Modelling sample proportion of underfive stunted children in Nigeria

Olopha Paul Omoh
Department of Statistics, Federal University of Technology, Akure

Abstract

Many epidemiological processes are measured as discrete or continuous and are spatially sampled. One of such processes is under-five child stunting where stunting is measured as z-score transformation of the age to height ratio. Stunting is then defined as HAZ<-2SD. Most analysis are carried out assuming a binomial distribution where a logistic model is fitted. However, such binomial data is prone to over-dispersion, which results in an incorrect quantification of the uncertainty when applying the binomial generalized linear model (GLM).Wilson and Hardy, 2002 proposed that the inclusion of a random intercept term using generalized linear mixed model may improve the assessment of uncertainty. In this study however, beta regression with the inclusion of a spatially correlated term within the Bayesian framework was used to analyse the spatial pattern and determinants of under-five stunting in Nigeria.

Keywords: Modelling proportion, Beta regression, Stunting

1. Introduction

One health outcome established by various researchers to be a major cause of infant mortality in low and middle income countries particularly in sub-Saharan Africa is under-five child stunting(WHO, 2010; UNICEF, 2018). According to UNICEF 2019, when it interacts with infections, it is very lethal and is the cause of about half of the deaths among children worldwide, and about 37% of such deaths in Nigeria alone. The major cause of child stunting (low height for age) has been commonly traced to malnutrition or under-nutrition in the first 1,000 days of a child, after which the after effect becomes very difficult to correct through-out the child's life-time. Stunting, which has been said to be the best indicator of overall well-being among children, is associated with poor health, impaired cognitive function, poor school performance and limited economic productivity (Puntis, 2010; De Onis and Branca, 2016).

Child malnutrition has been identified to be related to lack of quality and quantity food intake, and this has been traced partly to the lack of mothers' knowledge of the right type of food intake necessary for child growth and development, aside economic ability (Adedeji et al, 2018; Hamel et al, 2015). But educating pregnant women on issues of child growth and development is one of the cardinal objectives of antenatal care services (UNICEF, 2019). Various researchers have however established the poor attitude of women in the developing world to antenatal care utilization. For example in Nigeria, only 61% of women is reported to attend antenatal care at least once (ANC1+), while a worse figure i.e. 57% of Nigerian women made the World Health Organisation (WHO) recommended minimum number of at least four antenatal visits (ANC4+) (WHO, 2002).This is known as "Focused or Basic antenatal care" (FANC).The focused ANC (FANC) model, also known as the basic ANC model, includes four ANC visits occurring between 8 and 12 weeks of gestation, between 24 and 26 weeks, at 32 weeks, and between 36 and 38 weeks. It is expected that a woman who completes at least four antenatal visits would have had the basic knowledge required for her child's growth and development, feeding inclusive, and thereby avoiding under-five child stunting, among other benefits (WHO).

Some studies have reported a linkage between maternal care during pregnancy and childhood malnutrition. These include Yimmer (2000) who reported in a study carried out in southern Ethiopia that an inverse relationship was found to exist between the number of antenatal visits a mother had during pregnancy and stunting in the child. In a related study, Hammel et al (2015) used multivariate analysis and crossvalidation techniques to analyse pregnant women data from two states in Nigeria, and provided evidence of a link between poor care during pregnancy and childhood malnutrition in the two states. Nigeria currently ranks 2nd in the world in infant death due to stunting.

