International Journal of Mathematics Trends and Technology
ISSN: 2231 - 5373/doi:10.14445/22315373/IJMTT-V68I1P515
Volume 68 Issue 1, 143-157, January, 2022

A Geo-Classification Model for Mapping Mixed Discrete and Continuous Response Data: An Application to Poverty Mapping

Mark Adjei[1,⋆], Elphas Okango[2], Richard Puurbalanta[3], Henry Mwambi[4], Naiga Babra Charlotte [5]

[1,5] Department of Mathematics, Pan African University, Institute for Basic Science, Technology and Innovation, Nairobi, Kenya

[2] Department of Mathematics, Strathmore University, Nairobi, Kenya

[3] Department of Statistics, C.K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

[4] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Abstract

Welfare data for monitoring poverty are usually gathered over a wide geographical area, and as such proximal observations are more likely to be affected by common environmental elements and therefore share similar characteristics than distant observations. This is known as spatial dependence. However, poverty analysts have largely ignored this spatial property in welfare data. This work seeks to quantify relationships between poverty-severity and potential covariates while accounting for spatial dependence using a geo-classification model. The source of data for this study, is the seventh round of the Ghana Living Standards Survey (GLSS). We asserted that social and economic characteristics which are bounded in socially constructed spaces affect poverty-generating process. To investigate the interactive association, we use a statistical regime which has the benefit of parsimoniously analyzing all location-specific circumstances simultaneously, thus yielding a broad view of the processes generating poverty in Ghana. Bayesian estimation was adopted in our model computation. This was due to the hierarchical and highly parameterized nature of our model. Evident from our preliminary results, spatial effect and variation is empirical in the GLSS 7 data and cannot be ignored in the bid to understand poverty and its correlates in the study region. In general, the posterior means and 95% credible intervals show that fixed effect estimates (household size, income level of householder, ecological zone and location/area of residence) and spatial effects significantly influence poverty levels and distribution patterns in Ghana.

Keywords: Household Expenditure, Posterior Densities, Poverty-Severity, Kriging, Spatial Variation.

1. Introduction

More recently, new prospects of measuring poverty have been proposed to enrich the understanding of poverty construct and to better reflect the evolving concept of poverty [1, 2, 3, 6, 9, 28]. Poverty is identified as a condition in which an individual or household is unable to sustain socially acceptable standard of living due to a lack of necessary funds and resources for food, housing, clothing, as well as the absence of basic infrastructures such as roads, transportation, water, health, proper sanitation, and lack of basic education [12]. Poverty has been conceptualized in a variety of ways over the years, but the general opinion is that it is multidimensional, encompassing both monetary and non-monetary aspects, and that it is non-static. Despite these understanding, poverty remains an issue in the majority of developing countries. The world bank argued that Sub-Sahara Africa has over the years lagged in the fight against poverty [30]. For example, it is reported that the absolute poverty index in Africa increased from 278 million in 1990 to 710 million in 2020 and the majority of the world's poor live in Sub-Saharan Africa as of 2020 [30].

Poverty is becoming overly rooted in many sub-Sahara African countries including Ghana which is experiencing worsening poverty incidence rates. Half of Ghana's 10 regions have their poverty rates high above the national average of 24.5%. Accordingly, more Ghanaians are becoming extremely poor as the number of people living in extreme poverty increased from 2.2 million in 2013 to 2.4 million in 2021 [6].

Moreover, in poverty estimation, there are salient issues that call for consideration. That is, establishing an indicator of welfare and a minimum acceptable standard of that indicator to separate the poor and the non-poor. This is called the poverty line. The international poverty line is fixed at $1.90 per day [30]. When estimating poverty, scholars have a choice between using income or expenditure as the indicator of welfare. Some researchers argue that the later may better show poverty, since income may be erratic during the year [14]. In order to compare different household sizes and composition, it is highly recommended to utilize equivalent scale of household expenditure. Poverty and vulnerability analysis in Ghana has been driven by high poverty rates and low per capita incomes, as well as increased vulnerability to multiple shocks. Government in Ghana is investing more in poverty monitoring through welfare monitoring surveys called the Ghana Living Standards Survey in order to inform policy decisions and poverty reduction strategies. Consequently, these policy proponents rely heavily on results and predictions of small area poverty estimates. In particular, varying statistical methods have been developed and applied over the past few years that generate comparable poverty outcomes in terms of poverty incidence as measures of consumption expenditures.

Currently, there are two main ways to modeling poverty and its correlates particularly in Ghana: using consumption expenditure per adult equivalent and regressing it against potential covariates. The other way is to use household income [12]. The common drawback to these approaches is that, since data is collected over a large geographical location, proximal observations are more likely to be influenced by common environmental factors and therefore have similar characteristics when compared to distant observations. This causes spatial dependence, which means that global parameters are not enough to describe the site-specific poverty conditions. It is well-known that when covariates are spatially dependent, coefficient estimates can be biased and variances inflated when the traditional Generalized Linear Models (GLM) are employed. To the best of our knowledge, the effect of spatial confounding on household poverty-severity estimation in Ghana has not been investigated.

A work by [25] introduced spatial dependence into the model residual structure, however, he decomposed poverty into a multi-category ordered random variable and described the overall state of a household's poverty-severity, using multivariate ordered regression. The drawback is that, too much categorization of all continuous variables leads to a loss of information [24] and the scaling of all categorical variables introduces an unknown degree of subjectivity. Also, when the continuous expenditure variable is overly censored, modeling becomes much more difficult and requires complex inferential methods to successfully produce parameter estimates. Additionally, non-technical end users of the model are unable to apply model outcomes in policy design. That is, adding more categories to the underlying latent scale might simply serve theoretical purposes only, but not for practical interpretation.

