

# SCIENTIFIC REPORTS



OPEN

## Diarrhea Morbidities in Small Areas: Accounting for Non-Stationarity in Sociodemographic Impacts using Bayesian Spatially Varying Coefficient Modelling

F. B. Osei<sup>1,2</sup> & A. Stein<sup>2</sup>

Model-based estimation of diarrhea risk and understanding the dependency on sociodemographic factors is important for prioritizing interventions. It is unsuitable to calibrate regression model with a single set of coefficients, especially for large spatial domains. For this purpose, we developed a Bayesian hierarchical varying coefficient model to account for non-stationarity in the covariates. We used the integrated nested Laplace approximation for parameter estimation. Diarrhea morbidities in Ghana motivated our empirical study. Results indicated improvement regarding model fit and epidemiological benefits. The findings highlighted substantial spatial, temporal, and spatio-temporal heterogeneities in both diarrhea risk and the coefficients of the sociodemographic factors. Diarrhea risk in peri-urban and urban districts were 13.2% and 10.8% higher than rural districts, respectively. The varying coefficient model indicated further details, as the coefficients varied across districts. A unit increase in the proportion of inhabitants with unsafe liquid waste disposal was found to increase diarrhea risk by 11.5%, with higher percentages within the south-central parts through to the south-western parts. Districts with safe and unsafe drinking water sources unexpectedly had a similar risk, as were districts with safe and unsafe toilets. The findings show that site-specific interventions need to consider the varying effects of sociodemographic factors.

Disease indices, such as the relative risk, of common morbidities are important criteria for comparison of neighborhood health status, neighborhood health planning, and health budgetary allocations. With the changing population redistribution and deformation of sociodemographic characteristics, up to date statistic is required for planning. For diarrhea, reducing morbidity levels is closely linked to the attainment of the Millennium Development Goal 4 of reducing child mortality by two-thirds of the 1990 estimates. Globally, over 1.7 billion episodes are recorded every year with the majority of these occurring in low and middle-income countries<sup>1-5</sup>. Five important pathogens, rotavirus, enteropathogenic *E. coli*, enterotoxigenic *E. coli*, calicivirus, and *shigella*<sup>6,7</sup> cause the majority of diarrhea cases. The pathogens can spread from feces through multiple exposure pathways (water, food, flies, soil, fingers, fomites) which have a complex web of interactions<sup>8</sup>. The past years have seen much developmental efforts on therapeutic treatments such as oral rehydration therapy, zinc, and nutrient supplementation, leading to dramatic reductions in mortalities<sup>9,10</sup>. Morbidities, however, have declined only moderately<sup>4</sup>. Reduction of poverty and malnutrition, improving access to safe water and living conditions, provision of adequate sanitation and health care remain the formidable approach to reducing diarrhea in low and middle-income countries<sup>8,11</sup>. The inadequacy of resources demands prioritization of relatively high risks areas during interventions. The often available national and regional level estimates result in a major setback to prioritization. Local estimates are therefore necessary. These should be unbiased and reliable to inform good practices. Controlling for

<sup>1</sup>Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana.

<sup>2</sup>Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, Netherlands. Correspondence and requests for materials should be addressed to F.B.O. (email: [f.b.osei@utwente.nl](mailto:f.b.osei@utwente.nl)) or A.S. (email: [a.stein@utwente.nl](mailto:a.stein@utwente.nl))

sociodemographic effects could lead to precise risk estimates. Also, understanding diarrhea risk relationship with sociodemographic characteristics opens up the pathway for developing effective strategies to combat this menace.

Mapping diarrhea risk can provide a better understanding of the geographical variation of neighborhood health status. Ecological regression models have commonly been used to account for relevant confounding socio-demographic covariates to provide (1) reliable estimates of risk, and (2) estimates of the association between risk and covariates. When geographic information on neighborhoods is available, the inclusion of independent Gaussian and conditional autoregressive (CAR) processes as spatially varying intercepts could be adequate to account for residual spatial effects. These inclusions have the advantage to account for variance instabilities due to heterogeneous populations, unobserved influential factors, spatial interactions induced by similar sociodemographic conditions, and improve prediction accuracy. Wakefield<sup>12</sup> provides discussions on disease mapping using ecological regression models. These, however, are fixed effect models and estimate a single coefficient for each covariate based on the implicit assumption of stationarity in the effect parameters. For most diseases, given the complexity of the relationship with sociodemographic factors, non-stationarity in the impacts of neighborhood sociodemographic factors is plausible<sup>13</sup>. For diarrhea, the assumption of stationarity in neighborhood sociodemographic effects is difficult to meet because of differences in neighborhood specific characteristic and unobserved factors that can locally influence disease outcomes. A mix of attributable socioeconomic inequalities such as low-income level, illiteracy, inadequate water and sanitation, urbanization<sup>14–18</sup> are expected to exhibit spatially varied impacts. For instance, the effect of urbanization on diarrhea has two opposing facets; on the one hand, urbanization fueled by economic growth along with improvements in amenities such as availability safe water sources and good sanitation practices will reduce diarrhea morbidities. On the other hand, unplanned rapid urbanization fueled by rural-urban migration can enhance substantial increases in diarrhea morbidities due to stress on existing amenities which do not meet the demands of the rising population. If the neighborhood socio-demographic effects are allowed to vary by location, then we can remedy the preceding concerns.

The methodology in this paper is an alternative to the common ecological regression approach for disease mapping where covariates effects are assumed to be homogenous across neighborhoods. Non-stationarity impacts of the covariates may be accommodated by geographically weighted regression (GWR)<sup>13</sup>, or spatially varying coefficient (SVC) models<sup>19</sup>. GWR focuses on metric outcomes. Those, however, are less frequently encountered in disease mapping. The exception, however, is directly observed outcomes like incidence or prevalence rates. The robustness of the GWR relies on the selection of appropriate bandwidth for the kernel function which is dependent on the geographic locations of point reference data<sup>20</sup>. In this study, our interest lies in modeling the relative risk as an unobserved latent random variable. SVC models can easily be extended to latent random variable model and formulated within a hierarchical Bayesian framework. Gelfand *et al.*<sup>19</sup> provide extensive theoretical discussions on SVC models using Gaussian random fields on point referenced data, and extensions to the spatio-temporal domain. There are still specification and implementation concerns for discrete disease data. Hence its application is rarely found in disease mapping literature. Morbidity data often consist of aggregates for administrative areas. For these reasons, we focus on modeling the spatially varying coefficients as realizations of Conditional Autoregressive Processes (CAR). Our empirical application is motivated by diarrhea epidemiology which remains as one of the top 5 out-patient morbidities in Ghana. The manuscript intends to demonstrate the methodological significance and the substantive epidemiological implications.

We adopt a hierarchical Bayesian modeling framework and consider a variety of hierarchical models that account for spatial random effects, temporal random effects, and spatio-temporal interaction effects. Markov Chain Monte Carlo (MCMC) simulations of numerically evaluating complex integrals can be time consuming and discouraging. We used Integrated Nested Laplace Approximation (INLA) for approximate Bayesian inference<sup>21,22</sup>. The study has twofold objectives; (1) map area specific relative risk estimates, and (2) estimate the spatially varying association between relative risk and potential risk factors. In what follows, we present the model specifications and estimation methods, the data description and empirical applications. We finally end with the results and some discussions.

## Methods

We consider the double  $\{y_{it}, n_{it}\}$  representing the spatio-temporal outcomes of diarrhea and population data disaggregated by districts  $i = 1, \dots, m$  and temporal periods  $t = 1, \dots, T$ . Such sampling models are typically realizations from the Poisson process  $Y_{it} \sim P(E_{it}\zeta_{it})$  with likelihood

$$y_{it} | \zeta_{it} = \prod_{it} \frac{e^{[E_{it}\zeta_{it}]} (E_{it}\zeta_{it})^{y_{it}}}{y_{it}!} \quad (1)$$

where  $\zeta_{it}$  is the relative risk and  $E_{it}$  is the expected number of cases. Of particular interest to policy makers and health strategists are inequalities in the relative risk  $\zeta_{it}$ . The conventional estimate of area specific relative risk is the ratio of observed cases to population,  $\zeta_{it} = y_{it}/E_{it}$ , where the expected number of cases  $E_{it} = r_t n_{it}$  is defined, in the absence of covariates, as the number of cases defined in an epidemiologic “null model” of incidence. The risk variable  $r_t$  is the individual level constant baseline risk estimated from the aggregated population via  $\hat{r}_t = \sum_{i=1}^m y_{it} / \sum_{i=1}^m n_{it}$ . Population heterogeneity would influence variation in the relative risk estimates, and consequently, any estimated fixed effect parameter for areas of relatively small populations will show the highest variability<sup>23</sup>. Imposing independent normal prior distribution on the log-relative risk,  $\log \zeta_{it} \sim N(\alpha, \sigma_v^2)$ , results to a log-linear regression with exchangeable random intercepts that corrects for population heterogeneity amongst areas. Thus  $\log \zeta_{it} = \alpha + v_i$ , implying  $\zeta_{it} = \exp(\alpha + v_i)$ , where  $v_i \sim N(0, \sigma_v^2)$  and  $p(v) \propto \sigma_v^{-n} \exp\{-0.5\sigma_v^{-2}\sum v^2\}$ . Here,  $\alpha$  denotes the overall level of the relative risk on a log scale and  $v_i$  denotes the exchangeable random intercepts across districts. It is conceivable that the relative risks over a set of contiguous areas are likely to demonstrate

spatial correlation. Unmeasured confounders are potentially continuous in space and can exhibit spatial correlation. Such confounders are often accommodated by introducing a spatially correlated random effect term  $u_i$  which specifies the distribution of  $u_i$  conditional on the set  $u_{-i} = u_{j \neq i} = \{u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n\}$ . A widely used scheme of representation under irregularly shaped areas is the intrinsic conditional autoregressive (ICAR) prior<sup>24</sup> where the conditional distribution of  $u_i$  is given by

$$u_i | u_{-i=j} \sim N\left(\bar{u}_i, \frac{\sigma_u^2}{w_{i+}}\right) \tag{2}$$