Stunting of underfive children is one of the many epidemiological processes that are spatially sampled and measured as binary data. The traditional approach in epidemiology has been to first generate Height for age z-score (HAZ). Stunting is then defined as HAZ< -2SD. Specifically, a child is said to be stunted if the HAZ is less than -2SD and not stunted otherwise. The distribution commonly used to analyse such data is therefore the binomial distribution. The model commonly adopted is the logistic regression, an analytical method designed to deal with binomial proportional data [Steel et al., 1997, Wilson and Hardy, 2002, Warton and Hui, 2011, Paradini et al, 2018], i.e. proportions measured as x out of n. The logistic regression provides an interpretative analysis and is not sensitive to sample size. Nonetheless, such binomial data is prone to overdispersion, resulting in an incorrect quantification of the uncertainty when applying the proposed binomial generalised linear model (GLM)(Paradini et al, 2018). In these cases, the inclusion of a random intercept term using generalised linear mixed models (GLMMs) may improve the assessment of uncertainty [Wilson and Hardy, 2002]. When data are non-binomial, that is, observations do not follow the x out of n pattern, as in the raw z-scores, the logistic regression is no longer applicable. As an alternative approach, Warton and Hui [2011] suggested the logit transformation of the data, which overcomes the problems of interpretability and range of the confidence/credible intervals using the arcsine square root transformation. However, any transformation of the data (y_t) implies that regression parameters are only interpretable in terms of the transformed mean of y_t and not the mean of the original data. In this study, the beta distribution was applied to the proportion data (x out of n) derived from the number of U5 stunted children out of total number of U5 children residing in an area or community. The beta distribution is appropriate as it satisfies the characteristics of proportions, bounded to the [0, 1] interval with asymmetric shapes. The use of Beta distribution in various applications involving proportions, probabilities and linear regression modeling is not new among many researchers (Gupta and Nadarajah, 2004). One major advantage over the transformation approach is that it allows bounded estimates and intervals with model parameters that are directly interpretable in terms of the mean of the response (Paradini et al, 2018). Researches have also shown that stunting among underfive children are driven by a set of factors and interactions. Understanding these drivers is very often the ultimate goal among scientists seeking to understanding the dynamics of underfive child stunting for decision making. This is with the intent of proper and effective intervention by appropriate agencies for distribution of scarce resources. However, most of the times, the spatial variability of the data has been found to exceed the variability explained by the explanatory variables due to the increasingly complexity of epidemiological spatial processes. This phenomenon usually results in spatially autocorrelated model residuals that can yield incorrect results and a restricted predictive capacity of the models [Paradini et al, 2018, Fortin and Dale, 2009, Legendre et al., 2002]. Consequently, the need to introduce spatial terms in our models cannot be over-emphasised. This is to improve model fit and prediction. Spatial terms are based on the principle that close observations have more in common than distant observations [Rue et al., 2009]. The introduction of spatial terms is capable of improving predictions and identifying hidden spatial hot and/or cold spots that may be important for intervention purposes. In addition, it is crucial to address the uncertainty associated with our predictions and estimates. The Bayesian hierarchical approach is one of the most commonly used because of its ability to improve on the estimates through the use of prior distributions. The remainder of this article goes as follows. First, we summarise the characteristics of the hierarchical spatial beta regression. The Integrated Nested Laplace Approximation (INLA) method and the Stochastic Partial Differential Equations (SPDE) approach were utilized as an effective way to deal with the spatially sampled proportional data.

2. Methodology

2.1 Data

The data used in this study involved a cross-section of children U5 years extracted from the Nigerian Demographic Health Survey (NDHS) data conducted in the year 2013. The survey was conducted by the Demographic Health Survey (DHS) and contains a total of 904 communities or localities into which Nigeria was divided. In the selected households, 23,220 women in the childbearing age (15 – 49years) were found eligible for the interview. Using a two stage sampling approach, the women were sampled from the communities or localities across the 37 states of Nigeria. In this study, focus was on proportion of children in a locality who had chronic malnutrition (insufficient height for age) referred to as stunting. A response variable of U5 child stunting proportion was created as the fraction of U5 stunted children of the total number of children sampled in the community. The variable was assumed as propensity for a child to be stunted in the community, while the proportion of mothers in the community who attended ANC at least four times was assumed to be the disposition of mothers to complete ANC4+ in the community. The use of proportion instead of the binary observations created from the computed z-score was to avoid the ambiguity that may arise due to the varying sample sizes of the localities which may introduce some bias into the estimates. The use of proportions is independent of sample size and over-dispersion. It could therefore be regarded as introducing weights to the responses in the communities (Paradini et al, 2018). Furthermore, the use of proportion instead of total stunted underfive children will be an indicator of whether or not U5 child stunting is disproportionate to the population of underfive children in the community, a proxy of number of households with underfive children.