The assessment of human well-being is complex in and of itself, and raises many methodological issues. Over the years, poverty has been measured, analysed, and described by varying techniques with varying degrees of sophistication. Aboagye (2019) examined the socioeconomic factors that correlate with poverty in Ghana using the sixth and seventh rounds of the Ghana Living Standard Survey dataset.The study used binomial probit regression to analyse the degree of correlation of poverty with selected socioeconomic variables. Also, Mahama et al. (2018) investigated the correlates of poverty in the Northern Region of Ghana. The study used data from 1,702 households from the sixth round of the Ghana Living Standard Survey (GLSS6). The study used households per capita consumption as the response variable and employed Ordinary least Squares regression to estimate the correlates of poverty. A large number of methods have been developed to estimate the poverty indicators generally, such as unit level models [15], empirical best method [23], temporal and spatiotemporal area level models [22], hierarchical Bayes estimation [22], unit level logit mixed model [20], M-quantile model [8] and others.

In this work, we assert that these social and economic characteristics which are bounded in socially constructed spaces affect the poverty-generating process. Because the factors are not evenly distributed across the nation, the extent of poverty is not evenly distributed; hence, spatial differentiation in poverty can be understood by differences in the underlying distribution of factors generating the distribution. To investigate the interactive association, we use a statistical regime which has the benefit of parsimoniously analyzing all location-specific circumstances simultaneously, thus yielding a broad view of the processes generating poverty in Ghana.

The main objective of this study is to geo-classify and map household poverty using mixed discrete and continuous response data. Unlike previous studies we employ a censoring mechanism with cut points that discretize the real line $(-\infty, +\infty)$ that underlie the continuous expenditure variable to produce the categorical scale with little or no loss of continuity in the expenditure variable. This way, we estimated the fixed-effects and spatial parameters associated with the differences in poverty-risk in Ghana using Bayesian inference and developed a prediction map of poverty distribution in Ghana identifying poverty "hotspots".

2. Methodology

2.1. Data

Data used in this study is from the seventh round of the Ghana Living Standards Survey (GLSS 7). The GLSS is a multipurpose household survey that collects a wealth of information to determine Ghanaians' living conditions. The survey offers useful information on Ghanaian households' socioeconomic characteristics. It also offers evidence for tracking progress against national policies and international commitments, such as the UN Sustainable Development Goals (SDGs). The GLSS 7 dataset contained the following variables: Details of households (household size, age, sex, and educational level of household head), socio-economic conditions of household (consumption expenditure, and availability of social amenities), and neighbourhood related spatial information (ecological zone, rural/urban residence, and other environmental factors). The GLSS 7 survey is nationally representative with a sample size of 15,000 households selected from 1,000 enumeration areas. In all 14,009 households responded to the survey, resulting in a response rate of 93.3% [19].

2.2. The Probit Regression Framework

In this section, we motivate our model by first appealing to the probit regression framework by illustrating latent variable representation in context. Let Y be a binary response variable and X a vector of regressors which are assumed to influence the outcome Y. Suppose that there is an unobserved or latent variable Y^* on the interval $(-\infty, \infty)$ that generates the observed Y. The probit model is specified as:

$$P(Y_i = 1) = \Phi\left(\sum_{i=1}^{n} \beta X_i\right) \tag{1}$$

$$= \int_{-\infty}^{\sum_{i=1}^{n} \beta X_i} exp(-\frac{\mu^2}{2})/\sqrt{2\pi} d\mu \tag{2}$$

where Φ represents the standard normal cumulative distribution. The probit model is motivated as a latent variable model [7] and has representation using latent variables given as:

$$Y_i^* = X_i\beta + \varepsilon_i \tag{3}$$

$$\varepsilon_i \sim N(0, 1)$$

Y^* is linked to the observed binary variable Y with the measurement equation given as:

$$Y_i = \begin{cases} 1 & \text{if,} \quad Y_i^* > \alpha \\ 0 & \text{if,} \quad Y_i^* \leq \alpha \end{cases} \tag{4}$$

Where α is the threshold or cut point.

2.3. Model Specification and Estimation

For the purpose of this study, we let $\{(Y_i, x_i);\ i = 1, \cdots, n\}$ be paired observations at locations $s_i = \{s_1, s_2, \cdots, s_n\} \in S \subset \mathbb{R}^2$ in the study region D, S being a continuous subspace of $\mathbb{R}^2$. Define Y_i as n spatially-dependent multivariate binary response variable.

Then, for the geo-classification Model, we introduce latent variables $Y_i^* = (Y_1^*, \cdots, Y_n^*)$, which are realizations of a Gaussian spatial process, and assigning values to y_i according to a regression function:

$$Y_i^* = \Phi(X_i'\beta) + (\eta_i)_s \tag{5}$$

$$\eta_i = \varepsilon_i + \mu_i \tag{6}$$

$$\varepsilon_i \sim N(0, \sigma^2 I_n)$$

$$\mu_i \sim N[0, \Sigma]$$

Where X_i' are $n \times p$ dimensional covariates, β is a $p \times 1$ matrix of fixed effect coefficients and μ_i a $n \times 1$ spatially-dependent residual that capture all unobserved errors arising from the influence of common features for observations within certain proximal distances. Our aim is to model the likelihood of a household (y_j) at location s_i being in a particular poverty-severity category.

