

Bayesian Spatial Ordinal Models for Regional Household Poverty-Severity

Richard Puurbalanta^{1,*}, Atinuke O. Adebajji²

¹University for Development Studies, Faculty of Mathematical Sciences, Department of Statistics, Navrongo Campus, Ghana

²Kwame Nkrumah University of Science and Technology, Faculty of Physical Sciences, College of Science, Department of Mathematics, Kumasi, Ghana

Abstract The use of discrete spatial-statistical methods for poverty analyses is important, especially in light of the fact that living standards surveys are generally dominated by categorical observations made at several locations. The proximity of these observations imposes geographical structure on the data. This study presents ordinal geo-statistical models for household poverty analyses that recognize the ordinal nature of poverty severity. For all models, Bayesian inference via Markov Chain Monte Carlo (MCMC) was used. Precision of the models, understood in terms of ease of implementation and accuracy of estimation, is compared. The objective was to quantify spatial associations, given some household features, and produce a map of poverty-severity for Ghana. The Clipped Gaussian Spatial Ordinal Probit (CG-SOP) Model was identified as best for describing spatial poverty. Positive correlation with respect to the distribution of extreme poverty was observed. We see evidence of this in the map of predictions. Significant variables include household size, education, and residency of household head. This approach to poverty analysis is relevant for policy design and the implementation of cost-effective programmes to reduce (category and site)-specific poverty, and monitoring changes in both category and geographical trends thereof. Analysis was based on the Ghana living standards survey (GLSS) 2012 data.

Keywords Household poverty-Severity, Latent Variables, Ordinal models, Bayesian inference, Markov Chain Monte Carlo, Prediction maps

1. Introduction

If a variable is random, being independent and identically distributed (i.i.d.), it can be analyzed using standard statistical techniques such as the generalized linear models (GLMs). When the variable, however, is distributed in a manner that subjects it to systematic spatial variation, where observations at proximal locations are more similar than observations further apart, the independence assumptions of standard statistical methods fail. Parameters cannot easily be estimated via simple or straightforward statistical techniques such as the GLMs. Methods that allow for spatial dependency in the data are instead required for parameter estimation.

When the variable is also categorical, and ordered, modelling is further complicated, requiring complex inferential techniques to effectively handle the multiple categories, and also account for the ordering.

Premising the discourse on poverty this way is important, especially in light of the fact that welfare data are often

dominated by categorical observations collected (via living standards surveys) at several locations across space, and proximity of these observations imposes geographical structure on the data. This means that global parameters (obtained via independence assumptions of standard statistical methods) cannot adequately describe site-specific conditions of the variable.

Previous works to investigate poverty is abundant [1-5], but have largely ignored spatial dependence of observations and used methods that rely on independence assumptions, albeit with varying degrees of sophistication. This is problematic because when a regression model incorrectly assumes homoscedasticity in the errors of a spatial variable, outcome of the analysis can be biased [6].

The existing literatures [7-9] that have used spatial tools to study poverty have also largely ignored the multi-categorical and ordinal nature of the variable. Binary models for poverty analysis mask the effect of important intermediate information during the binary transformation of the variable [10], but also excludes the severest form of the variable; extreme poverty. This has the potential for incomplete and inaccurate description of poverty. The method for poverty analysis must be careful to explain why some population groups are non-poor, poor, or extremely poor.

* Corresponding author:

pubarichard@yahoo.com (Richard Puurbalanta)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

So, what we do differently is to first discretize poverty into a multi-category ordered random variable using the severity levels as thresholds, and employ multivariate ordered analysis to describe the overall condition of a household as either being non-poor, poor, or extremely poor. Ordinal models solve the problem of loss of information during discretization by appealing to the concept of latent variables [11, 12], where a Gaussian random variable is assumed to be latent, and assigning values to the ordered categories according to a regression function.

Second, we assume that poverty, like all spatial phenomena, is not only a function of covariates, but also of unobservable large-scale interaction of geographical and agro-climatic forces.

Thus, we model poverty-severity risk as a spatially-varying ordinal phenomenon. Three models were proposed: The simple Ordinal Probit (OP) model [11], ordinal Spatial Generalized Linear Mixed Model (SGLMM) [12-14], and Clipped Gaussian Spatial Ordinal Probit (CG-SOP) model [15-17]. All the models were implemented in R version 3.1.3 [18], and ArcGIS [19].

To solve the inferential problems associated with spatial ordered models, we follow [20] to use the Bayesian decision paradigm, implemented via Markov Chain Monte Carlo (MCMC) sampling techniques.

Overall, our method can be considered a (category and site)-specific report that identifies all segments of the poor for easy targeting, thus avoiding the blanket approach that prefers the one-fits-it-all solution to the problem of poverty.

2. Measurement of Poverty

Poverty has many dimensions, but is generally measured in non-monetary or monetary terms [18]. We adopt the monetary approach to capture and report information in a form which facilitates comparison of our findings with similar international research.

Estimation of monetary poverty requires a choice between disposable income and total consumption expenditure as the indicator of wealth, the latter being the preferred choice in most developing countries [21], and in this study. Moreover, to measure monetary poverty, we need to set minimum standards of the poverty indicator to separate the various categories of the poor. These are called poverty lines [21]. The three categories of poverty lines adopted by [21] are based on family daily expenditure per adult equivalent. The thresholds are set at GHC 3.60 per day- non-poor, between GHC3.60 and GHC2.17-poor, and less than GHC2.17-extremely poor.

3. Data Sources and Covariates

The study used secondary data from the Ghana Living Standard Survey (GLSS). Included in the dataset of 16772 sampled households are details of households (household

size, age, sex, and educational level of household head), socio-economic conditions of household (availability of social amenities), and neighbourhood information (ecological zone, rural/urban residence, and other environmental factors). The dataset of the GLSS collected in 2012 were not geo-coded. To perform spatial analysis, we geo-located each sampled household randomly (assuming a uniform distribution) at the district level in order to be consistent with existing spatial units.

4. Ordinal Random Variables and Computations

4.1. The Ordinal Probit (OP) Model

The traditional OP GLM [12] derives from the multinomial distribution, albeit with ranked categories. The density function for the multinomial distribution is:

$$Pro(P|Y) = \prod_{i=1}^n \prod_{j=1}^J p_i^{y_{ij}} \quad (1)$$

where y_i is the vector of $n \times 1$ responses, and $p_i^{y_{ij}}$ constitutes the probability of observation y_i being in category j . Here, the order of the J outcomes is not important.

