Bayesian Factor Analysisof a UnidimensionalUrbanSprawlIndex in Ibadan, Nigeria.

Abstract

Urban sprawl has been a phenomenon on increase over the past years, people migrate into the core part of a major city, due to different factors like available housing, land availability, infrastructural development or population growth among others, making it a major cause of concern. Therefore, this study aimed at investigating and understanding the factors influencing the rate of sprawl in Ibadan, Oyo state, Nigeria. The research considered multivariate and Bayesian techniques to characterize and understand the sprawl for a given period.The multivariate techniques were Independent Component and the Principal Component Analysis (ICA & PCA) while, the Bayesian technique was Bayesian Factor Analysis (BFA). These three techniques deal with extraction of underlying factors affecting sprawl indices. The result showed that the BFA identified the built-up density, openness, frag-saturation, and proximity as the major factors with the loadings (posterior values):1.000, 0.412, 0.423, & 0.657 respectively. The ICA reduced to three underlying factors that could have effect on four variables (fragmentation, built-up density, urban density, cohesion). The PCA shared similar result of identified variables' reduction (built-up dinsity, urban density, frag-saturation, proximity, and cohesion) to three factors with the ICA. Thus, the BFA extracted a single(unidimentional) index by reduction of the factors gotten, thereby making it easier to compute. Therefore, the major elementsof sprawl are built-up density, fragmentation, and openness and could all be related to the Intensity of use of the land area. These four variables are the most significant agent of sprawl and should be looked into economically, of which built-up density is the highest, with the most significant impact, which is the major call of concern.

Keywords: Urban sprawl, Bayesian Factor Analysis, Independent Component Analysis, Principal Component Analysis, Factor loadings

1.0 Introduction

Urban Ecology is the study of ecosystems that includes the interaction of living things and their ecosystem or their interaction with their environment. It is a discipline that claims to help understand how people interact with the ecological factors and how human systems have made coexisting in the environment more sustainable for human activities (Akanbi O. B., 2018). Urban ecology is a term that has been employed in describing the relationship of humans in the city, natures in the city and that of human and nature. It is a study that has brought about the understanding of human activities and their physical environment plus other living things they cohabit together so as to make the cohabiting quite easy (Akanbi and Nurudeen, 2024).The dependence of human for their basic needs and the day to day activities of man in their interaction to the environment have led to the rapid growth and development of land uses and its modification over time, this growth comes with either negative or positive effect. An aspect of this growth is what led to “Urbanization”.

Urbanization has been one of the major players in human ecology, which is the movement of people to the major part of a city, as people migrate in search of better source of living has posed some challenges to the environment and natural resources at large. Man’s instinct for survival through urbanization has led to some environmental hazards like deforestation, traffic congestion, climate alterations, air, water and land pollution (Ohwo and Abotutu, 2015). This rapid increase in urbanization has now caused a spillover to the outskirts, which is where urban sprawl comes into play. The major consequence of urbanization is urban sprawl based on the fact that as urbanization of an area increase the need for some basic infrastructure and amenities also increases such as housing, health centres, recreation and the likes and this is mostly what brings about urban sprawl (Chettry, 2023; Enoh et al., 2023).

Urban sprawl is the movement of people from the major urban settlement to the outlying or outskirts of the urban settlement that might sometimes be seen as the relocation to a rural settlement far away from developed settlements (Rafferty P. O., 2024; Salam et al., 2023; Tagnan et al., 2022). This settlement moved to might be a settlement with lower population density, low infrastructural services and lack of major road network. Urban sprawl has been seen as a major aspect of analysis over the years and has migration and human activities have been progressing (Mourguiart B., 2022). People relocate from their place of residence for diverse reasons this could be due to some personal reasons or work related reasons. Urban sprawl has been a topic that has been addressed globally and has been a center of research for quite a number of authors. Urban sprawl can be analyzed either by using socio-economic census indicator or by spatial metrics. Researchers have either made use of these means to derive a unidimensional factor or a multidimensional factor with different methods that are not statistically inclined. In this work, spatial metrics would be employed. The metrics are used to characterize and understand urban sprawl in a particular area by means of measuring it.

Nigeria as a developing country with a population size of 140.5million as at 2006 census carried out by the National Population Census, with a land mass of 923,768km^2, with over 64% of the population living in rural areas. Nigeria is a centre of development because of the fast rising industrial build up and the industrial growth in the urban areas.Ibadan is a major city located in the south-eastern part of Oyo State, Nigeria and is one of the most populous and fast developing cities of the country. Ibadan is located on seven hills and about 160km away from the Atlantic coast. Ibadan is a centre for urban activities due to its fast rising population and has a commercial centre with diverse market structure ranging from local marketers to industrial sectors like Agricultural, commerce andmanufacturing industries (Dada et al., 2023; Taiwo, 2022; Otokiti et al., 2021). There has been a rapid decline over the years in the city's Agricultural activities (Kasim et al., 2014; Olubi, 2019).Over the years, the Lagos-Ibadan expressway generated the highest urban sprawl, both east and North of the city. The city has almost every corner with a market square or stall which makes exchange an easy one.Ibadan has eleven local government areas which areIbadan North, Ibadan North-East, Ibadan North-West, Ibadan South-East, Ibadan South-West, Akinyele, Egbeda, Ido, Lagelu, Ona-Ara and Oluyole.

The population of Ibadan as at 2006 census carried out by the National Population Commission was estimated to be 2,567,000 million and has a projected population to be 3,552,000 million by 2020 and a projected population of 4,004,000 by 2024 population projection by the United Nations.Ibadan is stated to be the third largest city by population in the country and has since experienced urban development. As at 1952, the city has an approximate total area of 103.8km^2 and a built-up area of 36.2km^2 which means the remaining area can be attributed to non-urban uses like farming, forest reserves and water bodies. Fast forward to the year 2000, this city has since experienced a rapid increase in its land use where it was estimated that Ibadan covered a 400km^2 and has a built up area of over 250km^2. This clearly shows that the urbanization of this city can lead to an urban sprawl which makes the city a very good area for the study of urban sprawl (Akinrinola, 2019; Owolabi, 2018).

Many authors have analyzed urban sprawl through different techniques, quite a number of articles have been reviewed on Bayesian statistics (Fusco and Tettamanzi, 2017; Bosch et al., 2019). The Bayesian statistical analysis would be employed in analyzing urban sprawl in the city of Ibadan (Akanbi O. B., 2022) through the help of GIS (Geographic Information System) to get the adequate data needed for the analysis of this work.Bayesian analysis deals with calculating different values of the probability of a parameter given the data. The basis of Bayesian statistics is the ability to incorporate a prior knowledge to influence the output of the posterior distributions. Bayesian analysis is flexible in the sense that it enables you influence your prior input based on how rigid the information gathered by the variable to be analyzed is vetted (Akanbi and Oladoja, 2019). Researchers sometimes want the data to speak and really influence the analysis that is why weak or non-informative priors is made use of, while sometimes want their knowledge of the variables to influence the analysis. There are various aspect of Bayesian statistics, but in this study we will make use of the Bayesian Factor Analysis.Bayesian factor analysis is a statistical technique that incorporates Bayesian methods in parameter estimation to extract latent variables influencing the observed data by its dimensionality reduction process (Luan and Fuller, 2022)

Conservation of land use for agricultural purposes is kind of going extinct as there is high rate at which development of urban areas and even outskirts of major cities. The rate at which urbanization is taking place can give rise to the depletion of some major environmental factors land use conservation has always been a major focus for the government, because of the preservation of wildlife, agricultural produce which helps the growth of the economy. Over the years it has been noted that urbanization has been fast rising, people no longer feel the need for land conservation, different estate companies rising and purchasing the lands that could have been conserved to balance the environmental scale, all in the name of development. Thus, this study hereby shows the patterns and attributes of sprawl in urbanized locations by obtaining a one-dimensional index using Bayesian Factor Analysis (BFA).

2.0 Methodology

The urban sprawl is a phenomenon that can be analyzed using the geographical information system (GIS). The first step taken is to get the indicators available for analyzing sprawl of a particular community. The indicators used are built-up density, urban extent density, fragmentation which is divided into saturation and openness, compactness which is also divided into proximity and cohesion, and the last is leapfrog. This analysis will be carried out with the Independent Component Analysis, Principal Component Analysis and Bayesian Factor Analysis technique. Below is a table showing the variables to be used and how they are estimated.

Table 1: Variable Indicators and their formula.