$\bar{u}_i = \sum_j w_{ij} u_j / w_{i+}$  and  $w_{i+} = \sum_{-i=j} w_{ij}$ . In practice,  $w_{ij}$  is an  $n \times n$  weight matrix which captures the spatial proximity structure such that  $w_{ij} = 1$  if  $i$  and  $j$  are neighbors  $i \sim j$  (share a common boundary), and 0 otherwise. Setting  $w_{ii} = 0$  ensures that a location will not be used to predict itself. The above specification leads to the prior joint distribution

$$p(u_i | \sigma_u^2) \propto \sigma_u^{-n} \exp\left\{-0.5 \sigma_u^{-2} \sum_{i=1}^n \sum_{j < i} w_{ij} (u_i - u_j)^2\right\} \tag{3}$$

This prior is not proper as it is based on paired differences. Imposing a sum to zero constraint  $\sum u_i = 0$  ensures identifiability. The mean of the joint prior distribution  $p(u_i | \sigma_u^2)$  is therefore set to zero, where its precision matrix  $\Sigma_u$  has diagonal elements  $\sum_{-i=j} w_{ij} \sigma_u^{-2}$  and off-diagonal elements  $-w_{ij} \sigma_u^{-2}$ . For brevity, we write  $u_i \sim ICAR(w, \sigma_u^2)$ . To overcome the difficulty of choosing between the spatially structured and unstructured effects, one can sum up the two priors, resulting in the so-called *convolution priors* of the two independent components. Thus

$$\log \varsigma_{it} = \alpha + v_i + u_i \tag{4}$$

The term  $v_i + u_i$  can be interpreted as a random spatial adjustment to the overall intercept, and  $\alpha + v_i + u_i$  as varying intercepts across areas. To account for time and varying space-time effects, further parameterization of the model results in

$$\log \varsigma_{it} = \alpha + v_i + u_i + \rho_t + \varphi_{it} \tag{5}$$

where  $\rho$  accounts of the temporal processes and  $\varphi_{it} = \{\varphi_{1t}, \dots, \varphi_{nt}\}$  are space-time random interactions. Considering the years  $t = 1, \dots, T$  as Gaussian vectors, one can specify either first or second-order random walk processes for  $\rho_t$ . First and second-order random walk processes are specified as  $\rho_t = \rho_{t-1} + \Delta\rho$  and  $\rho_t = 2\rho_{t-1} - \rho_{t-2} + \Delta\rho$ , receptively, where  $\Delta\rho \sim N(0, \sigma_\rho^2)$ . The first-order random walk penalizes abrupt jumps, while a second-order random walk penalizes deviations from the linear trend  $2\rho_{t-1} - \rho_{t-2}$ . The space-time interaction effects  $\varphi_{it}$  may be specified either of the four ways: unstructured temporal and unstructured spatial effects, structured temporal and unstructured spatial effects, unstructured temporal and structured spatial effects, or structured temporal and structured spatial effects. Details of these specifications can be found in Knorr-Held<sup>25</sup>.

The model can further be adjusted to accommodate the joint effects of area-level continuous covariates  $x_{ip}, p = 1, \dots, P$ , and categorical covariates  $z_{ik}, k = 1, \dots, K$

$$\log \varsigma_{it} = \alpha + v_i + u_i + \rho t + \varphi_{it} + \sum_p \beta_p x_{ip} + \sum_k \gamma_k z_{ik} \tag{6}$$

where  $x_{ip}$  is the  $p$ th continuous regressor at location  $i$  with coefficients  $\beta_p$ , and  $z_{ik}$  is the  $k$ th categorical regressor at location  $i$  with coefficients  $\gamma_k$ .

The model can be reparameterized to accommodate spatial variability of the covariates effects by varying the coefficients across areas. A spatially varying coefficient version is

$$\log \varsigma_{it} = \alpha + v_i + u_i + \rho t + \varphi_{it} + \sum(\beta_p + \delta_{ip})x_{ip} + \sum(\gamma_k + \delta_{ik})z_{ik} \tag{7}$$

where  $\delta_{ip}$  and  $\delta_{ik}$  are differential spatially varying effects which account for varying effects of the covariates. Thus,  $\beta_{ip} = \beta_p + \delta_{ip}$  and  $\gamma_{ik} = \gamma_k + \delta_{ik}$  can be viewed as random slope processes. The common specifications for  $\delta_i = \{\delta_{ip}, \delta_{ik}\}$  are either ICAR processes  $\delta_i \sim ICAR(w, \sigma_\delta^2)$  or exchangeable Gaussian processes  $\delta_i \sim N(0, \sigma_\delta^2)$ .

**Bayesian Inference.** Let  $\psi_1 = \{\varsigma, \alpha, \beta, \gamma, \delta, v, u, \rho, \varphi\}$  be the full Gaussian latent (unobservable) field and  $\psi_2 = \{\sigma_\delta^2, \sigma_v^2, \sigma_u^2, \sigma_\rho^2, \sigma_\varphi^2\}$  be a vector of hyper-parameters. The components of  $\psi_1$  are conditionally independent with the sparse precision matrix  $\Omega_{ij} = 0$  for  $i \neq j$ . Hence we assume a multivariate normal prior  $\psi_1 \sim MVN(0, \Omega^{-1})$  with conditional density

$$p(\psi_1 | \psi_2) = (2\pi)^{n/2} |\Omega|^{1/2} \exp\left(-\frac{1}{2} \psi_1^T \Omega \psi_1\right) \tag{8}$$

The conditional density is a Gaussian random field of which the sparsity of  $\Omega$  gives rise to computational benefits. Consequently, we adopted a Bayesian hierarchical specification through the INLA to estimate the model parameters. INLA provides accurate estimates of the integrals through Laplace approximation, a deterministic algorithm proposed by Rue and Martino<sup>21</sup>. By the Bayesian paradigm, we require calculating the posterior distribution of the unknown parameters given the data. Hence, the model parameters are assumed random with

prior distributions assigned at each stage of the hierarchy. The modeling can be summarized under a three stage hierarchical framework; the data model, process model and parameter model. Thus:

Stage 1:  $y|\psi_1, \psi_2 \sim p(y|\psi_1, \psi_2)$

Stage 2:  $\psi_1|\psi_2 \sim p(\psi_1|\psi_2)$

Stage 3:  $\psi_2 \sim p(\psi_2)$

The joint posterior distribution of  $\psi_1$  and  $\psi_2$  given the data likelihood is

$$p(\psi_1, \psi_2|y) = \frac{p(\psi_1, \psi_2, y)}{p(y)} = \frac{p(y|\psi_1, \psi_2)p(\psi_1|\psi_2)p(\psi_2)}{\int_{\psi_1} \int_{\psi_2} p(y|\psi_1, \psi_2)p(\psi_1|\psi_2)p(\psi_2)d\psi_1d\psi_2} \tag{9}$$

The joint posterior is written in shorthand as  $p(\psi_1, \psi_2|y) \propto p(y|\psi_1, \psi_2) \times p(\psi_1, |\psi_2) \times p(\psi_2)$  since the denominator is integrated over the latent field parameters and the hyper-parameters. Using INLA, the integrals are numerically evaluated through nested Laplace approximations. The implementation of INLA can be summarized as follows:

Computation of the joint posterior of the hyper-parameters through nested approximations

$$p(\psi_2|y) = \frac{p(\psi_1, \psi_2|y)}{p(\psi_1|\psi_2, y)} \approx \frac{p(y|\psi_1, \psi_2)p(\psi_1|\psi_2)p(\psi_2)}{\tilde{p}(\psi_1|\psi_2, y)} \Bigg|_{\psi_1\psi_2=\psi_2^*(\psi_1)} \tag{10}$$

Next, a simplified Laplace approximation approach based on Taylor’s series expansion is used to approximate the posterior marginal  $p(\psi_1|y)$ . Thus

$$\tilde{p}(\psi_1|y) \approx \frac{p(\psi_1, \psi_2|y)}{\tilde{p}(\psi_{-i1}|\psi_{i1}, \psi_2, y)} \Bigg|_{\psi_{-i1}=\psi_{-i1}^*(\psi_{i1}, \psi_2)} \tag{11}$$

where  $\tilde{p}(\psi_{-i1}|\psi_{i1}, \psi_2, y)$  is the Laplace-Gaussian approximation to  $p(\psi_{-i1}|\psi_{i1}, \psi_2, y)$  and  $\psi_{-i1}^*(\psi_{i1}, \psi_2)$  is its mode. Lastly, the marginal posteriors are computed as  $\tilde{p}(\psi_{i1}|y) \approx \int \tilde{p}(\psi_{i1}|\psi_2, y)\tilde{p}(\psi_2|y)d\psi_2$ .

### Application

**Study area and data.** Ghana is centrally located on the west coast of Africa. It has a total land area of 238,589 km<sup>2</sup> and bordered by Cote d’Ivoire to the west, Togo to the east, Burkina Faso to the north, and the Atlantic Ocean to the south. The country consists of ten administrative regions which are subdivided into 216 districts. Population projection by the Ghana Statistical Service (GSS) at the end of 2014 puts Ghana’s population at 27,043,093. The disease data used for this study consist of yearly diarrhea morbidity records of outpatient departments (OPD) from 2010 to 2014. We obtained the data from the Centre for Health Information and Management (CHIM) of the Ghana Health Services (GHS). These data exist in aggregated format per administrative districts. The geographical scale of analysis was restricted to the 170 administrative districts of which data had been recorded. We obtained population and sociodemographic data from the Ghana Statistical Service (GSS). We constructed four sociodemographic indicators as risk factors of diarrhea. These included *unsafe drinking water* (*uw*), *unsafe toilet* (*ut*), and *unsafe liquid waste disposal* (*ud*), and *urbanization* (*ur*). We estimated *uw* as the percentage of the district’s population who do not have access to pipe-borne water (either in dwellings, outside dwellings, or public standpipes). We estimated *ut* as the percentage of the district’s population who do not have access to flush toilet, and *ud* as the percentage of the district’s population who dispose liquid waste either on the streets or the compound. The variable *ur* indicates the percentage of the district’s population who live in urban communities. We classified *ur* as rural (less than 30% urban population), peri-urban (30–70% of urban population) and urban (greater than 70% urban population).

**Model implementation.** The case study for model specification involved diarrhea outcomes disaggregated by  $i = 1, \dots, 170$  districts over  $t = 1, \dots, 5$  years. We fitted three separate models. We specified  $v_i \sim N(0, \sigma_v^2)$  for the unstructured spatial effects,  $u_i \sim ICAR(w, \sigma_u^2)$  for the structured spatial effects. For the temporal effects, we specified first-order random walk prior  $\rho_t = \rho_{t-1} + \Delta\rho, \Delta\rho \sim N(0, \sigma_\rho^2)$  because of the short temporal span of the data. Since  $u_i$  and  $\rho_t$  capture the spatially and temporally structured variabilities, we specified  $\varphi_{it} \sim N(0, \sigma_\varphi^2)$ .