To derive the proportional odds cumulative OP GLM from (1), let the responses $j = 1, \dots, J$ be arranged in order of magnitude, and α_j be the corresponding thresholds associated with the ordering. Further let Y_i^* be a Gaussian random variable assumed to be latent, and assigning values to α_j according to the regression function:

$$y_i^* = X_i \beta + \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \quad (2)$$

where X is $n \times p$ design matrix, and β is a $p \times 1$ unknown vector of regression coefficients, and ε is the $n \times 1$ vector of independently and identically distributed (*i. i. d*) measurement errors: $\varepsilon_i \sim N(0,1)$. Though the values of y_i^* cannot be directly observed, the rule that assigns y_i^* to α_j is that if y_i^* exceeds a given threshold, then, an observation falls in the j^{th} category. This culminates in cumulative multiple binary outcomes:

$$y_i = \begin{cases} j, & \alpha_{j-1} < y_i^* \leq \alpha_j \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where y_i is the vector of $n \times 1$ response, $\alpha_j \in \mathbb{R}$, and $\alpha_1 < \alpha_2 \dots < \alpha_J$.

Clearly, Y_i^* , in our application, refers to the Gaussian expenditure line, and is asymptotic of the ordinal variable Y_{ij} when $J \rightarrow \mp\infty$.

Our objective is to predict the probability of a household falling in the j^{th} category given the observed covariates $x = (x_1, \dots, x_n)^T$. This probability is determined by the values of the latent variable Y_i^* , and is given by

$$p(y_i = j|x_i) = p(\alpha_{j-1} < y_i^* \leq \alpha_j) \quad (4)$$

Since Y_i^* is Gaussian, and $\varepsilon_i \sim N(0,1)$, the outcome is a probit model, implying that the probability of falling in the j^{th} category is:

$$p(y_i = j | x_i) = p(\alpha_{j-1} - X_i \beta < \varepsilon_i \leq \alpha_j - X_i \beta) \\ = \Phi(\alpha_j - X_i \beta) - \Phi(\alpha_{j-1} - X_i \beta) \quad (5)$$

where $\Phi(\cdot)$ is the cumulative distribution function (*cdf*) for the normal ε_i . Thus, the likelihood function for the parameters is

$$L(\beta, \alpha | y = j) = \int_{\Theta} \prod_{i=1}^n \prod_{j=1}^J [\Phi(\alpha_j - X_i \beta) - \Phi(\alpha_{j-1} - X_i \beta)] d(\beta, \alpha) \quad (6)$$

Several normalizations are needed to identify the ordered model parameters. For details see [12].

4.1.1. Bayesian Estimation via MCMC

Following the Bayesian criteria, we set priors for the parameters, and build the posterior as:

$$f(\beta, \alpha | y_i^*, y, x) = \frac{\prod_{i=1}^n [\Phi(\alpha_j - X_i \beta) - \Phi(\alpha_{j-1} - X_i \beta)] p(\beta, \alpha)}{\int_{\Theta} \prod_{i=1}^n [\Phi(\alpha_j - X_i \beta) - \Phi(\alpha_{j-1} - X_i \beta)] d(\beta, \alpha)} \quad (7)$$

where $\Phi(\alpha_j - X_i \beta) - \Phi(\alpha_{j-1} - X_i \beta)$ is the data likelihood, $p(\beta, \alpha)$ is the prior density of model parameters, and $\int_{\Theta} \prod_{i=1}^n [\Phi(\alpha_j - X_i \beta) - \Phi(\alpha_{j-1} - X_i \beta)] d(\beta, \alpha)$ is the integrated likelihood, with Θ being the parameter space.

Posterior estimation is done by setting up the Gibbs sampler [14], which requires us to derive full conditionals for all parameters. We impose, following [12], non-informative priors on the regression coefficients. Then

$$p(\beta | y^*, \alpha, y, x) = p(y^* | \beta) p(\beta) \quad (8)$$

$$p(y_i^* | \alpha, \beta, y, x) = p(y | \alpha) p(y_i^*) \quad (9)$$

$$p(\alpha | y^*, y, \beta) = p(y | \alpha) p(\alpha) \quad (10)$$

where y is the vector of observed ordinal outcomes, while y^* is latent, $p(y | \alpha)$ is an indicator variable, leading to:

$$p(\alpha | y^*, y, \beta) = \left[\sum_{j=1}^J \mathbf{1}(y_i = j) \times \mathbf{1}(\alpha_{j-1} < y^* \leq \alpha_j) \right] \\ = \left\{ \mathbf{1}(y_i = j) \times \mathbf{1}(\alpha_{j-1} < y^* \leq \alpha_j) + \right. \\ \left. \mathbf{1}(y_i = j+1) \times \mathbf{1}(\alpha_j < y^* \leq \alpha_{j+1}) \right\} \\ = \mathbf{1}(\alpha_j^{inf} < \alpha_j < \alpha_j^{sup}) \quad (11)$$

4.2. Spatial Measurement Framework

If, for example, ordinal observations are made at sites $s = \{s_1, \dots, s_i\}$ under geo-statistical assumptions, then the data is defined as $\left[(y_{ij})_{s, j=1}^J \right]_{i=1}^n$, where $i = \{1, \dots, n\}$ and $j = 1, \dots, J$ respectively are observations and response levels.

Let X_i be the vector of associated covariates. Intuitively then,

$$(y_{ij})_s = X_i \beta + (\varepsilon_i)_s \quad (12)$$

spatial regression equations are written, where X_i are $n \times p$ -dimensional covariates, β a $p \times 1$ -dimensional vector of coefficients, and $(\varepsilon_i)_s$ a $n \times 1$ spatially-dependent error. Let $Y = \left[(y_{11})_1, \dots, (y_{ij})_s \right]^t$, $\varepsilon = [(\varepsilon_1)_1, \dots, (\varepsilon_i)_s]^t$, and $X = (x_1, \dots, x_{ip})^t$, then

equation (12) is written in matrix form as:

$$Y = X\beta + \varepsilon. \quad (13)$$

Spatial dependence is imposed by assuming that $cov(\varepsilon_i, \varepsilon_k) = \Sigma_{ik}$, $i, k = 1, \dots, n$, for $i \neq k$ where Σ_{ik} is a non-diagonal symmetric and positive definite matrix.

