Supplementary material to *Risk estimation and boundary detection in Bayesian disease mapping*

1 Introduction

This supplementary material accompanies the paper entitled "Risk estimation and boundary detection in Bayesian disease mapping" and has the following sections. Section 2 describes the $(\text{MC})^3$ estimation algorithm used to fit the model proposed in the main paper. Section 3 performs a sensitivity analysis assessing the robustness of the proposed methodology to changing the prior distribution for the spatio-temporal random effects variance parameter τ^2. In Section 4 we test the performance of the proposed model under different sample sizes of the candidate neighbourhood matrices which are generated by the graph-based optimisation algorithm in the first stage. Section 5 presents the convergence diagnostics of the posterior distributions for a selection of parameters of the proposed model in the motivating application. Section 6 presents the boundary detection performance of the proposed model under a set of model-free scenarios, with the different numbers of expected cases taken into account.

2 Inference

We note that the posterior probability distribution of $\widetilde{\boldsymbol{W}}$ contains multiple modes during model development, thus a Metropolis-coupled Markov chain Monte Carlo algorithm ($(\text{MC})^3$) used by Napier et al. [1] is adopted to perform model inference.

2.1 $(\text{MC})^3$ algorithm

Suppose the $(\text{MC})^3$ algorithm runs V Markov chains in parallel, where each chain is labeled by $v \in (1,2,\ldots,V)$. The temperature level for chain v is denoted by T_v, and we have $0 < T_V < T_{V-1} <, \ldots, < T_2 < T_1 = 1$. The first chain with $T_1 = 1$ is also known as the cold chain, and the posterior samples from the cold chain are used for model inference. $\boldsymbol{\Omega}_{vl}$ denotes the collection of model parameters at the l_{th} iteration of the Markov chain v and in our context $\boldsymbol{\Omega}_{vl} = (\boldsymbol{\beta}_{vl}, \boldsymbol{\phi}_{vl}, \widetilde{\boldsymbol{W}}_{vl}, \tau^2_{vl}, \alpha_{vl})$. The $(\text{MC})^3$ algorithm is presented as follows.

1. Set starting values $\boldsymbol{\Omega}_{v0} = (\boldsymbol{\beta}_{v0}, \boldsymbol{\phi}_{v0}, \widetilde{\boldsymbol{W}}_{v0}, \tau^2_{v0}, \alpha_{v0})$ in each chain for $v = 1,2,\ldots,V$.

2. Repeat the following steps for each sampling iteration $l = 1, 2, \ldots, L$.

   (a) At iteration l repeat the following steps for each Markov chain for $v = 1, 2, \ldots, V$, and each model parameter $\omega_{vl} \in \mathbf{\Omega}_{vl}$.

      i. Propose a new value for ω_{vl}, called ω^*_{vl}, from a proposal distribution $g(\omega^*_{vl}|\omega_{vl})$.

      ii. Accept ω^*_{vl} with probability p_1,

$$p_1 = \min\left\{\frac{f(\omega^*_{vl}|\boldsymbol{Y})^{T_v}/g(\omega^*_{vl}|\omega_{vl})}{f(\omega_{vl}|\boldsymbol{Y})^{T_v}/g(\omega_{vl}|\omega^*_{vl})}, 1\right\},$$

      where $f(\cdot)$ represents the full conditional distribution of ω_{vl} or ω^*_{vl}.

      iii. Generate a random variable, U_1, that is uniformly distributed on the interval $[0,1]$; If $U_1 \leq p_1$, accept ω^*_{vl} as the next value in the chain v, i.e., $\omega_{v,l+1} = \omega^*_{vl}$. Otherwise, $\omega_{v,l+1} = \omega_{vl}$.

