

Chapter 10

A Comparison of Bayesian Spatial Models for Cancer Incidence at a Small Area Level: Theory and Performance



Susanna Cramb, Earl Duncan, Peter Baade, and Kerrie L. Mengersen

Abstract The increase in Bayesian models available for disease mapping at a small area level can pose challenges to the researcher: which one to use? Models may assume a smooth spatial surface (termed global smoothing), or allow for discontinuities between areas (termed local spatial smoothing). A range of global and local Bayesian spatial models suitable for disease mapping over small areas are examined, including the foundational and still most popular (global) Besag, York and Mollié (BYM) model through to more recent proposals such as the (local) Leroux scale mixture model. Models are applied to simulated data designed to represent the diagnosed cases of (1) a rare and (2) a common cancer using small-area geographical units in Australia. Key comparative criteria considered are convergence, plausibility of estimates, model goodness-of-fit and computational time. These simulations highlighted the dramatic impact of model choice on posterior estimates. The BYM, Leroux and some local smoothing models performed well in the sparse simulated dataset, while centroid-based smoothing models such as geostatistical or P-spline models were less effective, suggesting they are unlikely to succeed unless areas are of similar shape and size. Comparing results from several different models is recommended, especially when analysing very sparse data.

S. Cramb (✉) · P. Baade

Viertel Cancer Research Centre, Cancer Council Queensland, Fortitude Valley, QLD, Australia
e-mail: susanna.cramb@qut.edu.au; peterbaade@cancerqld.org.au

E. Duncan · K. L. Mengersen

ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology (QUT), Brisbane, QLD, Australia

10.1 Introduction

Bayesian spatial modelling continues to increase in popularity, offering a suite of models with a range of strengths in various contexts. Modelling spatial effects through a Bayesian hierarchical model has many advantages, such as being able to include a range of functions to represent outcomes over space and time, as well as the capacity to incorporate data characteristics such as rare outcomes, missing information, misclassifications, measurement error and known biases [9, 47]. Moreover, direct probabilistic statements can be made, such as the probability that an area has a higher disease risk than a comparison area [20].

A popular form for a Bayesian spatial model for disease mapping uses data aggregated by area and specifies the likelihood as:

$$Y_i \sim \text{Poisson}(E_i e^{\mu_i}) \quad \text{for } i = 1, \dots, N \text{ areas}$$

where $\{Y_1, \dots, Y_N\}$ are count data for a relatively uncommon disease, making a Poisson distribution appropriate. Other distributions are possible, including variants of Poisson such as negative binomial. The expected counts (E_i) are commonly defined using indirect standardisation to account for population size and age structure. The modelled log standardised incidence ratio (SIR) μ_i , also called log-relative risk, is often expressed as a regression equation and typically includes an overall fixed effect (intercept, denoted α), covariate effects (β) where \mathbf{x}_i denotes a vector of covariates relating to area i , and spatial random effect(s) R_i , as follows:

$$\mu_i = \alpha + \mathbf{x}_i^T \boldsymbol{\beta} + R_i.$$

Much of this chapter shall discuss options for modelling the spatial random effect(s), R_i . Prior distributions are then specified for each of the unknown parameters:

$$\begin{aligned} \alpha &\sim p(\cdot | \boldsymbol{\theta}_\alpha) \\ \boldsymbol{\beta} &\sim p(\cdot | \boldsymbol{\theta}_\beta) \\ R_i &\sim p(\cdot | \boldsymbol{\theta}_R). \end{aligned}$$

The spatial random effects are given a spatial prior, which may be assumed to follow a conditional autoregressive (CAR) or alternative prior to enable spatial correlation and smoothing [8, 10]. If the parameters $\boldsymbol{\theta}_\alpha$, $\boldsymbol{\theta}_\beta$, or $\boldsymbol{\theta}_R$ are unknown, then the hyperpriors represent an additional stage of the hierarchy.

Many different Bayesian spatial models have been proposed, most of which vary the representation of the spatial prior. Understanding the theoretical assumptions and appropriateness of different models is important. It is also necessary to consider how models perform in different circumstances. Therefore, this chapter discusses the theoretical underpinnings of key spatial models. Where possible and pertinent,

these models were applied to typical cancer incidence mapping scenarios obtained by simulating rare and common cancer incidence data across Australia. This nation has more than 2100 small areas, with large differences in population size, demographic structure, land area size and shape.

10.2 Bayesian Spatial Models

Fourteen Bayesian spatial models used in disease mapping are considered. These can be divided into two broad types, namely ‘global’ spatial smoothing models that have a common spatial correlation term across the region, and ‘local’ spatial smoothing models that allow for differential spatial correlation depending on neighbourhood characteristics.

10.2.1 Global Spatial Smoothing

Global spatial smoothing means that the same correlation parameters are applied consistently across the entire region [26]. Although the global CAR-based models are relatively easy to implement in a range of software, disadvantages of global models include the potential to obscure genuine deviations in the underlying spatial patterns (i.e. to over-smooth), as discontinuities between adjacent areas are smoothed over.

