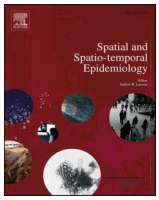
Spatial and Spatio-temporal Epidemiology 47 (2023) 100617

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste



Evolution of spatial disease clusters via a Bayesian space-time variability modelling

Frank Badu Osei

Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands

ARTICLE INFO

Keywords:
Bayesian
Clusters
Hot-spots
Space-time
Spatial
Temporal
Intestinal parasite infections

ABSTRACT

This study proposes to use exceedance posterior probabilities of a space-time random-effects model to study the temporal dynamics of clusters. The local time trends specified for each area is further smoothed over space. We modelled the common spatial and the space-varying temporal trend using a multivariate Markov Random field to incorporate within-area correlations. We estimate the model parameters within a fully Bayesian framework. The exceedance posterior probabilities are further used to classify the common spatial trend into hot-spots, cold-spots, and neutral-spots. The local time trends are classified into increasing, decreasing, and stable trends. The results is a 3 × 3 table depicting the time trends within clusters. As a demonstration, we apply the proposed methodology to study the evolution of spatial clustering of intestinal parasite infections in Ghana. We find the methodology presented in this paper applicable and extendable to other or multiple tropical diseases which may have different space-time conceptualizations.

1. Introduction

Spatial disease mapping has been vital in public health disease surveillance. The goal has been to investigate the geographical distribution of the diseases. In the end, one achieves the purpose of providing smoothed estimates of the risk or standardized rate ratios (SIR) and the exploration of the association between disease outcomes and environmental or demographic exposures. Findings from disease mapping may give clues to etiology and generate hypotheses and also inform public health resource allocation. The posterior probabilities from Bayesian estimates of disease mapping parameters can be used to detect and map higher than expected (hot-spots) and lower than expected (cold-spots) clusters of the SIR (Richardson et al., 2004). The recent availability of routine health information management systems has necessitated the development of space-time methods that accounts for spatial, temporal, and space-time interactions for dynamic disease mapping. Space-time disease mapping goes a step further to evaluate the space-time variations in the SIR; whose findings can further inform public health officials about the effectiveness of applied or existing interventions.

Space-time disease clustering and mapping have a long history and rich literature. Scan statistic cluster detection methods like the spatial scan statistic (Kulldorff, 2001, 1997) and the flexible scan statistic (Takahashi et al., 2008; Takahashi and Tango, 2005) have been utilized in many epidemiological studies (Azage et al., 2015; Hjalmars et al., 1996; Odoi et al., 2004). However, they have some limitations. First, their measurement of spatial proximity for area data is based on distances between the centroids. This obscures the effects of the varying shapes and sizes of areal features. Also, scan statistic methods cannot directly accommodate explanatory variables or confounders. Model-based space-time methods that utilize contiguity-based spatial proximity can present a comparable opportunity to investigate the evolution of small-area disease clusters. Bernardinelli et al. (1995) proposed a hierarchical model-based space-time variation of disease risk, where the time variation of each specific area is assumed linear with each time effect spatially dependent on those in neighboring areas. The spatial autocorrelation smoothing terms were specified via the conditional autoregressive (CAR) *prior* proposed by Besag et al. (1991). Their approach can be called separable ***parametric space-time*** modeling. A separable space-time ***nonparametric space-time interaction*** effect has been proposed by Knorr-Held (2000). Here, four different space-time interaction terms can be defined. Type I: unstructured temporal and unstructured spatial effects, Type II: structured temporal and unstructured spatial effects, Type III: unstructured temporal and structured spatial effects, and Type IV: structured temporal and structured spatial effects. The focus, as demonstrated in many space-time epidemiological studies, has been on providing space-time variation estimates for the diseases. An important aspect of disease surveillance is monitoring and evaluation of the space-time risk within hot-spots. Within the

E-mail address: f.b.osei@utwente.nl.

https://doi.org/10.1016/j.sste.2023.100617
Received 29 November 2022; Received in revised form 22 May 2023; Accepted 26 August 2023
Available online 27 August 2023

Bayesian framework, disease hot-spots can additionally be deduced from the residual risk using exceedance posterior probabilities with set cutoffs (Richardson et al., 2004). This can reduce the number of areas to monitor especially in resource-burdened countries. We argue that, if the evolution of the hot-spots through time is evaluated, then the number of areas needed to be monitored can further be reduced to optimize scarce resources. Also, our understanding of the disease cluster evolution can inform policymakers on what can become of the disease clusters in the future.

In a previous study, we adopted the Bayesian space-time model developed by Bernardinelli et al. (1995) to analyze the spatial clustering of the differential time trends and applied it to diarrhea incidences in Ghana. Our focus was to detect and map clusters of areas with increasing, decreasing, and stable time trends using Richardson and colleagues' cutoffs of posterior probabilities. In this study, we seek to answer the following questions: First, are the area-specific risks diverging from the same level, or converging to the same level over time? Second, how do the identified hot-spots/cold-spots evolve? Third, do areas that are neither hot-spots nor cold spots (neutral-spots) show the tendency to become hot-spots? The answers to these questions are of strategic importance to epidemiologists and public health managers. The last two questions are adopted from Li et al. (2014) who asked and answered similar questions for the space-time variability of burglary risk mapping. They described a two-step univariate classification which proceeds as follows. First, areas are classified as hot-spots, cold-spots, or neutral-spots. Next, the temporal behaviors of these areas are classified as decreasing trend, increasing trend, or neither. Like in our previous study, we will adopt the Bayesian space-time model developed by Bernardinelli et al. (1995). Afterward, we will adopt the two-stage classification described above to study the evolution of the clusters. We apply our models to study the space-time variability of intestinal parasite infections in Ghana. Intestinal parasites remain a major public health burden in developing countries, especially in sub-Saharan Africa. This disease infects more than 1 billion people, although there are concerns about the actual number of infections (Bethony et al., 2006; de Silva et al., 2003). Intestinal parasites can greatly comprise the quality of life of children of school-going age. Understanding the evolution of the hot-spots/cold-spots and the potential for neutral-spots to turn into hot-spots plays a fundamental public health role to guide monitoring and intervention. The structure of the paper is as follows. First, we describe the statistical modeling as applied to intestinal parasite infection data in Ghana. Next, we present the findings and discuss their interpretation. We conclude with some implications of the findings.

2. Statistical modeling

Consider the space-time disease counts y_{ij}, over the finite set of $i = 1, \ldots, M$ discretized spatial entities (districts in this case) and $j = 1, \ldots, J$ time steps. We assume that the counts are conditionally independent realizations from the Poisson distribution with mean $\lambda_{ij} = E_{ij}r_{ij}$ and likelihood

$$L(y_{ij}|r_{ij}) = \prod_{i=1}^{M}\prod_{j=1}^{J}\left(\frac{e^{-\lambda_{ij}}\lambda_{ij}^{y_{ij}}}{y_{ij}!}\right) \tag{1}$$

where $\lambda_{ij} = E_{ij}r_{ij}$ is the mean of the Poisson distribution and r_i is the unknown relative risk. The Poisson data generating process enforces mean and variance equality; $\mathbb{E}(y_{ij}|r_{ij}, E_{ij}) = E_{ij}r_{ij}$ and the variance $var(y_{ij}|r_{ij}) = E_{ij}r_{ij}$. The expected number of cases is $E_{ij} = n_{ij}\tilde{r}_j$ after accounting for district-level population differentials n_{ij}, where $\tilde{r}_j = \sum_i y_{ij} / \sum_i n_{ij}$ is the overall risk for time j within the study area. We write $y_{ij}|r_{ij} \sim P(E_{ij}r_{ij})$.