   (b) Randomly select two of the chains to couple the chains, e.g., chains j and k, and exchange their values.

      i. Swap chains j and k with probability p_2, where

$$p_2 = \min\left\{\frac{f(\mathbf{\Omega}_{kl}|\boldsymbol{Y})^{T_j} f(\mathbf{\Omega}_{jl}|\boldsymbol{Y})^{T_k}}{f(\mathbf{\Omega}_{jl}|\boldsymbol{Y})^{T_j} f(\mathbf{\Omega}_{kl}|\boldsymbol{Y})^{T_k}}, 1\right\}.$$

      ii. Generate a uniform random sample $U_2 \sim \text{Uniform}(0,1)$; If $U_2 \leq p_2$, then the proposed swap is accepted and chains j and k exchange their values.

The $(\text{MC})^3$ algorithm is not applied to the parameters that are sampled using Gibbs sampling. The temperatures are determined by a geometric progression, which is a common choice in the literature [2, 3]. The geometrically spaced temperatures are given by $T_{v+1} = c * T_v$, with a scale factor $c \in (0,1)$. The value of c is altered within the algorithm to ensure the swaps of two chains are accepted between 20 % and 30% of the time, thereby providing a sufficient amount of mixing [1]. The number of chains needed for adequate mixing depends on the complexity of the data [4]. Exploratory model runs suggested that $V = 5$ coupled chains appear to result in good mixing for both the simulated and real application data. The $(\text{MC})^3$ algorithm is written and implemented in R [5] and C++ via the R package Rcpp [6, 7].

3 Sensitivity analysis to changing the prior distribution for τ^2

In the main paper we use an Inverse-Gamma $(1, 0.01)$ prior for the variance parameter τ^2 in the proposed model. To assess the impact of the prior for τ^2 on model performance, we re-run part of the simulation study by fitting the proposed model separately with both Inverse-Gamma$(0.001, 0.001)$ and Inverse-Gamma$(0.5, 0.0005)$ priors. Specifically, one hundred simulated data sets are generated as described in Section 4 of the main paper, where we consider $Z = 1, 0.5, 0.25$. In generating the data $\rho = 0.9$ is used to simulate the random effects $\boldsymbol{\phi}_t$ at each time period t and the expected numbers of disease cases are taken from the motivating data (i.e., SF $= 1$). The proposed model is fitted to each data set using the three different choices of Inverse-Gamma (IG) prior distribution for τ^2, and the results are summarised in Figures S1, S2 and S3, which display boxplots of RMSE, 95% coverage probabilities for risk estimates, DIC, p_d and the AUC over all simulated data sets. The figures show that changing the hyperparameters of the Inverse-Gamma prior for τ^2 does not seem to have any substantial effect on the ability of the proposed model in terms of both risk estimation and boundary detection, because the differences in the values of the performance metrics are very minimal when the prior varies. Therefore the proposed methodology appears to be robust to the choice of the hyperparameters of the prior Inverse-Gamma distribution for τ^2.