2.2 Modelling underfive children stunting proportion per locality

Variable	Description	Effect
Total population	Total population of underfive children in each locality	Linear
Location	Geolocation in longitude and latitude	Geo-statistical effect
Proportion of ANC4+	Proportion of women who attended ANC at least four times in the locality	Non-linear
Locality/community	Locality ID	Random noise effect

Table 1: List of covariates included in the analysis and the effect assigned to them.

The analysis of stunted U5 child proportions in the locality included the total population of children in each locality, the proportion of women who attended ANC at least four times in the locality, a geostatistical term and a community effect as predictors (Table 1)

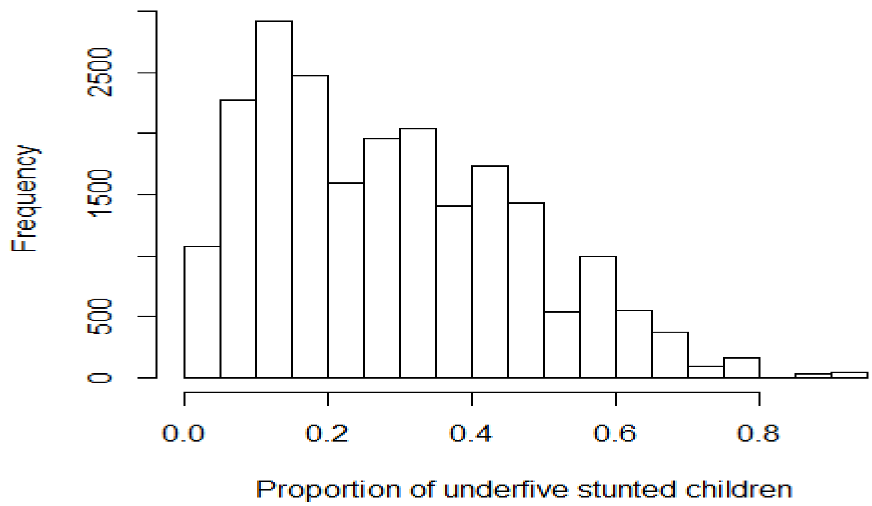

Fig. 1 The histogram of the proportion of underfive stunted children

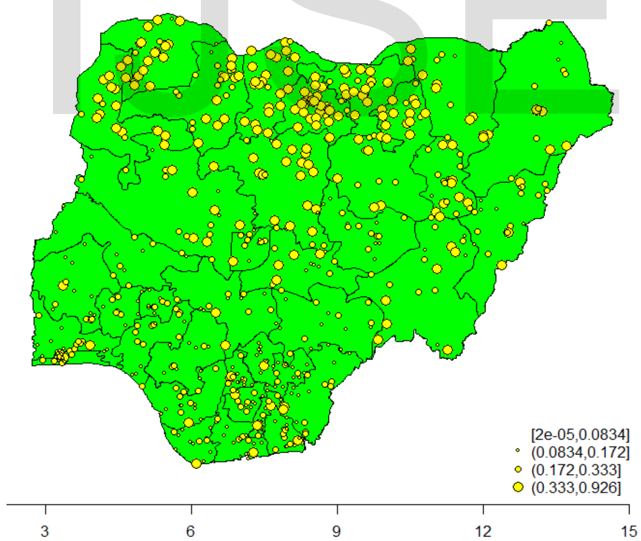

Fig. 2: The map showing the localities together with the proportion of under-five stunted children.

The map in Fig. 2 shows that the northern part of Nigeria is fraught with localities that has high proportion of underfive children that are stunted. This is especially true of Kebbi, Sokoto, Zamfara, Katsina, Kano, Kaduna, Jigawa, Bauchi, and Gombe. Yobe and Borno seem to have very few localities with high proportion of stunting. This might be due to the ravaging militant activities going on there as at the time of survey which might have grossly limited survey activites in those areas. The southern part is populated with localites with very low proportion of stunted underfive children .