Next, following Bernardinelli and colleagues (Bernardinelli et al., 1995), we decompose the log of the unknown space-time relative risk into a spatial pattern common to all the time steps, a linear temporal trend common to all areas, and a space-time interaction term that allows different time trends for different areas. Thus, $\log r_{ij} \sim N(\eta_{ij}, \sigma_j^2)$, where σ_j^2 is the variance and

$$\eta_{ij} = \beta_0 + (\overline{\beta}_1 + u_{1i})t_j^* + \sum_q x_{iq}\gamma_q + u_{0i} \tag{2}$$

Here, the parameter β_0 denotes the overall relative risk on the log scale, and $\overline{\beta}_1$ is a fixed effect parameter for the overall time trend. The parameter u_{1i} accounts for district-specific spatially structured differential time trends. Inferentially, $\beta_{1i} = \overline{\beta}_1 + u_{1i}$ specifies the district and disease-specific temporal trends. The time steps are centered at the mid-observation period $t_j^* = j - N_j^{-1}\sum j$. The coefficients γ_q, $q = 1, \ldots, Q$ are the fixed effects of Q vulnerability or exposure variables x_{iq}. Here, we have the spatially structured intercepts u_{0i} which account for residual spatial variations. The spatially structured intercepts also account for unobserved ecological factors which may give rise to spatial clustering. The additive common spatial and temporal trends $\overline{\beta}_1 t_j^* + u_{0i}$ of the model can be termed as the stable component. The component allowing space-time interaction or local deviations from the common trends $u_{1i}t_j^*$ can be termed as the unstable component. We refer to (1) as a log-Gaussian parameterization. The alternative, yet common structure, is the log-linear parameterization $y_{ij}|r_{ij} \sim P(E_{ij}r_{ij})$ with $\log r_{ij} = \beta_0 + (\overline{\beta}_1 + u_{1i})t_j^* + u_{0i} + v_{0i}$, where v_{0i} is the unstructured spatial random variation. The log-Gaussian and log-linear models are inferentially the same. But the log-linear parameterization has Markov chain Monte Carlo (MCMC) convergence issues. The spatial random effects, when modelled as the sum of structured and unstructured variations, have only their sum identifiable. This can result in MCMC convergence problems as observed in a previous study (Osei et al., 2022).

We now look at modeling the random effects u_{0i} and u_{1i}. Each of the random effects could be modelled separately with the univariate conditional autoregressive (CAR) *prior*. The incorporation of correlation between the random slopes and intercepts will provide the answer to our first research question; i.e. are the area-specific risks diverging from the same level or converging to the same level over time? A previous study (Osei and Stein, 2019) has explored different mechanisms for inducing correlations between random intercepts and slopes. The bivariate conditional autoregressive (BCAR) model was found to be competitive over the others. The univariate CAR model extends naturally to the BCAR specification by replacing the univariate normal conditional distribution with a bivariate normal (*BN*) conditional distribution. Let $\boldsymbol{u_i} = (u_{0i}, u_{1i})'$ be a vector of the random intercepts and slopes, and $\boldsymbol{u_{-i}} = (u_{0,-i}, u_{1,-i})'$ be the vector of all random effects except $\boldsymbol{u_i}$. The conditional distribution is a bivariate normal distribution $\boldsymbol{u_i}|\boldsymbol{u_{-i}} \sim BN\left(\overline{\boldsymbol{u}}, \Gamma / \sum_{l\neq i} w_{il}\right)$. The mean vector $\overline{\boldsymbol{u}} = (\overline{u}_0, \overline{u}_1)'$, where $\overline{u} = \sum_l w_{il}u_{0l} / \sum_{l\neq i} w_{il}$ and $\overline{u}_1 = \sum_l w_{il}u_{1l} / \sum_{l\neq i} w_{il}$. The spatial weights w_{il} are fixed constants that measure the proximity of districts i and l. Let the set of boundary points on a district say i be denoted by (i). We defined w_{il} as a binary connectivity spatial weight matrix such that $w_{il} = 1$ if $(i) \cap (l) \neq \varnothing$, and $w_{il} = 0$ otherwise. The covariance matrix Γ is of dimension 2×2 and captures the within-area correlation between the two random effects (random intercepts in this case). The diagonal elements equal to their conditional variances. The off-diagonal elements capture the correlation between them. Thus, the off-diagonal elements $\Gamma_{12} = \Gamma_{21}$ capture the within-district association between the intercepts and slopes with their correlation coefficient measured by $\rho = \Gamma_{12}/\sqrt{\Gamma_{11}\Gamma_{22}}$. Since BCAR is translational invariant, we add the constraints $\sum_i u_{0i} = 0 = \sum_i u_{1i} = 0$ for identifiability of the mean and to ensure proper posteriors.

3. Data application

We demonstrate the proposed method by applying it to study the evolution of spatial clustering of intestinal parasite infections in Ghana. Globally, intestinal parasites are estimated to infect more than 1 billion people (Bethony et al., 2006; de Silva et al., 2003). Intestinal parasite infection occurs mainly through contact with infected environments, hand-to-hand contact, or contaminated food or water (fecal-oral) (Alum et al., 2010). Infection can lead to blood loss and the development of iron deficiency anemia thereby retarding the growth of children and cognitive deficiencies (Stephenson et al., 1993, 1990, 1989). Alternative intervention using the best available evidence is therefore crucial. In our previous study (Osei and Stein, 2017), we found evidence of spatial clustering via the global Moran's Index. In this study, we demonstrate this methodology to evaluate the evolution of intestinal parasites clustering in Ghana. This is useful to understand the dynamics of the disease hot-spots. We use yearly intestinal parasites morbidities from 2010 to 2014 from outpatient records. These are obtained from the District Health Information Management System (DHIMS) managed by the centre for Health Information and Management (CHIM) of the Ghana Health Services (GHS). The data exist in aggregated format per administrative district. The geographical scale of analysis is restricted to the 170 administrative districts of which data had been recorded for this period. The population information for the districts was obtained from the Ghana Statistical Service (GSS). Since the focus has been on studying the space-time random effects, we avoided the inclusion of covariates in the model. Thus, $\eta_{ij} = \beta_0 + (\overline{\beta}_1 + u_{1i})t_j^* + u_{0i}$.