Indicators	Formulas
Built-up Density	$\frac{total\ built-up\ area}{total\ land\ area}$
Urban Extent Density	$\frac{total\ urban\ area}{total\ land\ area}$
Fragmentation: saturation	$\frac{total\ lenght\ of\ fragmented\ boundaries}{total\ lenght\ of\ perimeter}$
Openness	$\frac{total\ area\ of\ open\ space}{total\ urban\ area}$
Proximity	$\frac{total\ number\ of\ links\ between\ urban\ patches}{total\ urban\ area}$
Cohesion	$\frac{total\ no\ of\ links\ between\ urban\ patches}{total\ number\ of\ patches}$
Leapfrog	$\frac{n_i - m_i}{n_i + m_i}$

Where,

$n_i$is the number of neighboring locations with values greater than the focal locations' value.

m_i is the number of neighboring locations with values lower than the focal location's value.

2.1 Independent Component Analysis

Independent component analysis is a statistical technique that tends to expose the underlying factors that influence a dataset. It is a tool for getting out useful information from a set of observed data.

Assumptions of Independent Component Analysis

Independence: it is assumed that each observed variable value is statistically independent of the other, that is the observed value of a variable does not affect the observed value of another.

Normality: the observed values are expected not to be Gaussian because of central limit theorem, that if two of the values are Gaussian then their linear combination will become more Gaussian which makes it hard to recover the original data.

Linear Mixture: each observed variable is assumed to be linear which makes the combination of the observed variables a linear algebra of the mixing matrices to be estimated.

At most one Gaussian source: it is assumed that at most one of variable is Gaussian, if more than one it affects the result and makes the analysis impossible.

Mathematical Model Formulation

The Whitening Matrix: this is the transformed observed data with unit variance and now uncorrelated. That is:

$$E\{XX^{I}\} = 1 \quad (1)$$

$$X_{\text{white}} = \text{whitemat} \cdot X \quad (2)$$

Where,

X is the observed variable data

X_{white} is result of the product whitening matrix and the observed data.

Whitemat is the whitening matrix

To estimate the mixing matrix:

Assumming n set of individual linear mixtures $x_{1,}$ x_2, …x_n

$$X_{\text{j}} = a_{\text{j1}} s_1 + a_{\text{j2}} s_2 + a_{\text{j3}} s_2 + \ldots + a_{\text{jn}} s_{\text{n}} \quad (3)$$

let x denote a random vector whose elements are a mixture of $x_{1,}$ x_2 …x_n

and s be a random vector with elements $s_{1,}$ s_2 … $s_{n.}$

let A denote the matrix with elements a_{ij}

Using this vector–matrix notation, the above mixing model is written as:

$$x = As \quad (4)$$

$$\text{x} = \sum_{i=1}^{n} a_i \ s_i \quad (5)$$

Then after estimating the matrix A, we compute the inverse W.

$$S = Wx \tag{6}$$

Where, W = transformation matrix.

2.2 Principal component analysis(PCA)

This is a statistical technique that maximizes the variance of linear combination of variables.

Let $Y_1, Y_2, \dots, Y_n$ be an n-observation vector in a p-dimensional space. The variables $y_1, y_2, \dots, y_p$ are uncorrelated, the mean vector is $\bar{Y}$.

Let A be an orthogonal matrix, then

$$Z_i = AY_i \tag{7}$$

Since A is orthogonal then A'A = I

$$Z_i'Z_i = (AY_i)'(AY_i) = Y_i'A'AY_i = Y_i'Y_i \tag{8}$$

the orthogonal matrix has transformed the point Y_i to a point Z_i. thus the resulting principal components $Z_1, Z_2, \dots Z_P$ in $Z = A$.

The sample covariance matrix Z is $S_z = ASA' = \begin{pmatrix} s_{z_1}^2 & 0 & 0 \\ 0 & s_{z_2}^2 & 0 \\ 0 & 0 & s_{z_p}^2 \end{pmatrix}$ (9)

Where S is the sample covariance matrix $y_1, y_2, \dots, y_n$.

Proportion of variance estimated $= \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$

$$= \frac{\lambda_1 + \lambda_2 + \dots + \lambda_k}{\sum_{j=1}^{p} s_{jj}} \tag{10}$$

Algebraically, z = a'y = a'Sa (11)

$$\boldsymbol{\lambda} = \frac{a'Sa}{a'a} \tag{12}$$

$$(S - \lambda I)a = 0 \tag{13}$$

Consider the matrix S whose variables are uncorrelated

$$S = \begin{bmatrix} S_{11} & 0 & \dots & 0 \\ 0 & S_{22} & \dots & 0 \\ & & & \end{bmatrix}$$

$\vdots\vdots \quad \dots \quad \vdots$

$0 \qquad 0S_{pp}$

The characteristic equation will be

$$0 = |S - \lambda I| = \prod_{i=1}^{p}(S_{ii} - \lambda) \qquad (14)$$

Which has the solutions $\boldsymbol{\lambda}_i = S_{ii}$, $i = 1, 2, \dots, p$

Thus, the ith component is $z_i = a_i'y = y_i$

Assumptions of Principal Component Analysis

- Linearity: it is assumed that the underlying associationsamong variables are linear.
- Orthogonality: it assumes that the principal components are orthogonal to each other.
- Variance: it maximizes the amount of variance of the data.
- Normalization: PCA assumes the data is normalized that is they have zero mean and unit variance.

Detection of components to be retained

By retaining the component that explains a substantial amount of the total variance, like 80%. Components whose eigenvalues are greater than the average of the eigenvalues should be retained i.e,$\sum_{i=1}^{p}\frac{\lambda_i}{p}$.Making use of the scree plot, of the eigenvalues against the number of variables, and detect the bet of the curve between the larger eigenvalues and the smaller eigenvalues.Components corresponding to the larger eigenvalues should be tested for it's significance level.

2.3 Bayesian Factor Analysis

Bayesian factor analysis is a statistical technique that is used for dimensionality reduction and extracting latent factors underlying the variables. The statistical model for BFA is

$$Y = \lambda F + \varepsilon \qquad (15)$$

Where, Y is the matrix of observed variables; λ is the matrix of factor loading

F is the matrix of latent factors; ε is the error term

The Bayesian factor model makes use of hierarchical models. Let Q_{ij} be the value of the indicator $j, j = 1,2, \ldots, 7$ and i be the areas covered $i = 1, 2, \ldots, 10$.

$$Q_{ij} \sim \text{Normal}(\mu_{ij}, \sigma)$$

$$\mu_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + + \beta_k x_{ki} \qquad (16)$$

where,

$$\beta_0 \sim \text{Normal}(\mu, \sigma^2)$$

$$\beta_j \sim \text{Normal}(\mu, \sigma^2) \quad \text{for } j = 1,2, \ldots, k$$

$$\sigma^2, \sigma \sim Gamma\ (mean,\ variance)$$

x_{ij} represents the value of predictor variable j for the i-th observation.

β_0 is the intercept.

β_j are the coefficients for the predictor variables x_{1i} through x_{ki}.

μ_i is the mean of the normal distribution for the i-th observation.

σ is the standard deviation of the normal distribution.

The model assumes that the observed variable Q_{ij} is normally distributed.

The likelihood of a Bayesian factor analysis is assumed to be a multivariate normal distribution and can be estimated below:

$$\Phi(x) = \left[\frac{1}{2\pi}\right]^{p/2} |\Sigma|^{-1/2} \exp\{ -\frac{1}{2}(x - \mu)' \Sigma^{-1}(x- \mu)\} \qquad (17)$$

$|\Sigma|$ denotes the determinant of the variance-covariance matrix Σ

Σ^{-1} is the inverse of the variance-covariance matrix.

The multivariate normal density function is given below:

$$\Phi(x) = \prod_{j=1}^{p} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \{\frac{-1}{2\sigma^2} (x_j - \mu_j)^2\} \qquad (18)$$

Assumptions of Bayesian Factor Analysis

Normality: it is assumed that the data is normally distributed; this ensures our posterior inference is proper.

Prior Information: The Bayesian approach makes use of prior knowledge of the observed data, which makes it have the best knowledge of the data to be analyzed.

Model structure and Uncertainty: Bayesian model accounts for uncertainty in flexible model estimation.

Sample size: the sample size should be greater than the number the number of observed variables.

Multicolinearity: there should be no perfect multicollinearity among the observed data.

Linearity: it is assumed the relationship among the observed variables is linear.

Markov Chain Monte Carlo

The mathematical process of a state space model undergoing transition is regarded as a markov property and is represented below:

$$P(X_{t+n} = x | X_t, X_{t-1}, \dots, X_{t-k}) = P(X_{t+n} = x | X_t) \quad (19)$$

This property tells us that the process is a memoryless one and future parameters are estimated based on the current state of the process.

The conditional probability is given by $X_{t+n} = x_{t+n}$ given that $X_t = x_t$ and is denoted by $f(x_{t+n}|x_t)$.

The unconditional probability is given by $X_t = x_t$ denoted by $f(x_t)$.

The optimization technique employed in this study for posterior inference is the Hamiltonian Monte Carlo Markov Chain.