4.2.1. Spatial Generalized Linear Mixed Model (SGLMM)

The framework of the ordinal SGLMM is:

$$\left. \begin{aligned} y_i^* &= X_i \beta + \xi_i, \quad \xi_i = u_i + \varepsilon_i, \\ y_i &= j \text{ if } \alpha_{j-1} \leq y_i^* < \alpha_j, \\ \varepsilon_i &\sim N(0, 1) \\ u_i &\sim N_p[0, \Sigma_{ik}(\Theta)] \end{aligned} \right\} \quad (14)$$

where X is an $n \times p$ design matrix, β is a $p \times 1$ matrix of fixed effects coefficients, and u_i spatially-dependent errors that capture all unobserved errors arising from the influence of common features for observations within certain proximal distances. The $(i, k)^{th}$ element of the $n \times n$ spatial covariance matrix $\Sigma_{ij}(\Theta)$ is $Cov(u_i, u_k)$, parameterized by

$$\Sigma_{ik}(\Theta) = \sigma_i^2 \rho(|i - k|; \phi_i), \quad (15)$$

where σ_i^2 represents variability of the spatial process, and $\rho(|i - k|; \phi_i) = \exp(-d_{ik} \phi_i)$ is a monotonic correlation function with a correlation decay parameter ϕ_i measuring the strength of spatial dependence over the Euclidean distance d_{ik} between locations i and k [22].

By maintaining the ordinal probit form, and assuming, following [22], that the spatial error term u_i also represents random measurement error in unexplained explanatory variables (the nugget), then

$$p(y_i = j | x_i) = p(\alpha_{j-1} < y_i^* \leq \alpha_j) \\ = \Phi\left(\frac{\alpha_j - X_i \beta - u_i}{\sqrt{\sigma_i^2 \rho(|i - k|; \phi_i)}}\right) - \Phi\left(\frac{\alpha_{j-1} - X_i \beta - u_i}{\sqrt{\sigma_i^2 \rho(|i - k|; \phi_i)}}\right) \quad (16)$$

Thus, the likelihood function is

$$f(\beta, u, \alpha, \sigma^2, \phi | y_i) =$$

$$\prod_{i=1}^n \sum_{j=1}^J I_{\{\alpha_{j-1} < y_i^* \leq \alpha_j\}} \Phi \int_{\Theta} \left(\frac{\alpha_j - X_i \beta - u_i}{\sqrt{\sigma_i^2 \rho(|i - k|; \phi_i)}} \right) - \Phi \left(\frac{\alpha_{j-1} - X_i \beta - u_i}{\sqrt{\sigma_i^2 \rho(|i - k|; \phi_i)}} \right) \\ d(\beta, \alpha, \sigma^2, \phi), \quad (17)$$

where $\sum_{j=1}^J I_{\{\alpha_{j-1} < y_i^* \leq \alpha_j\}}$ is an indicator function equalling 1 when $y_i = j$ and 0 otherwise.

4.2.1.1. Bayesian Estimation via MCMC

Using the Bayesian approach [17], the posterior likelihood is:

$$f(\beta, u, \alpha, \sigma^2, \phi | y) p(y) = f(y | u, \beta, \alpha) f(\beta, \alpha, \sigma^2, \phi) \\ = \prod_{i=1}^n f(y | u, \beta, \alpha) f(y_i^* | \beta, u) f(u | \sigma^2, \phi) f(\beta, \alpha, \sigma^2, \phi), \quad (18)$$

where $f(y | u, \beta, \alpha)$ is the data vector, $f(y_i^* | \beta, u)$ is latent regression function assigning values to observable ordered data, $p(u | \sigma^2, \phi)$ is distribution of the latent errors

incorporated via data augmentation, and $p(\beta, \alpha, \sigma^2, \phi)$ are prior densities for the model parameters.

The use of MCMC sampling, especially Gibbs techniques [20], requires us to derive full conditionals for all parameters. We assume following [22] non-informative priors for the regression coefficients and inverse gamma for the spatial parameters. Thus, the full conditionals of all parameters have the following functional forms:

$$f(\beta|y^*, y) = f(y^*|\beta, u, \sigma^2, \phi)f(\beta) \quad (19)$$

$$f(u|y^*, \sigma^2, \phi, y) = f(y^*|\beta, u, \sigma^2, \phi)f(u|\sigma^2, \phi) \quad (20)$$

$$f(\phi|u, \sigma^2, y) = (u|\sigma^2, \phi)f(\phi) \quad (21)$$

$$f(\sigma^2|u, \phi, y) = (u|\sigma^2, \phi)f(\sigma^2) \quad (22)$$

$$f(\alpha|y^*, y, \beta) = f(y|y^*, \alpha, \beta)f(\alpha) \quad (23)$$

$$f(y^*|u, \beta, \sigma^2, \phi, y) = f(y|y_i^*, \alpha)f(y^*|\beta, u, \sigma^2, \phi) \quad (24)$$

4.2.2. The CG-SOP Model

To create a workable geo-statistical framework, we define an ordinal random field $(Y_{ij})_s$ by assuming that the spatially-dependent errors u_i are latent, but together with some covariates, are assigning values to α_j according to the multivariate regression function:

$$U_i \sim N_p[X_i\beta, \Sigma_{ik}(\Theta)] \quad (25)$$

where variables and parameters are as previously defined. The $(i, k)^{th}$ element of $\Sigma_{ik}(\Theta)$ is $cov(u_i, u_k)$, parameterized by $\Sigma_{ik}(\Theta) = \sigma_i^2 \rho(|i - k|; \phi_i)$, where parameters are as previously defined.

The link between U and $(Y_{ij})_s$ is seen by assuming that for a vector of thresholds α_j , the ordinal random field, $(Y_{ij})_s$, is obtained by quantizing U at levels $j = 1, \dots, J$, thus, culminating in multiple binary outcomes:

$$(Y_i)_s = \sum_{j=1}^J j I_{\{\alpha_{j-1} < U \leq \alpha_j\}} \quad (26)$$

where $\alpha_j \in \mathbb{R}$.

Our objective is to model the probability of an observation $(y_i)_s = \{(y_1)_s, \dots, (y_n)_s\}$ at site s falling at the j^{th} category. This probability is determined by U :

$$P[(y_i)_s = j|x_i] = p(\alpha_{j-1} < u_i \leq \alpha_j) \quad (27)$$

Now, $(Y_{ij})_s = [(y_{11})_1, \dots, (y_{nj})_s]^T$ is n -dimensional, and determined by the density of U . Therefore, the likelihood of the data is the integral of an n -dimensional multivariate normal distribution taking values in the interval $[\alpha_{y_{nj}-1}, \alpha_{y_{nj}}]$:

$$f(y|\beta, \alpha, \phi, \sigma^2) = \int_{\alpha_{y_{1j}-1}}^{\alpha_{y_{1j}}} \dots \int_{\alpha_{y_{nj}-1}}^{\alpha_{y_{nj}}} (2\pi)^{-\frac{n}{2}} |\Sigma(\phi)|^{-\frac{n}{2}} \times \exp\left\{-\frac{1}{2}[u - X\beta]^T \Sigma(\phi)^{-1}[u - X\beta]\right\} d(u) \quad (28)$$

4.2.2.1. Bayesian Estimation via MCMC

Following the Bayesian methodology [20], the full joint posterior distribution for all unknown parameters is completely defined by the posterior likelihood:

$$f(u, \beta, \alpha_j, \sigma^2, \phi|y) = f(y|u, \beta, \alpha_j, \sigma^2, \phi)f(u|u_{i \neq k}, \beta, \alpha_j, \sigma^2, \phi)f(\beta, \alpha_j, \sigma^2, \phi), \quad (29)$$

where $f(y|u, \beta, \alpha_j) = \prod_{i=1}^n I_{\{\alpha_{y_{i-1}} < u_i \leq \alpha_{y_i}\}}$ is an indicator variable, and $f(\beta, \alpha_j, \sigma^2, \phi)$, being the joint prior density, specifies uncertainty on the respective estimated parameters. We complete the specification by imposing non-information priors on the regression parameters and inverse gamma on the spatial parameters [16, 17].