Following after Paradini et al, 2018, we assume that the underfive child stunted proportion Y_i at location i follows a beta distribution, the final model can be expressed as:

$$Y_i \sim \mathrm{Be}(\mu_i , \emptyset_i) , \quad i = 1 \cdots n$$

$\mathrm{Logit}\ (\mu_i) = \beta m_i + c_i + W_i$

$\beta \sim \mathrm{N}(0, 0.001)$

$\Delta^2 c_j = c_j - c_{j+1} + c_{j+2} \sim \mathrm{N}\ (0, \rho_c), \mathrm{j} = 1, \cdots, \mathrm{k}$

$\log\rho_c \sim \mathrm{LogGamma}(0.1, 0.00001)$

$\mathrm{W} \sim \mathrm{N}(0, \mathrm{Q}(k, \tau)$

$2\log k \sim \mathrm{N}(\mu_k, \rho_k)$

$\mathrm{Log}\ \tau \sim \mathrm{N}(\mu_\tau, \rho_\tau)$

Where the mean of U5 child stunting proportion enters the model through the logit link, i indexes the location of each population of underfive children and j indexes the different locality or community. In the last two rows μ stands for the mean of the normal distributions while ρ denotes its corresponding precision. Assuming the proportions of U5 stunted children are not fully proportional to the total population of children in the locality, total population was introduced as a linear effect with a vague normal prior distribution as implemented by default in R-INLA. The exploratory analysis revealed non-linear relationships between ANC4+ proportion and U5 stunting proportion, so a second order random walk (RW2) latent model was applied based on constant increase in the stunting proportion. These RW2 models perform as Bayesian smoothing splines (Fahrmeir and Lang, 2001) and can be expressed as a computationally efficient GMRF [Rue and Held, 2005], and are therefore applicable in INLA.

The two dimensional geostatistical latent model W, used to identify the hot spots, depends on two hyperparameters k and T that define the variance and the range of the spatial effect. The smoothing parameter of the Matern fixed (v=1), the range of the spatial terms is approximately $\sqrt{8}/k$ and the variance $1/(4\pi k^2\tau^2)$. The priors for k and τ are specified over the logτ and 2logk.

2.3 Model diagnostics

	DIC	CPO
M1=Spatial effects only	-448.50	5876.60
M2=Spatial + covariates	-534.32	12161.61
M3=Spatial + covariates(linear and nonlinear) + community (random)	-616.93	16483.73

Model diagnostics was based on deviance information criterion (DIC). The model with the lowest DIC value was regarded as the best fit (Spiegelhalter et al., 2002), and was consequently discussed. This model was the full model M3 where spatial, linear, non-linear and random covariates were controlled. The significance of the lower DIC value when M2 was adjusted for community suggests that there exists a strong community effect or influence in the proportion of stunted underfive children in Nigeria.

3. Results

Variable	mean	sd	0.025quant	0.975quant
Intercept	0.6699	0.3457	-0.0059	1.3508
ANC4+ proportion	-3.2521	0.2833	-3.8039	-2.6926
Population of women	0.0257	0.0256	-0.0247	0.0757
τ	0.1861	0.0191	0.1527	0.2274
σ^2	0.8378	1.3249	0.0462	4.2550
range	0.4651	0.3628	0.0624	1.4158

As expected, the effect of total population of women in a community on the proportion of stunted underfive children in the community was positive but not significant i.e. the propensity of an underfive child getting stunted in the community is not sufficiently influenced by the population of women in that community. However, the proportion of women who attended ANC at least four times (ANC4+ proportion) during pregnancy had a strong negative relationship with proportion of underfive children getting stunted suggesting that a high proportion of stunted underfive children are located in areas where pregnant women rarely attended antenatal care up to four times, and increases as the proportion of mothers who attended ANC at least four times decreases.