The Hamiltonian Monte Carlo technique makes use of the derivative of the density function being sampled to generate transitions and get the posterior and makes use of the metropolis acceptance step.

The joint density of a Hamiltonian is:

$$H(\rho, \theta) = -logp(\rho, \theta)$$

$$= -logp\,(\rho|\theta) - log\,p(\theta)$$

$$= T(\rho|\theta) + V(\theta) \quad (20)$$

where

$T(\rho|\theta) = -logp\,(\rho|\theta)$ is called the "kinetic energy"

and the term$V(\theta) = -logp(\theta)$is called the "potential energy." The potential energy is specified by the Stan program through its definition of a log density.

To generate transitions in Hamiltonian,

$$\rho \sim MultiNormal(0, \Sigma)$$

where ρ is an auxiliary momemtum variable.

The joint system (θ,ρ) becomes,

$$\frac{d\theta}{dt} = +\frac{\partial H}{\partial \rho} = +\frac{\partial T}{\partial \rho} \quad (21)$$

$$\frac{d\rho}{dt} = -\frac{\partial H}{\partial \theta} = -\frac{\partial T}{\partial \theta} - \frac{\partial V}{\partial \theta} \quad (22)$$

The metropolis acceptance is given by:

$$P_t = \min\left(\frac{f(y_t)}{f(x_{t-1})}\frac{q(x_{t-1}|y_t)}{q(y_t|x_{t-1})}, 1\right) \quad (23)$$

The marginal log-likelihood and posterior predictive p-value are used to access the model fit of the data. For a model M, with likelihood function $\{f_\theta(y): \theta \in \varphi\}$, parameterized by θ, a prior $\pi(\theta)$ and observations $y_{1:n} \in y^n$, the marginal log-likelihood is:

$$P_m(y_{1:n}) = \int f_\theta(y_{1:n})d\pi(\theta) \quad (24)$$

The posterior predictive with a prior $\pi_G(\theta)$ is given as:

$$s_G(\widetilde{y}|y_{1:n}) = log \int g\{1(\theta, \check{y})\}d\pi_G(\theta|y_{1:n}) \quad (25)$$

The extraction method each analysis type will make use of explicitlyformulas related to each methoddiscussed above.

3.0 Analysisand Discussion of results

3.1 Presentation of Data

The data used was gotten from spatial interpretation of Landsat imagery within the period of 1984-2020 through the help of Geographical Information System(GIS). Adequate data needed for the analysis was recorded. Below is the imagery data used for the analysis:

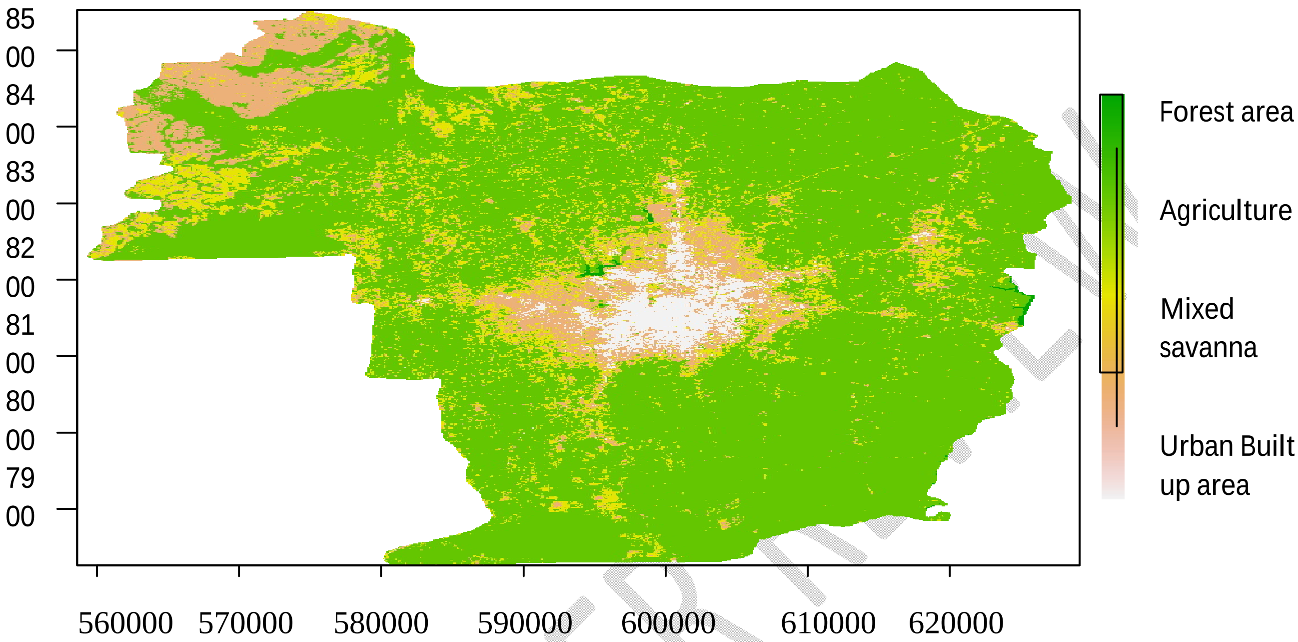

Figure 1:The general land cover of the city of Ibadan for the year 1984.

.

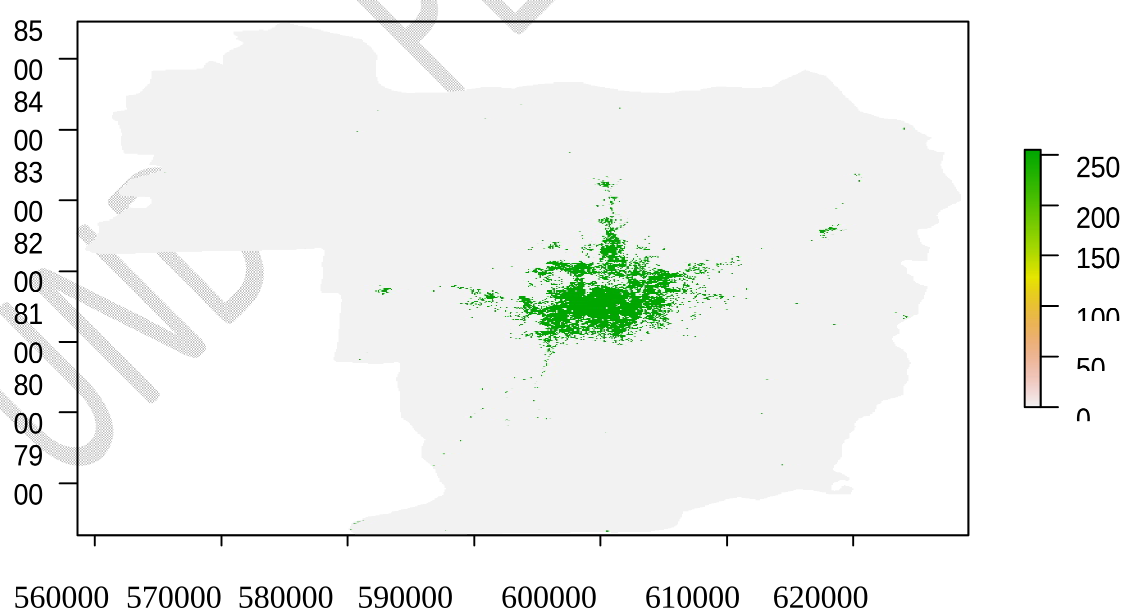

Figure 2: **Built-up area of Ibadan for the year 1984**.