4. Model inference

We estimated the model parameters by generating samples from the posterior density using MCMC sampling within a Bayesian framework. Let $\psi_1$be the full Gaussian latent (unobservable) field and let ψ_2 be a vector of hyper-parameters. Then the process layer is $\psi_1|\psi_2 = \{\beta_0, \beta_{1i}, \overline{\beta}_1, u_{i1}, u_{i2}\}$ and the hyper-*priors* is $\psi_2 = \{\sigma_j^2, \Gamma\}$. Then the model can be summarized under a three-stage hierarchical framework; the data model, process model, and parameter model. Stage 1 is the data model where the likelihood $\mathrm{y}|\psi_1,\psi_2 \sim p(\mathrm{y}|\psi_1,\psi_2)$. Stage 2 is the process model where $\psi_1|\psi_2 \sim p(\psi_1|\psi_2)$. Stage 3 is the parameter model $\psi_2 \sim p(\psi_2)$. The posterior density is then $p(\psi_1, \psi_2|\mathrm{y}) \propto p(\mathrm{y}|\psi_1, \psi_2) \times p(\psi_1, |\psi_2) \times p(\psi_2)$. Before estimation, we assign *prior* distributions to all variance parameters σ_j^2, the fixed effects $\beta_0, \overline{\beta}_1$ and the covariance Γ. We assigned a non-informative flat distribution for the intercept, $p(\beta_0) \propto 1$ to ensure that the data drive inference. For the global time trend, $\overline{\beta}_1$, we assigned vague normal *prior* $\overline{\beta}_1 \sim N(0, 10^5)$. We assigned a Wishart *prior* for the inverse of the covariance matrix, thus $\Gamma^{-1} \sim \mathrm{W}(\Omega, df)$. This *prior* is a conjugate of the inverse of the covariance parameters of a multivariate normal distribution (Gelman et al., 2013; Press, 2005). The scale matrix is Ω and the degrees of freedom *df* equal to the number of random effects for a weakly informative distribution. We set Ω as a scaled identity matrix of dimension equal to the number of random effects, a specification Moraga and Lawson (Moraga and Lawson, 2012) utilized to run simulation studies of multivariate CAR modeling. We used the MCMC samplers employed in the WinBUGS software to generate samples from the posterior distribution. The Gibbs sampler is automatically used to sample values of the unknown parameters from their conditional posterior distribution if the *prior* is a conjugate of the likelihood. The software uses an expert system to select the appropriate MCMC sampler if a closed-form conditional posterior is not available, especially if the *prior* is not a conjugate of the likelihood. For our study, which is a form of a Generalized linear model, for instance, the full conditional posterior of the fixed effects is log-concave; hence the derivative-free adaptive-rejection sampling is used (Gilks and Wild, 1992). We fitted the models via the R2WinBUGS package (Spiegelhalter et al., 2008) of the R software (R Core Team, 2016).

5. Model evaluation and comparison

We used the Chi-square goodness-of-fit statistic based on the discrepancy function, $\chi^2_{obs} = \sum_{ij}[(y_{ij} - n_{ij}r_{ij})^2 / n_{ij}r_{ij}]$ (Marshall and Spiegelhalter, 2003) to evaluate the adequacy of the model. By generating predicted observations y_{ij}^{pred}using the predictive distribution of the model, we as well compute the discrepancy for the predicted data as $\chi^2_{pred} = \sum_{ij}[(y_{ij}^{pred} - n_{ij}r_{ij})^2 / n_{ij}r_{ij}]$. These two quantities are compared using the Bayesian *p*-values, $\Pr(\chi^2_{obs} \geq \chi^2_{pred})$. A Bayesian *p*-value close to 0.5 suggests that the generated data are compatible with the model.

6. Spatial clustering of time trends

To answer our second and third research questions, we adopt the two-stage classification approach proposed by (Li et al., 2014) for burglary risk mapping. This proceeds as follows. First, areas are classified as hot-spots, cold-spots, or neither (neutral-spots). Next, the temporal behaviors of these areas are classified as decreasing or increasing trends or neither (stable trend). To evaluate the first step, we adopt the spatial clustering classification rule and cutoffs by Richardson and colleagues (Richardson et al., 2004). We classify districts as either hot-spot, cold--spot, or neutral-spot (i.e. neither hot-spot nor cold-spot) based on the exceedance posterior probabilities of the residual risks. We define a district as a hot-spot if the posterior probability $p(\exp(u_{0i}) > 1|y_{ij}) > 0.8$. Likewise, a district is defined as a cold-spot if the posterior probability $p(\exp(u_{0i}) > 1|y_{ij}) < 0.2$. Neutral-spots are those districts with posterior probabilities $0.2 \leq p(\exp(u_i) > 1|y_{ij}) \leq 0.8$. For the second stage, which is the classification of time trends into decreasing, stable, and decreasing trends, we use the same probability cutoffs but on the local slopes, β_{1i}. Thus, we define districts with increasing time trends as those with $p(\beta_{1i} > 0|y_{ij}) > 0.8$, decreasing time trends as those with $p(\beta_{1i} > 0|y_{ij}) < 0.2$, and stable time trends as those with $0.2 \leq p(\beta_{1i} > 0|y_{ij}) \leq 0.8$.

7. Results

We report our results based on posterior samples from 2 MCMC chains, each with 100,000 iterations with a burn-in of 40,000. We checked convergence visually and also via the Brooks and Gelman plots and shrink factor (Brooks and Gelman, 1998). Fig. 1 shows the Brooks-Gelman plots for the fixed parameters. We observed the shrink factor to approach one as the number of iterations increased. For the alternative parameterization, i.e. the log-linear model, there were severe convergence issues. None of the model parameters converged under the same *prior* densities and initial values as the log-Gaussian model. Fig. 2 shows the time series plots of the MCMC chains of the fixed parameters for the log-linear parameterization. There is a clear visual impression of a non-convergence of the MCMC iterations for this parameterization. Hence, in what follows, only results from the log-Gaussian parameterization are described. We checked for model fit using the chi-square goodness of fit discrepancy statistics. Fig. 3 shows the Bayesian *p*-values of the observed and predicted discrepancies for each area. There are only a few areas with extreme *p*-values that indicate a lack of fit. With an average Bayesian *p*-value of 0.65 (see Table 1), there is an agreement that the model well fits the data since this is within the suggested interval $[0.1, 0.9]$.

The posterior means of the fixed effects are shown in Table 1. The parameter $\exp(\beta_0) = 0.763$ represents the overall relative risk after accounting for the spatial and temporal effects. This is epidemiologically interpreted as the ratio of the observed to the expected number of cases being lower than expected. Fig. 4 shows the details of the spatially