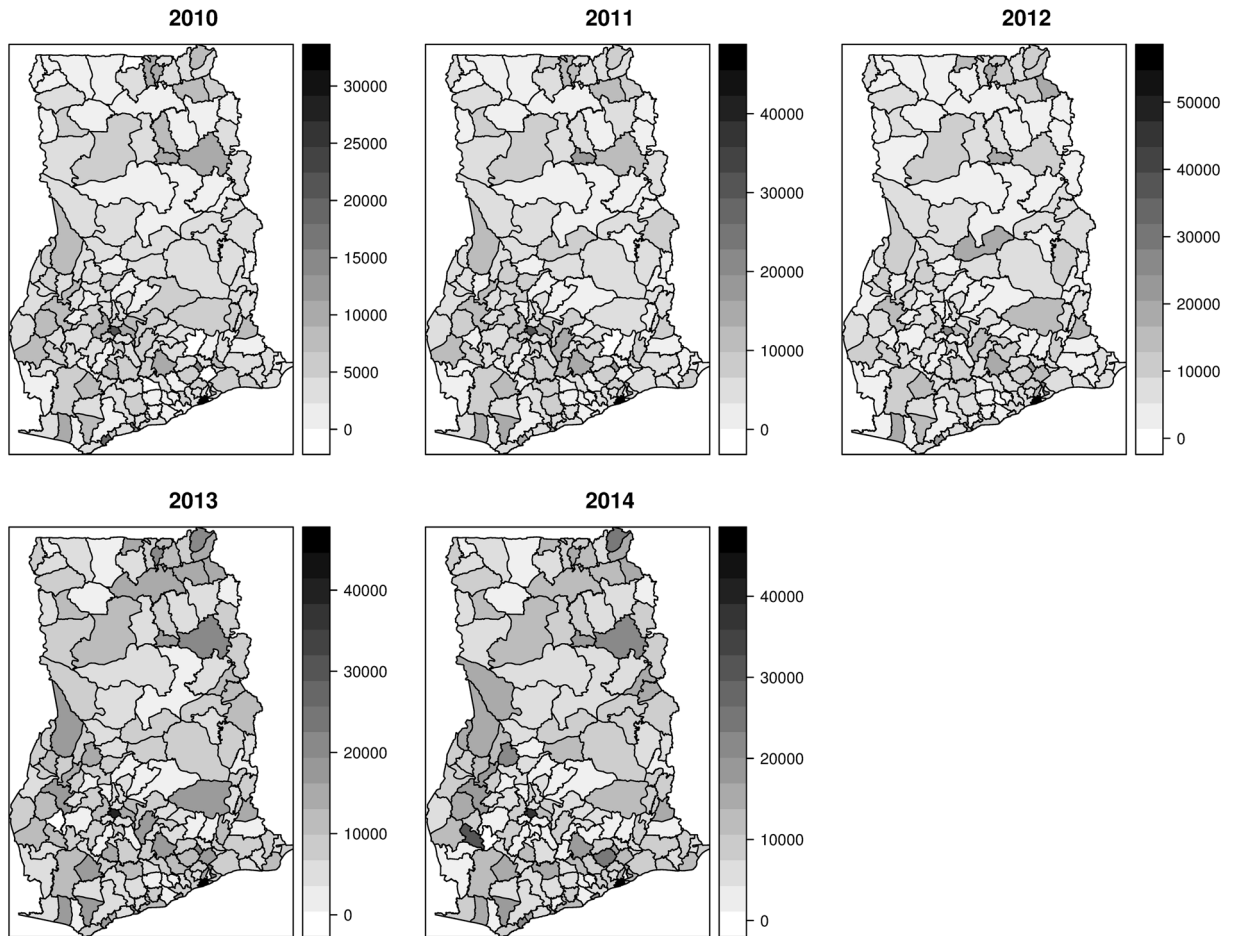
**Model 1:** Spatio-temporal fixed effect regression model.

$$\log s_{it} = \alpha + v_i + u_i + \rho_t + \varphi_{it} + \beta_{uw}uw + \beta_{ut}ut + \beta_{ud}ud + \gamma_{ur}ur \tag{12}$$

**Model 2:** This is a spatially varying coefficients model where  $\{\delta_{uw,i}, \delta_{ut,i}, \delta_{ud,i}, \delta_{ur,i}\} \sim N(0, \sigma_\delta^2)$ . The premise of this model derives from the assumption that the varying coefficients are Gaussian random processes.

$$\log s_{it} = \alpha + v_i + u_i + \rho_t + \varphi_{it} + (\beta_{uw} + \delta_{uw,i})uw + (\beta_{ut} + \delta_{ut,i})ut + (\beta_{ud} + \delta_{ud,i})ud + (\gamma_{ur} + \delta_{ur,i})ur \tag{13}$$

**Model 3:** The premise of Model 3 derives from evidence of spatial autocorrelation of the sociodemographic risk factors *uw*, *ut*, *ud*, *ur*. This will suggest ICAR process  $\delta_i \sim ICAR(w, \sigma_\delta^2)$  for the covariates with significant



**Figure 1.** Mapped cases of diarrhea, 2010 to 2014. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

spatial autocorrelation. We tested the risk factors for evidence of spatial autocorrelation using the Moran's index. We then fitted Model 3 as:

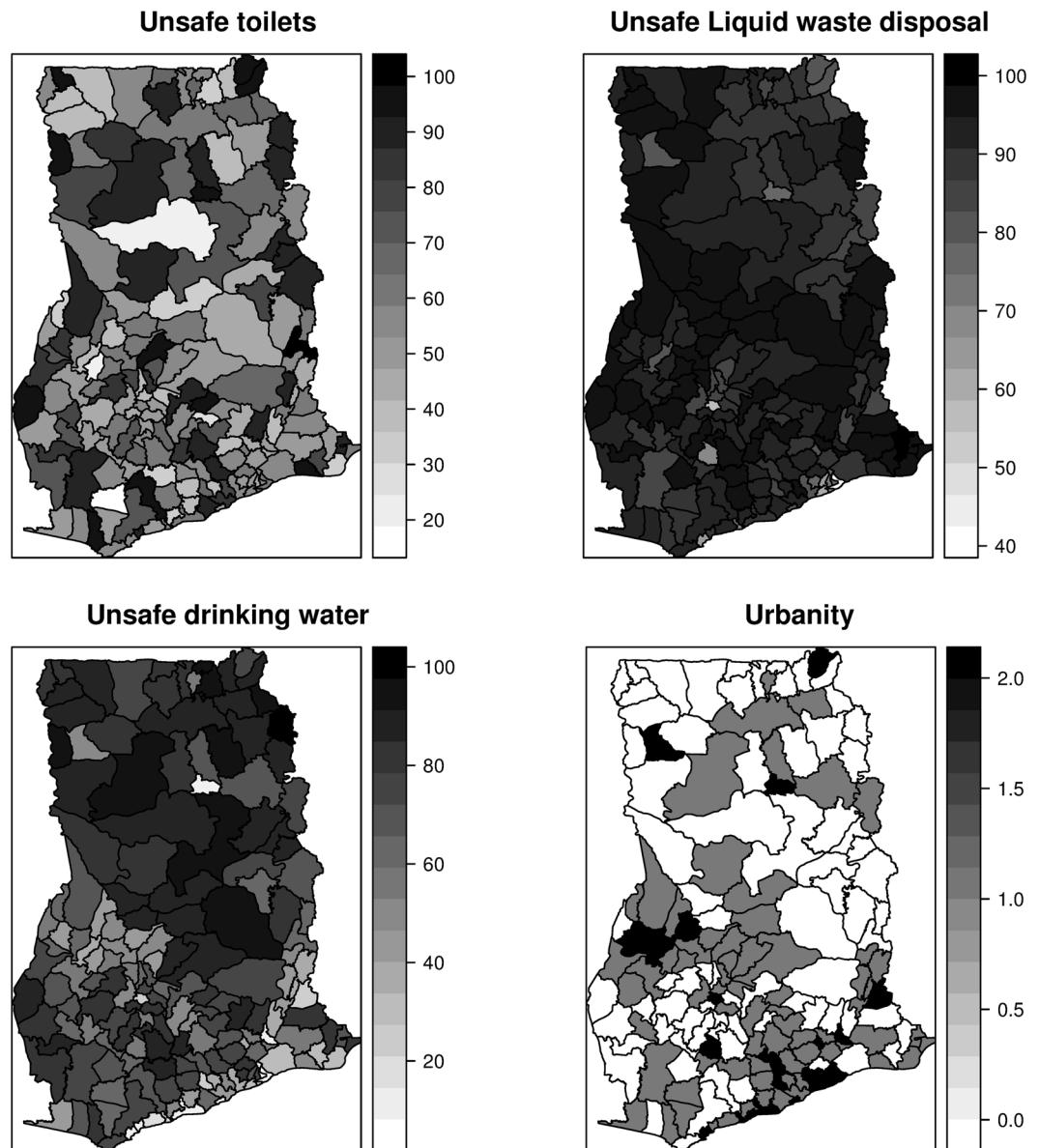
$$\begin{aligned} \log s_{it} = & \alpha + v_i + u_i + \rho_i + \varphi_{it} + (\beta_{uw} + \delta_{uw,i})wat \\ & + (\beta_{ut} + \delta_{ut,i})toi + (\beta_{ud} + \delta_{ud,i})ud \\ & + (\gamma_{ur} + \delta_{ur,i})ur \end{aligned} \tag{14}$$

where

$$\{\delta_{uw,i}, \delta_{uw,i}, \delta_{ud,i}, \delta_{ur,i}\} \sim \begin{cases} ICAR(w, \sigma_\delta^2) & \text{if Moran's index is significant} \\ N(0, \sigma_\delta^2) & \text{Otherwise} \end{cases}$$

For the prior parameters, we assigned non-informative Gaussian prior distribution with zero mean and precision  $10^{-5}$  for the fixed effects  $\{\beta_{uw}, \beta_{ut}, \beta_{ud}, \gamma_{ur}\} \sim N(0, 10^5)$ , and independent diffuse prior for the intercept  $p(\alpha) \propto const$ . For the precision parameters,  $\tau_l = 1/\sigma_l^2, l = \delta, u, v, \rho, \varphi$  which are at the lowest level of the hierarchy, the Gamma family are often chosen as conjugate priors. We assigned Gamma distribution of the form  $\log \tau_l \sim \log Gamma(1, 0.0005)$ .