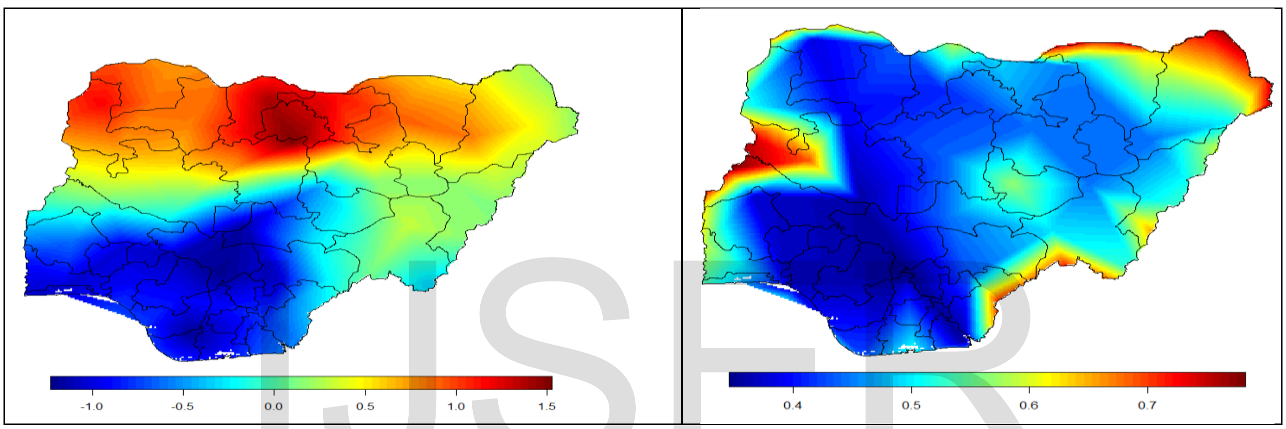

Figure 3: Posterior predictive mean and standard deviation maps of the spatial component of proportions of underfive stunted children

The posterior mean and standard deviation of the spatial component, which represents the intrinsic spatial variability of the data without the rest of the independent variables. This effect highlights in (red), high U5 stunting proportion areas or hot spots while cold spots (blue) highlights low U5 stunting proportion areas or cold spots. The result shows that there is high propensity of underfive child stunting in the core northern Nigeria with Kano state and eastern side of Katsina states as hot spots, and require urgent intervention by appropriate agencies. The middle belt shows average propensity for stunting while the south shows low propensity for underfive child stunting. Figure (b.) reveals low spatial variability in the distribution of underfive child stunting in northern Nigeria but a much lower variability in the South.

4. Predictive performance test for the model

With transformation into the natural scale, the root mean square (RMSE) was computed to be 1.8569 while the correlation between the observations and the predicted values gave 0.70, meaning that there exists 70% performance in the predictive ability of the model. The predictability of the spatial model was also validated using the cross-validation approach.

5. Prediction of the response variable at the locations

The map of the predicted proportion values at the locations is shown in Fig. 1.