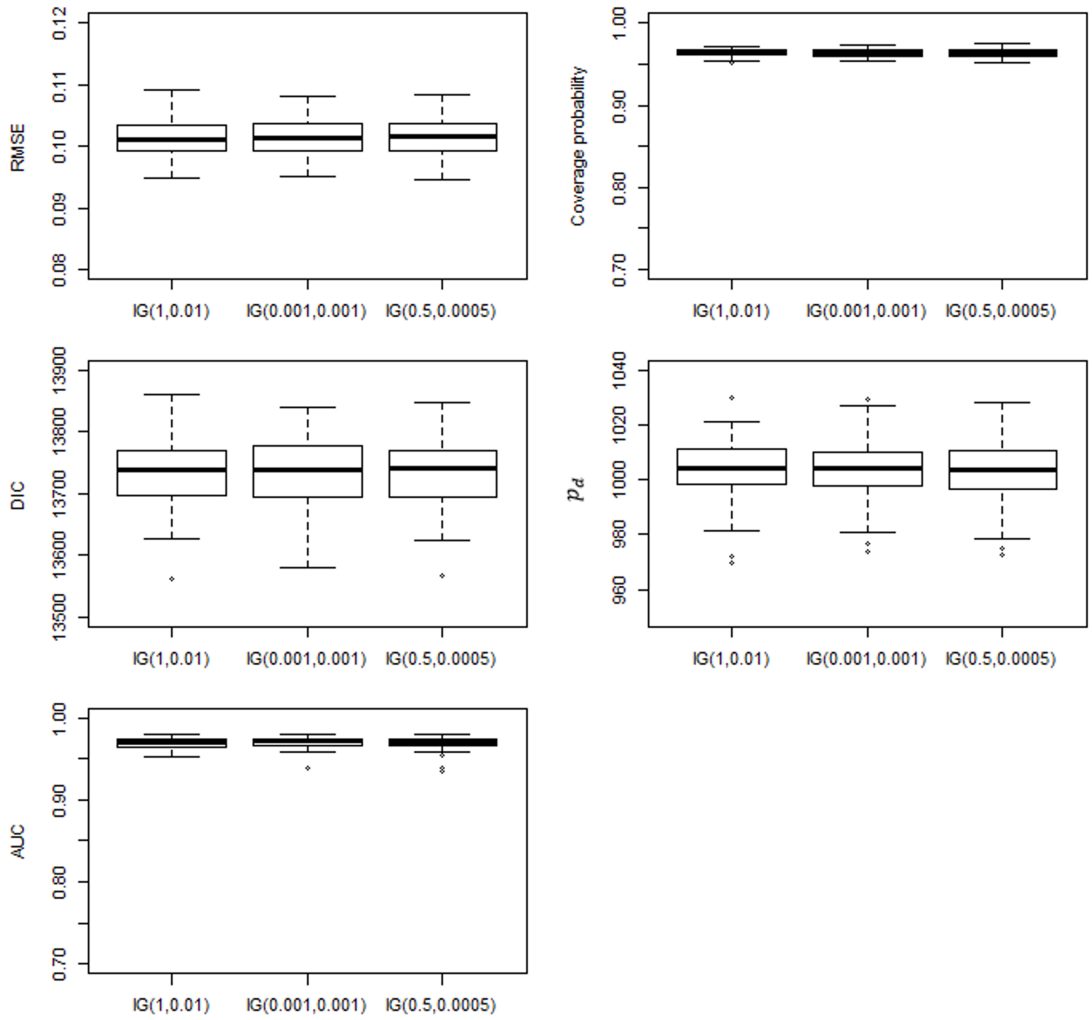

Figure S1: Summary of the simulation results from changing the hyperparameters of the Inverse-Gamma (IG) prior distribution for τ^2 when $Z = 1$.