We compared the predictive performances of the models using the deviance information criterion (DIC) and the mean square error (MSE). We estimated  $MSE = 0.5 \sum (y_{it} - y_{it}^*)^2$ , where  $y_{it}^*$  are the predicted counts of diarrhea. The  $DIC = \bar{D} + p_D$  is the sum of the model fit and model complexity<sup>26</sup>. Negative twice the log-likelihood of the deviance  $D(\psi_1|\psi_2) = -2 \log p(y|\psi_1, \psi_2)$  informs the model fit, while the effective number of parameter informs the model complexity. The effective number of parameters,  $p_D = \bar{D}(\psi_1, \psi_2) - D(\bar{\psi}_1, \bar{\psi}_2)$ , is measured through the difference between the posterior mean deviance  $\bar{D}(\psi_1|\psi_2) = E[D(\psi_1, \psi_2)]$  and the deviance of the posterior mean  $D(\bar{\psi}_1|\bar{\psi}_2) = D(E[(\psi_1, \psi_2)])$ . Like the MSE, the smaller the DIC value, the better the predictive performance of the model. We fitted all models using the R-INLA package<sup>27</sup> together with the R software<sup>28</sup> (R Development Core Team 2016).



**Figure 2.** Mapped estimates of the covariates. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

## Results and Analysis

**Distribution of covariates.** Mapped distribution of diarrhea counts from 2010 to 2014 are shown in Fig. 1. Cases ranged from 390 to 3140 in 2010, 590 to 45580 in 2011, 1345 to 54840 in 2012, 357 to 44670 in 2013, and 1069 to 45430 in 2014. Mapped estimates of the covariates are also shown in Fig. 2. The proportion of the population without safe toilets ranged from  $\approx 19\%$  to  $\approx 98\%$ ; the proportion without safe drinking sources ranged from  $\approx 8\%$  to  $\approx 98\%$ ; the proportion without access to safe liquid waste disposal ranged from  $\approx 42\%$  to  $\approx 99\%$ . Per our classification, 76 out of the 170 districts were dominated by rural dwellers, 67 out of the 170 districts were peri-urban, and urban dwellers dominated 27 out of 170 districts.

**Model selection.** We fitted multiple models and systematically increased the complexity from Model 1 to Model 3. Model 1 is a spatio-temporal regression model with fixed effects. Model 2 and Model 3 rather highlight the spatially varying effects of the sociodemographic factors, an indication of epidemiological benefits. Model 2 is also a spatio-temporal regression model that includes spatially unstructured varying coefficients of the covariates. Moving from Model 1 to Model 2 saw a change in fit as the DIC and the MSE values were reduced from 10601.95 to 10580.32 and from 209.55 to 201.42, respectively (See Table 1). Table 2 shows the results of the test of spatial autocorrelation on the sociodemographic risk factors. We found significant spatial autocorrelation for all the sociodemographic factors except for *ut*. Subsequently, we fitted Model 3 by specifying spatially structured varying coefficients  $\delta_i \sim ICAR(w, \sigma_\delta^2)$  for *uw*, *ud*, *ur*, and spatially unstructured varying coefficients  $\delta_i \sim N(0, \sigma_\delta^2)$

Variables	Model 1	Model 2	Model 3
$e^\alpha$	1.283 (0.249–3.980)	1.648 (0.140–7.051)	1.648 (0.140–7.051)
<b>Continuous</b>			
$e^{\beta_{uw}}$	0.996 (0.991–1.001)	0.996 (0.994–1.001)	0.996 (0.991–1.001)
$e^{\beta_{ut}}$	0.998 (0.994–1.002)	0.998 (0.991–1.001)	0.998 (0.993–1.002)
$e^{\beta_{ud}}$	1.117 (0.998–1.257)	1.095 (0.961–1.240)	1.115 (0.961–1.240)
<b>Categorical</b>			
$e^{\gamma_{ur}}$ (Rural) (Reference)	1.00	1.00	1.00
$e^{\gamma_{ur}}$ (Peri-Urban)	1.129 (0.936–1.347)	1.146 (0.953–1.362)	1.132 (0.985–1.254)
$e^{\gamma_{ur}}$ (Urban)	1.114 (0.773–1.550)	1.104 (0.760–1.545)	1.108 (0.759–1.547)
DIC	10601.95	10580.32	10567.47
MSE	209.55	201.42	198.53

**Table 1.** Results of the various models fitted.

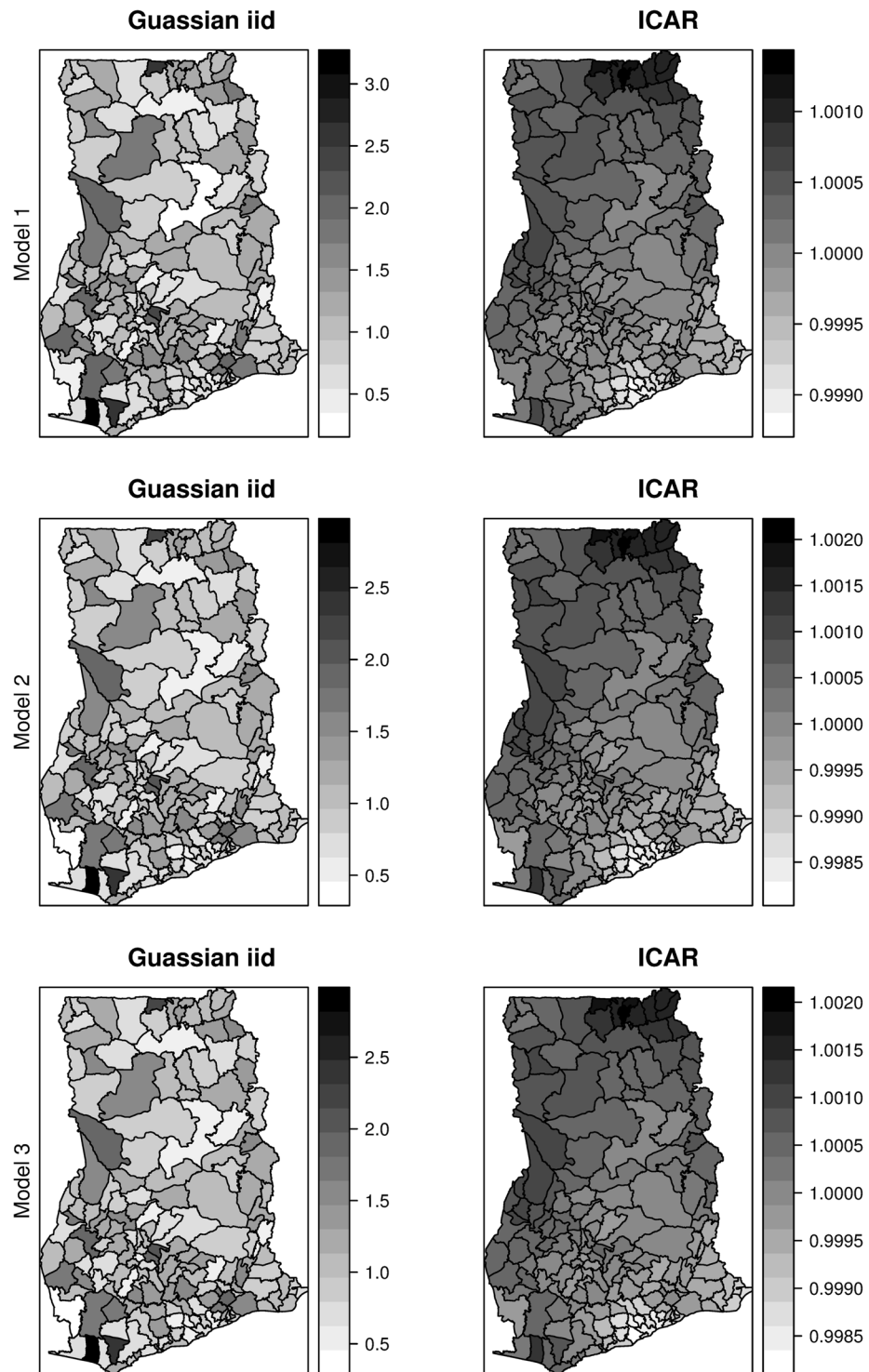
Variable	Moran's I	p-value
$uw$	0.364	<0.0001
$ut$	−0.092	0.0964
$ud$	0.388	<0.0001
$ur$	0.238	<0.0001

**Table 2.** Results of the Moran's Index of spatial autocorrelation for the varying coefficients in Model 2.

for  $ut$ . With the increased complexity, Model 3 adds some improvement regarding fit to Model 2 as the DIC and MSE values were slightly reduced. Both models 2 and 3 highlight a fundamental gain regarding the epidemiological implications on understanding the varying impacts of the covariates. Since the present analysis is concerned with the epidemiological implications of the varying regression coefficients, we focus our further analyses and discussions on Model 3, though the epidemiological implications of Models 2 and 3 will not be different. The relatively lower residual Gaussian random effects of Models 2 and 3 compared with Model 1 (Fig. 3) provide additional support for choosing the varying coefficients models over the fixed effect model.

**Distribution of posterior estimates of diarrhea risk.** All parameters have been transformed on the natural scale to facilitate easy interpretation of their impacts. In this respect, an additive change in the covariates has a multiplicative effect on the risk. Figure 4 shows the spatio-temporal distribution of the posterior estimates of the relative risks  $\varsigma_{it}$  from 2010 to 2014 after accounting for spatially random and structured effects, space-time interactions, temporal effects, and varying coefficient effects. We interpret these as model-based relative risks. The relative risks considerably contrast across space and appear consistent and notable over time with indistinguishable temporal trends. Areas with  $\varsigma_{it} > 1$  have higher than expected risk, while those with  $\varsigma_{it} < 1$  have lower than expected risk. We observed few isolated instances of exceptionally high risk and low risk which appear to emerge and disappear over time. The corresponding exceedance probabilities,  $p(\varsigma_{it} > 1)$  and  $p(\varsigma_{it} > 1.25)$ , are shown in Figs 5 and 6, respectively. The darker colors show areas of high probabilities, while the white colors show areas of low probabilities. Supported by the exceedance probability maps, patterns of districts with relative risk  $\varsigma_{it} > 1$  or  $\varsigma_{it} > 1.25$  seem spatially continuous. Increasing the exceedance from 1 to 1.25 still showed a majority of the districts with 25% higher than expected risks, except for 2011 and 2012 which had fewer districts. This is in coherence with the temporal patterns which showed a sharp decline in the risk from 2010 to 2012 and increased again from 2012 to 2014 (Fig. 7).