The full conditional posterior distributions of each parameter [20] have the following functional forms:

$$f(\beta|y, u, \alpha, \sigma^2, \phi) = f(u|\beta, \sigma^2, \phi)f(\beta), \quad (30)$$

$$f(\phi|y, u, \beta, \alpha) = f(u|\beta, \alpha, \phi, \sigma^2)f(\phi) \quad (31)$$

$$f(\sigma^2|y, u, \beta, \alpha) = f(u|\beta, \alpha, \phi, \sigma^2)f(\sigma^2) \quad (32)$$

$$f(\alpha|y, \beta, \sigma^2, \phi) = f(y|u, \alpha, \beta, \sigma^2, \phi)f(\alpha) \quad (33)$$

$$f(u|y, \beta, \alpha, \sigma^2, \phi) = f(y|u, \alpha, \beta, \phi, \sigma)f(u|\beta, \alpha, \sigma^2, \phi). \quad (34)$$

At this point, all parameters of interest are sampled sequentially via MCMC simulation until the desired number of draws is achieved.

4.3. Posterior Predictions and Mapping

A key goal of this study is to produce a smooth map of household poverty-severity risk in Ghana by predicting the outcome at new locations. We achieve this via kriging [23-25].

Define Y_* to be the map of predictive ordered responses corresponding to poverty-severity risk at unsampled locations $s_{*i}, i = 1, \dots, n_*$. The predictions are derived from the conditional probabilities that a new location falls in category j given the observed data. In ordered modelling, these conditional probabilities cannot be obtained via direct MCMC simulations; they are obtained by calculating $Pr(y_{*i} = j|y) = Pr(\alpha_{j-1} < \mathcal{A}_* \leq \alpha_j|y)$ using MCMC integration and data augmentation, where \mathcal{A}_* is a latent variable at new locations, being equal to y_* in both OP GLM and SGLMM, and u_* for the CG-SOP model.

Following the Bayesian approach, the predictive posterior distributions y_{*i} is augmented by \mathcal{A}_* to give:

$$f(y_{*i}|y) = \int_{\Theta} f(y_{*i}|\mathcal{A}_*)f(\mathcal{A}_*|y, \beta, \alpha)f(\beta, \alpha|y)d(\mathcal{A}_*, \beta, \alpha) \quad (35)$$

in the case of the aspatial OP GLM,

$$f(y_{*i}|y) = \int f(y_{*i}|\mathcal{A}_*, u_*)f(\mathcal{A}_*|y, \beta, \alpha, u)f(u_*|u, \sigma^2, \phi)f(\beta, \alpha, u, \sigma^2, \phi|y)d(\mathcal{A}_*, \beta, \alpha, u, \sigma^2, \phi) \quad (36)$$

for the SGLMM, and for the CG-SOP case, we have

$$f(y_{*i}|y) = \int f(y_{*i}|u_*)f(u_*|u, \beta, \alpha, \sigma^2, \phi)f(\beta, \alpha, u, \sigma^2, \phi|y)d(\beta, \alpha, u, \sigma^2, \phi) \quad (37)$$

4.4. Model Evaluation and Validation Measures

The posterior predictive densities are the distribution of possible future observations arising from the current model.

We determine a checking function for both predicted data and actual observations to assess predictive model fit. In line with this, we proceed by first randomly splitting the dataset into two equal parts: one part (8375) as training set and the other (8375) as a validation set [26].

In standard ML analyses, a repertoire of fit measures includes R^2 and likelihood-ratio χ^2 statistic. In Bayesian analyses, predictive model selection is checked within the framework of Bayesian decision theory [26, 27, 20], where prediction accuracy is based on the Bayesian expected loss (BEL). The loss function $L(y_{*i}, \hat{y}_{*i})$, represented here by the additive loss function estimates the loss incurred in predicting y_{*i} by \hat{y}_{*i} at location; assigning zero loss to correct predictions. For the multi-categorical application, the additive loss function in respect of n_* new locations is expressed as:

$$L(y_{*i}, \hat{y}_{*i}) = \frac{1}{n_*} \sum_{i=1}^{n_*} \left[\sum_{j=1}^J \sum_{r \neq j} d_{jr} I_{\{y_{*i}=j, \hat{y}_{*i}=r\}} \right] \quad (38)$$

where $d_{jr} \geq 0$, $\forall j, r$, penalizes the loss incurred for mis-predicting category r when the true category is j . If d_{jr} is set to 1, equation (38) translates to the well-known mis-prediction rate (MPR) (defined as the proportion of predictions that are incorrect), and optimal Bayes' prediction is attained by choosing the value for r that minimizes (38), leading to the expected value of the loss function:

$$BEL = \frac{1}{n_*} \sum_{i=1}^{n_*} \sum_{j \neq \hat{y}_{*i}} Pr[y_{*i} = j | y]. \quad (39)$$

Equation (39) corresponds to the value of j that reports the highest estimated probability.

In assessing the models, it is important to note the difference between the estimated BEL (BEL_{Est}) (using the training data) and validation BEL (BEL_{Val}) (using the validation data) for each model. A comparatively large

difference shows a model is making incorrect predictions [26, 20].

For the ordered data, it is also possible for the model to suffer greater loss for predictions that are farther from the truth. Thus, we assign loss coefficients $d_{jr} = |j - r|$, and get the absolute mis-prediction rate (AMPR) defined as:

$$AMPR = \frac{1}{n_*} \sum_{i=1}^{n_*} |y_{*i} - \hat{y}_{*i}| \quad (40)$$

where y_{*i} is the true value and \hat{y}_{*i} is the prediction. In principle, the AMPR accounts for the size of the mis-prediction.

We apply all the fit measures (MPR, AMPR, and BEL) to determine the models' ability to correctly forecast future values.