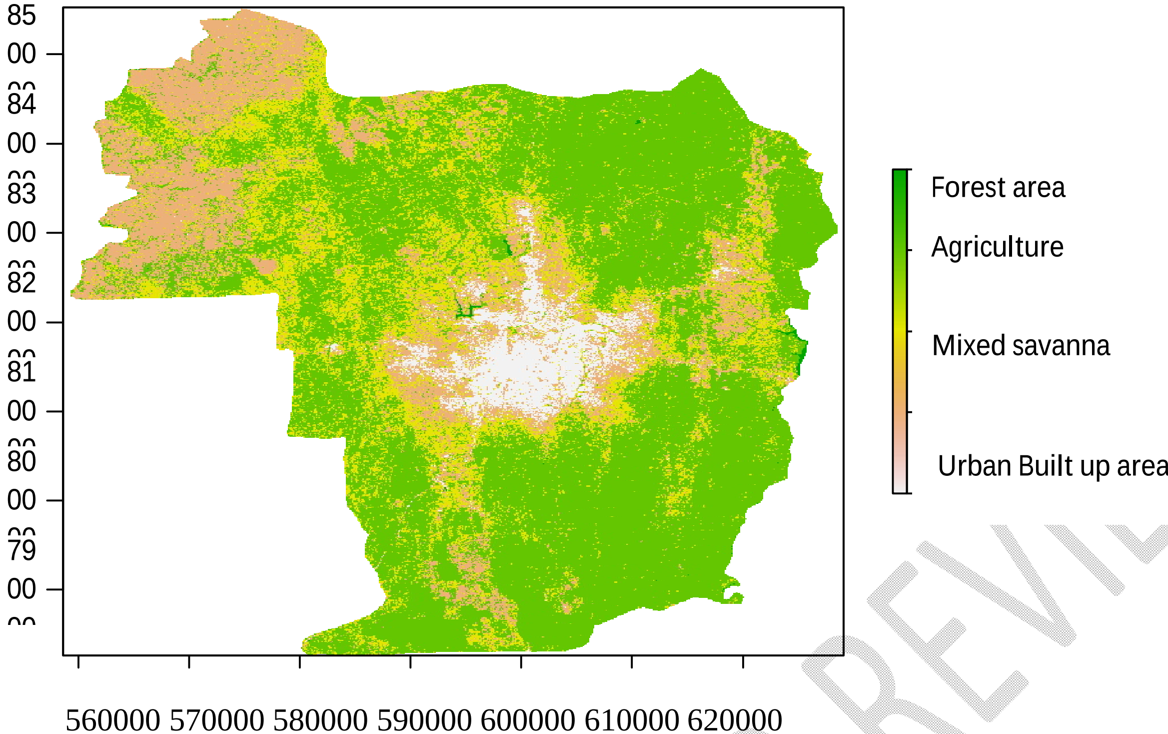

Figure 3:The general land cover for the year 2000.

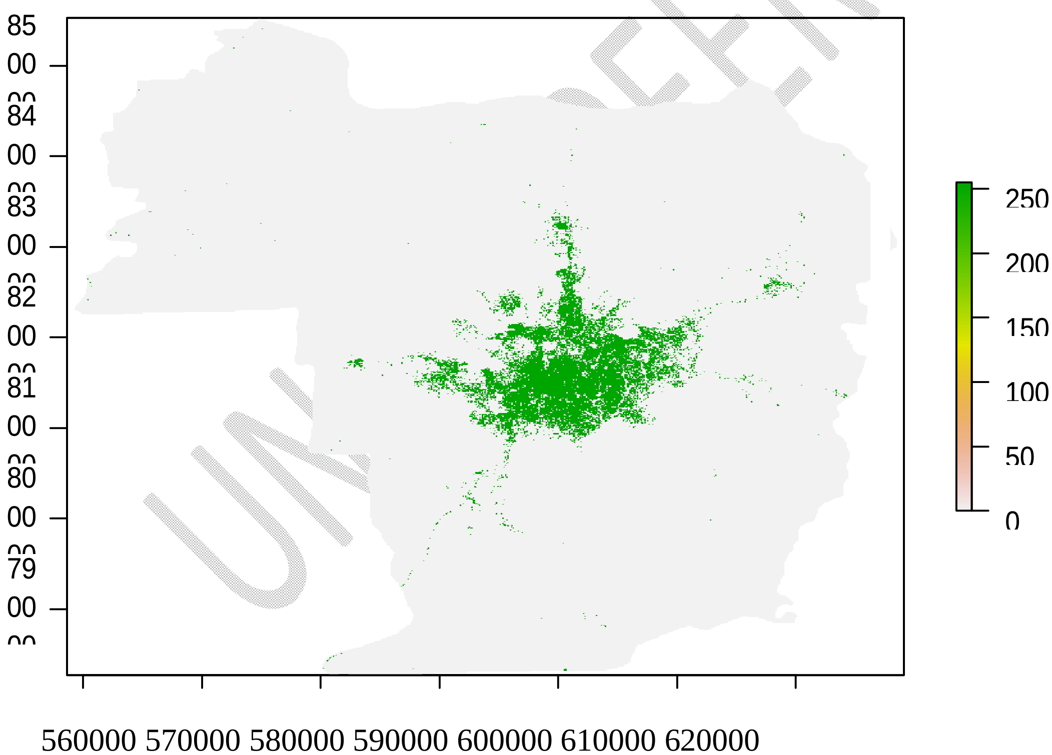

Figure 4 :Urban built up area for year 2000.

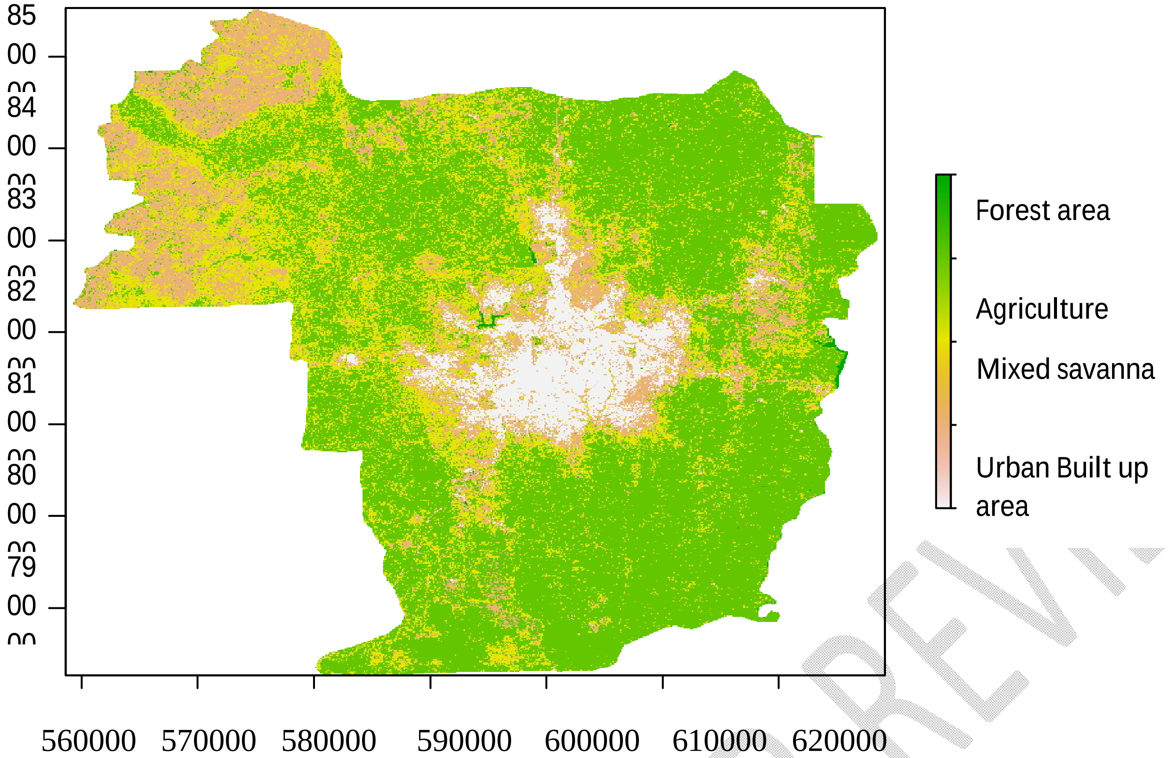

Figure 5:**Land cover area for the year 2006.**

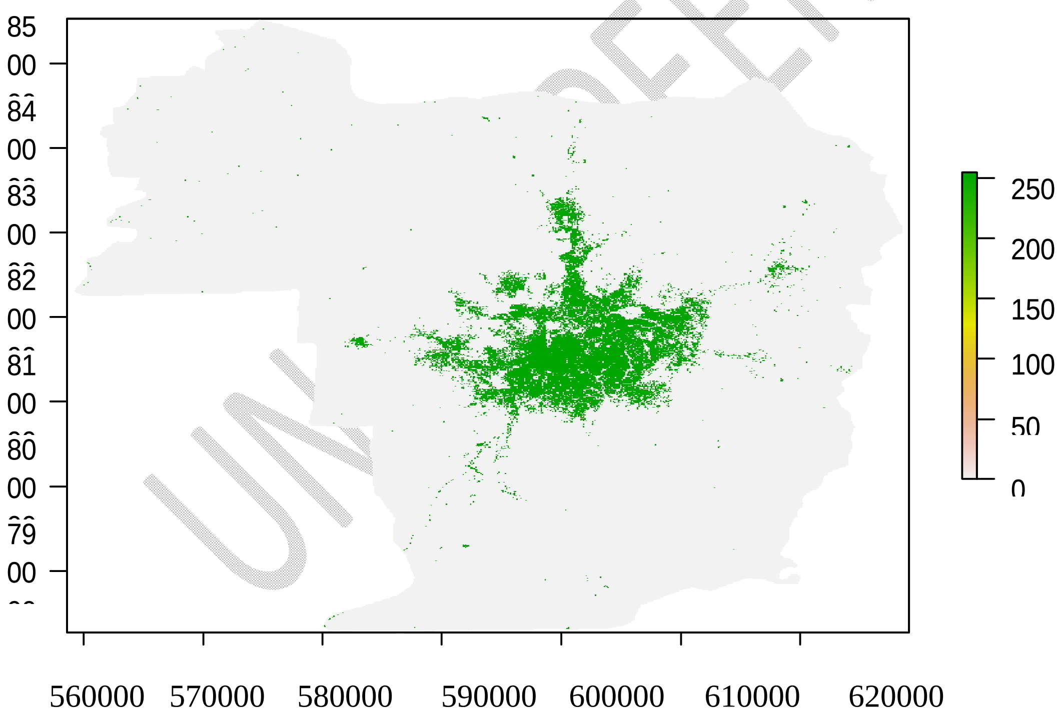

Figure 6: Built-up area for the year 2006.