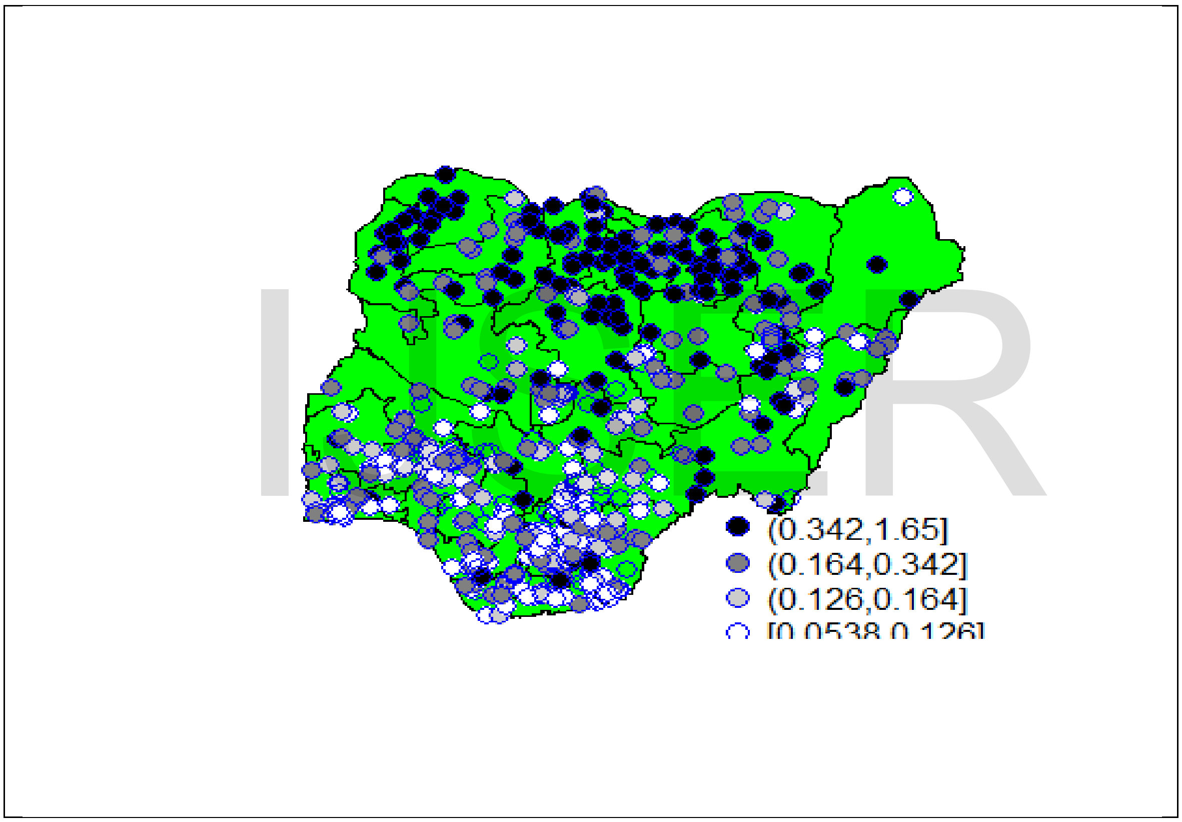

Fig. 1: Map showing the distribution of the predicted response (proportion of stunted underfive children)

Discussion

One health outcome established by various researchers to be a major cause of infant mortality in low and middle income countries particularly in sub-Saharan Africa is under-five child stunting. According to UNICEF 2019, when it interacts with infections, it is very lethal and is the cause of about half of the deaths among children worldwide, and about 37% of such deaths in Nigeria alone. The major cause of child stunting (low height for age) has been commonly traced to malnutrition or under-nutrition in the first 1,000 days of a child, after which the after effect becomes very difficult to correct through-out the child's life-time. Stunting, which has been said to be the best indicator of overall well-being among children, is associated with poor health, impaired cognitive function, poor school performance and limited economic productivity (De Onis and Branca, 2016).

Child malnutrition has been identified to be related to lack of quality and quantity food intake, and this has been traced partly to the lack of mothers' knowledge of the right type of food intake necessary for child growth and development, aside economic ability. It is expected that a woman who completes at least four antenatal visits during pregnancy would have acquired the basic knowledge required for her child's growth and development, feeding inclusive, and thereby avoiding under-five child stunting. However, research has shown that only 57% of Nigerian women made the World Health Organisation (WHO) recommended minimum number of at least four ANC visits. This study used Bayesian SPDE approach to generate continuous spatial maps of underfive child stunting while accounting for the influence of mothers' ANC attendance at least four time during pregnancy. Our results confirmed that stunting among children under five years old varied spatially with higher concentrations in locations where pregnant women are less likely to attend ANC at least four times. High propensity for underfive child stunting was shown in the core northern Nigeria with Kano state and eastern side of Katsina state becoming hot spots and require urgent intervention. The south-south region is also shown to have localities with high propensity for underfive child stunting. These areas with high propensity for stunting are noted for women with low economic capability as well as poor educational background (Ogbe, 2020; Aja-Okorie, 2013), both of which have been found to be strong determinants of stunting among children especially those who are under the age of five. Results further show that propensity of underfive child stunting continues to decrease as one goes down south until when one gets to the south-south where the contrary is the case. This pattern tends to authenticate our result that shows a negative significant relationship between the proportion of underfive stunted children in the community and the proportion of mothers who attended ANC at least four times during pregnancy.

