysis could be ecologically more meaningful. For a test Ivimey-Cook and Proctor (1967) had recommended rotation of the component to the simple structure while Ferrari and Mol (1967) recommended rotation according to the caumax model. In the present study, Kaiser's (1958) varimax rotation model, being the most popular was used to maximally group the factor variables into distinct clusters. Hence all ordination diagrams presented in this study have been on the rotated components of a PCA.

The loadings on the more important principal components that account for the interpreted trends have been included in the text. The loadings themselves, being adjusted, are the coefficients of correlation between attributes and the principal components. The proportion of attributes' variances have been indicated also. The mathematical details of PCA can be found elsewhere e.g Seal (1964), Cooley and Lhones (1971), Ter Braak and Prentice (1988).

2.0 Study Area

The Stubbs Creek Forest Reserve is geographically located in the coastal zone of Akwa Ibom State (Fig. 1), at longitude 8°16'E and latitude 4°37'N. The reserve covers an area of 310.8km^2 between the Kwa Iboe River and Cross River estuary. Lying below 30m above sea level, the geology relates to recent deposition of sand, clay and mud. The rocks are made up of sandstones and shales. Stubbs Creek flow west from the centre of the reserve into the Kwa Iboe River while Widenham Creek flows east into the Cross River (Fig. 1). The two creeks have a common headwater which permits navigation across the reserve at high tides. Douglas Creek flows from the centre of the

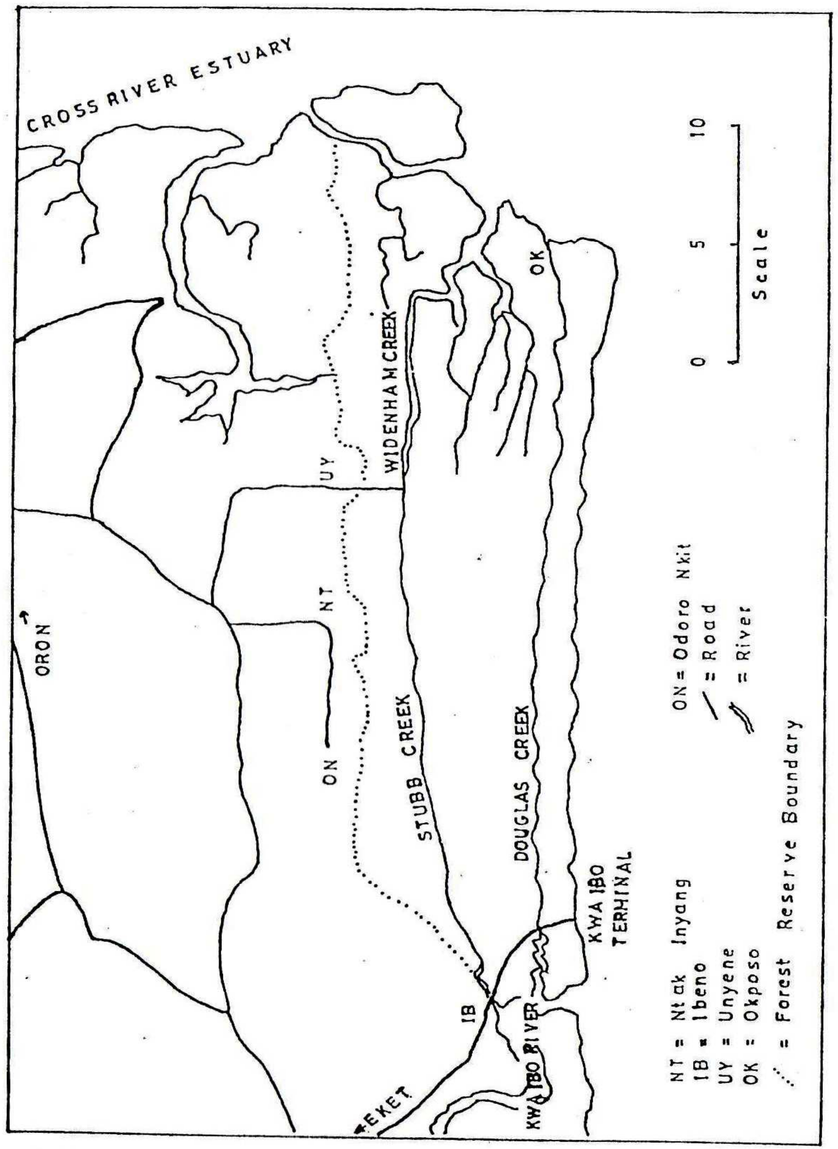

Figure 1. Map of the Stubbs Creek Forest Reserve, site of the study.

reserve into the Kwa Iboe River south of Stubbs Creek. The forest reserve, being so linked, constitute a homogeneous ecological unit.