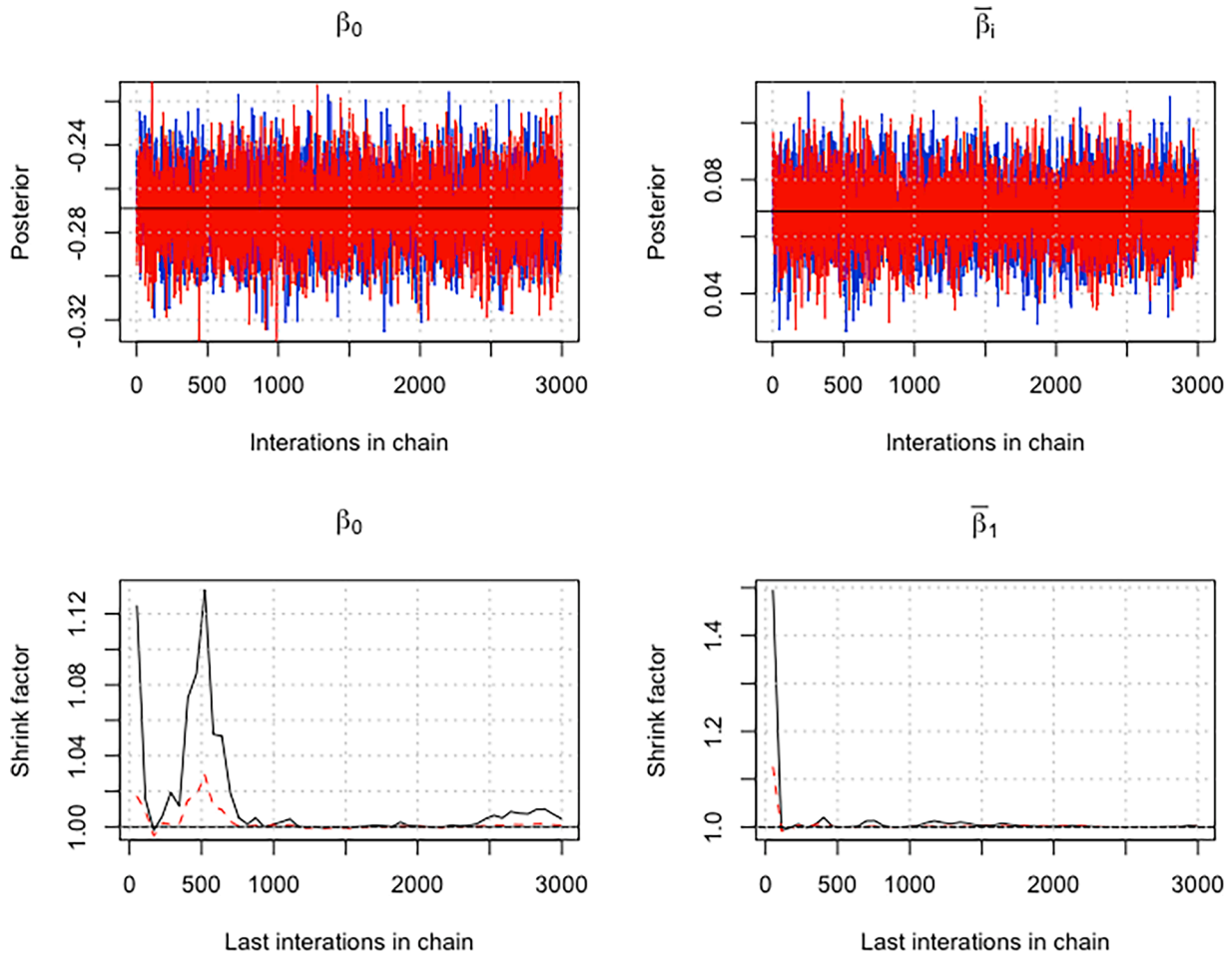


Fig. 1. This figure shows the time series and Gelman plots of the iterations of the two MCMC chains for the log-Gaussian parameterization.

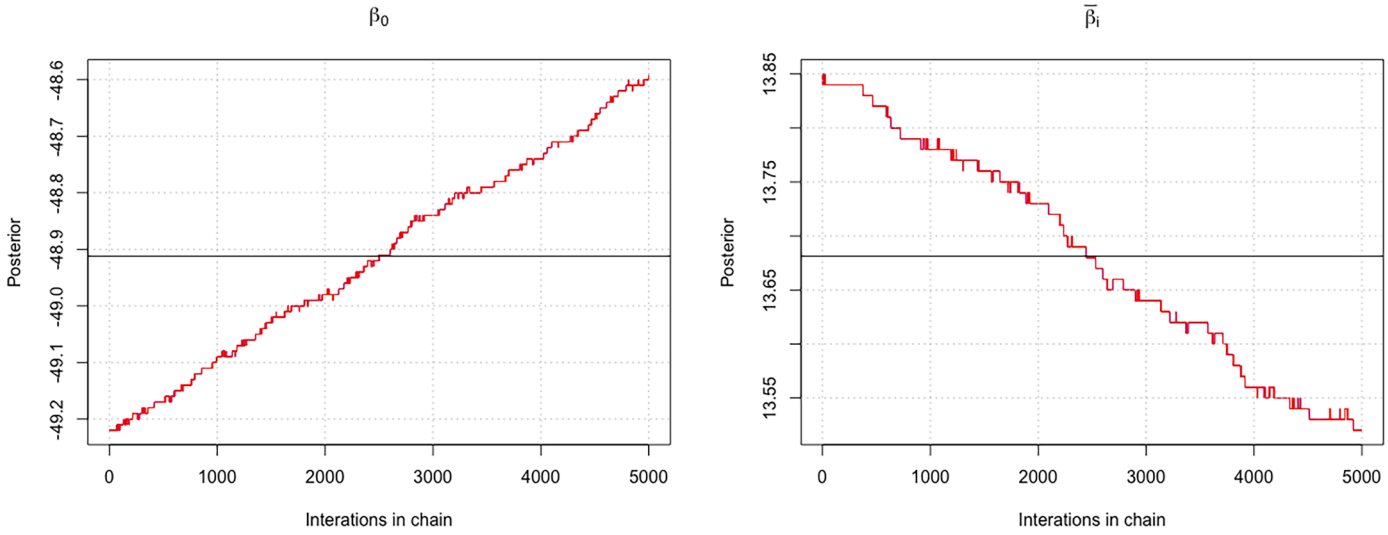


Fig. 2. This figure shows the time series of the MCMC chains of the fixed parameters for the log-linear parameterization.

varying residual relative risks and time trends. These have been exponentiated for easy comparative interpretation. The common spatial trends $\exp(u_{0i})$ measure the residual relative risks (left panel) while the estimated local time trends (right panel) $\exp(u_{1i})$ measure the local departures from the overall time trend. Thus, those districts with $\exp(u_{0i}) > 1$ indicate that they have higher than expected relative risk. The districts with $\exp(u_{1i}) > 1$ also indicate those with increasing time trends, and vice versa. There is obvious north-south trend variation in the residual relative risks; the risks in the northern parts are lower than those in the southern parts. The overall time trend $\overline{\beta}_1 = 0.069$ is positive and reflects a general yearly 7% increment in the risk. The residual time trends u_{1i} (Fig. 4: Right panel) reflect the local departures from the