Though the values of y^* cannot be directly observed, the rule that assigns y^* to y_i is that if y^* exceeds the threshold value of zero, then, for instance, a household falls in the non-poor category of poverty. The conditional likelihood of the observed data given the unobserved latent variable and the underlying model's parameters is

$$Y_i = \begin{cases} 1 & \text{if,} \quad Y_i^* > 0 \\ 0 & \text{if,} \quad Y_i^* \leq 0 \end{cases} \tag{7}$$

We assume that the marginal variance of the Y_i are known up to a multiplicative constant σ^2 and that the matrix $\Sigma(\lambda)$ captures the spatial dependencies between sites in the data. By assuming that the diagonal elements of $\Sigma(\lambda)$ are fixed and known constants, we allow for the possibility of heteroskedasticity. We will apply σ^2 as the variance of Y_i^* and $\Sigma(\lambda)$ as a spatial correlation matrix in our model.

2.4. Spatial Correlation Matrix Parameterization

Spatial correlation may be defined as the existence of a functional relationship between what happens at one point in space and what happens elsewhere [11]. Mathematically, spatial correlation implies that an observation associated with a location in space labelled s_i depends on other observations at locations $s_j \neq s_i$, that is

$$Cov\{(y_{is}), (y_{js})\} = E\{(y_{is})(y_{js})\} - E(y_{is}) \times E(y_{js}) \neq 0, \quad fori \neq j \tag{8}$$

It is critical to ensure that $\Sigma(\lambda)$ is a valid correlation matrix when modeling spatial dependency in a Bayesian setup. The correlation of a Gaussian spatial process at sites i and j is modeled as a map of the distance between the two locations corresponding to the assumption of second-order stationarity and isotropy. In this study, we employ the Matérn, a class of parametric spatial correlation function:

$$\Sigma(\lambda) = \frac{1}{2^{w-1}\Gamma(w)} \left(\frac{2\sqrt{w} d_{ij}}{\lambda} \right)^w B_w \left(\frac{2\sqrt{w} d_{ij}}{\lambda} \right) \tag{9}$$

The covariance matrix $\sigma^2\Sigma(\lambda) = Cov\{(y_i)_s, (y_j)_s\}$, is parametrized by the Matérn correlation function. The decay parameter λ, measures the strength of spatial dependence over the Euclidean distance $d_{ij} = \|s_i - s_j\| = \left(\sum_{i=1}^{2} (s_i - s_j)^2 \right)^{\frac{1}{2}}$ between sites i and j, using the longitude and latitude of data points present in GLSS 7 dataset.

2.5. Model Estimation

Estimation of latent variable spatial models, such as equation 5, requires the researcher to be quite more careful because of the geographical structure of the data. Each observation in spatial data is assumed to be related to neighbourhood observations, contrary to maximum likelihood (ML) estimation theory, which only permits the multiplication of independent terms to compute the likelihood function. Additionally, these models being hierarchical are highly parameterized. So the traditional maximum likelihood (ML) process may not be able to estimate all parameters simultaneously.

A Bayesian Markov Chain Monte Carlo (MCMC) method is utilized for estimating equation 5.

Let $\theta = (\beta, \lambda, \sigma^2, \Sigma)$ denote the parameters of the model. Then in accordance with the Bayesian methodology [18], the joint posterior density is denoted by

$$p(\theta, Y_i^*|Y_i) = \frac{p(Y_i|Y_i^*, \theta) \times p(Y_i^*|\theta) \times p(\theta)}{\int p(Y_i|Y_i^*, \theta) \times p(Y_i^*|\theta) d(\theta)} \tag{10}$$

Where $\int p(Y_i|Y_i^*, \theta) \times p(Y_i^*|\theta) d(\theta)$ denotes the integral likelihood, ensuring that equation 10 integrates to 1, $p(Y_i|Y_i^*, \theta)$ is an indicator variable such that each y_i^* remains within a given interval of the observed binary category, the function $p(Y_i^*|\theta)$ is the density of the latent variable Y_i^*, and $p(\theta)$ is the prior of the parameters. Posterior estimation was done by setting up the Gibbs sampler [18], which required a derivation of the full conditionals for all parameters. Samples of β, λ, σ^2, and Σ were drawn from their respective full conditional distributions for inference. But first, following the Bayesian criteria, prior distributions for all parameters are assigned.

Prior for $\boldsymbol{\beta}$:

$p(\beta) \sim N(g_o, G_o)$ where g_o and G_o respectively are the mean and variance.

Prior for $\boldsymbol{\sigma^2}$:

$p(\sigma^2) \sim \mathcal{IG}(\frac{a_1}{2}, \frac{b_1}{2})$, where $\frac{a_1}{2}$ and $\frac{b_1}{2}$ are shape and scale parameters respectively.

Prior for $\boldsymbol{\lambda}$:

We assigned the uniform density prior on the interval $(-\frac{1}{w}, \frac{1}{w})$, that is $U(-\frac{1}{w}, \frac{1}{w})$

Prior for $\boldsymbol{\Sigma}$:

$p(\Sigma) \sim \mathcal{IG}(\frac{a_2}{2}, \frac{b_2}{2})$, where $\frac{a_2}{2}$ and $\frac{b_2}{2}$ are shape and scale parameters respectively.