The Stubbs Creek forest lies over alluvium and mangrove soils (Group 1992). North of the mangrove alluvium occurs the coastal plain sands. The climatic regime fluctuates between wet and relatively dry. Dry season occurs from mid October to April while wet season occurs from May to October, with 125-180 rainy season days. The highest temperatures (30°C) occur between March, and April while the lowest (26°C) occurs in August. Average relative humidity is about 81.5% while mean annual rainfall is 3418.5mm.

The forest reserve is surrounded by relics of natural lowland rainforest vegetation and farmland fallows of about 0-5 years of age. On the swampy coasts and creeks salt and freshwater mangroves occur. Where soil water is salt-free, acid tolerant palms and groves of tropical hardwood are formed.

3.0 Methods

Considering the aim of study, the first sampling option was to adopt a plotless technique across the reserve but this was hampered by inaccessibility of certain portions of the swamp. Consequently 0.5km^2 grids were overlain on the reserve based on air-photo interpretation. Thirty accessible grids were selected out of which ten were located within the saline/brackish swamp zone. From each grid square, one quadrat of 150m^2 carrying relatively undisturbed vegetation was sampled by the point-quarter method. The vegetation was stratified into A, B and C strata. Plant species were identified and their abundance measured, including coverage, density and basal area (Kershaw 1980). Tree height was measured with the foresters' Hagar. For the C stratum (groundlayer), the point-quarter method was supplemented by transect survey in order to facilitate estimates for creepers and ferns in the undergrowth.

The vegetational variables were transformed in order to meet the requirements for parametric statistics (Gregory 1973). The skewness of each variable was determined for the raw data after standard deviation checks were used to assess the scatter. Crown cover values for the A and B strata and tree height being slightly negatively skewed approximated to normal distribution after square-

root transformations. Raw density values showed normal distribution.

4.0 Results

4.1 General

The thirty vegetation quadrats sampled could be classified into lowland and tidal according to soil moisture relations. Ten quadrats occurred in areas affected by diurnal tides across the Stubbs Creek. Connected to the Kwa Iboe River estuary, Atlantic tides inundate parts of the forest reserve with saline/brackish water which enable the growth of mangroves and associates e.g. *Rhizophora* spp., *Avicinnia* spp, *Nypa, Pandanus, Phoenix* and *Acrostiohum* species. Beyond the brackish/freshwater interface, lowland species e.g *Selaginella spp., Costus afer, Vossia cuspidata, Anthonota* spp, *Marantochloa* spp, *Arachis* spp and *Fimbristylis* spp. are common occurrences among the remaining twenty quadrats.

There was low variability (CV 11.9-18.0%) in tree density across all strata in the lowland areas (Table 1). But as expected, stem density increased from 8/150m^2 in the A stratum to 24.150m^2 in the C stratum. The canopy layer was high and relatively open which permitted a thick understorey and groundlayer vegetation to be formed. Barkham and Norris (1970) had regarded overstorey canopy cover as an environmental factor that correlates with groundlayer performance. In the Stubbs Creek, despite a large cover (75.1%), the relatively high structure (25.0m) made the A stratum ineffective in retarding C stratum performance. The wood volume was high in the A stratum (19.6m^2/150m^2) clearly dominating over the B and C strata as is characteristic of tropical rainforests (Table 1).

The vegetation in the tidal zone (Table 2) was similar in characteristics to the lowland subsystem. Density and coverage were comparable particularly in the A and C strata. Shorter overstorey species (17.2m) were restricted to the tidal areas. The tidal species had lower wood volume (0.2m^2 - 13.1m^2) indicating lower productivity than the lowland species.

Table 1: Summary of Vegetation Characteristics in Twenty Lowland Quadrats

Characteristic	A Stratum	B Stratum	C Stratum
Density (s/150m²)	8.1±1.4(17.2)	13.3±2.4(18.0)	24.2±2.9(11.9)
Height(m)	25.0±2.5(10.0)	6.1±3.3(54.1)	3.2±0.3 (9.4)
Basal area (m²/150m²)	19.6±3.9(19.8)	2.5±0.5(20.0)	0.02±0.01(50.0)
Cover (%)	75.1±7.3 (9.7)	13.4±4.8(35.8)	4.7±1.4(29.8)

( ) = Coefficient of variation.