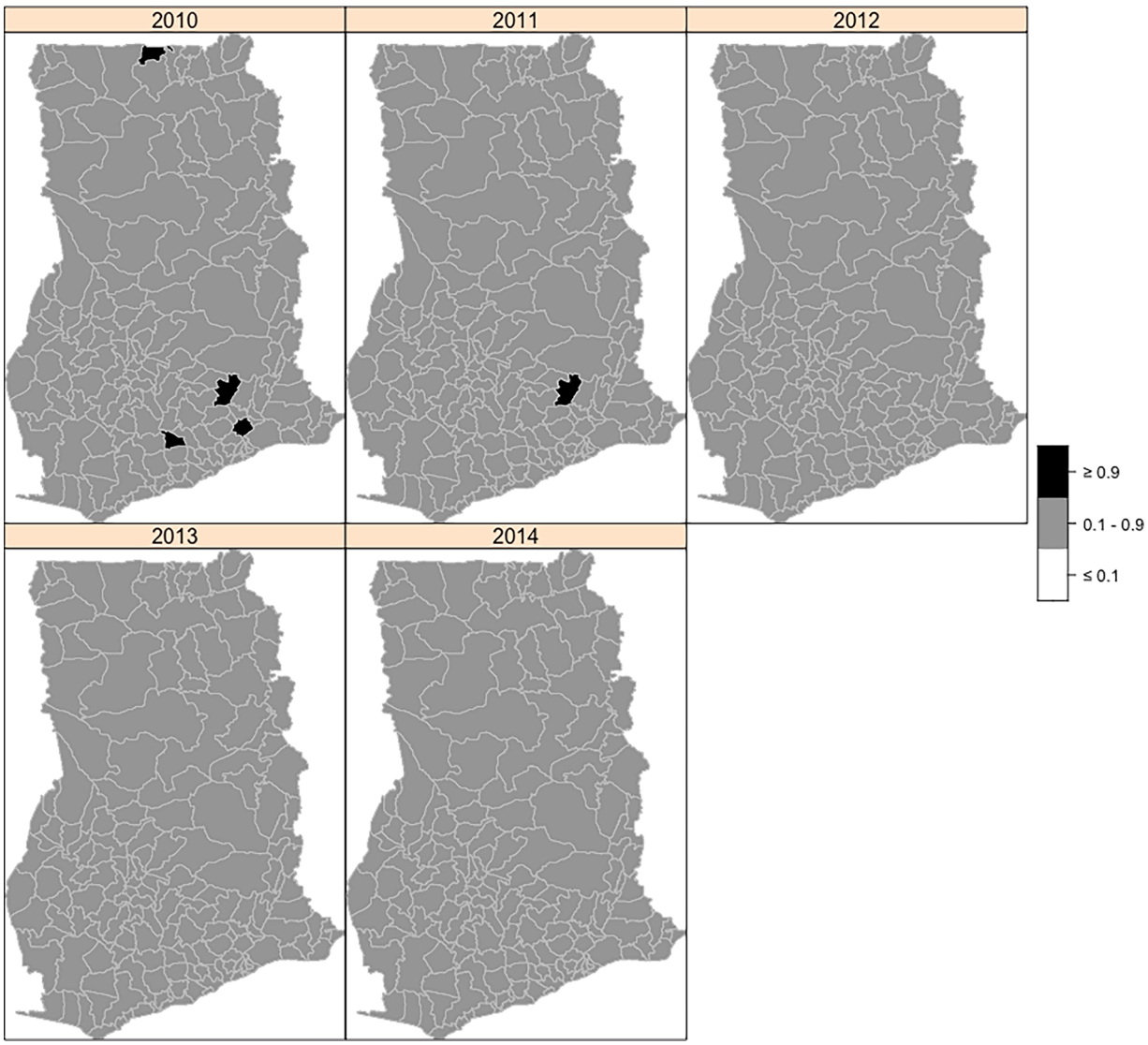


Fig. 3. Bayesian *p*-values of the discrepancy statistic between the observed and predicted cases of intestinal parasite infections in Ghana. The dark areas are those with the observed number of cases markedly higher than the numbers predicted while those in white are when the converse is true.

Table 1
The posterior means of fixed effects and model fit parameters.

Parameter	Estimates (CI)
β_0	−0.27 (−0.304, - 0.237)
$\overline{\beta}_1$	0.069 (0.047, - 0.093)
χ^2_{obs}	892.35
χ^2_{pred}	838.001
$\Pr(\chi^2_{obs} \geq \chi^2_{pred})$	0.65

overall time trend. The south-central part is characterized by a large cluster of decreasing time trends. Moderate increasing time trends dominate with isolated areas of faster time trends.

The within-district correlation between the residual risks and the time trends answers our question of whether the district-specific risks are diverging from the same level or converging to the same level over time. This correlation is observed to be negative (see Table 2), suggesting that the district-specific disease rates are converging to the same level.

Table 3 shows the results of the two-stage classification of clusters into increasing, stable, and decreasing time trends. For this, 47.6% of districts were classified as hot-spots, 34.7% as cold spots, and 17.6% as neutral-spots (Neither hot-spots nor cold-spots). Figs. 5–7 (left panels) show the spatial distribution of the time trends within the hot-spots, cold-spots, and neutral-spots, respectively. The plots (right panels) also show the trends of the relative risks for districts with increasing time trends. We have focused on the relative risks for districts with increasing time trends because this is of particular interest in disease monitoring. Among the hot-spots, only five districts are observed to show increasing time trends while 40 districts show decreasing time trends. There are 36 hot-spot districts with stable time trends. Among the neutral-spots, there are seven districts found to have increasing time trends. These districts have relative risks exceeding the expected ($r_{ij} > 1$) since 2013 (Fig. 7, right panel). The implication is that these districts have a higher tendency to migrate into hot-spots. The cold-spots with increasing times trends, however, have a moderate tendency of becoming hot-spots since their relative risks fall below the expected, except few districts which have relative risks that exceed the expected.

8. Discussion

Here we provide the answers to our reach questions and their interpretation, methodologically and epidemiologically. The

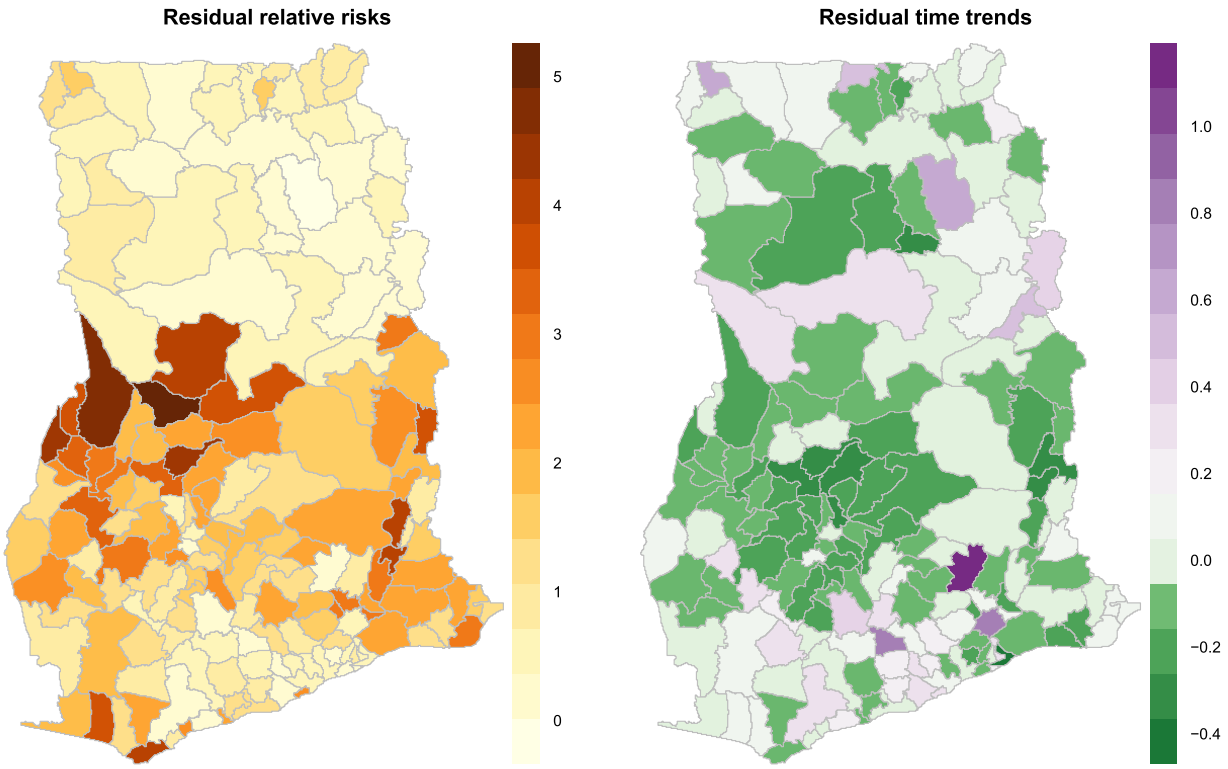


Fig. 4. This figure shows the posterior means of the common spatial components for the residual relative risks $\exp(u_{0i})$ (left panel) and time trends u_{1i} (right panel).

Table 2
The covariance and correlation matrix for the random effects.

	u_{01}	u_{11}
	Covariance matrix: Γ^{-1}	
u_{01}	3.026	−0.449
u_{11}	−0.449	0.304
	Correlation matrix: R	
u_{01}	1.000	−0.469
u_{11}	−0.469	1.000

Table 3
Results from the cross-classification (Clustering and time trends) of the 170 districts.