**Structured, unstructured, and temporal effects.** The proportion of variance explained by  $u_i$  out of the total spatial random effects is  $\approx 1\%$ , suggesting that unstructured heterogeneity  $v_i$  dominates the spatial variability. Additionally, maps of the residual spatial effects indicate unstructured heterogeneity outweighed the structured spatial effects (Fig. 3). The plots of the space-time interaction terms are shown in Fig. 8. These are viewed as residual effects after the spatially structured and unstructured, temporal, and covariates effects have been accounted for in the model. We observed notable spatial patterns as areas of similar values cluster. The temporal patterns appear random, which is in agreement with the modeling assumption. Districts with elevated estimates indicate

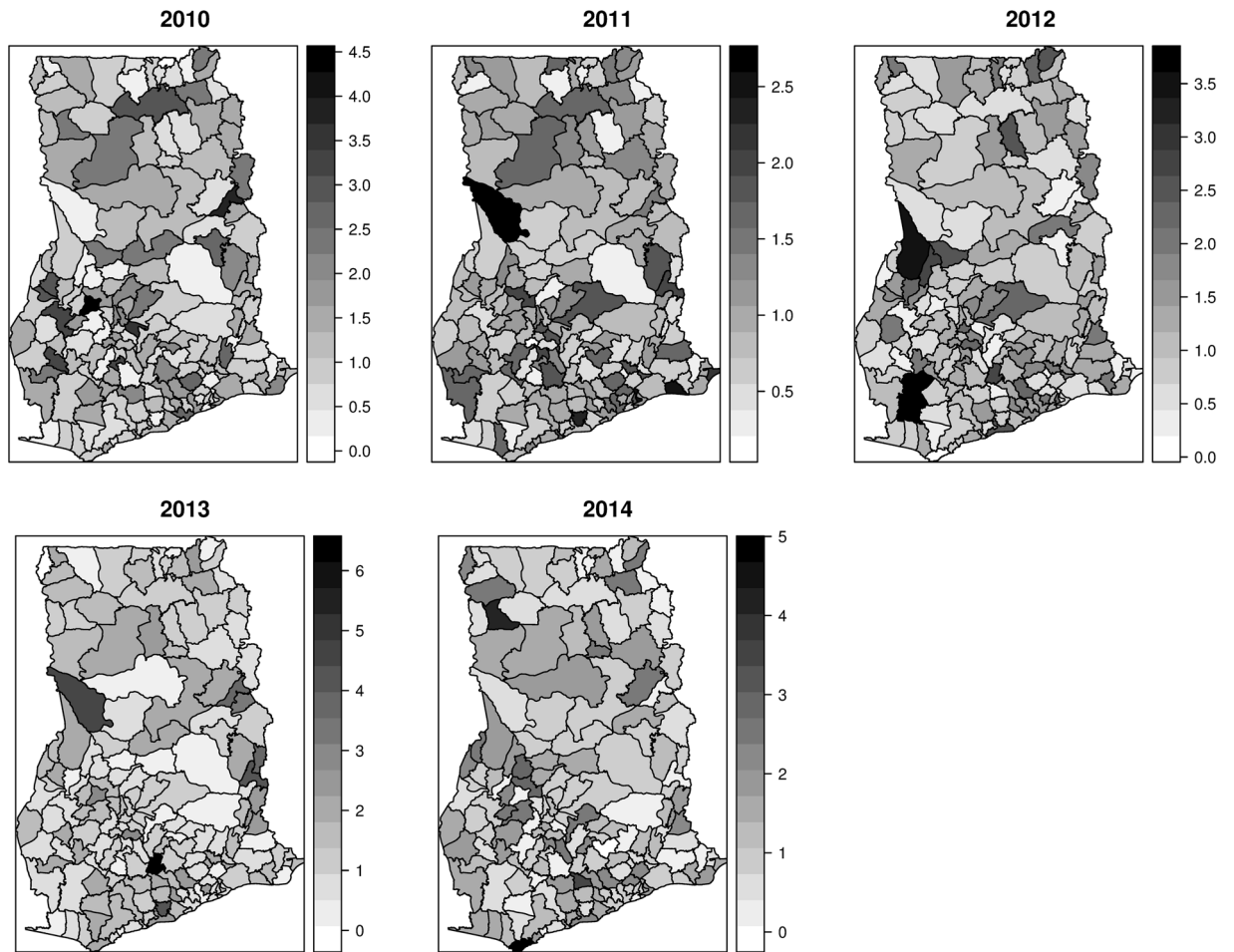


**Figure 3.** Structured and unstructured spatial effects for Models 1, 2, and 3. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

sporadic or short-term increases in the risk. This appears common amongst many districts in the study area, indicating signs of unstable or unusual temporal trend.

**Spatially varying effects of sociodemographic factors.** We observed that a unit increase in the proportion of people with unsafe liquid waste disposal increases diarrhea risk by 11.5%. This predisposes that diarrhea risk for inhabitants with unsafe liquid waste disposal is 11.5% higher than those with safe liquid waste disposal. The risk of diarrhea on average is 10.8% higher in urban ( $\gamma_{ur} = 1.108$ ) and 13.2% higher in peri-urban





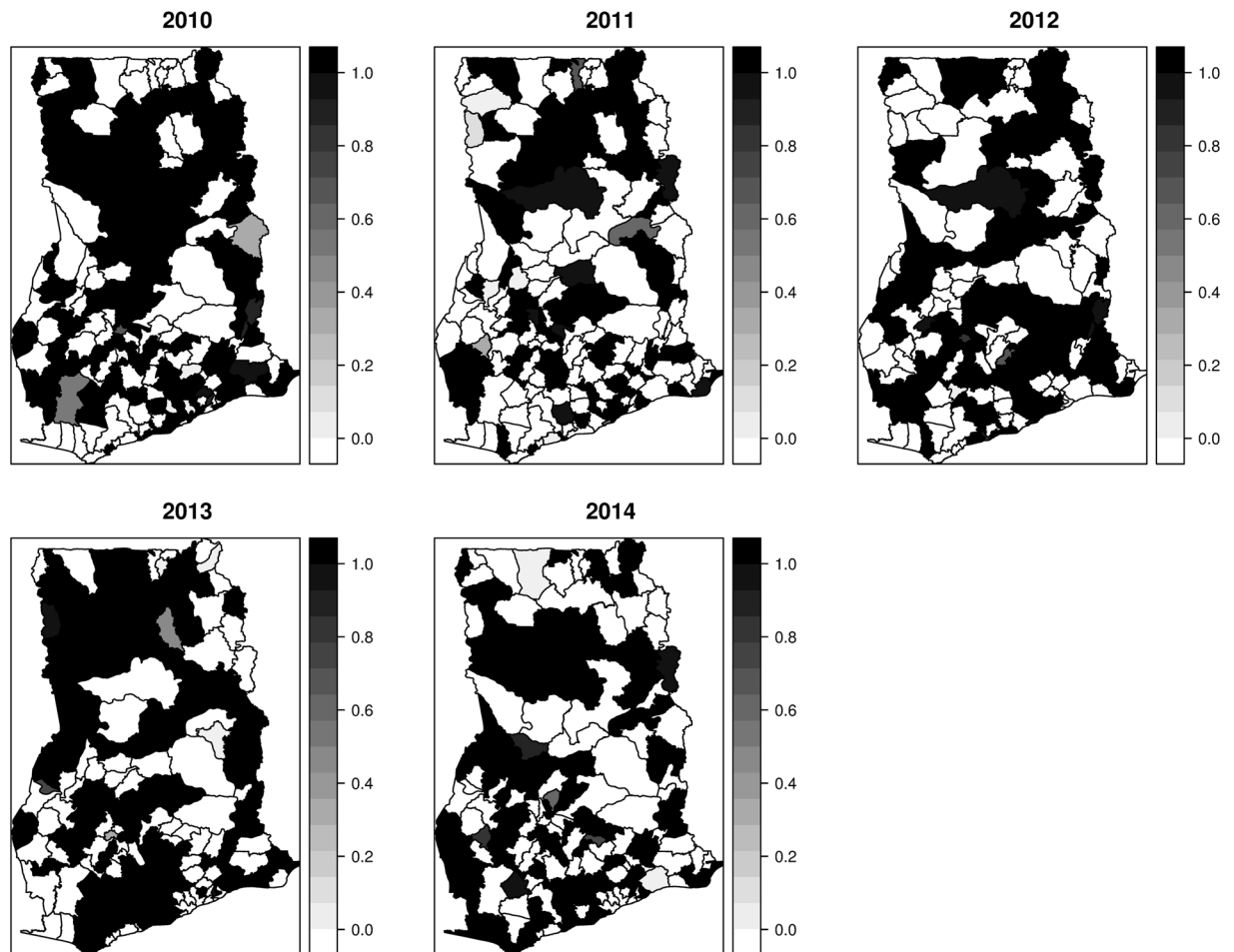
**Figure 4.** Spatio-temporal distribution of the posterior estimates from 2010 to 2014. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

( $\gamma_{ur} = 1.132$ ) districts than rural areas districts. The coefficients, however, varied marginally amongst districts (Fig. 9). For peri-urban districts, relatively higher coefficients occurred within northern and south-western parts, while low and no coefficients occurred within the south-eastern parts. The risk in urban districts relative to rural districts ranged varied geographically. We observed lower effects dominating within the southern parts and higher effects within the western parts (Fig. 9).