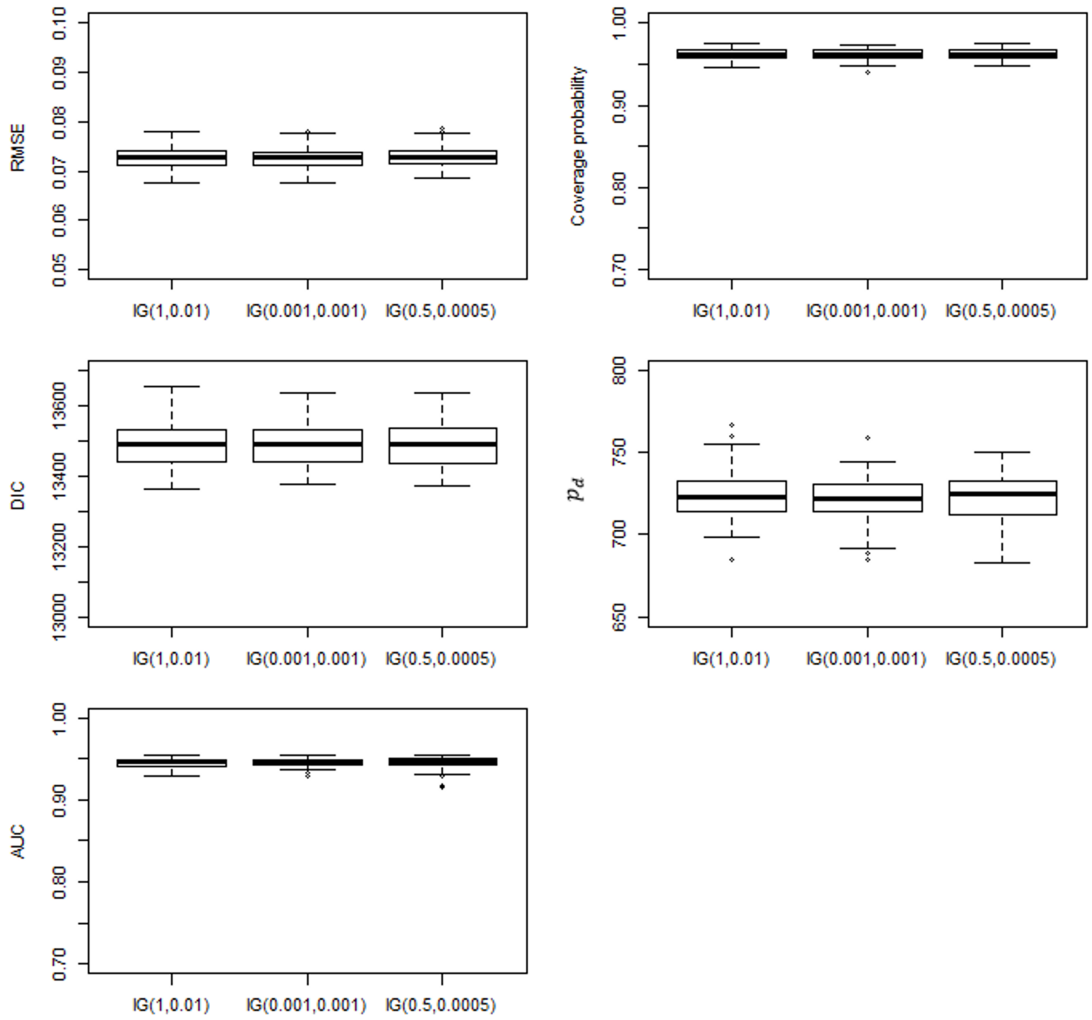

Figure S2: Summary of the simulation results from changing the hyperparameters of the Inverse-Gamma (IG) prior distribution for τ^2 when $Z = 0.5$.