	Time trends			
Clustering	Increasing	Stable	Decreasing	Total
Hot-spots	5	36	40	81 (47.6%)
Neutral-spots	7	16	7	30 (17.6%)
Cold-spots	24	24	11	59 (34.7%)
Total	36 (21.2%)	76 (44.7%)	58 (34.1%)	170

contribution to disease mapping literature is based upon the imposition of correlation between the residual relative risks and time trends to ascertain whether the area-specific risks are diverging from the same level or converging to the same level over time. Regarding our case study of intestinal parasites in Ghana, we found that the district-specific disease rates are converging to the same level. We ascertained this via the consolidation of the two random effects into a discrete bivariate Markov random field with an unknown within-area correlation parameter. A bivariate Gaussian random field is the alternative if the data were point referenced. The Gaussian random field can still be sampled over a discrete spatial domain if the areas are collapsed into points, but one should not lose sight of the positional uncertainties and computational burden related to evaluating high dimensional matrices at each step of the MCMC iterations. In a related study, we compared the different mechanisms of exploring the correlation between random intercepts and slopes and found the bivariate Markov random field competitive over the others (Osei and Stein, 2019).

The greater divergence of this study from existing studies is based upon the posterior analysis of the random effects, i.e. the residual relative risk and time trends. Quite frequently, spatial epidemiologists have been concerned with estimating and mapping spatial, temporal, and space-time disease variations. On some occasions, exceedance probabilities of the relative risks have been used to classify spatial clusters as inspired by the work of Richardson et al. (2004). In this study, we have adapted the two-stage classification proposed by Li et al. (2014). In their study, they decomposed the latent relative risk into structured and unstructured spatial, temporal, and space-time effects. With greater flexibility and compromise between structured and unstructured spatial variation, the use of such convolution *priors* (Besag et al., 1991) for modeling spatial random effects has attracted a lot of space in disease mapping literature notwithstanding their identifiability issues. The identifiability issue arising from modeling the spatial random effects as the sum of structured and unstructured variations has long been observed (Eberly and Carlin, 2000). In this study, we have bypassed the possible identifiability and MCMC convergence issues by first sampling the log of the relative risk from Gaussian distribution before the decomposition into structured spatial, temporal, and space-time random effects. We set the variance parameter of the Gaussian distribution to impose spatial exchangeability for each time step. In this regard, the variance parameters account for the unstructured spatial effects for each time step. The empirical advantage is that we observed faster MCMC convergence.

The modeling structure presented here is similar to the Bayesian space-time model developed by (Bernardinelli et al. (1995). The space-time interaction utilized can be likened to the type IV interaction proposed by Knorr-Held (2000) where the local time trends specified for each spatial entity are further smoothed over space. Here, we have

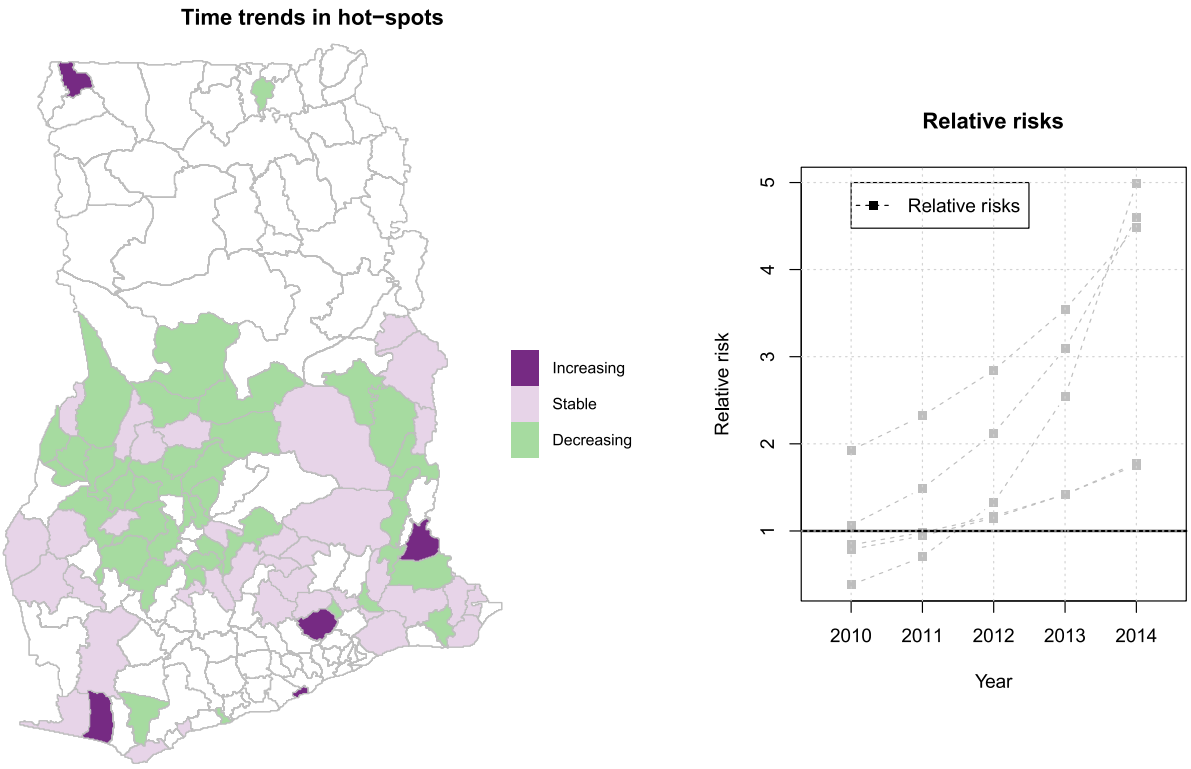


Fig. 5. This figure (left panel) shows the spatial distribution of increasing, stable, and decreasing time trends within hot-spots. The figure (right panel) also shows the trends of the relative risks for hot-spot districts with increasing time trends; the horizontal dotted line shows the expected relative risk.