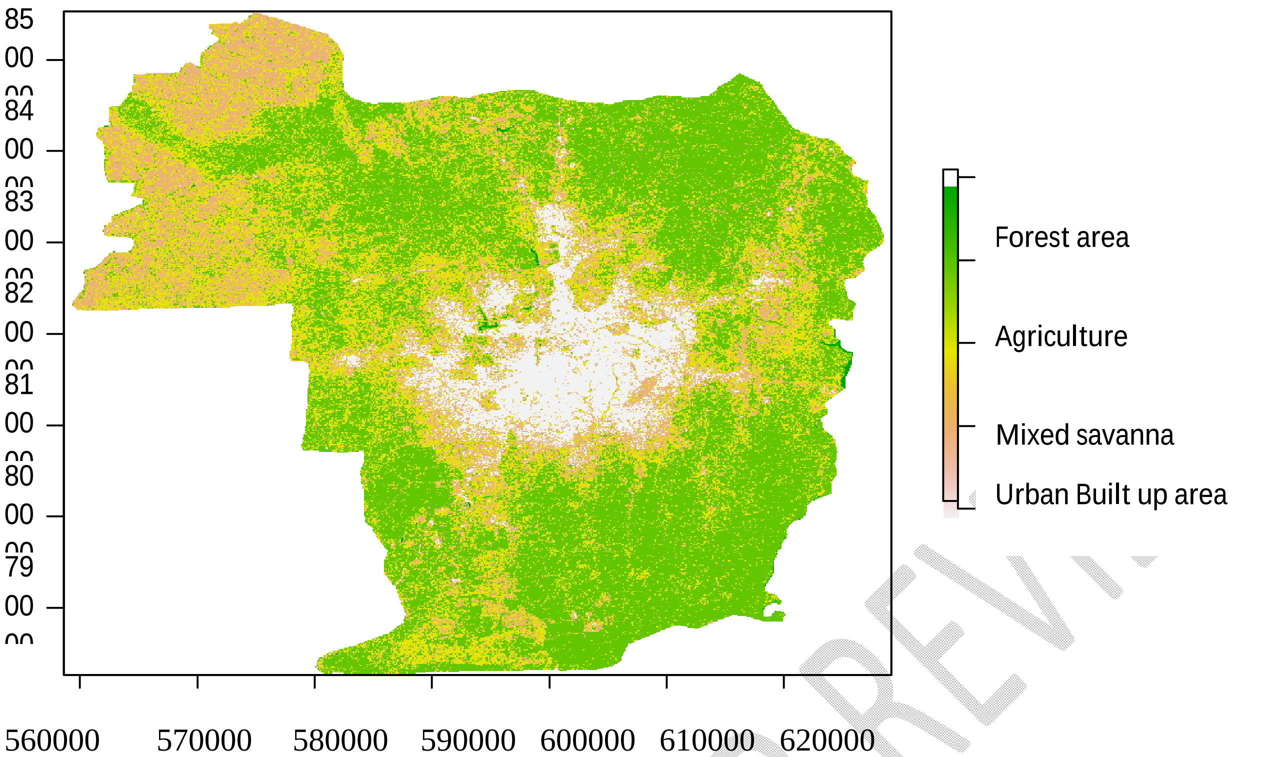

Figure 7: Land cover area for year 2013.