The multiplicative effect of *dis* was observed to be  $e^{\beta_{dis}} = 1.115$ , suggesting that diarrhea risk for inhabitants with unsafe liquid waste disposal is 11.5% higher than those with safe liquid waste disposal. The spatially varying coefficient model rather highlighted variation in these effects (Fig. 9). Higher effects were observed within the south-central parts through to the south-western parts. We observed no multiplicative effect for *uw*,  $e^{\beta_{uw}} \approx 1$  and that of *ut*,  $e^{\beta_{ut}} \approx 1$ . This suggests that differences in sources of drinking water and types of toilets could not account for variation in diarrhea risk. The varying coefficients  $e^{\beta_{uw,i}} \approx 1$  and  $e^{\beta_{ut,i}} \approx 1$  similarly showed no multiplicative effects across districts.

## Discussion

This study presented and illustrated the application of a Bayesian hierarchical varying coefficient model to study the sociodemographic effects on diarrhea morbidities, and developed model-based maps of the relative risks. We accounted for local variations in neighborhood covariate effects through spatially varying coefficients. The overriding consequence of our findings is that sociodemographic factors are spatially continuous, hence fixed effects models are inadequate to quantify covariate impacts. Neighborhood morbidities are often aggregated over discrete administrative units such as regions, districts, or census tracts and do not reasonably match the distribution processes of the disease. Disease outcomes misreported across neighborhood boundaries can consequently lead to spatial spillovers. Population heterogeneities also cause variance instability in risk estimates. Adopting spatially structured and unstructured hierarchical random effects model which pools strength over neighborhoods using ICAR and exchangeable intercepts account for the mentioned confounders. An intuitively convenient alternative is a hierarchical Poisson-Gamma model where hyper-priors are assigned to the gamma parameters through the hierarchy  $c_{ii}|a, b \sim G(a, b)$ , with  $a|\omega \sim h_a(\omega)$ ,  $b|\omega \sim h_b(\omega)$ , where  $h_a(\cdot)$  and  $h_b(\cdot)$  are the hyper-priors distribu-



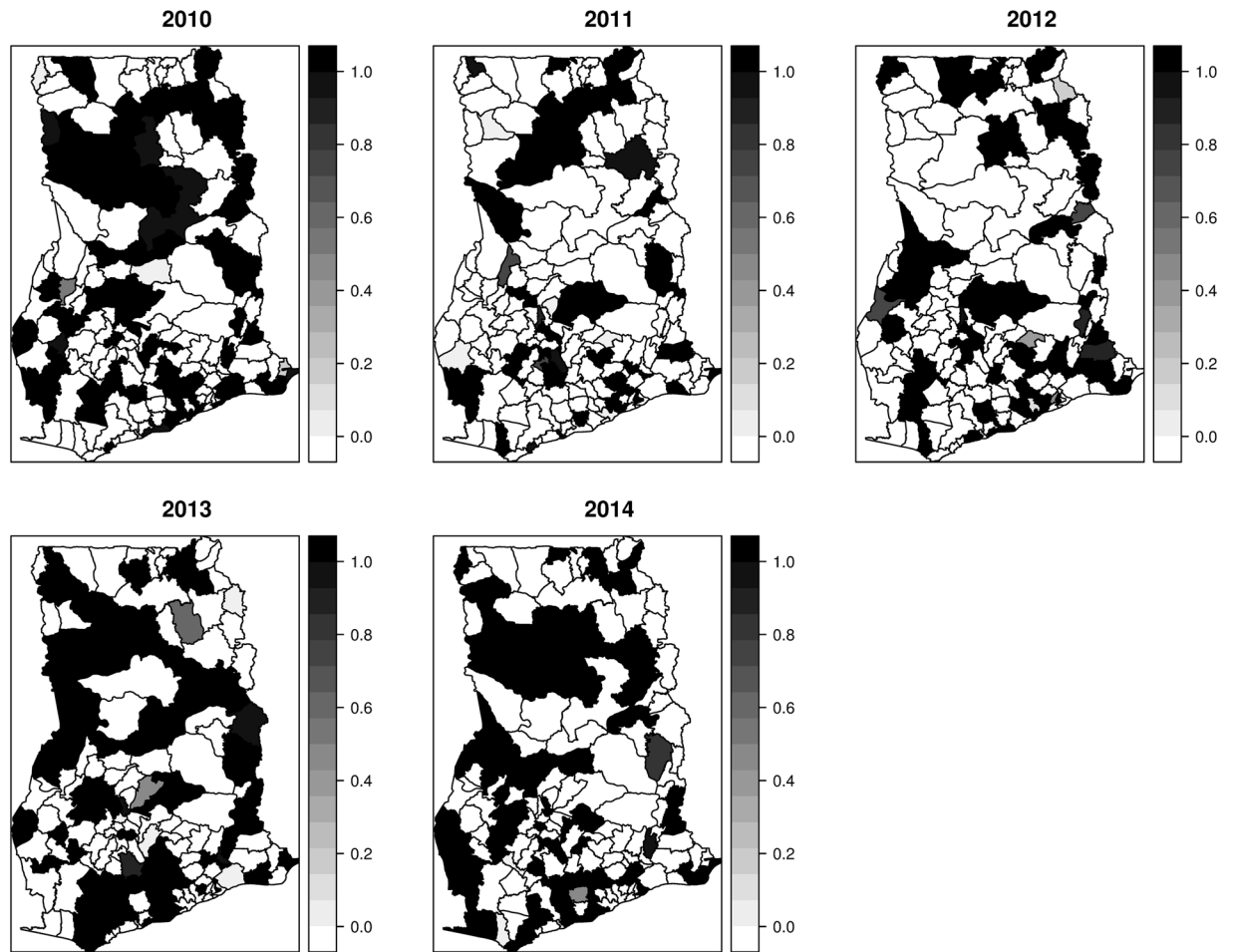
**Figure 5.** The exceedance probabilities of  $p(\varsigma_{it} > 1)$ , 2010 to 2014. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

tions for  $a$  and  $b$ , respectively. The major disadvantage of the Poisson-Gamma model lies in its inability to be easily generalized to cope with spatial correlation and covariate adjustment<sup>29</sup>.

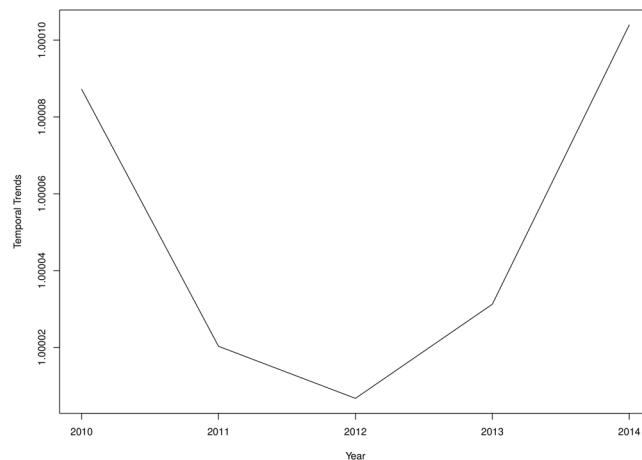
Literature on models which seek to account for varying covariate effects on dependence variables remains scarce, especially in epidemiological applications where non-Gaussian outcomes are common. A pioneering methodology is Fotheringham *et al.*'s<sup>13</sup> GWR which estimates varying coefficients by fitting  $n$  repeated regression models for separate likelihoods. The departure of our study is that we considered the relative risk as a latent random variable within a single likelihood, and adopted a random effects approach to accommodate all possible confounders. This approach is easily extendable to jointly accommodate independent Gaussian and Gaussian Markov random processes, temporal, and spatio-temporal processes.

Our empirical study indicates evidence of modeling benefit of the spatially varying coefficient models over the fixed effect model regarding fit and epidemiological significance. The relatively large number of districts in our study,  $n = 170$ , adds to the advantage of adding varying coefficients. The results suggest spatially varying effects of sociodemographic factors on diarrhea risk. There was minimal contrast in the varying coefficients, especially for the continuous covariates, probably because the sociodemographic data were based on population samples. Even though the improvement in fit and variations in the effects seem marginal, the epidemiological advantage of fitting a varying coefficient model cannot be overlooked. Thus, the varying coefficient models can provide added advantage in terms of understanding the substantive epidemiological implications of the varying effects of the sociodemographic factors. We argue that estimates of the varying effects could have revolutionary implications for guiding and prioritizing interventions.

The area specific relative risk maps can be interpreted as model-based relative risks (Fig. 4). These maps suggest considerable contrast across space and appear consistent and notable over time with indistinguishable temporal trends. This may be explained by the varying effects of sociodemographic factors. Site-specific interventions need to consider the relative importance of targeted multiple transmission routes<sup>8</sup>. The model-based risk maps suggest the importance of specific sociodemographic factors at specific locations necessary for reducing morbidities. We observed spatially varying associations between diarrhea risk and sociodemographic factors. Specifically, we found evidence of peri-urban and urban disadvantage as the risk in peri-urban and urban districts were higher than the risk in rural districts. Peri-urban areas are mostly transitional zones often neglected by urban planners. Constant pressure by increasing populations from rural population influx coupled with the high cost of