5. Results

Before proceeding to perform spatial analysis, we need to confirm that the data exhibits significant spatial correlation. In this regard, the simple OP GLM was fitted to the data. Although spatial correlation in the raw data is of interest, we were primarily interested in whether there was correlation in model residuals once any correlation explained by the explanatory variables has been accounted for by the simple OP GLM. Figure 1 shows the classical semi-variogram estimator by [24], which is based on the residuals obtained after fitting the GLM.

It is clear from Figure 1 that there is a spatial trend beyond what was described by the mean structure alone. The semi-variogram shows an increasing trend from the origin, indicating lag-dependent variation. Therefore we expect high variability in the covariance parameters.

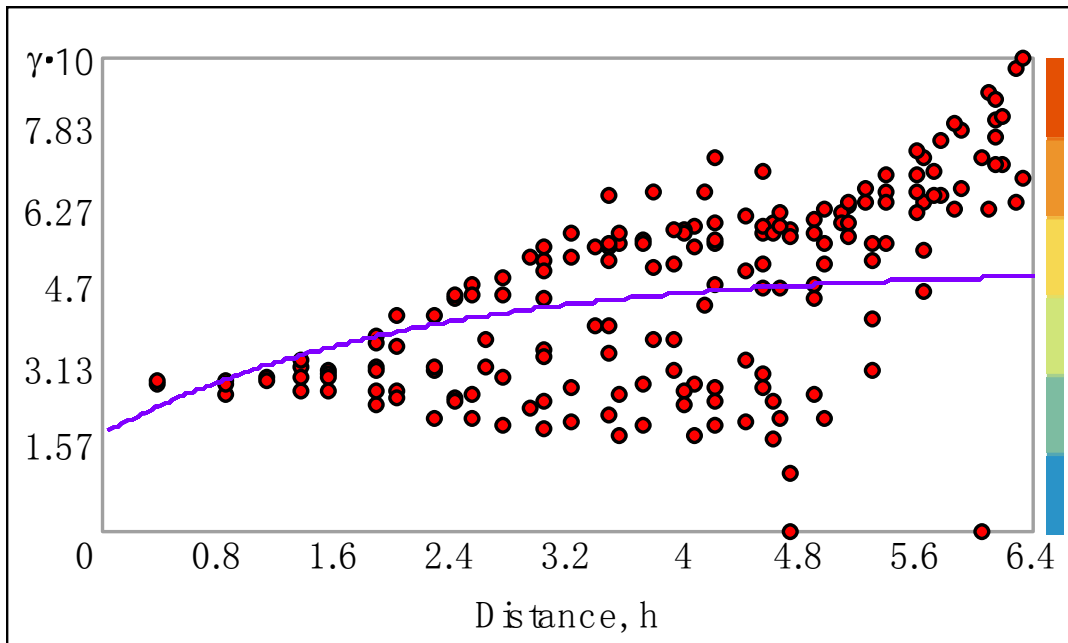


Figure 1. Classical Semi-Variogram

Table 1 presents estimation results for posterior means and 95% Bayesian credible intervals (BCIs) to compare performance of the three different OP models. The value 0 in the range of the BCI means that the parameter is not significant.

Estimates of the beta coefficients differ significantly among the three models. The posterior means for the spatial models (SGLMM and CG-SOP) appear smaller than those from the aspatial OP GLM. Perhaps, this confirms that some of the spatial dependence is explained by the correlation functions whereas these functions are absent in the aspatial model.

Differences were also observed between the competing spatial models. While the SGLMM estimates ϕ to be 0.035, the CG-SOP model estimate of ϕ is 0.041. These disparities are largely due to the differential parameterizations used for each model.

We use the ϕ estimate to determine the effective range of spatial correlation, commonly defined as the distance beyond which the correlation reduces to less than 5% of variance. For the exponential correlation function, the effective range is $\frac{2.3}{\phi}$ for the SGLMM, and $\frac{3}{\phi}$ for the CG-SOP model [20]. Thus, the estimated effective range for the response surface of the SGLMM is 66 kilometers, suggesting lower spatial variation compared to the CG-SOP model with an effective

range of 73 kilometers. Though both models appear to exhibit large ranges relative to the maximum inter-site distance of $d_{Max} = 67$ kilometers, the CG-SOP model does better at larger lags (see Figures 2 and 3).

Regarding our application, the estimation results (Table 1) show that, for the spatial models, all the risk factors are well within the 2.5% and 97.5% percentiles; being significantly different from 0. This shows clearly that these models were very successful at capturing key features of the observed data, thus, providing correct conclusions about the association between poverty-severity and the predictor variables.

The β_1 estimate in all three models shows a positive link between household size and poverty severity, showing that poverty severity levels tend to increase from non-poor to extremely poor for sites with higher household sizes. Age was found not statistically significant in the OP GLM. Higher education was found to reduce the risk of extreme poverty in all three models. Extreme poverty-risk was also related to the residency of household head, with urbanites being at lower risks of extreme poverty. The impact employment has on poverty-risk in this study is typical; showing lowest risk for wage employed householders in the SGLMM and CG-SOP models. Its effect in the non-spatial OP GLM was however mixed. Unemployed was found not to be statistically significant in this model.

Table 1. Comparison of Posterior Estimates using the Different OP Model Specifications

					Household
Model	Estimate	σ^2	ϕ	β_0 (Inter)	β_1 (hhsiz)
OP GLM	Mean	-	-	2.56	0.16
	95% BCI	-	-	(1.43, 3.74)	(0.11, 0.17)
SGLMM	Mean	0.61	0.035	1.80	0.14
	95% BCI	(0.23, 2.61)	(0.001, 0.015)	(1.42, 2.37)	(0.09, 0.25)
CG-SOP	Mean	0.92	0.041	1.72	0.11
	95% BCI	(0.21, 3.88)	(0.007, 0.25)	(0.41, 3.05)	(0.03, 0.30)

		Age	Education		
Model	Estimate	β_4 (age)	β_5 (Basic Edu)	β_6 (sec edu)	β_7 (ter edu)
OP GLM	Mean	0.12	-0.65	0.22	0.77
	95% BCI	(-0.13, 0.36)	(-0.78, -0.50)	(0.01, 1.68)	(0.67, 0.87)
SGLMM	Mean	0.12	-0.62	0.33	0.61
	95% BCI	(0.07, 0.40)	(-0.83, -0.43)	(0.18, 0.42)	(0.53, 0.67)
CG-SOP	Mean	0.09	-1.094	0.36	0.46
	95% BCI	(0.01, 0.36)	(-1.29, -0.89)	(0.31, 0.43)	(0.06, 3.31)

		Sex	Location	Employment	
Model	Estimate	β_2 (Female)	β_3 (Urban)	β_8 (unemp)	β_9 (wage)
OP GLM	Mean	-0.02	0.79	-0.14	0.27
	95% BCI	(-0.07, -0.01)	(0.67, 3.37)	(-0.15, 0.19)	(0.13, 0.29)
SGLMM	Mean	-0.28	0.39	-1.94	2.31
	95% BCI	(-0.30, -0.23)	(0.33, 0.42)	(-2.51, -1.40)	(1.74, 2.85)
CG-SOP	Mean	-0.15	0.30	-2.01	2.25
	95% BCI	(-0.17, -0.04)	(0.23, 0.37)	(-2.47, -1.60)	(1.79, 2.66)

Table 2 summarizes and compares results for predictive abilities while Table 3 does same for Bayesian expected loss (BEL) for each specification.