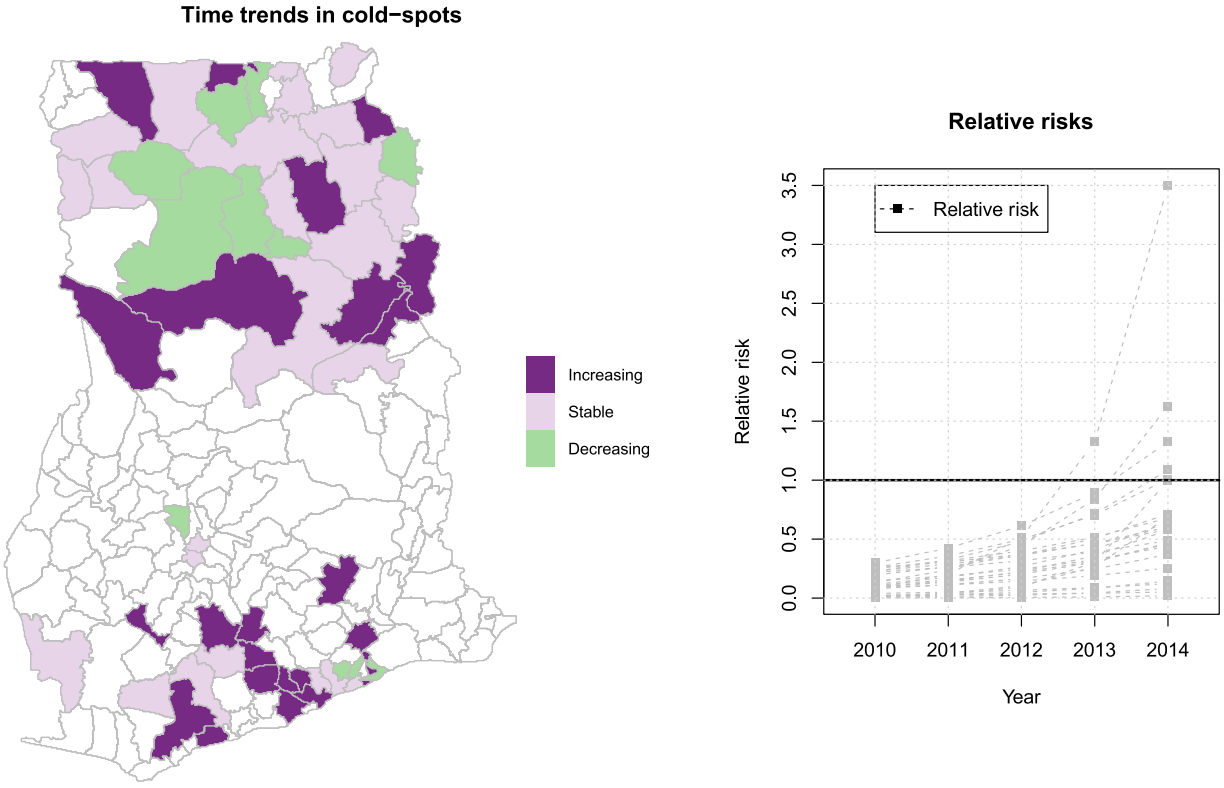


Fig. 6. This figure (left panel) shows the spatial distribution of increasing, stable, and decreasing time trends within cold-spots. The figure (right panel) also shows the trends of the relative risks for cold-spot districts with increasing time trends; the horizontal dotted line shows the expected relative risk.

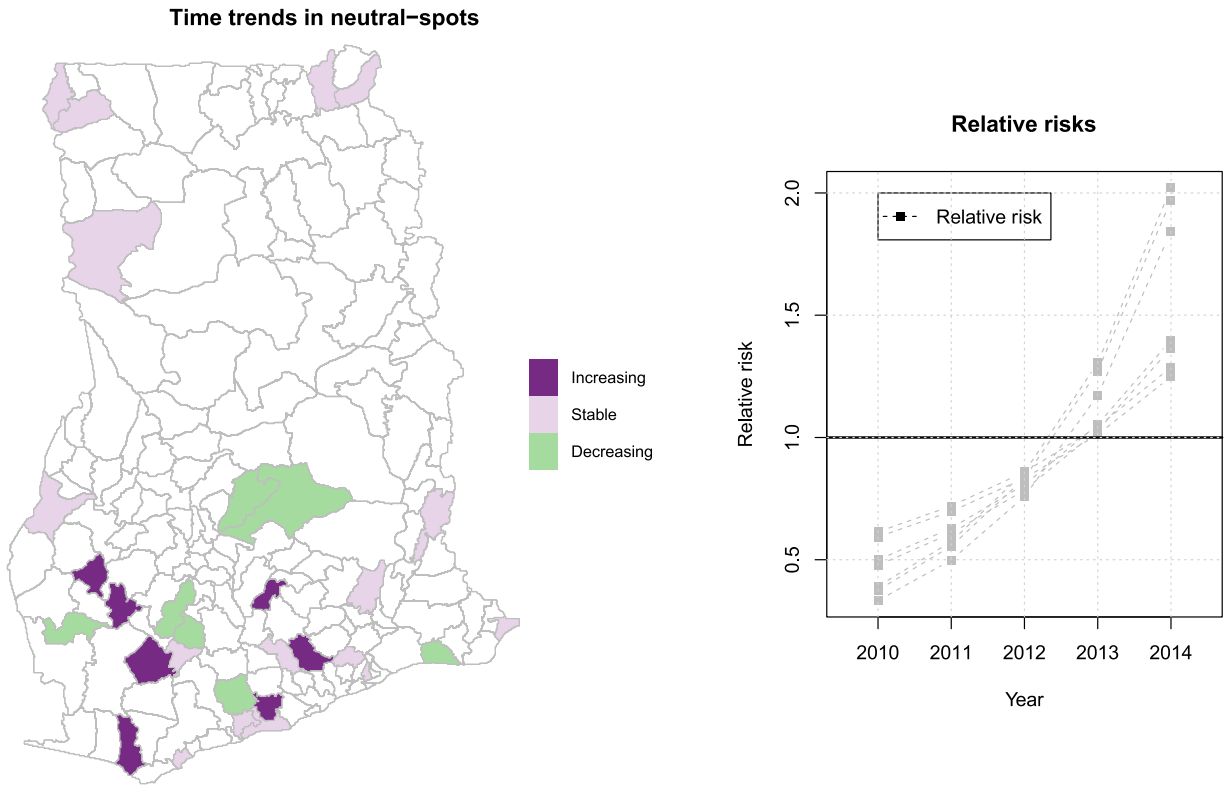


Fig. 7. This figure (left panel) shows the spatial distribution of increasing, stable, and decreasing time trends within neutral-spots. The figure (right panel) also shows the trends of the relative risks for neutral-spot districts with increasing time trends; the horizontal dotted line shows the expected relative risk.

specified local linear time trends for each spatial entity due to the fewer number of time steps of the data. This is, in a nutshell, a dynamic Markov Random field that evolves linearly over time. The limitation of the linear time trend is that it can over smooth and obscure possible departures from linearity. Alternatives and extensions are possible for the space-time interaction structure (Quick et al., 2013), but some data requirements must be met. For instance, if the exact locations of the spatial entities are known, then a Gaussian random field can be used for the spatial smoothing and paired with either discrete or continuous time domain models. For areal unit spatial entities like our case, one may suggest collapsing the districts into points represented by their centroids. The varying shapes and sizes can be a setback but a workaround to overcome such limitations has been presented by Goovaerts (2006). The disadvantage, however, is the computational burden imposed by inverting high dimensional matrices when there are many spatial entities.

We now turn to the interpretation of the results directly related to our research questions. First, are the area-specific risks diverging from the same level or converging to the same level over time? The findings have indicated that the district-specific disease rates are converging to the same level. The implication is that environmental determinants of intestinal parasites, which varied widely in the past, are becoming similar as time passed. The model has not included environmental determinants which might have played a role in the observed variations of the time trends. Further studies to elucidate the role of unknown environmental determinants will be worthwhile. Second, how do the identified hot-spots/cold-spots evolve? On the evolution of the clusters, the result is a 3×3 cross-categorization table that indicates the time trends within each cluster. For the case study, the increasing, stable, and decreasing time trends within each of the clusters have been detected. It is critical to highlight that the hot-spots with increasing time trends should be given immediate public health attention. The additional observation that deserves mention is the fact that 24 out of the 59 cold-spots have increasing time trends and therefore have a higher tendency to migrate into hot-spots if not given attention. Lastly, do areas that are neither hot-spots nor cold-spots (neutral-spots) show the tendency to become hot-spots? We postulate that the neutral-spots that have increasing time trends have a higher tendency to become hot-spots if immediate attention is not given. We have found seven neutral-spots with increasing time trends, and therefore we conclude that these neutral-spots have a higher tendency to migrate into hot-spots.