10.2.1.1 Intrinsic CAR and BYM Models

The most commonly used prior for enabling spatial correlation within a Bayesian model is the intrinsic CAR distribution. This approach allows for smoothing of estimates over neighbouring areas, but it assumes a common variance for the smoothing term (and therefore a smooth spatial trend) over the whole region.

The intrinsic CAR (ICAR) model specifies the following set of conditional distributions for the spatial random effect parameter:

$$R_i = S_i$$

$$S_i | s_{\setminus i} \sim \mathcal{N}\left(\frac{\sum_j w_{ij} s_j}{\sum_j w_{ij}}, \frac{\sigma_s^2}{\sum_j w_{ij}}\right)$$

or in matrix notation

$$S_i | s_{\setminus i} \sim \mathcal{N}\left(\left\{\mathbf{D}^{-1} \mathbf{W} \mathbf{s}\right\}_i, \sigma_s^2 \left\{\mathbf{D}^{-1}\right\}_{ii}\right)$$

where w_{ij} is the element of a spatial weights matrix \mathbf{W} corresponding to row i and column j [6, 10], and \mathbf{D} is a diagonal matrix with elements $\text{diag} \left\{ \sum_j w_{ij} \right\}$. The term \mathbf{W} determines the spatial proximity between the random effects, and it is most commonly defined as a binary, first-order, adjacency matrix, whereby

$$w_{ij} = \begin{cases} 1 & \text{if areas } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases} \quad (10.1)$$

This model implies that the conditional expectation of S_i is equal to the mean of the random effects at neighbouring locations.

The S_i can be regarded as structured spatial random effects. If $R_i = S_i + U_i$, so that unstructured spatial random effects $U_i \sim \mathcal{N}(0, \sigma_U^2)$ are also included, the resulting model is referred to as the convolution model, or the BYM model in honour of Besag et al. [8]. However, the two separate random effects components cannot be individually identified—only their sum is identifiable [15]. Note that for all CAR-based models, the strength of the partial autocorrelation depends on the number of neighbouring areas rather than on any underlying relationship [27]. The BYM remains the most popular approach to incorporating spatial smoothing, in part due to its computational synergy with fairly standard MCMC approaches [47] and ease of implementation.

10.2.1.2 Proper CAR Model

The full conditionals for the ICAR prior are proper, but the joint distribution is improper since the precision matrix is singular [7]. The impropriety of the ICAR prior can be overcome by redefining the precision matrix

$$\mathbf{T} = \frac{1}{\sigma_s^2} (\mathbf{D} - \mathbf{W})$$

to

$$\mathbf{T} = \frac{1}{\sigma_s^2} (\mathbf{D} - \phi \mathbf{W})$$

such that the conditional distributions for the spatial random effect are:

$$S_i | s_{\setminus i} \sim \mathcal{N} \left(\frac{\phi \sum_j w_{ij} s_j}{\sum_j w_{ij}}, \frac{\sigma_s^2}{\sum_j w_{ij}} \right)$$

with the constraint $|\phi| < 1$, where ϕ represents the expected proportional ‘reaction’ of S_i to $\sum_j w_{ij} s_j / \sum_j w_{ij}$ [5]. This ensures that the covariance matrix \mathbf{T}^{-1} is positive definite and S has a proper joint distribution [19]. The proper CAR prior may have certain disadvantages, including potentially limiting the breadth of the

posterior spatial pattern. Moreover, ϕ will likely need to be very close to 1 for there to be a reasonable amount of spatial association [5].

10.2.1.3 Leroux CAR Model

Another variant of the BYM model was proposed by Leroux et al. [29],

$$S_i | S_{\setminus i} \sim \mathcal{N} \left(\frac{\rho \sum_j w_{ij} s_j}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\sigma_s^2}{\rho \sum_j w_{ij} + 1 - \rho} \right)$$

which only requires a single set of random effects [24]. This avoids the difficulties in identifiability, and also the selection of hyperpriors (given that in the BYM model, the S_i variance are conditional on neighbouring areas, while the U_i have a marginal variance term) [41].

The precision matrix can be expressed as

$$\mathbf{T} = \frac{1}{\sigma_s^2} [\rho (\mathbf{D} - \mathbf{W}) + (1 - \rho)].$$

This mixture representation consists of correlated smoothing of the neighbouring random effects (weighted by ρ) as well as uncorrelated smoothing to a global mean of zero (weighted by $(1 - \rho)$) [26]. Thus S_i has a conditional expectation based on a weighted average of both the independent random effects and the spatially structured random effects. The ICAR prior is therefore a limiting case of both the proper CAR and Leroux CAR models when ρ is set to 1. The spatial autocorrelation parameter ρ is typically given either a continuous [19, 25] or a discrete [24] uniform prior

$$\rho \sim \text{Uniform}(0, 1),$$

where the discrete case offers gains in computational efficiency [24], although other priors have been suggested such as a diffuse Gaussian prior on the logit scale [27].