Table 2: Summary of Vegetation Characteristics in Ten Tidal Quadrats

Characteristic	A Stratum	B Stratum	C Stratum
Density (s/150m²)	7.6±1.4(18.4)	13.9±3.9(28.1)	22.9±2.1(9.2)
Height(m)	17.2±2.1(12.2)	7.1±1.7(23.9)	3.4±0.5 (14.7)
Basal area (m²/150m²)	13.1±5.8(44.3)	4.6±3.6(78.2)	0.02±0.01(50.0)
Cover (%))	76.9±9.8 (12.7)	9.1±3.7(40.7)	3.9±0.5(12.8)

( ) = Coefficient of variation.

Stand Ordination

Using presence-absence data of species in the sampled quadrats, the thirty stands were ordinated using principal components analysis. Presence-absence data was preferred because of high biodiversity of taxon particularly in the lowland areas. This enabled more species to influence the analysis than if only the identified species were used (Aweto 1978). Therefore a large matrix of 30 x 26 was achieved for the A stratum, 30 x 38 for the B stratum and 30 x 67 for the C stratum. The analysis utilized the R-technique based on a correlation matrix of the quadrat data. The eigenvalues extracted are exemplified by the A stratum and are presented in Table 3a. Five components are presented but much of the variations in the data were explained by the first three components. The eigenvalues for the B and C strata, being similar to Table 3a, need not be elaborated further.

Table 3(a): Eigenvalues of the Principal Components of Stand Ordination

Vegetation Components (VC)	Eigenvalue	Percentage variation	Cumulative percentage
VC 1	17.56	40.4	40.4
VC 2	8.32	19.2	59.6
VC 3	5.94	13.7	73.3
VC 4	4.02	9.3	82.6
VC 5	3.11	7.2	89.8

Table 3(b): Vegetation Characteristics of the Abstract Community Types, from the PCA of Quadrat Data (Mean Values are Reported)

Community Types	A Stratum		B Stratum		C Stratum	
	Density (S/150m²)	Cover (%)	Density (S/150m²)	Cover (%)	Density (S/150m²)	Cover (%).
1	7.4	35.2	6.8	5.5	20.4	5.2
2	6.3	58.6	10.5	5.8	18.7	4.4
3	9.8	77.3	11.8	12.5	28.0	4.8
4	11.9	70.1	14.3	12.8	15.2	3.1
5	8.1	60.8	8.4	7.4	12.0	2.1
6	7.2	65.5	9.2	8.8	9.4	2.0
7	5.0	40.0	7.0	6.5	8.0	1.3

Table 4(a): Principal Components of Species Ordination

Species Components	Eigenvalue	Percentage variation	Cumulative percentage
SPC 1	24.58	48.3	68.5
SPC 2	11.92	20.2	87.1
SPC 3	8.30	18.6	96.0
SPC 4	7.64	8.9	98.0
SPC 5	3.33	2.0	98.0

Table 4(b): Loadings from the PCA (Q-analysis) of Species Data on Components 1 To 4.

Species	Varimax rated Component scores (SPC)*				Communality (h²)
	SPC 1	SPC 2	SPC 3	SPC 4	
1. *Costus afer*	-0.19	0.76	-0.36	0.11	0.76
2. *Acrostichum aureaum*	0.22	0.92	0.04	0.10	0.91
3. *Fiscus exasperata*	0.95	0.10	0.14	0.11	0.94
4. *Avicennia africana*	0.88	0.10	0.20	0.13	0.84
5. *Aspilia africana*	0.38	0.72	0.28	0.11	0.75
6. *Rhizophora racemosa*	0.95	0.01	0.02	0.12	0.92
7. *Solomum toruum*	0.45	0.36	0.60	0.13	0.71
8. Pandamus candelabra	0.83	0.50	0.12	0.11	0.96
9. Pteris burtoni	0.56	0.50	0.53	0.04	0.60
10. Ipomoea cairica	0.26	0.23	0.15	0.19	0.19
11. Sida alba	0.60	0.49	0.26	0.19	0.70
12. Hibiscus tilaceus	0.51	0.17	0:57	0.25-	0.68
13. Spondias mimbin	0.34	0.25	0.43	0.11	0.38
14. Raphia hookeri	-0.08	0.58	-0.15	0.18	0.40
15. Emilia coccinea	-0.20	-0.21	-0.45	0.22	0.33
16. Philoxerus spp	0.12	0.31	-0.03	-0.45	0.31
17. Cyperus distins	-0.18	0.34	-0.01	0.33	0.26
18. Paspalum vaginatum	0.90	-0.07	-0.05	0.13	0.83
19. Finbristylis dichotoma	-0.13	0.06	-0.10	-0.32	0.13
20. Selaginella spp	0.23	-0.91	0.28	-0.13	0.98
21. Alchonea cordifolia	-0.18	-0.14	-0.33	0.21	0.21
22. Vossia cuspidata	0.82	0.18	-0.14	0.34	0.84
23. Anthonata nigerica	0.30	0.27	-0.14	0.17	0.21