Clearly, results of Table 2 identified the CG-SOP as the preferred model for prediction. Its predicted values come closest to the truth compared to the aspatial cumulative OP GLM and the SGLMM. Closely following the CG-SOP in predictive performance is the SGLMM. Thus, incorporating spatial dependence in the modelling framework improves estimation and prediction as evidenced by the performance of the two spatial models.

Table 2. Validation Test on the Three OP Models using Measures of MPR and AMPR

Model	1	2	3	MPR	AMPR
Truth	6407	1166	802	-	-
OP GLM	3293	4832	250	0.56	0.75
SGLMM	4702	3332	341	0.41	0.44
CG-SOP	6598	909	868	0.37	0.40

For ordered models, values of AMPR provide information about how far the mis-predictions are from the true category. The OP-GLM is the worst performing in this regard, followed by the SGLMM (Table 2). The proportion of mis-predictions exceeding one category is 0.75 for the OP-GLM, reducing to 0.40 for the SG-SOP model.

A further comparison of predictive performance of the models was made using Bayesian expected loss (BEL). Calculations were based on training (BEL_{Train}) and validation (BEL_{Val}) data. A significantly large difference between BEL_{Train} and BEL_{Val} indicates a model is mis-predicting the outcomes [26]. Table 3 presents results of the analysis, showing the OP-GLM and SGLMM as culprits in this regard. The CG-SOP has the smallest difference between BEL_{Train} and BEL_{Val} .

Table 3. Validation Test on the Three OP Models using Measures of BEL

Model	$d_{jr} = 1$	
	BEL_{Train}	BEL_{Val}
OP GLM	0.50	0.66
SGLMM	0.28	0.47
CG-SOP	0.34	0.47

Overall, all the fit statistics suggest that the CG-SOP model more accurately estimates the parameters, with a better fit to the observed data, and should be preferred.

Predictions of poverty-severity risk, using the spatial models (SGLMM and CG-SOP), were made at unobserved locations across all ten regions of Ghana. Each thresholding level of the models was represented by a specific colour, thus thresholding at three levels produced a three-colour poverty severity map shown in Figures 2 and 3 respectively for the SGLMM and CG-SOP.

The predicted map exposes relatively large regional differences. The map shows higher concentration of extreme poverty in the northern part of the country, extending to the transitional belt of the savannah ecological zone, along the

border with Brong Ahafo Region. Large fluctuations are however observed within the Ashanti and Greater Accra regions and environs.

6. Discussion

Geo-statistical OP models that ordered the population into three distinct categories (non-poor, poor, and extremely poor) were employed. The models were built to forecast the risk of regional household poverty-severity in Ghana.

Inference compared MCMC simulations from the aspatial OP GLM and ordinal SGLMM with the CG-SOP model. The large difference in predictive ability between the aspatial and the spatial models confirms the limitation of the simple OP GLM, and by extension, all aspatial models when dealing with spatial correlation in geo-referenced data.

Bayesian inference via MCMC produced estimates of spatial parameters and subsequently, posterior predictive probabilities at unobserved sites. It is difficult to see how traditional aspatial methods such as the ordinary least squares (OLS) would estimate spatial parameters. The OP GLM for instance, though easy to formulate and estimate, does not sufficiently exploit sample information, thus producing biased results. Although the 95% confidence intervals (CIs) for the betas of the OP GLM (Table 1) were consistently narrower, this was hard to believe. Of course, as sample size increases, the CIs correspondingly decrease. However, if the increase in sample size results from the inclusion of more heterogeneous observations, narrowing of the CIs is anomalous.

The CG-SOP model results, summarized by the 95% credible intervals, and associated spatial parameters σ^2 and ϕ , show a relatively stronger spatial dependence. Results of the ordinal SGLMM are similarly distributed. However, the SGLMM overestimates ϕ suggesting a lower spatial variation than the CG-SOP model, especially at large lags (Figures 2 and 3). The results thus suggest that the CG-SOP is the preferred model for estimation and prediction.

Regarding our application, higher levels of extreme poverty is predicted in most of the northern half of the country, extending to the borders with Togo, Cote d'Ivoire, and Burkina Faso. This region lies in the dry savannah ecological zone with short seasonal rainfalls, rendering agricultural lands unproductive for the predominantly farming populations in the area.

Low poverty is predicted in much of the forested south and coastal parts of the country. The Greater Accra region is predicted with the lowest risk of extreme poverty, as expected. The seaports and heavy industrial plants (along with the recently discovered crude oil) are found predominantly across the middle and coastal belts of the country, where both the rich agricultural lands and tropical rain forest coincide. Heavy mining of gold, diamond and other mineral resources over the last half-century have all contributed significantly to the low levels of poverty predicted at various locations in the south of the country.

However, pockets of poverty observed at different scales in the southernmost parts of the country reflect urban and peri-urban deprivations. These could be due to disturbances linked to short range environmental factors, reiterating the

fact that even over much smaller distances, local disturbances can have a distinct effect on the distribution of the poor.