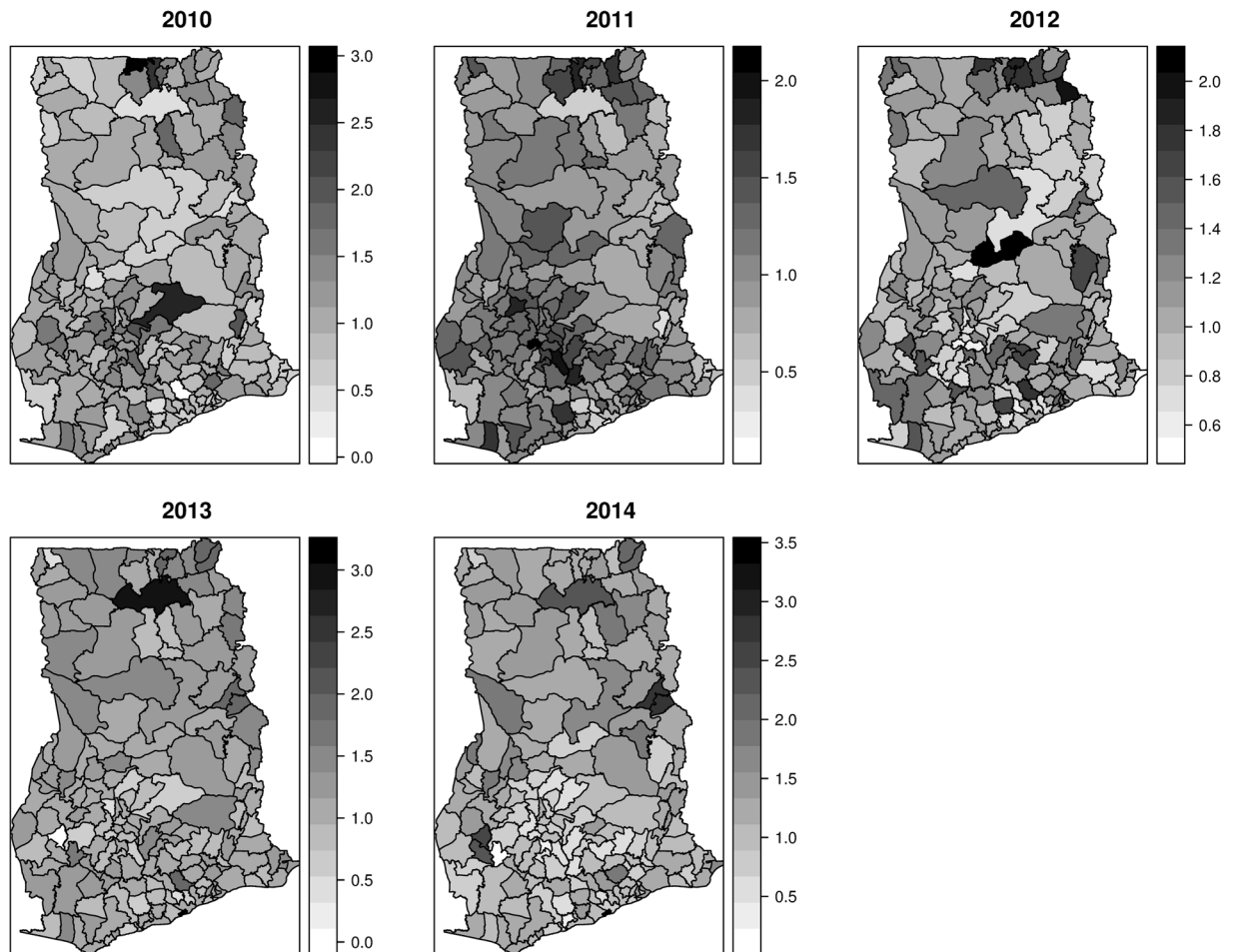


**Figure 6.** The exceedance probabilities of  $p(\varsigma_{it} > 1.25)$ , 2010 to 2014. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.



**Figure 7.** Temporal patterns of diarrhea risk from 2010 to 2014.

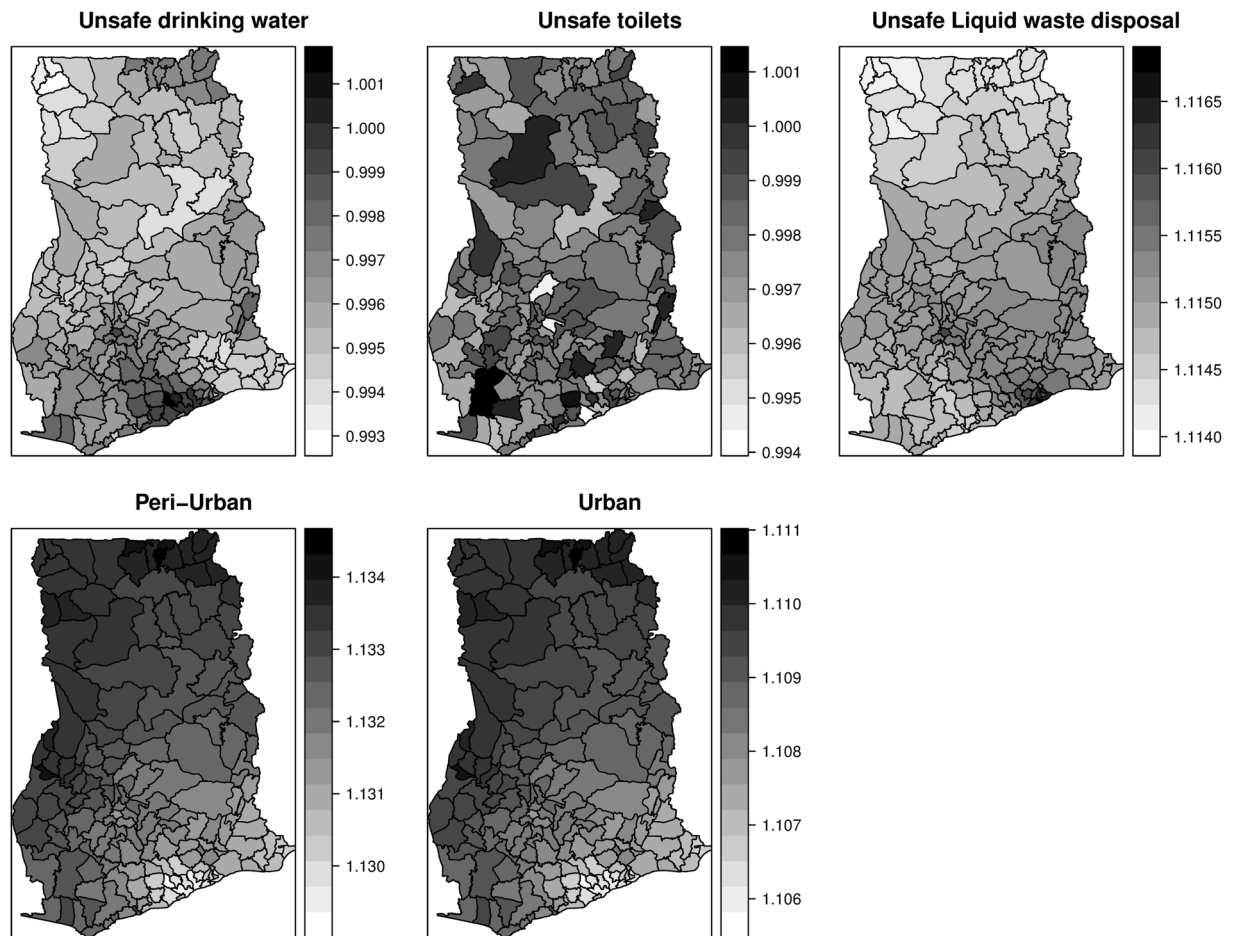
housing in urban districts heightens the potential for creating slums and informal settlements in both urban and peri-urban areas. Such settlements are often plagued with poor water and sanitation which are the well-known driving forces of diarrhea<sup>30, 31</sup>. This aspect of rural-urban differences in diarrhea risk has also been indicated by Kumi-Kyereme and Amo-Adjei<sup>32</sup> for childhood diarrhea in Ghana using Multiple Indicator Cluster Survey (MICS) data.



**Figure 8.** Space-time interaction terms. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

A percentage increase in the proportion of the population with unsafe liquid waste disposal was observed to increase diarrhea risk by 11.5%. The implication is that increasing access to safe liquid waste disposal could drastically reduce diarrhea morbidities. This finding was expected as improper disposal of liquid waste increases the potential for fecal contamination and increases the soil moisture content in the environment which plays an intermediate role in diarrhea transmission<sup>33</sup>. Some studies have evaluated this role as an important transmission concern for all diarrhea pathogens, especially in children<sup>34,35</sup>. Mika *et al.*<sup>36</sup> noted that with adequate moisture content, some diarrhea-causing fecal indicator bacteria diarrhea decays less rapidly in soil. The varying coefficients showed relatively lower coefficients for this effect for districts within the northern parts. This is probably because the northern parts are dominated by Guinea and Sudan Savannah ecological zones which have much drier soil and high temperature, and likely provide unfavorable environmental and ecological conditions for the survival of diarrhea-causing pathogens. In fact, high temperature has been shown to have a reduced effect on diarrhea<sup>37</sup>.

The observance of similar diarrhea risk amongst districts with safe and unsafe drinking water does not follow the predictable pattern. In the same manner, a recent study in Ghana observed that the incidence of diarrhea was not significantly associated with the quality of household water<sup>32</sup>. The attributable reason could be due to practices associated with fetching, storage, and handling which could contaminate even safe sources of water<sup>38</sup>. In Ghana, the availability of safe water sources in urban populations does not guarantee constant safe water availability due to intermittent supply prompted by high demands. Such urban dwellers, though have access to safe water, constantly resort to alternative water sources such as tanker-supply, and streams and rivers which may have higher levels of bacterial and viral counts. These alternative water sources can easily compromise the immune system due to their sporadic usage, increasing the susceptibility to infection. On the other hand, these alternative water sources for urban and peri-urban dwellers are mostly the main water sources for most rural dwellers who might have, due to their continuous usage, developed immunity. Comparable to differences in drinking water, differences in toilets also had no effect on diarrhea risk. The plausibility of this finding might be explained by the fact that flush toilet, which is considered as the safest, is most common amongst urban dwellers. It is, however, the worse in terms of sanitation when water for flushing is unavailable. Intermittent supply of piped water amongst urban dwellers with flush toilet may thus increase diarrhea risk. The implication is that the provision of safe water sources may not profoundly reduce diarrhea unless other mediating factors such as ensuring a constant supply of the safe water and household sanitation are taken into consideration.



**Figure 9.** Spatially varying effects of sociodemographic factors. These maps were created using R software (R Development Core Team 2016)<sup>28</sup>.

Our study diverges from other studies both in method and scope. Previous diarrhea related studies in Ghana and elsewhere have predominantly focused either on single geographic units or the characteristics of the affected individuals<sup>38–42</sup>. Unlike the prior studies, our study explains the spatial patterns and the varying effects of socio-demographic factors on diarrhea. In addition to the methodological significance, our study also highlights several noticeable epidemiological findings. While urbanization is an advantage to reducing water-borne diseases, its implication in developing countries can be daunting if not driven by the necessary requirements, such as availability of continuous safe water supply and access to safe toilet facilities. Although diarrhea risk was found to be high amongst peri-urban and urban districts than rural districts, these relativities varied in space. We observed comparable diarrhea risks amongst districts with safe and unsafe drinking water, indicating that accessibility does not necessarily mean availability. We, however, recognize that these results could be attenuated if all theoretically relevant variables were included. For instance, some studies have indicated that the poor in urban areas are advertently or inadvertently pushed to marginal areas where environmental health conditions are unsuitable for health<sup>30, 31</sup>. Such unsuitable areas are often associated with the creation of urban slums which lack safe water supply and sanitation which potentially combine to increase the risk of diarrhea. Further studies with interaction effects of urbanization and locality type (slums or non-slums) could demonstrate additional reasons for these findings.