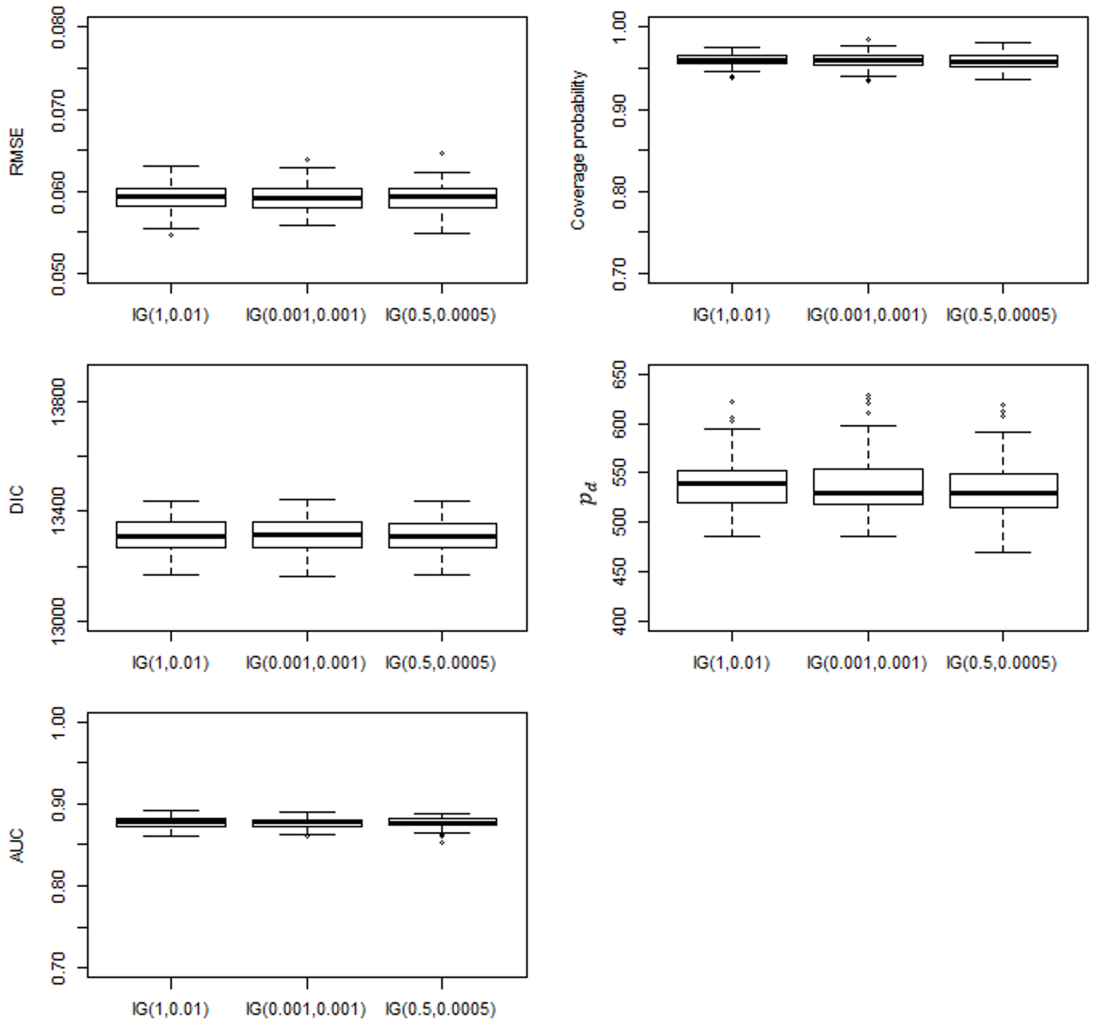

Figure S3: Summary of the simulation results from changing the hyperparameters of the Inverse-Gamma (IG) prior distribution for τ^2 when $Z = 0.25$.

4 Sensitivity analysis to changing the value of M

In the main paper we generate $M = 100$ candidate neighbourhood matrices in stage one of the proposed approach for both the simulation study and the real data analysis. To examine the impact of the value of M on model performance, we re-run part of the simulation study by fitting the proposed model separately with both $M = 20$ and $M = 50$. Specifically, one hundred simulated data sets are generated as described in Section 4 of the main

paper, where we consider $Z = 1, 0.5, 0.25$. In generating the data $\rho = 0.9$ is used to simulate the random effects $\boldsymbol{\phi}_t$ at each time period and the expected numbers of disease cases are taken from the motivating data (i.e., SF $= 1$). The proposed model is fitted to each data set using the three different choices of M and the results are summarised in Tables S1 and S2, which display median values of the RMSE, 95% coverage probabilities for risk estimates, DIC, p_d and the AUC for each model and scenario.

Tables S1 shows that the ability of our model to estimate disease risk appears to be robust under different values of M, as the differences in RMSE and 95% coverage probabilities are minimal when M varies. Both the proposed model and the LM model perform better than the RL model in terms of their lower RMSE, DIC and p_d values across three values of M. The RL model has an increased number of effective parameters (p_d), and this is because it enforces spatial smoothing between all pairs of neighbouring areas, even those that have very different data values, which hence inflates the variance of the random effects and increases p_d. In contrast, in our model and the LM model the spatial random effects are only allowed to smooth towards their neighbours that no boundary is present between them, i.e., those neighbours that have similar random effects values, which leads to a reduction in the random effects variance τ^2 and p_d. Compared with the LM model, our model has slightly higher coverages and smaller RMSE values, with a reduction of around 1% in RMSE on average. In addition, the proposed model has a higher DIC and p_d than the LM model in a number of scenarios. This is because the neighbourhood matrix is fixed when estimating the disease risk in the LM model, whereas it is treated as an additional parameter to be estimated by our approach.