The full conditional posterior distribution for each parameter is as follows:

The full conditional posterior distribution of β:

$$\begin{aligned} p(\beta|y, y^*, \lambda, \sigma^2, \Sigma) &= p(y^*|\beta, \lambda, \sigma^2, \Sigma) \times p(\beta) \\ p(\beta|y, y^*, \lambda, \sigma^2, \Sigma) &= exp\{-\frac{1}{2}(y^* - X\beta)'\Sigma^{-1}(y^* - X\beta)\} \times \{(\beta - g_o)' G_o^{-1}(\beta - g_o)\} \end{aligned} \tag{11}$$

The full conditional posterior distribution of σ^2 is given by

$$\begin{aligned} p(\sigma^2|y^*, y, \beta) &= p(y^*|\beta, \lambda, \sigma^2, \Sigma) \times p(\sigma^2) \\ p(\sigma^2|y^*, y, \beta) &= exp\{-\frac{1}{2}(y^* - X\beta)'\Sigma^{-1}(y^* - X\beta)\} \times \sigma^{2^{-(a_1+1)}} exp(-\frac{b_1}{\sigma^2}) \end{aligned} \tag{12}$$

Thus the full conditional distribution of y^* is given by:

$$p(y^*|y, y_i^*, \beta, \lambda, \sigma^2, \Sigma) = I_{\{}y^* < 0 \; or \; y^* > 0\} \times exp\{-\frac{1}{2}(y^* - X\beta)'\sigma^{-1}(y^* - X\beta)\} \tag{13}$$

The full conditional distribution for the spatial decay parameter λ is given by

$$\begin{aligned} p(\lambda|y^*, y, \beta, \sigma^2) &= p(y^*|\beta, \lambda, \sigma^2, \Sigma) \times p(\lambda) \\ p(\Sigma|\beta, \sigma^2) &= exp\{-\frac{1}{2}(y^* - X\beta)'\Sigma^{-1}(y^* - X\beta)\} \times \frac{1}{w_2 - w_1} \end{aligned} \tag{14}$$

Similarly, the full conditional distribution for Σ is given by

$$\begin{aligned} p(\Sigma|\beta, \sigma^2) &= p(y^*|\beta, \lambda, \sigma^2, \Sigma) \times p(\Sigma) \\ p(\sigma^2|y^*, y, \beta) &= exp\{-\frac{1}{2}(y^* - X\beta)'\Sigma^{-1}(y^* - X\beta)\} \times \sigma^{2^{-(a_2+1)}} exp(-\frac{b_2}{\sigma^2}) \end{aligned} \tag{15}$$

2.6. Predictions and Mapping

Given the Bayesian predictive distribution $p(y|y^*, \beta, \lambda, \sigma^2, \Sigma(\theta))$ in section 2.5, let $Y_0 = (Y_{i0}, \cdots, Y_{in})'$ and $Y_{i0} = Y(s_{i0})$, where s_{i0} is an unsampled location, be an arbitrary predictor of the binary response variable. To predict poverty situation and provide full description of poverty severity at unobserved locations, we adopted the optimal Bayes predictor [13] of Y_i (i.e., the predictor that minimizes the Bayesian expected loss (BEL)) given as

$$\hat{Y}_{i0} = \begin{cases} 1 & \text{if} \quad p\{Y_{i0} = 1|y\} > \frac{l_0}{l_0 + l_1} \\ 0 & \quad Otherwise \qquad\qquad\qquad i = 1, \cdots, n \end{cases} \tag{16}$$

Where l_0 and l_1 are losses for mispredicting Y_i as a 0 and 1, respectively. To estimate $p\{Y_{i0} = 1|y\}$ in the predictor, we used indicator kriging, basing on the assumption that the posterior mean and covariance function are well-known. Indicator Kriging predicts values of the binary random variable at unsampled locations based only on nearby observation [29]. Indicator kriging is based on the kriging method of spatial prediction. Indicator Kriging uses a function $\gamma(d_{ij}) = \frac{1}{2}Var\{(y_i)_s - (y_j)_s\}$. Where $\gamma(.)$ is an isotropic semivariogram and $d_{ij} = \|s_i - s_j\|$ are the distance between locations. Employing this function, prediction of $Y(s_{i0})$, is based on $p\{Y_{i0} = 1|y\}$. An estimate of $p\{Y_{i0} = 1|y\}$, is its best linear unbiased estimator [10, 13] given by

$$\hat{p}\{Y_{i0} = 1|y\} = \left(\gamma + \frac{(1 - \mathbf{1}'\Gamma^{-1}\gamma)}{1 - \mathbf{1}'\Gamma^{-1}\mathbf{1}}\mathbf{1}\right)' \Gamma^{-1} Y \tag{17}$$

Where $\Gamma_{ij} = \gamma(\|s_i - s_j\|)$, $\gamma_i = \|s_0 - s_i\|$, $\mathbf{1}$ is a $n \times 1$ vector of 1 and $ij = 1, \cdots, n$. The binary map is estimated using equations 17 with 16.

2.7. Deviance Information Criterion (DIC)

Deviance Information Criterion (DIC) proposed by [27] has largely been used in Bayesian estimations and modeling. It is made up of two parts: the model's fit measure and the model's complexity. The posterior expectation of the deviance for the data Y and parameter vector β is used to determine how well the model fits the data. This goodness-of-fit and model selection measure is define as

$$DIC = Dbar + pD = Dhat + 2pD, \quad pD = Dbar - Dhat \tag{18}$$

The first term in equation 18 measures the model fit via the posterior mean deviance and the second term measures model complexity known as the effective number of parameters. Smaller values of the DIC, following [27], mean a better model fit supported by the data.

3. Results and Discussion

3.1. Spatial Distribution and Patterns in GLSS 7 Dataset

Various spatial techniques and tools, were used to describe the spatial characteristics and test of spatial dependency of poverty distribution in the GLSS 7 data, as well as the poverty generation process in Ghana. Figure 2 shows sample site locations of the GLSS 7 together with the observed poverty risk. As depicted by Figure 2, the observations indicate clustered patterns.