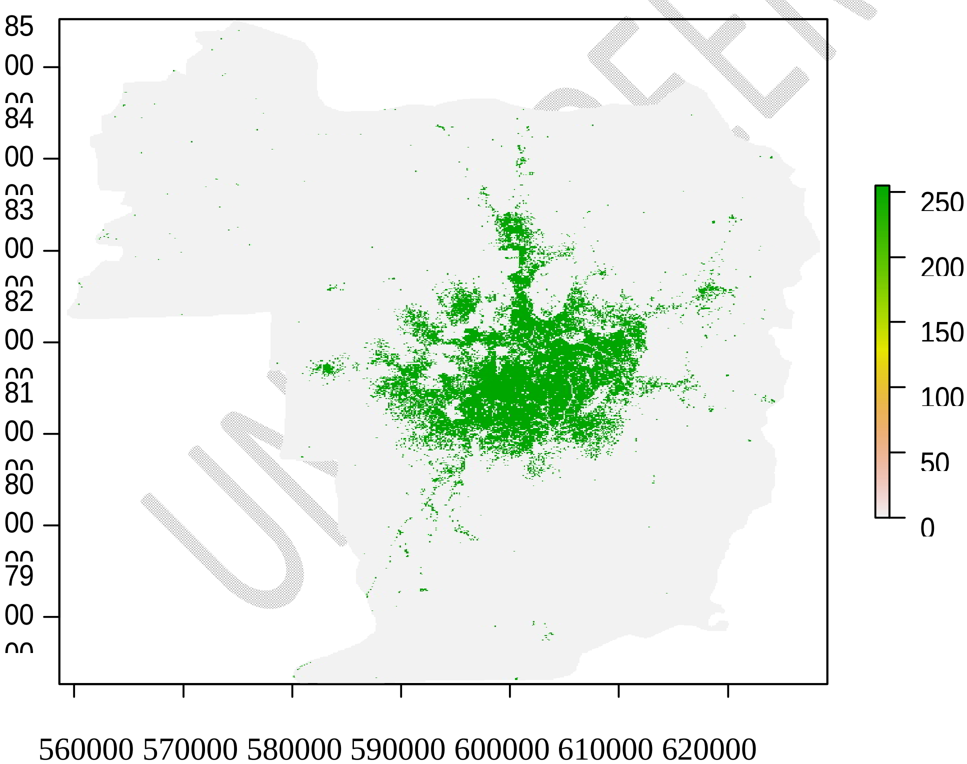

Figure 8: **Built-up area of Ibadan for the year 2013.**

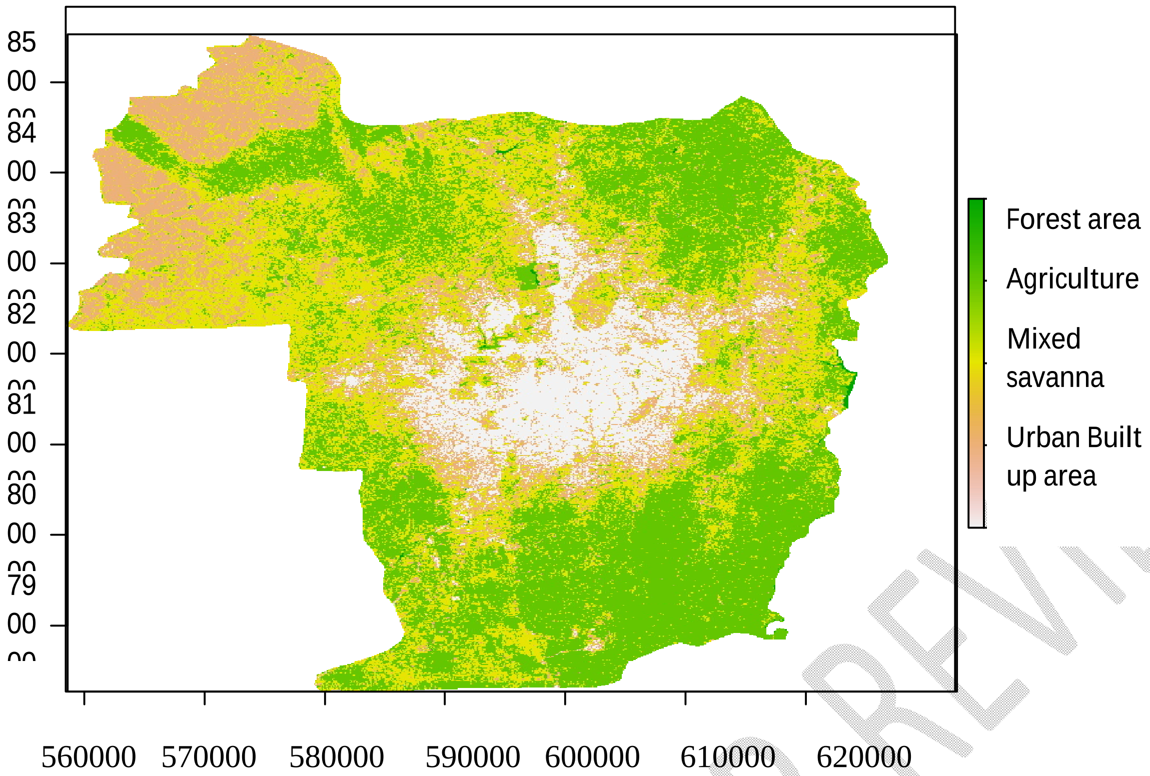

Figure 9: **Land cover area of Ibadan for 2020.**

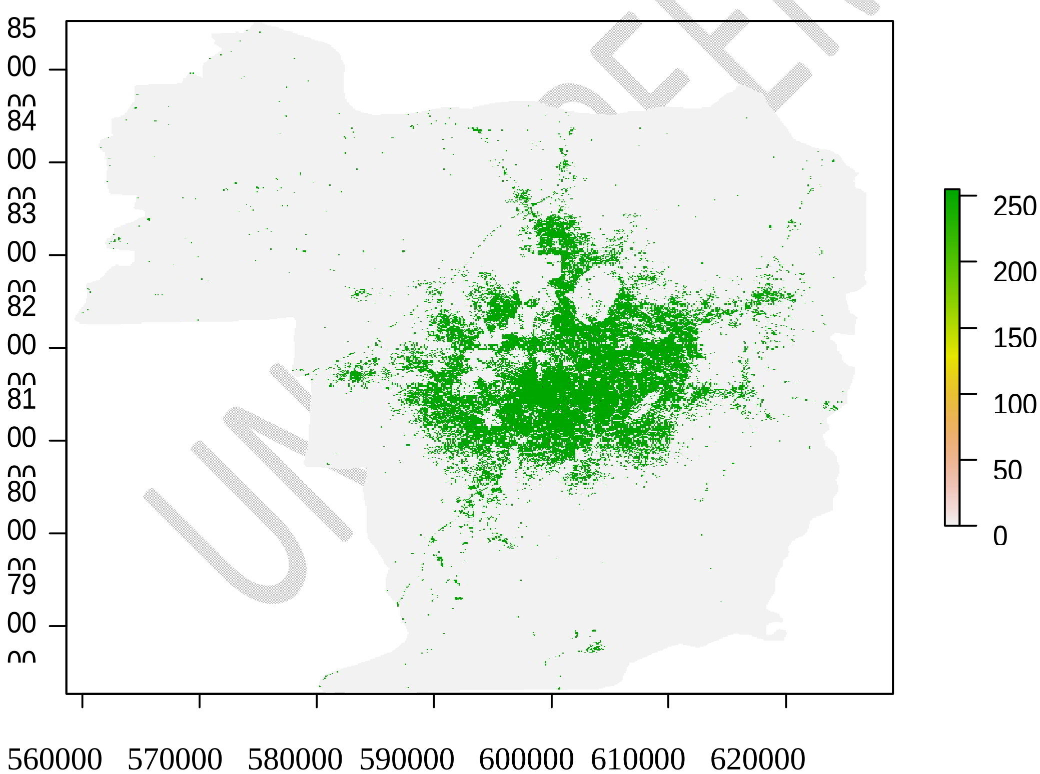

Figure 10: **Built-up area of Ibadan for the year 2020.**

3.2 Presentation of the Analysis Results

3.2.1 Independent Component Analysis

Table 2 shows the factor loadings of the ICA which represents how each variable contributes to the independent component.Factor 1: Has a positive loading with urban density, frag-saturation and leapfrog while negative with built-up, openness, proximity and cohesion. Thus, Factor 1 represents a strong association with frag-saturation and urban density.Factor 2: Has a positive loading with built up density, frag-saturation and cohesion while negative with urban density, openness, proximity and leapfrog. Thus,Factor 2 represents a strong association with cohesion.Factor 3: Has a positive loading with built-up, urban density, openness and cohesion while negative with frag-saturation, proximity and leapfrog. Factor 3 depictsa strong association with built-up density and urban density.

Table 2: Independent Component Analysis Result

Variables	**DIM 1**	**DIM 2**	**DIM 3**
Builtup density	-0.269	0.318	0.616
Urban density	0.313	-0.032	0.463
Frag-saturation	0.469	0.193	-0.068
Openness	-0.431	-0.253	0.032
Proximity	-0.094	-0.293	-0.316
Cohesion	-0.335	0.472	0.105
Leapfrog	0.155	-0.124	-0.194

3.3.2: Principal Component Analysis

The values of these loadings represent how much each variable contribute to each principal component. It can also be seen that the variables with high loadings on each factors are built-up density, urban density, fragmentation, proximity and cohesion. This verifies the result of the ICA done.

Table 3: Principal Component Analysis Results

	Dim1	**Dim2**	**Dim3**
Built-up density	0.111	0.297	0.900
Urban density	-0.104	0.828	0.153
Frag-saturation	0.777	0.434	-0.315
Openness	-0.852	-0.421	0.224
Proximity	0.717	-0.461	-0.057
Cohesion	-0.672	-0.319	0.581

Leapfrog	-0.255	0.386	-0.009

Table 4 shows the variance explained by each factor, the proportion of explained variances by each factor and the cumulative variance explained by all the factors. It can be seen that factor one explained 34% of the total variance, factor two explained 22%, factor three 18%, factor four 13%, factor five 5%, factor six 3% and the seventh factor explained 1%. This can be seen as the reason three factors were picked as having underlying information about our data.

Table 4: Amount of Variance Explained

	Dim1	**Dim2**	**Dim3**	**Dim4**	**Dim5**	**Dim6**	**Dim7**
Variance	2.383	1.603	1.324	0.956	0.417	0.261	0.056
% of var.	34.043	22.896	18.920	13.654	5.963	3.727	0.796
Cum. % of var.	34.043	56.939	75.860	89.514	95.477	99.204	100.00

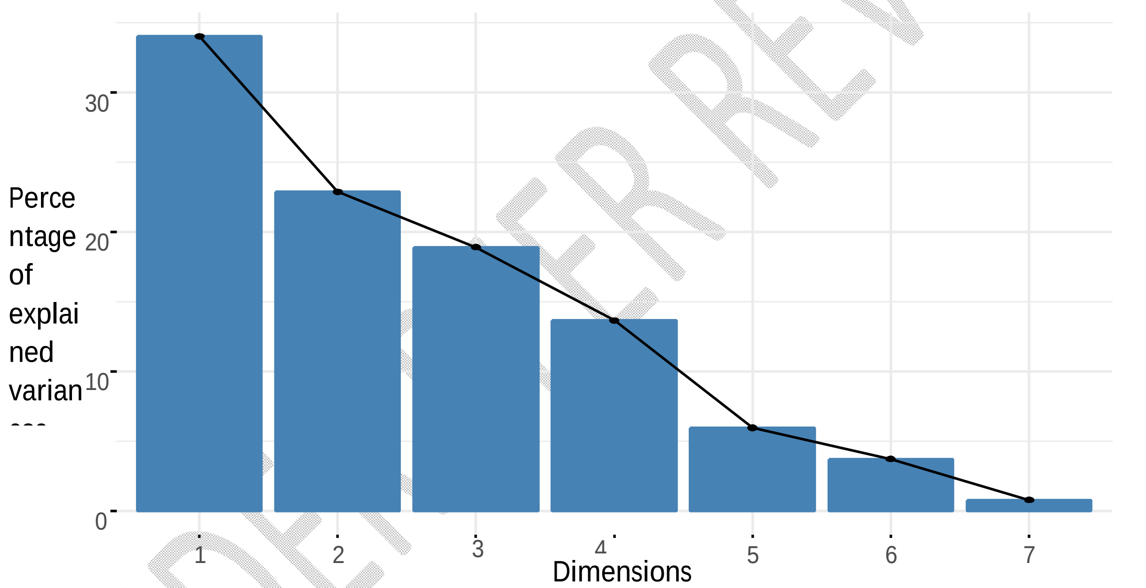

Figure 11: Screen plot showing the percentage of explained variances.