Table S2 indicates that the ability of the proposed model to detect the true boundaries and non-boundaries shows almost no sensitivity to the value of M, as the AUC values do not change greatly when M varies. When $Z = 1$, the model produces high AUC values close to 1. As the magnitude of the boundaries gets smaller, the AUC statistic decreases because smaller boundaries are more difficult to correctly identify, with values varying between 0.90 and 0.96 for the scenario $Z = 0.5$, and between 0.82 and 0.91 for the scenario $Z = 0.25$. In addition, we found that reducing the value of M results in a slight reduction in AUC, with reductions in median AUC of 0.4% ($Z = 1$), 1.1% ($Z = 0.5$) and 2.9% ($Z = 0.25$) from the scenario $M = 50$ to $M = 20$, and 0.93% ($Z = 1$), 0.74% ($Z = 0.5$) and 1.3% ($Z = 0.25$) from the scenario $M = 100$ to $M = 50$.

Table S1: Median values of the RMSE, 95% credible interval coverages associated with the estimated risks, Deviance Information Criterion (DIC), and the effective number of parameters (p_d) for each model and scenario. Here LM and RL refer to the models proposed by Lee et al. [8] and Rushworth et al. [9].

Metric	Z	M	Model		
			Proposed	LM	RL
RMSE	1	20	0.101	0.102	0.122
	0.5	20	0.074	0.075	0.089
	0.25	20	0.062	0.062	0.068
	1	50	0.101	0.102	0.122
	0.5	50	0.075	0.075	0.089
	0.25	50	0.062	0.062	0.068
	1	100	0.100	0.101	0.121
	0.5	100	0.074	0.075	0.089
	0.25	100	0.062	0.062	0.068
Coverage probability	1	20	0.962	0.960	0.953
	0.5	20	0.956	0.957	0.955
	0.25	20	0.949	0.942	0.948
	1	50	0.961	0.961	0.952
	0.5	50	0.957	0.957	0.954
	0.25	50	0.949	0.941	0.949
	1	100	0.963	0.960	0.954
	0.5	100	0.958	0.956	0.954
	0.25	100	0.951	0.944	0.948
DIC	1	20	13729.66	13711.89	14123.81
	0.5	20	13462.59	13456.34	13762.87
	0.25	20	13261.99	13252.31	13445.40
	1	50	13700.64	13698.46	14115.55
	0.5	50	13463.68	13455.32	13771.48
	0.25	50	13247.11	13230.65	13429.77
	1	100	13714.99	13701.59	14123.56
	0.5	100	13473.27	13456.07	13770.36
	0.25	100	13268.14	13259.73	13454.56
p_d	1	20	980.99	975.21	1347.79
	0.5	20	716.25	729.86	1005.98
	0.25	20	518.84	505.16	653.17
	1	50	982.00	988.15	1344.30
	0.5	50	721.53	729.47	1007.82
	0.25	50	516.92	502.42	650.05
	1	100	992.45	988.89	1345.70
	0.5	100	721.27	730.13	1007.26
	0.25	100	521.50	509.29	653.65

Table S2: Median values of the area under the ROC curve (AUC) for boundary identification for the proposed model for each scenario. Values in brackets correspond to the 95% credible intervals.

Z	M	AUC
1	20	0.964 (0.938, 0.981)
1	50	0.968 (0.947,0.983)
1	100	0.977 (0.960, 0.985)
0.5	20	0.934 (0.910,0.958)
0.5	50	0.944 (0.912, 0.961)
0.5	100	0.951 (0.933, 0.965)
0.25	20	0.841 (0.835, 0.868)
0.25	50	0.865 (0.836, 0.893)
0.25	100	0.876 (0.846, 0.902)

5 Convergence diagnostic

The convergence of the posterior distributions for the model parameters in the analysis of the real data is diagnosed both by examining parameter trace plots and by Gelman-Rubin diagnostic [10]. It is infeasible to check the convergence for all parameters in practice because there are a large number of parameters in the model. Therefore here we only select four parameters, which are $(\beta_0, \tau^2, \phi_{100,1}, \phi_{129,3})$, to check their convergence. Figure S4 shows trace plots of the posterior samples for $(\beta_0, \tau^2, \phi_{100,1}, \phi_{129,3})$, where each Markov chain is represented in a different color. The figure shows that there is no clear pattern in the trace plots for all selected parameters, suggesting that all the chains appear to have converged. In addition, the Gelman-Rubin diagnostic is also used to check the convergence for multiple chains, with a value less than 1.1 indicating convergence of the chains. Here the Gelman-Rubin statistics for the selected parameters are all smaller than 1.1 with a maximum value of 1.01, which means that the posterior samples have converged.