*SPC = Species Components

The abstract vegetation communities resulting from the PCA classification are presented as ordination diagrams (Figs. 2-4). Figure 1 shows that seven vegetation communities were abstracted from the A stratum data while four could be abstracted from the B stratum data. Although four groupings were also apparent in the C stratum, other related sub-classifications could also be discerned (Fig. 4).

The A stratum appeared to be less homogeneous than the understorey and groundlayer. The reason could be that the A Stratum species have been selectively exploited relative to distance from

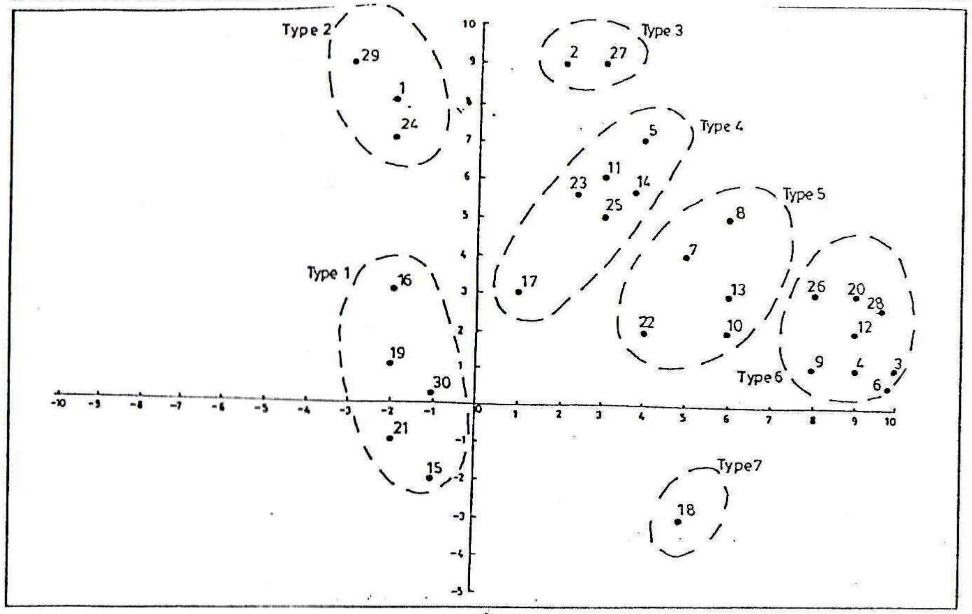

Figure 2. Ordination of 30 vegetation quadrats on factors 1 and 2 of the PCA.