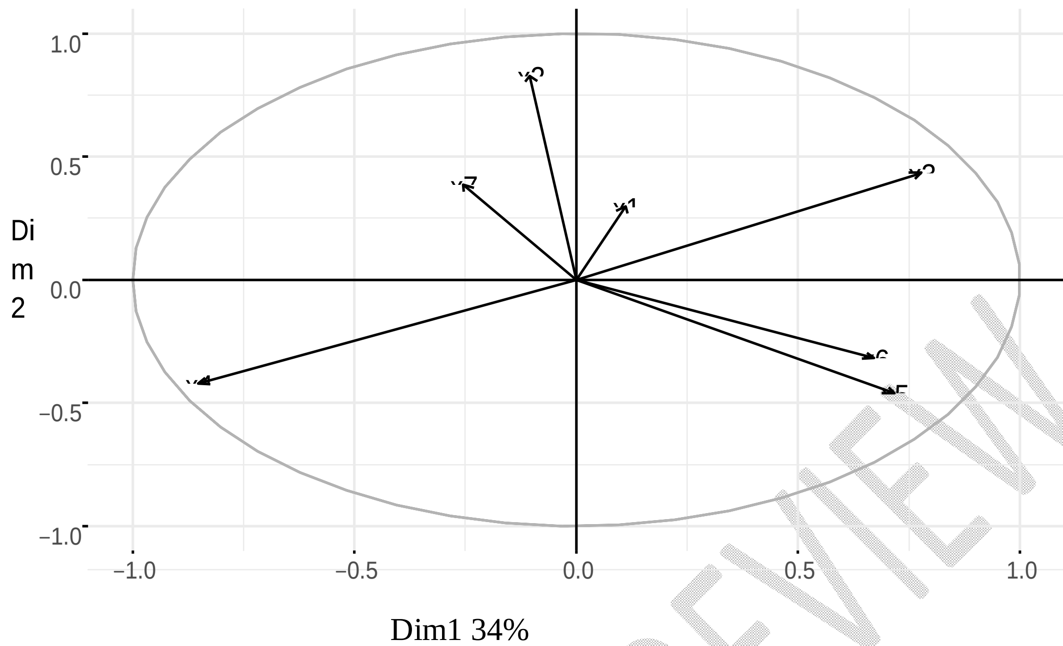

Figure 12: **Variables plot of the principal components (PCA).**

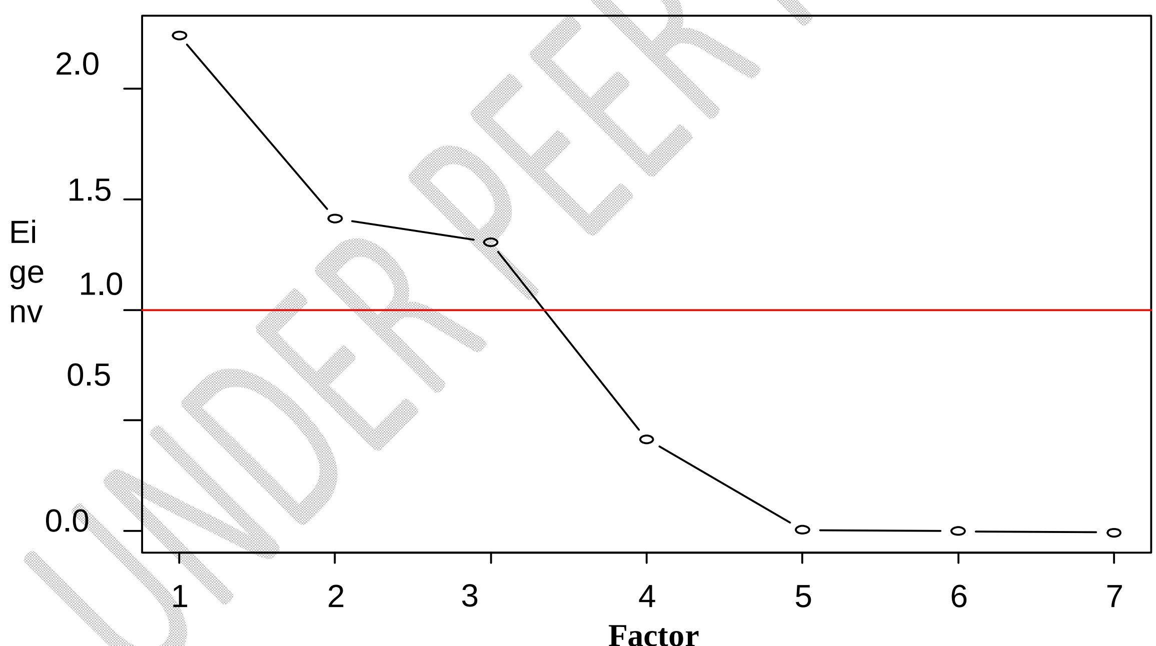

Figure 13: **Screen plot showing the components to retain.**

Comparing the results of both the ICA and PCA, it is observed that they are somewhat similar. Three factors were both extracted as the underlying latent factor affecting the observed variable. Each of the factors is expected to produce almost likely results, if at all they aren't all similar. The first factor produced by both the ICA and PCA results both has strong association on the variable frag-saturation.

The second factor from the ICA has a strong association with cohesion while from the PCA has with urban density and saturation. The third factor from the ICA has strong association on built-up density and urban-density while from the PCA has strong association from built-up and cohesion. With these results, it is obvious the ICA and PCA gives us the clearest output of each factors but the extent of variances explained wasn't recorded in the ICA technique and since the PCA is somewhat similar, it can be said that the proportion of variances explained by each factor are similar too. In summary, three factors were extracted and from the outputs of these factors, it can be concluded that the twoanalysis in the third factor have the most similar result and can be interpreted likely.

3.3.3 Bayesian Factor Analysis

The Bayesian factor model makes use of hierarchical models. Let Q_{ij} be the value of the indicator $j, j = 1,2, \ldots ,7$ and i be the areas covered $i = 1, 2, \ldots ,10$.

$$Q_{ij} \sim \text{Normal}\ (\mu_{ij}, \sigma)$$

$$\mu_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + + \beta_k x_{ki}$$

where, $\beta_0 \sim \text{Normal}\ (0,10\ )$

$\beta_j \sim \text{Normal}\ (0, 10) \quad \text{for}\ j = 1,2, \ldots , k$

$\sigma 2 = 1, \quad \sigma \sim Gamma\ (1,\ 0.5)$

x_{ij} represents the value of predictor variable j for the i-th observation.

$\beta_0$is the intercept ;$\beta_j$are the coefficients for the predictor variables x_{1i} through $x_{ki.}$

μ_i is the mean of the normal distribution for the i-th observation; σ is the standard deviation of the normal distribution.The model assumes that the observed variable $Q_{ij}$is normally distributed

In Table 5, the estimate column signifies the point estimate of the factor loading, it indicates the strength of the relationship of the observed variables and the latent factor.The Post SD column provides the standard deviation of the posterior distribution of the factor loading.The Pi lower and Pi upper gives the lower and upper bounds of the 95% credible interval for the factor loading.The Rhat column is related to convergence diagnostic in MCMC sampling, values close to one indicate good convergence.The result signifies good convergence, the model fit the observed data accurately.

Table 5: Bayesian Factor Analysis Result.

	Estimate	Post SD	Pi lower	Pi upper	Rhat
Built-up density	1.000				
Urban density	-0.214	9.971	-18.478	19.423	1.000
Fragmentation	0.423	9.734	-18.857	20.397	1.001
Openness	0.412	9.420	-19.348	18.558	1.009
Proximity	0.657	10.169	-19.460	20.238	1.012
Cohesion	-0.120	9.502	-18.343	18.920	1.008
Leapfrog	-0.583	4.273	-11.561	8.346	1.040

The Bayesian model fit for the data is

$$\boldsymbol{x_i = \lambda_i f_1 + \varepsilon_i}$$

Where x_i is the ith observed variable.

$\boldsymbol{\lambda_{i1}}$ = is the factor loading of the ith observed variable

F_1 = the latent factor

$\boldsymbol{\varepsilon_i}$ = error term

When i = 1, $x_1 = 1.000 \times f_1$

When i = 2, $x_2 = -0{,}214 \times f_1 + \boldsymbol{\varepsilon_i}$

When i = 3, $x_3 = 0.423 \times f_1 + \boldsymbol{\varepsilon_i}$

When i = 4, $x_4 = 0.412 \times f_1 + \boldsymbol{\varepsilon_i}$

When i = 5, $x_5 = 0.657 \times f_1 + \boldsymbol{\varepsilon_i}$

When i = 6, $x_6 = -0.120 \times f_1 + \boldsymbol{\varepsilon_i}$

When i = 7, $x_7 = -0.583 \times f_1 + \boldsymbol{\varepsilon_i}$

Table 6 shows the amount of variance explained by each observed variable. It was observed that variable X_1 explained 5.783. X_2 explained 6.882. X_3 explained 0.001. the post. SD explain the posterior standard deviation of the observed variable. The pi-lower and upper show the credible interval of the observed variable. The Rhat shows the convergence of the analysis, the values estimated showed good convergence of the estimated variables.

Table 6: Amount of variance estimated.