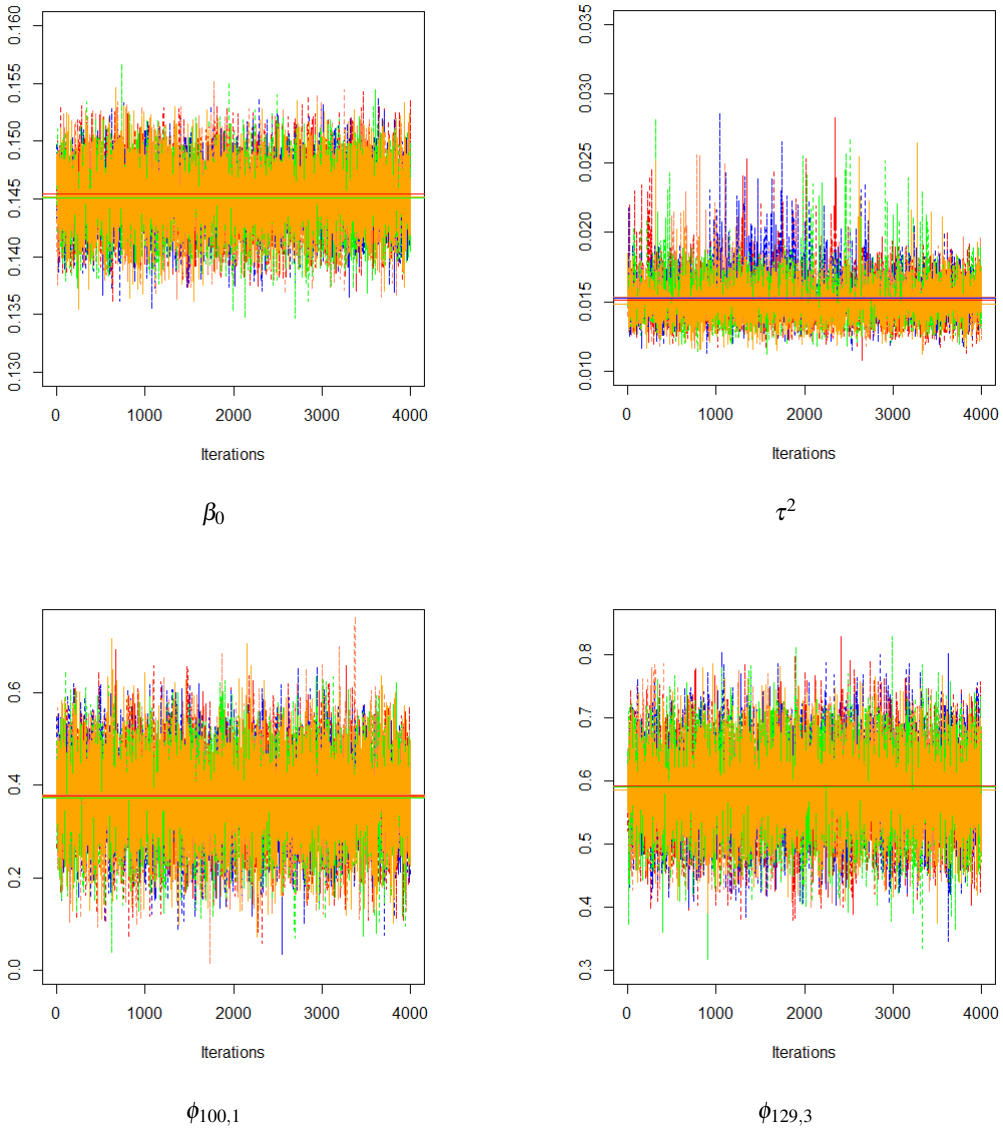

Figure S4: Trace plots of the posterior samples for selected parameters $(\beta_0, \tau^2, \phi_{100,1}, \phi_{129,3})$ from the proposed model.