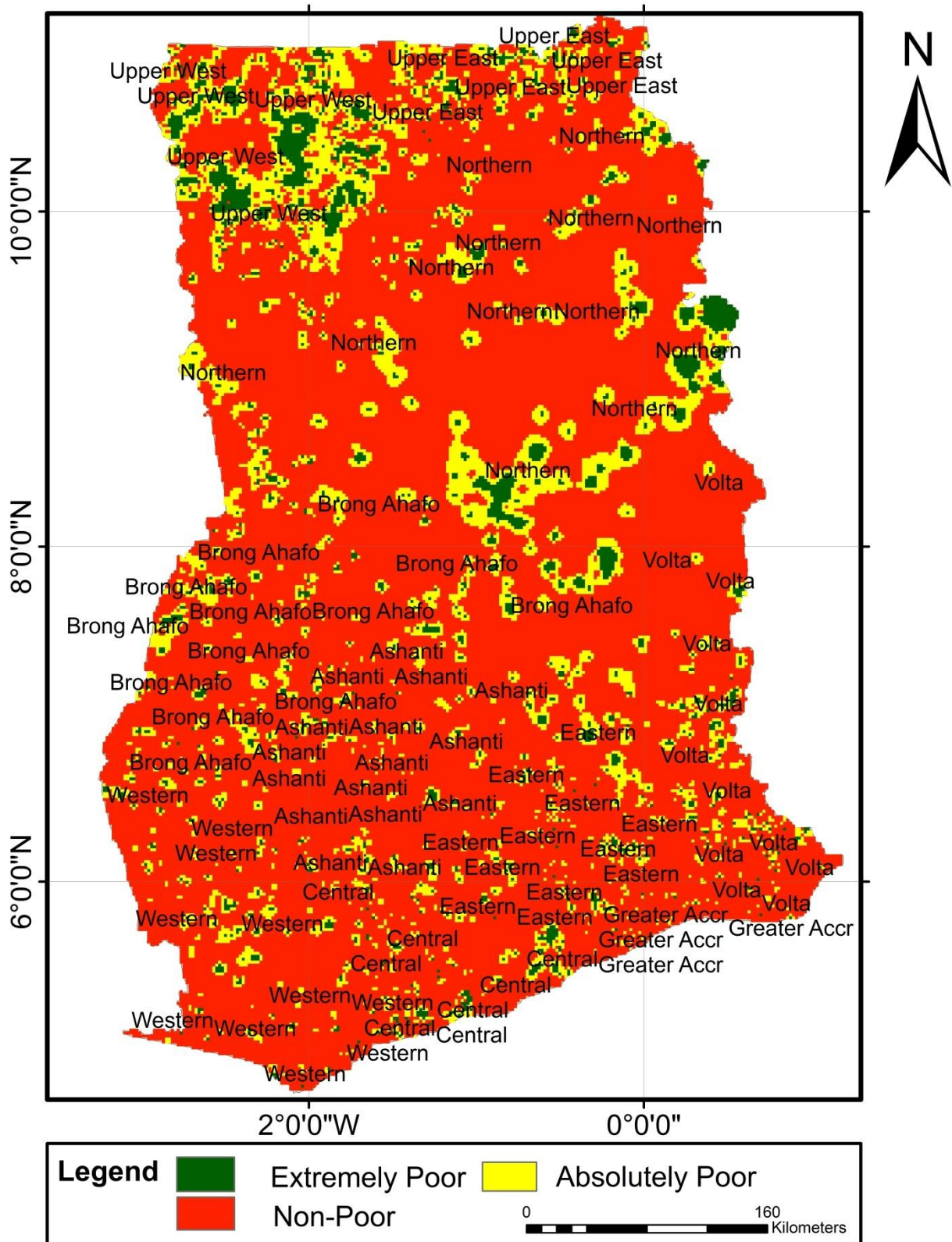


Figure 2. Poverty-Severity Map of Ghana Using the SGLMM

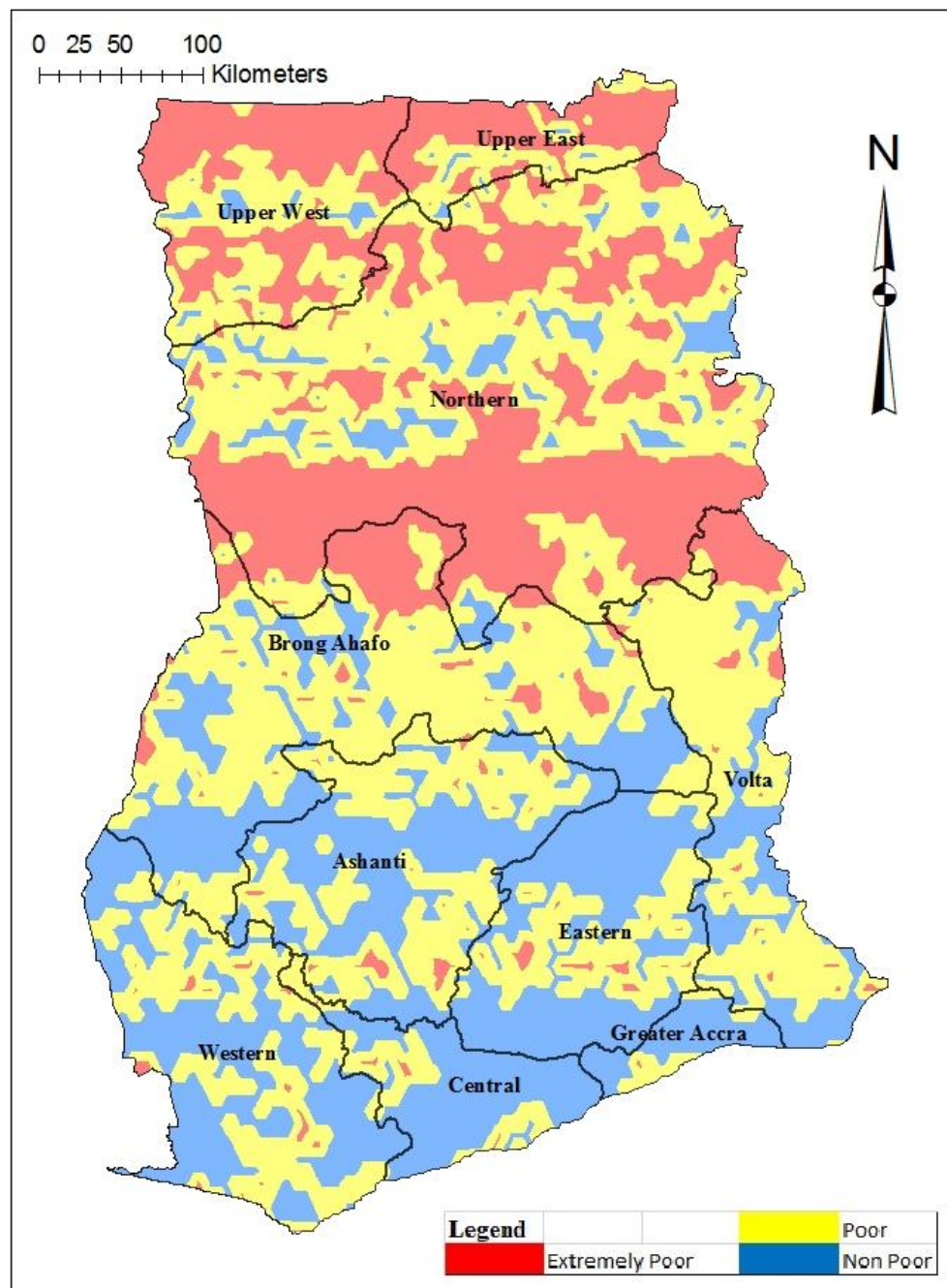


Figure 3. Poverty-Severity Map of Ghana using the CG-SOP Model

The spatial disparities of poverty-severity revealed by this study conform to expert opinion that in a geographic environment, there can be a dominant non-stochastic relationship between economic wellbeing and the spatial dynamics of a country [28, 29].