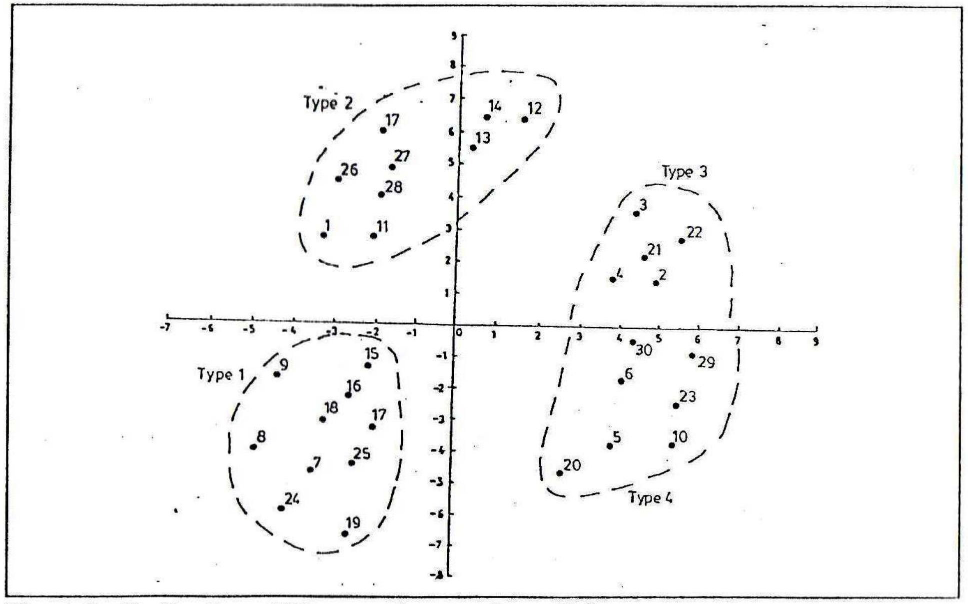

Figure 3. Ordination of 30 vegetation quadrats (B Stratum) on rotated components 1 and 2 of the PCA..

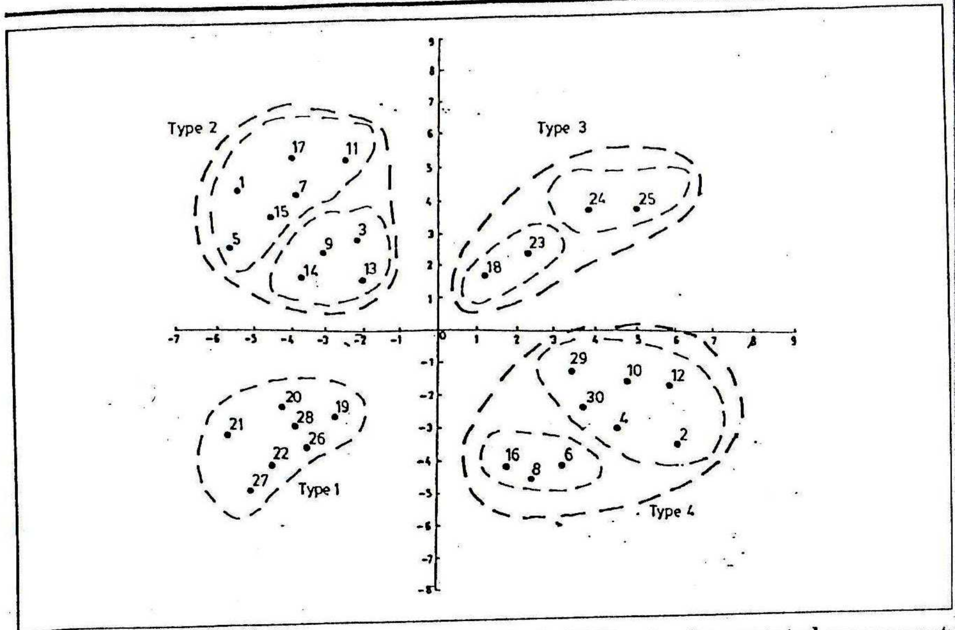

Figure 4. Ordination of 30 vegetation quadrats (C Stratum) on rotated components 1 and 2 of the PCA.

settlements and accessibility. Tree fall could not be considered a factor of heterogeneity although a few die-backs were observed in the tidal areas. Anthropogenic more than natural disturbance is reflected in the A stratum ordination which is probably why a larger number of communities are abstracted than in the B and C strata. In addition, the A stratum has fewer number of species with a high level of community similarity between the stands. This means that the axes in the correlation matrix from which the roots and vectors were factored were highly multicollinear. The effects of such matrix on the principal components and factor loadings were discussed by (Goodall 1954, Cattel 1965, and Ukpong 1994, 1995).

The greater homogeneity exhibited by the groundlayer vegetation is mainly due to the large number of lowland quadrats. The performance of ferns, sedges and creepers is enhanced by the relatively high and uneven canopy which enable sunlight to penetrate the understorey. This helps to produce favourable microclimate that

enhance competition at the groundlayer, resulting in increased biodiversity. The groundlayer species therefore tend to share same niche relations to changing microclimatic conditions irrespective of A stratum structure.