Table 1: **Spatial Distance**

Statistic	Value
Minimum	0.0051
1st Quartile	4.6603
Median	7.8933
Mean	8.6023
3rd Quartile	12.3275
Maximum	22.1015

The output in Table 1 reports that, there is an average distance of about 8.6km to nearest neighbours, a minimum distance to nearest neighbour of about 0.01km and a maximum distance of 22km between data points in GLSS 7. Results from the distance analysis helped in an appropriate consideration of lag increaments, tolerance and number of lags for the semivariogram model by [21] given in Figure 1.

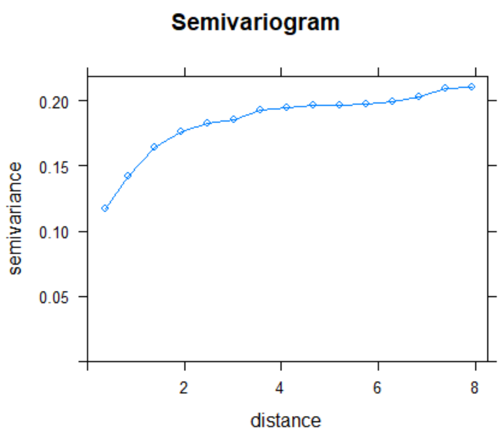

Figure 1: **Empirical Semivariogram**

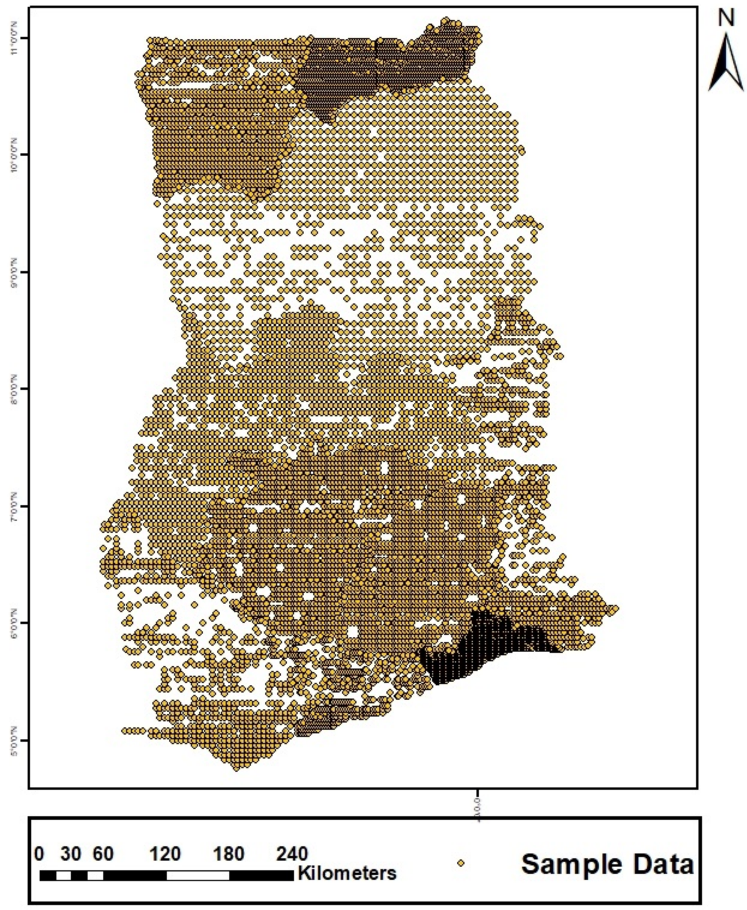

Figure 2: **Sample Site Locations and Observed Poverty Status**

The main objective of the variogram analysis is to best estimate the autocorrelation structure of the underlying spatial Gaussian process in the data. The semi-variogram provides a quantitative measure of the rate at which similarity between sample points decreases with separation distance [17]. Figure 1 was fitted using the Matérn covariance function. The semi-variogram shows an increasing trend from the start point, indicating lag-dependent variation, where closer observations have smaller semi-variances. This indicates the presence of extra variation in the binary response unaccounted for by covariates alone, underscoring the need to use spatial methods when dealing with poverty classification and prediction.

3.2. Parameter Estimation

The coefficient estimates (posterior means) for model parameters β, λ, σ^2 and Σ are shown in Tables 2 and 3. Markov Chain Monte Carlo (MCMC) estimation was applied for models based on the same dataset and covariates. Table 2 reports results for two models, the aspatial binary probit model (Model 1) discussed in section 2.2 and the results of our geo-classification model (Model 2) discussed in section 2.3.

In add

using the set of samples drawn from the MCMC estimation. The MCMC simulations were carefully tuned to obtain satisfactory convergence. Simulation results for the aspatial binary probit model were obtained based on one run of the MCMC algorithm for 10000 iterations after a burn-in of 2000. The geo-classification model results were based on two chains of 5000 iterations, after a burn-in of 2500 each because a single chain did not achieve adequate convergence.