Variances	Estimate	Post.SD	Pi.lower	Pi.upper	Rhat
X_1	5.783	3.587	2.022	4.473	1.006
X_2	6.882	3.713	2.557	4.821	1.004
X_3	0.001	0.001	0.000	0.005	1.001
X_4	0.000	0.000			1.005
X_5	0.000	0.000			1.006
X_6	0.000	0.000			1.007
X_7	0.000				1.034
F_1	0.000	0.000			1.029

Table 7 shows the table of relationship of covariance among the observed variances.

Table 7: Amount of covariance estimated.

Covariance	Estimate	Post.sd	Pi.lower	Pi.upper	Rhat
X_1 ~~					
X_2	1.057	1.734	-2.057	4.990	1.003
X_3	0.001	0.025	-0.046	0.056	1.004
X_4	0.003	0.014	-0.024	0.032	1.003
X_5	-0.002	0.011	-0.025	0.019	1.003
X_6	0.007	0.011	-0.014	0.030	1.024
X_7	0.000	0.000	-0.000		1.002
X_2 ~~					
X_3	0.002	0.031	-0.054	0.057	1.022
X_4	-0.009	0.015	-0.044	0.018	1.022
X_5	-0.008	0.012	-0.034	0.012	1.004
X_6	-0.005	0.010	-0.025	0.013	1.001
X_7	0.000	0.000	-0.000		1.001
X_3 ~~					
X_4	-0.000	0.000	-0.001		1.016
X_5	0.000	0.000	-0.000		1.003
X_6	-0.000	0.000		0.000	1.001
X_7	0.000	0.000	-0.000		1.001
X_4 ~~					
X_5	-0.000	0.000		0.000	1.007
X_6	-0.000	0.000		0.000	1.004
X_7	0.000	0.000	-0.000		1.003
X_5~~					
X_6	0.000	0.000	-0.000		1.009
X_7	-0.000	0.000		0.000	1.009

X_6~~					
X_7	-0.000	0.000		0.000	1.002

4 Summary and Conclusion

The purpose of this study was to understand and characterize urban sprawl in the city of Ibadan, and to obtain a unidimensional index variable using a combination of Independent Component Analysis (ICA), Principal Component Analysis (PCA), and Bayesian Factor Analysis (BFA). Based on the result of the analysis, the ICA produced three underlying factors that could have more effect on the variables, analyzing and checking the factors that have stronger association with the factors, it wasobserved thatfour variables; fragmentation, built-up density, urban density, cohesion, were identified. The PCA shared similar result of reduction to three factors with the ICA. Bayesian factor analysis was adopted to extract a unidimensional index variable, by defining it into a simpler model and extracting a single index by reduction of the factors gotten, thereby making it easier to compute.The Bayesian analysis result showed that the major variables being affected by the factors were the built-up density, fragmentation, openness and proximity. The built-up density with an estimate of 1.000, the fragmentation with an estimate of 0.432, the openness with an estimate of 0.412 and the proximity with an estimate of 0.652 all have significant role in the rate of sprawling. It could be inferred that the major elementsof sprawl are built-up density, fragmentation, and openness and could all be related to the Intensity of Use of the land area. The conclusion based on the result of the Bayesian factor analysis is that, the four variables are the most significant agent of sprawl and should be looked into economically, of which built-up density is the highest one with the most significant impact, which is the intensity of use of land areas is the major call of concern.

5 References

Akanbi Olawale Basheer (2018). Modelling non Evacuation of Waste Bin in North Western Part of Nigeria using Bayesian Approach. International Journal of Applied Sciences and Mathematics; 5(3):16-21.

Akanbi Olawale Basheer (2022).Spatial Analysis of Soil Radon Gas Concentration in Southwestern Nigeria: A Bayesian Approach. International Journal of Applied Science and Mathematics; 9(3)36–46.

Akanbi, O. B., and Nurudeen, T. S. (2024). "Determination of Heavy Metal Concentrations at Ewu-Elepe,Ikorodu Dumpsite, Lagos, Nigeria". Asian Journal of Environment & Ecology 23 (10):49-58.https://doi.org/10.9734/ajee/2024/v23i10608.

Akanbi O. B., and Oladoja O. M. (2019). Application of a Modified g-Parameter Prior in Bayesian Model Averaging to CO_2 Emissions in NigeriaMathematical Theory and Modeling; 9(11):57-71.

Olubi, Abiodun Rufus (2019). The Impacts Of Land Use on Urban Aesthetics: A Case of Ojoo, Ibadan, Nigeria" International Conference on Architecture and Urbanism.

Akinrinola,Akintunde John (2019). Spatial pattern of urban growth using remote sensing and landscape metrics. Journal of Geomatics, Vol. 13, No. 1, April 2019.

Chettry, V.,(2023).A Critical Review of Urban Sprawl Studies. *J geovis spat anal* **7**, 28 https://doi.org/10.1007/s41651-023-00158-w

Dada, J., Ajani, A., Alagbe, A., Fetuga, I., & Ilesanmi, O. (2023). Assessment of the impact of urban sprawl on indigenous agriculture land in Ibadan between 2009 and 2019: A case study of Akinyele lga. *ITEGAM-JETIA*, *9*(42), 61-76. https://doi.org/10.5935/jetia.v9i42.880.

Fusco,Giovanni, andTettamanzi, Andrea G. B. (2017). Multiple Bayesian Models for the Sustainable City: The Case of Urban Sprawl. 17th International Conference on Computational Science and Its Applications, Trieste, Italy. pp.392-407, ff10.1007/978-3-319-62401-3_29ff. ffhal-01567726.

Kasim, Oluwashinaayomi Faith, Agbola, Samuel Babatunde, and Coker, Mosunmola Olufunmilayo (2014). Dynamics of Land Use and Land Cover Change in Ibadan Region, Nigeria.

https://mc-stan.org/docs/reference-manual/mcmc.html#hamiltonian-monte-carlo
https://online.stat.psu.edu/stat505/

Tagnan, Jacob Nchagmado, Amponsah, Owusu, Takyi, Stephen Appiah, Azunre, Gideon Abagna, Braimah, Imoro, (2022). A view of urban sprawl through the lens of family nuclearisation, Habitat International, Volume 123, 2022, 102555, ISSN 0197-3975, https://doi.org/10.1016/j.habitatint.2022.102555.

https://www.sciencedirect.com/science/article/pii/S016920461730052X Manuscript_427b22531486e808727e0db24df95e3d

Luan, H., and Fuller, D. (2022). Urban form in Canada at a small-area level: Quantifying "compactness" and "sprawl" with bayesian multivariate spatial factor analysis. Environment and Planning B: Urban Analytics and City Science, 49(4), 1300-1313. https://doi.org/10.1177/23998083211062901

Bosch Martí, ChenalJérôme, and Joost Stéphane (2019). Addressing Urban Sprawl from the Complexity Sciences. Urban Sci., 3, 60; doi:10.3390/urbansci3020060

Enoh, Mfoniso Asuquo, Njoku, Richard Ebere, Okeke, Uzoma Chinenye, (2023). Modeling and mapping the spatial–temporal changes in land use and land cover in Lagos: A dynamics for building a sustainable urban city, Advances in Space Research, Volume 72, Issue 3, Pages 694-710, ISSN 0273-1177, https://doi.org/10.1016/j.asr.2022.07.042.

Mourguiart, Bastien (2022). Bayesian modelling of species-environment relationships for partially observed data.. General Mathematics [math.GM]. Université de Pau et des Pays de l'Adour, English. ffNNT : 2022PAUU3023ff. fftel-04085564f.

Ohwo, Odafivwotu, and Abel Abotutu. (2015). "Environmental Impact of Urbanization in Nigeria". Current Journal of Applied Science and Technology 9 (3):212-21. https://doi.org/10.9734/BJAST/2015/18148

Otokiti K. V., Collins A. A., Famewo A.S., & Raheem R. T. (2021). Monitoring peri-urbanization in Ibadan region: case study of Akinyele, Egbeda, Ido and Oluyole local government areas. Discovery, 2021, 57(301), 74-85.

Owolabi, Babatunde Oluwaseyi, (2018). " Assessment of Urban Sprawl Occurrence in Oyo State, Nigeria", International Journal of Research Studies in Science, Engineering and Technology, vol. 5, no. 12, pp. 23-50.

Rafferty, John P. (2024). "Urban sprawl". Encyclopedia Britannica.https://www.britannica.com/topic/urban-sprawl. Accessed 7 May 2024.

Salam, R.D., Oluwatimilehin, I.A. andAyanlade, A., (2023). Spatial analysis of urban expansion, land-use dynamics and its effects on land surface temperature in Oyo town, Southwestern Nigeria. *City Built Enviro* 1, 15. https://doi.org/10.1007/s44213-023-00017-w

Taiwo, O.J. (2022). Modelling the spatiotemporal patterns of urban sprawl in Ibadan metropolis between 1984 and 2013 in Nigeria. *Model. Earth Syst. Environ.* **8**, 121–140. https://doi.org/10.1007/s40808-021-01095-7