Disease cluster detection has a rich literature. Prominent is the space-time scan statistics (Kulldorff, 2001). However, in the scan statistics, the measurement of spatial proximity is based on distances between the centroids. This obscures the effects of the varying shapes and sizes of the areal data. The use of the proposed method in this study presents a comparable opportunity to investigate the evolution of small-area disease clusters. In future studies, we will undertake a simulation-based comparison between the proposed methods and the spatial scan statistics for cluster detection.

9. Conclusions

In this paper, we have presented a posterior analysis of a Bayesian space-time random-effects model to study the evolution of spatial disease clustering. We built upon a space-time random effect framework leading to in dynamic Markov Random field that evolves linearly over time. The model has been applied to study the evolution of intestinal parasite infections in Ghana. Several notable implications can be drawn. First, for the sake of MCMC convergence, the log-Gaussian parameterization is favored over the log-linear parameterization, although additional simulation research should be conducted to explain the reasons. Second, this study has shown the evolution of spatial clusters of intestinal parasite infection in Ghana. The cross-classification of clusters and time trends and the maps showing the spatial distribution of increasing, stable, and decreasing time trends within hot-spots, neutral-spots, and cold-spots is an asset to guide public interventions such as resource allocation. The methods presented in this paper are applicable and

extendable to other or multiple tropical diseases which may have different space-time conceptualizations.

Funding info

The study received no funding.

Ethics statement

The study was based on aggregated data, hence, no ethical approval was required.

Consent to participate and for publication

Not applicable.

CRediT authorship contribution statement

Frank Badu Osei: Visualization, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the staff of the District Health Information Management System (DHIMS) of the Ghana Health Services (GHS) for providing the data for this study.

References

Alum, A., Rubino, J.R., Ijaz, M.K., 2010. The global war against intestinal parasites–should we use a holistic approach? Int. J. Infect. Dis. 14, e732–e738. https://doi.org/10.1016/j.ijid.2009.11.036.

Azage, M., Kumie, A., Worku, A., Bagtzoglou, A.C., 2015. Childhood diarrhea exhibits spatiotemporal variation in Northwest Ethiopia: a SaTScan spatial statistical analysis. PLoS ONE 10, e0144690. https://doi.org/10.1371/journal.pone.0144690.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M., 1995. Bayesian analysis of space-time variation in disease risk. Stat. Med. 14, 2433–2443.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43, 1–20. https://doi.org/10.1007/BF00116466.

Bethony, J., Brooker, S., Albonico, M., Geiger, S.M., Loukas, A., Diemert, D., et al., 2006. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet 367, 1521–1532.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434–455. https://doi.org/10.1080/10618600.1998.10474787.

Eberly, L.E., Carlin, B.P., 2000. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Stat. Med. 19, 2279–2294. https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2279::aid-sim569>3.0.co;2-r.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, 3rd ed. Chapman and Hall/CRC, Boca Raton.

Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Appl. Stat. 41, 337. https://doi.org/10.2307/2347565.

Goovaerts, P., 2006. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. Int. J. Health Geogr. 5, 52. https://doi.org/10.1186/1476-072X-5-52.

Hjalmars U., Kulldorff M., Gustafsson G. Nagarwalla N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection 1996:15–707.

Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Stat. Med. 19, 2555–2567. https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-#.

Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic 2001;164:61–72.

Kulldorff, M., 1997. A spatial scan statistic. Commun. Stat. Theory Methods 26 (6), 1481–1496. https://doi.org/10.1080/03610929708831995.

Li, G., Haining, R., Richardson, S., Best, N., 2014. Space–time variability in burglary risk: a Bayesian spatio-temporal modelling approach. Spat. Stat. 9, 180–191. https://doi.org/10.1016/j.spasta.2014.03.006.

Marshall, E.C., Spiegelhalter, D.J., 2003. Approximate cross-validatory predictive checks in disease mapping models. Stat. Med. 22, 1649–1660. https://doi.org/10.1002/sim.1403.

Moraga, P., Lawson, A.B., 2012. Gaussian component mixtures and CAR models in Bayesian disease mapping. Comput. Stat. Data Anal. 56, 1417–1433. https://doi.org/10.1016/j.csda.2011.11.011.

Odoi, A., Martin, S.W., Michel, P., Middleton, D., Holt, J., Wilson, J., 2004. Investigation of clusters of giardiasis using GIS and spatial scan statistics. Int. J. Health Geogr. 3 https://doi.org/10.1186/1476-072X-3-11.

Osei, F.B., Stein, A., 2019. Bayesian random effect modeling for analyzing spatial clustering of differential time trends of diarrhea incidences. Sci. Rep. 9, 13217. https://doi.org/10.1038/s41598-019-49549-4.

Osei, F.B., Stein, A., 2017. Spatio-temporal analysis of small-area intestinal parasites infections in Ghana. Sci. Rep. 7, 12217. https://doi.org/10.1038/s41598-017-12397-1.

Osei, F.B., Stein, A., Andreo, V., 2022. A zero-inflated mixture spatially varying coefficient modeling of cholera incidences. Spat. Stat. 48, 100635 https://doi.org/10.1016/j.spasta.2022.100635.

Press, S.J., 2005. Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference, 2nd ed. Dover Publications, Mineola, N.Y.

Quick, H., Banerjee, S., Carlin, B.P., 2013. Modeling temporal gradients in regionally aggregated California asthma hospitalization data. Ann. Appl. Stat. 7, 154–176. https://doi.org/10.1214/12-AOAS600.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Richardson, S., Thomson, A., Best, N., Elliott, P., 2004. Interpreting posterior relative risk estimates in disease-mapping studies. Environ. Health Perspect. 112, 1016–1025. https://doi.org/10.1289/ehp.6740.

de Silva, N.R., Brooker, S., Hotez, P.J., Montresor, A., Engels, D., Savioli, L., 2003. Soil-transmitted helminth infections: updating the global picture. Trends Parasitol. 19, 547–551.

Spiegelhalter D.J., Thomas A., Best N.G. WinBUGS Version 1.4.3 2008.

Stephenson, L.S., Latham, M.C., Adams, E.J., Kinoti, S.N., Pertet, A., 1993. Physical fitness, growth and appetite of Kenyan school boys with hookworm, Trichuris trichiura and Ascaris lumbricoides infections are improved four months after a single dose of albendazole. J. Nutr. 123, 1036–1046.

Stephenson, L.S., Latham, M.C., Kinoti, S.N., Kurz, K.M., Brigham, H., 1990. Improvements in physical fitness of Kenyan schoolboys infected with hookworm, Trichuris trichiura and Ascaris lumbricoides following a single dose of albendazole. Trans. R. Soc. Trop. Med. Hyg. 84, 277–282. https://doi.org/10.1016/0035-9203(90)90286-N.

Stephenson, L.S., Latham, M.C., Kurz, K.M., Kinoti, S.N., Brigham, H., 1989. Treatment with a single dose of albendazole improves growth of Kenyan school children with hookworm, Trichuris trichiura, and Ascaris lumbricoides infections. Am. J. Trop. Med. Hyg. 41, 78–87.

Takahashi, K., Kulldorff, M., Tango, T., Yih, K., 2008. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. Int. J. Health Geogr. 7, 14. https://doi.org/10.1186/1476-072X-7-14.

Takahashi K., Tango T. A flexibly shaped spatial scan statistic for detecting clusters 2005; 4:4–11.