6 Summary of boundary detection performance under model-free scenarios

In this section we assess the boundary detection performance of the proposed approach under a set of model-free scenarios. In these scenarios the disease risks θ_{it} of the areas with high, medium and low risk levels are fixed at constant values of $\{\exp(Z), 1, \exp(-Z)\}$, rather than being generated by simulating random effects using a multivariate Gaussian distribution with a Leroux CAR covariance structure and a piecewise constant mean as described in Section 5.1 of the main paper. In order to explore the impact of the number of expected cases on boundary detection performance, the expected disease counts $\{E_{it}\}$ from the motivating study are divided by the scale factors (SF) of 1, 2 and 4. Thus SF = 1 corresponds to the motivating data, SF = 2 corresponds to having a smaller number of expected counts, while SF = 4 represents a rare disease that has very small expected counts. The observed disease counts $\{Y_{it}\}$ are then generated from a Poisson distribution with mean $E_{it}\theta_{it}$. One hundred simulated data sets are generated for each pairwise combination of $Z = 1, 0.5, 0.25$ and SF = 1, 2, 4.

The proposed model is applied to each data set, and the correctness of the boundary detection is presented in Table S3, which displays the median AUC values as well as the corresponding 95% credible intervals across the set of ROC curves calculated for each scenario. The boundary detection performance of the proposed model under the model-free scenarios is similar to that displayed in the simulation study in the main paper. When the size of the boundaries is not very small ($Z > 0.25$), the proposed model generally performs well in terms of boundary identification and is robust to different numbers of expected cases, with relatively high AUC values ranging between 0.876 and 0.978. The model obtains a lower AUC value when both the number of expected cases and the magnitude of the boundaries are very small (i.e., SF = 4 and $Z = 0.25$), with a AUC value of 0.755. The reason for this is that in this scenario the boundaries are difficult to be correctly identified based on their small size and small numbers of disease cases. Overall, these conclusions are consistent with those provided in the simulation study in the main paper.

Table S3: Median values of the area under the ROC curve (AUC) for boundary identification for the proposed model for each scenario. Values in brackets correspond to the 95% credible intervals.

Z	SF	AUC
1	1	0.978 (0.964, 0.986)
0.5	1	0.953 (0.928, 0.967)
0.25	1	0.878 (0.844, 0.903)
1	2	0.966 (0.931, 0.980)
0.5	2	0.922 (0.899, 0.944)
0.25	2	0.821 (0.793, 0.848)
1	4	0.948 (0.935, 0.961)
0.5	4	0.876 (0.856, 0.907)
0.25	4	0.755 (0.723, 0.771)

References

[1] Napier G, Lee D, Robertson C, Lawson A. A Bayesian space-time model for clustering areal units based on their disease trends. Biostatistics. 2019;20(4):681-97.

[2] Kofke DA. On the acceptance probability of replica-exchange Monte Carlo trials. The Journal of chemical physics. 2002;117(15):6911-4.

[3] Earl DJ, Deem MW. Parallel tempering: Theory, applications, and new perspectives. Physical Chemistry Chemical Physics. 2005;7(23):3910-6.

[4] Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics. 2004;20(3):407-15.

[5] R Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2013.

[6] Eddelbuettel D, François R, Allaire J, Ushey K, Kou Q, Russel N, et al. Rcpp: Seamless R and C++ integration. Journal of Statistical Software. 2011;40(8):1-18.

[7] Eddelbuettel D. Seamless R and C++ integration with Rcpp. Springer; 2013.

[8] Lee D, Meeks K, Pettersson W. Improved inference for areal unit count data using graph-based optimisation. Statistics and Computing. 2021;31(4):1-17.

[9] Rushworth A, Lee D, Mitchell R. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. Spatial and spatio-temporal epidemiology. 2014;10:29-38.

[10] Gelman A, Rubin DB, et al. Inference from iterative simulation using multiple sequences. Statistical science. 1992;7(4):457-72.