Table 2: **Posterior Means and 95% Bayesian Credible Intervals (BCIs) for Models**

Variable		Model 1		Model 2	
	β	Coefficient	[95% BCIs]	Coefficient	[95% BCIs]
Intercept	β_0	1.3615	[1.0955, 1.6301]	0.91334	[0.8253, 0.9868]
Household Size	β_1	-0.1541	[-0.1643,-0.1431]	-0.03321	[-0.0372, -0.0305]
Age	β_2	-0.0028	[-0.0048, -0.0005]	-0.0051	[-0.00106, -0.00002]
Income	β_3	0.0000081	[0.0000071, 0.0000091]	0.000001	[0.000001,0.000002]
Sex {*Ref:Male*}					
Female	β_4	-0.0403	[-0.1271, 0.0472]	-0.0083	[-0.0225, 0.0144]
EZ {*Ref:Coastal*}					
Forest	β_5	0.1416	[0.0312, 0.2520]	0.0296	[-0.0236, 0.0673]
Savannah	β_6	-0.2723	[-0.430, -0.115]	-0.0449	[-0.1089, 0.0555]
Accra	β_7	0.0030	[-0.2261, 0.7519]	-0.0156	[-0.0192, 0.0594]
AR {*Ref:Rural*}					
Urban	β_8	0.8982	[0.8305, 0.9664]	0.19197	[0.1706, 0.2193]
Edu {*Ref:None*}					
Basic	β_9	0.0039	[-0.2049, 0.2095]	0.0104	[-0.0668, 0.0573]
Secondary/High	β_{10}	0.0061	[-0.2128, 0.2110]	0.0089	[-0.0572, 0.0560]
Vocational/Tech	β_{11}	-0.0296	[-0.2930, 0.2323]	0.0285	[-0.0505, 0.0913]
Tertiary	β_{12}	0.0219	[-0.2484, 0.2021]	0.0087	[-0.0623, 0.0586]
ES {*Ref:Employed*}					
Unemployed	β_{13}	0.0595	[-0.0661, 0.1863]	0.0087	[-0.0202, 0.0431]
Not in labour force	β_{14}	-0.0201	[-0.0769, 0.0367]	0.0019	[-0.0123, 0.0922]
MS {*Ref:Married*}					
Consensual Union	β_{24}	0.0641	[-0.0617, 0.1908]	0.0024	[-0.0297, 0.0398]
Separated	β_{25}	0.1150	[-0.0885, 0.3219]	0.0211	[-0.0243, 0.0671]
Divorced	β_{26}	0.0193	[-0.0549, 0.0935]	-0.0041	[-0.0262, 0.0193]
Widowed	β_{27}	0.0325	[-0.1096, 0.1759]	0.0327	[-0.0187, 0.0779]
Never Married	β_{28}	0.0323	[-0.0417, 0.1062]	-0.0018	[-0.0207, 0.0217]
DIC		10394.0	-	8086.03	-

EZ= Ecological Zone, AR= Area of Residence, Edu=Educational Level, ES= Employment Status, MS= Marital Status. Model 1 is equation 3 and Model 2 is equation 5.

The results of Table 2 show that some of the fixed effect estimates from our models have 95% credible intervals that are significant. It is also observed that, the 95% credible intervals for other fixed effects are not statistically significant at 5% significance level, however the direction and magnitude of these effects on household poverty status should not be disregarded.

For instance, the household size has the 95% credible intervals in both model 1 and model 2 are statistically significant. This implies that household size significantly impact households poverty state in the region of study. What this means is that poverty levels tend to rise from non-poor to poor with increase in household size, thus decreasing the posterior likelihood of falling in the non-poor category, but increasing the likelihood of households sliding into the poor rank. A study in Ghana by [16] using logistic regression to estimate the probability of being in poverty based on the fourth and fifth rounds of GLSS data, reported that larger households negatively affected poverty levels in the country. Other research on poverty, for example, [1], [4] and [26], are congruent to this current work regarding the negative effect of large household sizes on poverty. Of great importance is the accute fall in the magnitudinal estimate of the variable in the spatial model. The physical interpretation of it is that, accounting for environmental and space compoundings in modeling the processes of poverty generation, improved posterior likelihood.

Households income level is seen to significantly influence poverty in both models at 5% significance probability. Increased levels of household income, result in a higher likelihood of the household residing in the non-poor category of poverty status as per the classification ascribed by this work. This result parallel poverty analysts such as [19], [30] and [3].

Whilst the forest and savannah ecological zone variables were statistically significant in model 1, the contribution of the Accra ecological zone variable was not significant. The posterior effect of ecological zone shows that living in the savannah ecological zone, as opposed to the coastal ecological zone (being the reference category), decreased the risk of being in the non-poor category, but increased the risk of being in the poor category. A similar and notable conclusion was reached in a study by [16]. They identified households in the savannah ecological zone of Ghana to be almost four times poorer than those living in the coastal and forest zones. [19] and [5] in separate studies in Ghana and Ethiopia respectively, also identified agro-ecological divisions as a key determinant of poverty.

Table 3: **Results of Spatial Parameters**

Parameter	Estimate	95% BCIs
Σ	0.1849	[0.1237, 0.2869]
σ^2	0.0453	[0.0449, 0.0786]
λ	66.40	[64.8747, 69.5436]

The parameter λ is used to determine the effective range of spatial correlation. The covariance function used in our estimation is the Matérn covariance model described in section 2.4. Based on our estimation, the posterior mean of the practical range of spatial dependency is approximately 66km, suggesting a strong spatial correlation in the data. What this means is that, Sample locations separated by distances equal to or closer than 66km are spatially autocorrelated, whereas locations farther apart than the practical range are not.

It is imperative to note that the posterior means for the geo-classification model are consistently smaller than the estimates from the non-spatial model. This indicates a stronger spatial structure that explained variation through its correlation function, rather than depending solely on the covariates. Moreover, the consistently narrower 95% posterior intervals for the spatial model further indicates better accuracy in the fixed effects estimation. It is important to bear in mind that the inherent nature of the geo-classification models is that they feature variation over spatial observations in the impacts arising from changes in the explanatory variables at each location. The fitted spatial model results show that environmental factors have statistical relevance. This implies that there is a link between poverty and the environment, and that environmental variables have an influence on the poor and poverty reduction initiatives.