Application of the modelling techniques to the socio-econometric problem considered here is novel, and demonstrates our contribution to the wider scope of spatial-statistical methods. We model multi-categorical socio-econometric data in a spatial measurement framework that recognizes the ordinal nature of the variable. In our application, we depart from the auto regressive (AR)

approach by directly embedding the ordinal variable within the distributional framework of a latent spatial GRF, and this marks a significant innovation.

7. Conclusions and Recommendations

An integral objective of this study was to improve prediction, especially when the data is collected over large geographical areas. The results suggest that the CG-SOP model is well suited to this objective.

The study has identified positive correlation with respect to the distribution of poverty. Individuals in specific

locations tend to uniformly experience specific categories of poverty regardless of their personal circumstances, thus, providing an empirical spatial character of poverty in Ghana. We see a pictorial evidence of this in the predictive maps (Figures 1 and 2).

Poverty eradication efforts should be directed towards areas with high posterior ranks. For example, extreme poverty incidence in the Upper West and some areas at the border between the Northern and Brong Ahafo regions, extending eastwards, may be of concern. Conversely, stakeholders may wish to preserve areas with low incidence of extreme poverty.

8. Future Work

While we are satisfied with the CG-SOP model per our stated objectives, it needs to be noted that our model did not account for non-stationarity in the underlying spatial process. We assumed a stationary spatial process, meaning that spatial correlation between locations of the same distance remains the same throughout the region [30]. Non-stationarity allows both the data and covariates to vary spatially. Addressing non-stationarity further improves both model fit and prediction.

REFERENCES

- [1] Ennin C.C., Nyarko P.K., Agyeman A., Mettle F.O., 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, 2011, ISSN: 2040-7505.
- [2] Dudek, H., & Lisicka, I. (2013). Determinants of poverty–binary logit model with interaction terms approach. *Ekonomika*, 3 (41), 65–77.
- [3] Elena (Tomori), M., Zyka, E., Bici, R. (2014). Identifying Household Level Determinants of Poverty in Albania Using Logistic Regression Model (SSRN Scholarly Paper No. ID 2457441).
- [4] Minh Son Le (2015). Determinants of Urban Poverty in Vietnam: A Micro Analysis of Households in Ho Chi Minh City and Hanoi.
- [5] Puurbalanta, R., Adebajji, A. O. (2016). Household Poverty-Risk Analysis and Prediction Using Bayesian Ordinal Probit Models. *International Journal of Statistics and Applications*, 6(6), 399–407.
- [6] Finley, A. O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence: Spatially varying coefficients models. *Methods in Ecology and Evolution*, 2(2), 143–154.
- [7] Joshi S., Gebremedhin T., (2012). A spatial analysis of poverty and income inequality in the Appalachian region. *Journal of Rural and Community Development*, 7(2), 118–130.
- [8] Akinyemi, F. O., & Bigirimana, F. (2012). A Spatial Analysis of Poverty in Kigali, Rwanda using indicators of household living standard. *Rwanda Journal*, 26(1), 3-22.
- [9] Jitka Bartošová, Marie Forbelská (2013). Poverty Rate in Czech Households Depending on the Age, Sex and Educational Level. The 7th International Days of Statistics and Economics, Prague, September 19-21, 2013.
- [10] Juvanol, L. L., Lana, R. M., Cabrelli, R., Bastos, L. S., Nobre, A. A., Rotenberg, L., & Griep, R. H. (2016). Factors associated with overweight: are the conclusions influenced by choice of the regression method? *BMC Public Health*, 16. <https://doi.org/10.1186/s12889-016-3340-2>.
- [11] Agresti, A. (2007). An introduction to categorical data analysis (2nd ed). Hoboken, NJ: Wiley-Interscience.
- [12] William H. Greene, David A. Hensher (2009). *Modeling Ordered Choices*.
- [13] Leon-Novelo, L. G., Zhou, X., Bekele, B. N., & Müller, P. (2010). Assessing Toxicities in a Clinical Trial: Bayesian Inference for Ordinal Data Nested within Categories. *Biometrics*, 66(3), 966–974.
- [14] Roy, V., Evangelou, E., & Zhu, Z. (2016). Efficient estimation and prediction for the Bayesian binary spatial model with flexible link functions: Estimation and Prediction for Binary Spatial Models. *Biometrics*, 72(1), 289–298.
- [15] Kozintsev, B. (1999). Computations with Gaussian random fields. Retrieved from <http://drum.lib.umd.edu/handle/1903/6044>.
- [16] De Oliveira, V., 2000. Bayesian prediction of clipped Gaussian random fields. *Computational Statistics and Data Analysis* 34, 299-314.
- [17] Berrett, C. (2010). Bayesian Probit Regression Models for Spatially-Dependent Categorical Data. The Ohio State University.
- [18] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [19] ArcGIS® software by Esri. ArcGIS® and ArcMap™ www.esri.com.
- [20] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin (2014). *Bayesian Data Analysis*. Third Edition. CRC Press, Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.
- [21] Ghana Statistical Service (GSS), 2014. Ghana Living Standards Survey Round 6 (GLSS6): Poverty Profile in Ghana (2005-2013), Main Report, Accra, Ghana.
- [22] Diggle, P. J., Moyeed, R. A., & Tawn, J. A. (1998). Model-based Geostatistics. *Applied Statistics*, 47, 299–350.
- [23] Krige, D.G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem. Metall. Min. Soc. S. Afr.*, Dec.1951, 119-139.
- [24] Matheron, G. 1963. Principles of geostatistics. *Economic Geology*, Vol. 58, 1963, pp. 1246-1266.
- [25] Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234.

- [26] Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3), 523–539. <https://doi.org/10.1093/biostatistics/kxm049>.
- [27] Dmitrij Schlesinger (2013). *Machine Learning Bayesian Decision Theory*. Computer Vision Lab. Dresden.
- [28] Jalan, J. and M. Ravallion, 1997, Spatial poverty traps?, World Bank Policy Research Working Paper 1862.
- [29] David Bigman and Hippolyte Fofack (2000). Geographical Targeting for Poverty Alleviation: An Introduction to the Special Issue. *The World Bank Economic Review*, VOL. 14, NO. li 129-45.
- [30] Gaetan C., Guyon X. (2010). *Spatial Statistics and Modeling* (2010 edition). New York; London: Springer.