Table 3 shows some characteristics of vegetation in the abstract community types (based on A stratum ordination). With reference to Fig. 2, stem density tend to increase from Community Type 1 to type 4 after which there is a decrease to the last community type. This trend is apparent in all strata except C stratum coverage which has been postulated as relating to a more or less uniform microclimatic condition. Obviously a strong environmental gradient has produced a trend in vegetation variation from the negative end of the first ordination axis, through the second positive axis, to the first positive axis (Fig. 2). (Stand locations for the B and C strata appear to be uniformly hyperspaced between the two positive and negative axes Figs. 3 and 4).

Considering the disparity in soil moisture between tidal and lowland areas, the prime gradient could be that of soil water variation. Although this has not yet been investigated beyond the hypothetical level, it is nevertheless valid to assume that the low topography, porous substrate and high rainfall result in rapid response of the soil to water inputs (including tides and subsurface seepage). Consequently vegetation communities vary according to water table or soil moisture gradients, which reflects in the measured characteristics of the vegetation.

4.3 Species Ordination

Ordination of species could be more meaningful than ordination of stands because while species may be known to the other reader, stands are often quite strange and unknown to him (Ukpong and Areola 1995). However, the species used for ordination may be subjectively selected, depending on the interest of research. While the PCA is used to generate a hypothesis, it may also be used to establish the validity of a previously generated hypothesis. This means that the species which constitute the variables must be selected in such a way as to conform with the basic assumption of a previously identified gradient (Swaine and Greig-Smith 1980).

Twenty-three species, some with known characteristics and occurrence elsewhere (Hopkins 1968) were identified and included in a PCA if their frequency was 4% and above. The coverage values of these species, treated as one item (Wikum and Wali 1974) were preferred to frequency or presence absence data in view of the occurrence of rare species in only a few quadrats. The coverage values were converted to Braun-Blanquet (1932) cover-abundance scale and modified by including a coverage of 0-5% to the class (+) in this scale. The values were standardized by sample totals and PCA (Q-technique) with varimax rotation of the factor matric was used to establish patterns of the 23 species (see Ukpong 1994).

The eigenvalues extracted in the analysis is shown in Table 4a. Much of the variation is explained by the first four components which account for 96.0%. Attribute loadings on the components are presented in Table 4b, including the communality values (h^2). Attribute loadings that contribute to the major trends observed have been underscored.

SPC 1 - Species that are weighted on this component are basically tidal e.g *P. vaginatum* (+0.90), *A. africana* (+.0.88), *P. candelabra* (+ 0.83). Other species are important associates in the hyperspace e.g *F. exasperata* (+ 0.95) , and *V. cuspidata* (+0.82). *Vossia* spp. occurs in freshwater swamps and reflect tolerance of high soil moisture (which may lie at a brackish/saline interface) (see Ukpong 1977).

SPC 2 - This reflects soil moisture variation e.g *A. aureum* (+0.92), *P. candelabra* (+0.50) and R. hookeri (+0.58). *Selaginella* spp. (-0.91), purely lowland in Stubbs Creek is an important indicator of the soil moisture gradient. The species, being an upland plant, probably marks a gradation of soil mositure as a continuum from poorly drained coastal marsh into relatively well drained lowland and uplands.

SPC 3 - This component reflects a lowland (deltaic) environment and perhaps human disturbance. *S. torvum* (+0.60), *H. tilaceus* (+0.25) and *E. coccinea* (-0.45) occur in thickets and may be indicative of previous human encouragement

SPC 4 - The component is basically strand (*Philoxerus* spp -0.45; *Cyperus* spp + 0.33) and occupy positions in the coarse sandy substrates. *V. cuspidata* (+0.34) on this component demonstrates a situation where a species can generate a substantial proportion of the total variation within an extracted factor but without reflecting an important trend in the species population.

High communalities are exhibited by *Fiscus* spp, *R. racemosa, A. aureum* and *P. candelabra.* These species should show a high level of occurrence compared to the other species, but this assumption may not hold for the entire forest reserve since a relatively small sample could be analysed.

5.0 Conclusion

The PCA is a tool for simplifying the multivariate vegetation subsystem to reveal the major interspace variation. The simplification has formed the basis for hypothesis generation about the causes of variation within the subsystem. Although the component of the PCA is a linear function of the data, the technique is robust since non-linear relationships can be found by examination of different components together (Barkham and Norris 1970). In this regard it could be more useful to relate variations in the vegetation to PCA gradients of changing soil (Ukpong 1994).