3.3. Predictions and Mapping under the Geo-Classification Model

A key objective of this study is to map household poverty risk by predicting the outcome at new locations across all the regions of Ghana. In the case of our binary mapping, for an unsampled location s_{i0}, we computed $p\{Y_{i0} = non{-}poor|y\}$. These probabilities were used to produce the binary map shown in Figure 3.

From Figure 3, we observed some levels of heterogeneity in poverty distribution accross the study area, but again, relatively homogeneous and clustering over same locations of the map.

The high- risk poverty "hotspots" included the areas around the northern region, volta region, parts of western and central regions and some clustering are also observed along the borders of upper west and upper east regions to Burkina Faso. We also see that poverty is sparsely distributed in some areas of the country, which points to the fact that poverty is a spatially lagged variable. The core of this work is to understand the spatial similarities present in the poverty generation process which was mostly implied in other poverty studies in Ghana. Households at spacific locations tend

Figure 3 shows the strong effect of spatial correlation, an indication that there is indeed a link between poverty and spatial characteristics in Ghana. Thus, to minimize or eliminate extreme poverty, stakeholder efforts should be directed towards areas with high posterior ranks. For example, extreme poverty incidence in the Upper West and some areas in the Northern, Volta, Western, Central and Brong-Ahafo regions, may be of concern to stakeholders.

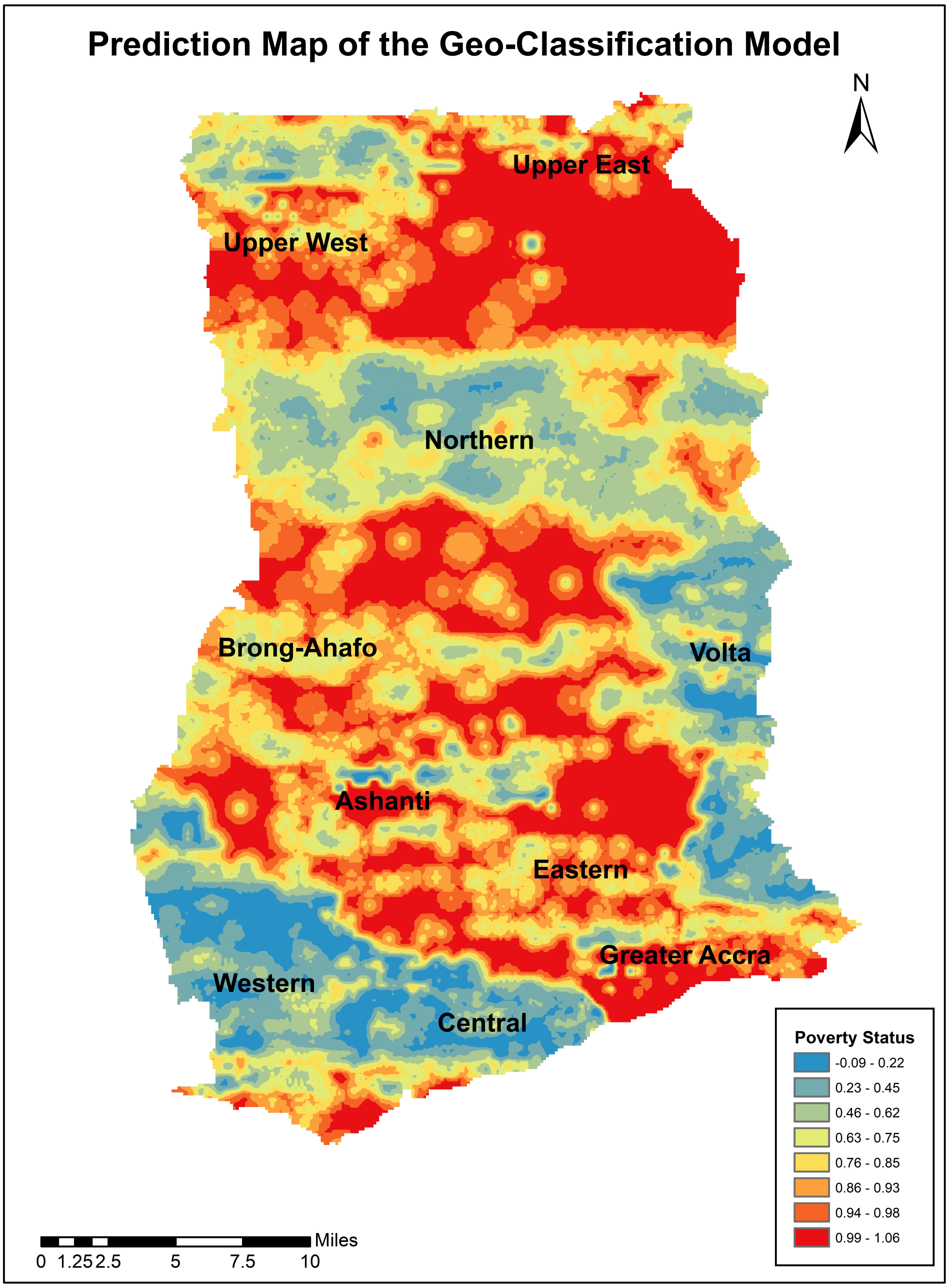

Figure 3: **Prediction Map of the Geo-Classification Model**

4. Conclusion

The main objective of this work was to model poverty-severity while accounting for spatial dependence using our geo-classification model and map household poverty using mixed discrete and continuous response data. According to the modeling technique used in this study, poverty categorization was performed in the most natural process, where households were classified into two distinct poverty regions; poor and non-poor.