The following limitations are identified in this study. The analysis was carried out on area basis. The individual effect was not considered. Also, other mitigating factors such as environmental and socio-economic factors could be considered in future studies.

5. Conclusion

The use of Bayesian hierarchical spatial beta model to analyse the spatial sampled proportion data has allowed the direct interpretation of model parameters. Also, the analysis is not sensitive to sample size; and lastly, posterior distributions are expected to concentrate well within the bounded range of proportions.

The SPDE approach has been used to study the spatial distribution of the propensity of underfive child stunting in Nigeria. It has allowed us to quantify the probability at every point of the spatial domain and thus allowing the proper view of the impact at every location including places where data were not collected.

Conflict of interest

The author declares no conflict of interest

References

Ogbe, A. (2020). The Travails of Women and Girls in Northern Nigeria: A Human Rights Issue. SSRN: https://ssrn.com/abstract=3599122 or http://dx.doi.org/10.2139/ssrn.3599122

Aja-Okorie, U. (2013). Women Education in Nigeria: Problems and Implications for Family role and Stability. European Scientific Journal. 9(28).

Delaunay, O.B. (1934) Sur la sphere vide. Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk, 7(793-800):1–2.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. Applied statistics, pages 201– 348 .

Fortin, M.J. and Dale, M. R. T. (2009). Spatial autocorrelation in ecological studies: a legacy of solutions and myths. Geographical Analysis, 41(4): 392–397.

A. K. Gupta and S. Nadarajah. Handbook of beta distribution and its applications.

Legendre, P., Dale, M. R. T., Fortin, M. J., Gurevitch, J., Hohn, M. and Myers, D. (2002). The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography, 25(5):601–615.

Warton, D.I. and Hui, F.K.C.. (2011).The arcsine is asinine: the analysis of proportions 428 in ecology. Ecology, 92(1):3–10.

Wilson, K. and Hardy, I.C.W. (2002). Statistical analysis of sex ratios: an introduction. Cambridge University Press.

Steel, R.G.D., Torrie, J.H. and Dickey, D.A. Principles and procedures of statistics: A biological approach. McGraw-Hill, 1997.

Yimmer, G. (2000). Malnutrition among children in Southern Ethiopia: levels and risk factors. Ethiopian Journal of Health Development. 14: 283-292

Hamel C., Enne J., Omer K., Ayara N., Yarima Y., Cockcroft A., and Andersson N. (2015). Childhood Malnutrition is Associated with Maternal Care During Pregnancy and Childbirth: A Cross-Sectional Study in Bauchi and Cross River States, Nigeria, J Public Health Res. , 4(1) 408.

Paradinas, L.,Pennino, M.G., Gonzalez, M.M., Lopez-Quilez, A., Bellido, J.M., Conesa, D. (2018). Modelling spatially sampled proportion processes. Revstat Statistical Journal 16 (1): 71-86

Rue,H., Martino, S. and Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the royal statistical society: Series b (statistical methodology), 71(2):319– 405, 2009.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A, (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B (Methodological), 64:583-639

Rue, H. and Held, L. Gaussian Markov random fields: theory and applications. CRC Press, 2005.

Adedeji, I.A., Bashir, M.F., Shwe, D.D. and John, C. (2018). Prevalence and Correlates of Stunting among the School age population in North Central Nigeria. Pan African Medical Journal. 31: 170