The results of PCA on vegetation data in this study, show that in theoretical hyperspace, the overstorey was the least homogeneous and consequently faces the danger of greater biodiversity loss than the understorey and groundlayer. Management practices should be geared towards the preservation of the trees. A soil moisture gradient is hypothesized as influencing vegetation structural development and this should be investigated in detail, particularly in relation to nutrient flows and cycling. Species composition of the forest reserve also reflect environmental gradient of soil moisture. In the context of indirect gradient analysis, PCA has discerned the pattern of species variation in Stubbs Creek Forest Reserve.

6.0 References

Austin, M. P. 1968. An ordination study of a chalk grassland community. *J. Ecol.* 56:739-757.

Aweto, A 1978. Secondary Succession and Soil Regeneration in a Part of the Forest Zone of Southwestern Nigeria. Ph.D. Thesis, University of Ibadan.

Barkham, J. P. and J. M. Norris 1970 Multivariate procedures in an investigation of vegetation and soil relations of two beech woodlands, Costworld Hills, England. *Ecology* 54: 630-639.

Braun-Blanquet, J. 1932. *Plant sociology: The study of Plant Communities.* McGraw-Hill, New York.

Cooley, W. W. and P. R. Lohnes 1971 *Multivariate Data Analysis.* J. Wiley, New York.

Cattel, R. B. 1865. Factor analysis: an introduction to essentials 1: The purpose of underlying models. *Biometrics* 21: 190-215.

Essien, J. N. 1998. The relationship between vegetation and soils in Stubbs Creek in Akwa Ibom State. M.Sc Thesis, University of Uyo.

Ferrari, Th. J. and J. Mol. 1967. Factor analysis of causal models. *Neth. J. Agric. Sci.* 15: 38-49.

Goodall, D. S. 1965. Objective methods for the classification of vegetation II: An essay in the use of factor analysis. *Aust. J. Bot.* 2: 304-324.

Group, F. O. A. 1992. Biological and socio-economic survey of Stubbs Creek Forest Reserve, Akwa Ibom State, Nigeria. NCF, Lagos.

Gregory, S. 1973. *Statistical methods are the geographer*, Longman, London.

Hobbs, R. J. and J. Grace 1981. A study of pattern and processes in coastal vegetation using principal components analysis - *Vegetatio* 44: 137-153

Hopkings, B. 1968, Vegetation of the Olakemeji Forest Reserve, Nigeria. *J. Ecol.* 56: 97-115.

Ivimey-Cook, R. B. and M. E. F. Proctor 1967. Factor analysis of data from an east Devon heath: A comparison of principal components and rotated solutions. *J. Ecol.* 55: 405-413.

Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrica* 23: 187-200.

Kershaw, K. A 1980. *Quantitative and Dynamic Plant Ecology.* Edward Arnold, 308 p.

Orloci, L. 1966. Geometric Models in ecology 1: The theory and application of some ordination methods. *J. Ecol.* 54: 193-215.

Swaine, M. D. and P. Greig-Smith 1980. An application of principal components analysis to vegetation change in permanent plots. *J. Ecol.* 68: 33-41.

Seal, H. L. 1964. *Multivariate Statistical Analysis for Biologists.* J. Wiley, N. Y.

Ter Braak, C. J. F. and C. I. Prentice 1988. A theory of gradient analysis. In: *Advances in Ecological Research,* Vol. 18, pp. 272-313. Academic Press, London.

Ukpong, I. E. and O. O. Areola 1995. Relationships between vegetation gradients and soil variables of mangrove swamps in Southeastern Nigeria. *Afr. J. Ecol.* 33: 14-24

Ukpong, I. E. 1994. Soil - vegetation interrelationships of mangrove swamps as revealed by multivariate analysis. *Geoderma* 64:167-181.

Ukpong, I. E. 1995. An ordination study of mangrove swamp communities in West Africa. *Vegetatio* 116: 147-159.

Ukpong, I., E. 1997. Mangrove swamp at a saline/fresh water interface near Creek Town, Southeastern Nigeria. *Catena* 29: 61-71.

Wikum, D. A. and M. K. Wali 1974. Analysis of North Dakota gallery forest: Vegetation in relation to topographic and soil gradients. *Ecol. Monogr.* 44: 441-464.