The estimation technique employed in this work handled the complexity of our model by decomposing it into sub-layers and estimating each parameter. The geo-classification model that clipped the population into two categories (poor and non-poor) compared results from the commonly used binomial probit model. Evident from our preliminary results, spatial effect and variation is empirical in the GLSS 7 data and cannot be ignored in the bid to understand poverty and its correlates in the study region.

In general, the posterior means and 95% credible intervals show that fixed effect estimates (household size, income level of householder, ecological zone and location/area of residence) and spatial effects significantly influence poverty levels and distribution patterns in Ghana.

To achieve a meaningful reduction in poverty levels in Ghana, stakeholders must work hard to mitigate the influence of covariates which were found to have significant effect on poverty severity in this study

References

[1] Aboagye-Attah, K. (2019). Socioeconomic correlates of poverty in Ghana using Ghana Living Standards Survey round 6 and 7. page 69.

[2] Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., and Alyaman, M. (2021). Poverty classification using machine learning: The case of Jordan. *Sustainability (Switzerland)*, 13(3):1–16.

[3] Beegle, K., Christiaensen, L., Dabalen, A., and Gaddis, I. (2016). *Poverty in a rising Africa.* The World Bank.

[4] Ben, R. and Kacem, H. (2012). The Determinants of Poverty by Cohort of Households : Evidence from Rural Tunisia. 36:33–37.

[5] Bogale, A. (2011). Analysis of poverty and its covariates among smallholder farmers in the eastern Hararghe highlands of Ethiopia. *Journal of Development and Agricultural Economics*, 3(4):157 – 164.

[6] Bukari, C., Essilfie, G., Aning-Agyei, M. A., Otoo, I. C., Kyeremeh, C., Owusu, A. A., Amuquandoh, K. F., and Bukari, K. I. (2021). Impact of COVID-19 on poverty and living standards in Ghana: A micro-perspective. *Cogent Economics and Finance*, 9(1).

[7] Cakmakyapan, S. and Goktas, A. (2013). a Comparison of Binary Logit and Probit Models With a Simulation Study. *Journal of Social and Economic Statistics*, 2(1):1–17.

[8] Chambers, R., Salvati, N., and Tzavidis, N. (2016). Semiparametric small area estimation for binary outcomes with application to unemployment estimation for local authorities in the UK. *Journal of the Royal Statistical Society. Series A: Statistics in Society.*

[9] Cooke, E., Hague, S., and McKay, A. (2016). The Ghana poverty and inequality report: Using the 6th Ghana living standards survey. *University of Sussex*, pages 1–43.

[10] Cressie, N. A. C. (1993). Statistics for Spatial Data. *Journal of the Royal Statistical Society. Series A (Statistics in Society).*

[11] Cressie, N. A. C. (1994). Statistics for Spatial Data, Revised Edition. *Biometrics.*

[12] Darmawan, D. (2019). *poverty and shared prosperity 2018: Piecing together the poverty puzzle.*, volume 53.

[13] De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics and Data Analysis*, 34(3):299–314.

[14] Dudek, H. and Lisicka, I. (2013). Determinants of poverty – binary logit model with interaction terms approach. *Ekonometria*, (41):65–77.

[15] Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, 71(1):355–364.

[16] Ennin, C. C., Nyarko, P. K., Agyeman, A., Mettle, F. O., and Nortey, E. N. N. (2011). Trend analysis of determinants of poverty in Ghana: Logit approach. *Research Journal of Mathematics and Statistics*, 3(1):20–27.

[17] Gaetan, C. and Guyon, X. (2010). *Spatial statistics and modeling*, volume 90. Springer.

[18] Gelman, A., Carlin, J. B. B., Stern, H. S. S., and Rubin, D. B. B. (2014). Bayesian Data Analysis, Third Edition (Texts in Statistical Science). *Book*.

[19] GSS (2019). Ghana Living Standards Survey round 7 (GLSS7), Main Report. *Ghana Statistical Service*, pages 1–343.

[20] Hobza, T. and Morales, D. (2016). Empirical best prediction under unit-level logit mixed models. *Journal of official statistics*, 32(3):661.

[21] Matheron, G. (1963). Principles of geostatistics. *Economic Geology*.

[22] Mdakane, B. P. (2019). Mapping of self-reported health among individuals between the ages of 15 – 49 years in South Africa .

[23] Molina, I., Nandram, B., and Rao, J. N. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical bayes approach. *Annals of Applied Statistics*, 8(2):852–885.

[24] Pleis, J. R. (2018). Mixtures of discrete and continuous variables: Considerations for dimension reduction. pages 1–144.

[25] Puurbalanta, R. (2019). Spatial Cumulative Probit Model: An Application to Poverty Classification and Mapping. *International Journal of Statistical Distributions and Applications*, 5(1):15.

[26] Puurbalanta, R. (2020). A Clipped Gaussian Geo-Classification model for poverty mapping. *Journal of Applied Statistics*, 0(0):1–14.

[27] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*.

[28] Sriyalatha, A. M. K. (2019). Determinants of Poverty Among Households in Monaragala District , Sri Lanka Africa. *IRE Journals*, 2(7):64–74.

[29] Switzer, P. (1977). Estimation of spatial distributions from point sources with application to air pollution measurement. Technical report No. 9. Technical report.

[30] World Bank (2020). Supporting Countries in Unprecedented Times. *Annual Report 2020*, pages 1–106.