Some limitations of this study deserve to be mentioned. First, this study relied on diarrhea data aggregated over districts level spatial units. Hence individual-level inferences would be inappropriate. Thus our study makes the implicit assumption of homogenous population and morbidity distribution within the districts which could lead to misspecification due to ecological bias<sup>43</sup>. Further studies using rigorous statistical estimations such as the log-Gaussian Cox processes<sup>44</sup> would be used in further studies to attenuate any possible misspecification. Secondly, the sociodemographic covariates were secondary data based on random sampling of individuals in the districts. Sampling size biases and the probability distribution of sampling outcomes of the covariates have not been accounted for in this study. Thirdly, only sociodemographic risk factors were included in this study, while environmental and climatic factors such as temperature, rainfall, and land use/land cover characteristics which might have an important impact on diarrhea have been absent. We have also not considered possible temporal changes in the sociodemographic risk factors due to data unavailability. Nonetheless, we believe this will have minimal effects on the results since most of the variability was captured by the spatial, temporal, and space-time

parameters. Our future studies seek to accommodate some of these limitations, especially augmenting sociodemographic risk factors with environmental factors extracted from remotely sensed images.

## Conclusion

Our study represents a contribution to spatial epidemiology literature by demonstrating both the methodological and epidemiological benefits of spatially varying coefficient modeling in estimating diarrhea risk. In contrast to the fixed effect model, the spatially varying coefficient model provided an added advantage to highlight varying effects of the sociodemographic risk factors. This has a practical implication of providing a scientific basis to facilitate precise targeting of site-specific risk factors for intervention. Our study disclosed the spatially varying nature of the relationship between diarrhea morbidities and urbanization, unsafe liquid waste disposal, unsafe toilets, and unsafe drinking water. The consequence of our results indicates that increased access to safe liquid waste disposal could reduce diarrhea drastically. Moreover, the results suggest that the provision of safe water and toilets will not assure reductions in diarrhea morbidities without ensuring a constant supply of the water. The maps of relative risk and the nature of the relationships provide empirical basis useful for guiding neighborhood health planning and resources allocation. Additionally, the study has provided a framework for health practitioners to (1) estimate, and map model-based area specific relative risk of diarrhea, (2) understand the relation between the relative risk and important sociodemographic factors. Further studies, identifying the spatially varying effects of climatic factors on diarrhea will be worthwhile. Moreover, another question worth further investigation is the sensitivity of the spatially varying intercepts and coefficients to complex definitions of proximity.

## References

- Black, R. E., Cousens, S., Johnson, H. L., Lawn, J. E. & Rudan, I. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* **375** (2010).
- Black, R. E., Morris, S. S. & Bryce, J. Where and why are 10 million children dying every year? *The Lancet* **361**, 2226–2234 (2003).
- Boschi-Pinto, C. Estimating child mortality due to diarrhoea in developing countries. *Bull. World Health Organ.* **86**, 710–717 (2008).
- Fischer Walker, C. L., Perin, J., Aryee, M. J., Boschi-Pinto, C. & Black, R. E. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC Public Health* **12**, 1–7 (2012).
- Parashar, U. D., Hummelman, E. G., Bresee, J. S., Miller, M. A. & Roger, I. Glass. Global Illness and Deaths Caused by Rotavirus Disease in Children. *Emerg. Infect. Dis. J.* **9**, 565 (2003).
- Lanata, C. F. *et al.* Global Causes of Diarrheal Disease Mortality in Children <5 Years of Age: A Systematic Review. *PLOS ONE* **8**, e72788 (2013).
- Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet* **382**, 209–222 (2013).
- Julian, T. R. Environmental transmission of diarrheal pathogens in low and middle income countries. *Env. Sci Process. Impacts* **18**, 944–955 (2016).
- Kosek, M., Bern, C. & Guerrant, R. L. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. **81**, 197–204 (2003).
- Bhutta, Z. A. *et al.* Interventions to address deaths from childhood pneumonia and diarrhoea equitably: what works and at what cost? *The Lancet* **381**, 1417–1429 (2013).
- Chan, M. & Lake, A. Integrated action for the prevention and control of pneumonia and diarrhoea. *The Lancet* **381**, 1436–1437 (2013).
- Wakefield, J. Disease mapping and spatial regression with count data. *Biostatistics* **8**, 158–183 (2007).
- Fotheringham, A. S., Brunsdon, C. & Charlton, M. E. *Geographically weighted regression: The analysis of spatially varying relationships.* (John Wiley & Sons, Chichester, 2002).
- Dasgupta, R. Exploring intra-household factors for diarrhoea diseases: a study in slums of Delhi. *India J Water Health* **6**, 289–299 (2008).
- Gupta, A., Sarker, G., Rout, A. J., Mondal, T. & Pal, R. Risk correlates of diarrhea in children under 5 years of age in slums of Bankura-West Bengal. *J Glob. Infect Dis* **2015**, 23–9.
- Mekasha, A. & Tesfahun, A. Determinants of diarrhoeal diseases: a community based study in urban south western Ethiopia. *East Afr. Med J* **80**, 77–82 (2003).
- Pande, S., Keyzer, M. A., Arouna, A. & GJS, S. B. Addressing diarrhea prevalence in the West African Middle Belt: social and geographic dimensions in a case study for Benin. *Int. J. Health Geogr.* **20087**, 17
- Woldemicael, G. Diarrheal Morbidity among Young Children in Eritrea: Environmental and Socio-Economic Determinants. *J. Health Popul. Nutr.* **19**, 83–90 (2001).
- Gelfand, A. E., Kim, H.-J., Sirmans, C. F. & Banerjee, S. Spatial Modeling With Spatially Varying Coefficient Processes. *J. Am. Stat. Assoc.* **98**, 387–396 (2003).
- Guo, L., Ma, Z. & Zhang, L. Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Can. J. For. Res.* **38**, 2526–2534 (2008).
- Rue, H. & Martino, S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *J. Stat. Plan. Inference* **137**, 3177–3192 (2007).
- Rue, H. & Held, L. *Gaussian Markov Random Fields: Theory and Applications.* (Chapman and Hall/CRC, 2005).
- Riggan, W. B., Manton, K. G., Creason, J. P., Woodbury, M. A. & Stallard, E. Assessment of spatial variation of risks in small populations. *Environ. Health Perspect.* **96**, 223–238 (1991).
- Besag, J., York, J. & Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* **43**, 1–20 (1991).
- Knorr-Held, L. Bayesian modelling of inseparable space-time variation in disease risk. *Stat. Med.* **19**, 2555–2567 (2000).
- Spiegelhalter, D. J., Best, N. G. & Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).
- Martins, T. G., Simpson, D., Lindgren, F. & Rue, H. Bayesian computing with INLA: New features. *Comput. Stat. Data Anal.* **67**, 68–83 (2013).
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria. <https://www.R-project.org/> (2016).
- Lawson, A. B., Browne, W. J. & Vidal Rodeiro, C. L. *Disease Mapping with WinBUGS and MLwiN.* (Wiley and Sons, Chichester, 2003).
- Goebel, A., Dodson, B. & Hill, T. Urban advantage or Urban penalty? A case study of female-headed households in a South African city. *Health Place* **16**, 573–580 (2010).

31. Fotso, J.-C. Urban–rural differentials in child malnutrition: Trends and socioeconomic correlates in sub-Saharan Africa. *Health Place* **13**, 205–223 (2007).
32. Kumi-Kyereme, A. & Amo-Adjei, J. Household wealth, residential status and the incidence of diarrhoea among children under-five years in Ghana. *J. Epidemiol. Glob. Health* **6**, 131–140 (2016).
33. Pickering, A. J. *et al.* Fecal Contamination and Diarrheal Pathogens on Surfaces and in Soils among Tanzanian Households with and without Improved Sanitation. *Environ. Sci. Technol.* **46**, 5736–5743 (2012).
34. Majo, F. *et al.* Formative Research on Hygiene Behaviors and Geophagy among Infants and Young Children and Implications of Exposure to Fecal Bacteria. *Am. J. Trop. Med. Hyg.* **89**, 709–716 (2013).
35. US EPA National Center for Environmental Assessment, W. D. & Moya, J. Exposure Factors Handbook 2011 Edition (Final Report). Available at: <https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252>. (Accessed: 2nd June 2017).
36. Mika, K. B. *et al.* Pilot- and bench-scale testing of faecal indicator bacteria survival in marine beach sand near point sources. *J. Appl. Microbiol.* **107**, 72–84 (2009).
37. Xu, Z. *et al.* Assessment of the temperature effect on childhood diarrhea using satellite imagery. *Sci. Rep.* **4** (2014).
38. Osumanu, I. K. Household environmental and behavioural determinants of childhood diarrhoea morbidity in the Tamale Metropolitan Area (TMA), Ghana. *Geogr. Tidsskr.-Dan. J. Geogr.* **107**, 59–68 (2007).
39. Benneh, G. *et al.* In *Environmental Problems and the Urban Household in the Greater Accra Metropolitan Area* (Stockholm Environment Institute, 1993).
40. Gyimah, S. O. *Interaction effects of maternal education and household facilities on childhood diarrhoea in sub saharan Africa, the case of.* (J Health Pop Dev Countries, 2003).
41. Krumkamp, R. *et al.* Gastrointestinal Infections and Diarrheal Disease in Ghanaian Infants and Children: An Outpatient Case-Control Study. *PLOS Negl Trop Dis* **9**, e0003568 (2015).
42. Shier, R. P. *et al.* Drinking water sources, mortality and diarrhoea morbidity among young children in Northern Ghana. *Trop. Med. Int. Health* **1**, 334–41 (1996).
43. Wakefield, J. & Lyons, H. Spatial Aggregation and the Ecological Fallacy. in *Handbook of Spatial Statistics* 541–558 (CRC Press, 2010).
44. Diggle, P. J., Moraga, P., Rowlingson, B. & Taylor, B. M. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Stat. Sci.* **28**, 542–563 (2013).

## Acknowledgements

We extend our sincere appreciation to the Disease Control Unit of the Ghana Health Service for providing all the necessary data and background information for this research.

## Author Contributions

F.B.O. conceived of the study and carried out the analysis and drafted the manuscript. A.S. conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017