10.2.1.4 Geostatistical Model

Here, the residual spatial structure is modelled as a Gaussian process using a geostatistical design [11]. Because this model incorporates distance, counts are assumed to be located in the centroid of an area.

$$R_i \sim \mathcal{N}(S_i, \sigma^2)$$

$$S_i = \exp(-(\lambda d_{ij})^k), \quad \lambda > 0$$

where λ controls the rate of decay, k is the “degree of spatial smoothing”, and d_{ij} is the distance between points (e.g. centroids of areas) i and j [11]. This expression is the exponential decay function with the addition of the power k . Rather than fix decay parameter λ *a priori*, a hyperprior is specified as a fourth stage of the hierarchy:

$$\lambda \sim \text{Uniform}(0.1, 6).$$

The justification for the bounds 0.1 and 6 were based on the minimum and maximum separating distance in decimal degrees between area centroids to ensure that the spatial correlation was able to be high at the minimum distance, and likely to be low at the maximum distance. This choice is also able to give near zero correlation for distances within the study region, which is vital to avoid non-identifiability of the mean and correlation parameters [10].

Alternative functions are possible, including the disc model [40] (a linear decrease with increasing distance, where two discs of common radius are centred on centroids, and the correlation is proportional to the disc intersection area), or combining two parametric functions to obtain different shapes of decrease, such as the Matern class [10]. Note that often limited information is available to guide the choice of functional form, or correlation parameters, especially as complexity increases [10]. Because the covariance matrix is inverted at each iteration, these models can be computationally intensive and slow to run in a naïve algorithm, although this can be mitigated to some extent with the use of sparse matrix algebra.

10.2.1.5 Global Spline Models

The spline model also assumes that the cases are all located at the centroid of each area [17].

There are two main methods: smoothing splines and P-splines [32]. Smoothing splines are penalised splines which have knots on all data points. P-splines allow for a smaller number of knots, and are commonly formulated as a penalised spline regression under a ‘difference penalty’ based on the coefficients of adjacent B-spline bases or other spline bases [32].

The correlation between areas i and j can be modelled by a two-dimensional smooth surface [17]. First, define the longitude and latitude pairs representing the centroid of each area, denoted (c_{1i}, c_{2i}) . Then

$$R_i = f(c_{1i}, c_{2i})$$

where the smooth function $f(\cdot)$ is expressed as

$$f(c_{1i}, c_{2i}) = \theta_1 B_1(c_{1i}, c_{2i}) + \dots + \theta_k B_k(c_{1i}, c_{2i})$$

which is estimated using P-splines with B-spline bases B_1, \dots, B_k . The terms $\theta_1, \dots, \theta_k$ are unknown coefficients which are penalised to control for “wiggleness” through a penalty matrix, and k depends on the number of knots and the degree of the B-spline bases.

Define $c_1 = (c_{11}, \dots, c_{1N})^T$ and $c_2 = (c_{21}, \dots, c_{2N})^T$ and univariate B-spline bases $\mathbf{B}_1 = \{B_{11}(c_1), \dots, B_{1k_1}(c_1)\}$ and $\mathbf{B}_2 = \{B_{21}(c_2), \dots, B_{2k_2}(c_2)\}$. The bivariate B-spline basis is then constructed as the row-wise Kronecker product (denoted by \boxtimes) of the marginal B-spline bases:

$$\begin{aligned} \mathbf{B} &= \mathbf{B}_2 \boxtimes \mathbf{B}_1 \\ &= \left(\mathbf{B}_2 \otimes \mathbf{1}_{k_1}^T \right) \odot \left(\mathbf{1}_{k_1}^T \otimes \mathbf{B}_1 \right). \end{aligned}$$

The basis \mathbf{B} is of dimension $N \times k$ where $k = k_1 k_2$, the symbols \otimes and \odot represent the Kronecker product and “element-wise” matrix product respectively, and $\mathbf{1}_{k_1}$ and $\mathbf{1}_{k_2}$ are column vectors of ones of length k_1 and k_2 [17].

Overall this model provides a relatively smooth surface, as the covariance structure is impacted by long distance effects that influence the smoothing. This is in contrast to the covariance structure of the CAR model where an area’s estimate depends on the mean of its neighbours [17].

The formulation of the P-spline model using the row-wise Kronecker product, or tensor product, is better suited to data which lie on a regular grid, or at least have similar distances between the centroids.

An alternative formulation [42] is to define the B-spline bases in terms of the distances,

$$z_{ik} = \exp\left(-\frac{d_{ik}}{\Delta}\right) \left(1 + \frac{d_{ik}}{\Delta}\right)$$

where d_{ik} is the distance between the i^{th} area and the k^{th} knot, and Δ is a constant used to normalise the distances so that the values of \mathbf{B} are more evenly spread between the lower and upper limits. This version of the P-spline uses a radial basis function which achieves rotational invariance [42].

10.2.2 Local Spatial Smoothing

In contrast to the global smoothing models, local smoothing is focused on allowing nearby areas to potentially have different amounts of spatial smoothing. Many of these are based on modifying the CAR prior to allow for discontinuous surfaces.

10.2.2.1 CAR Dissimilarity Models

Lee and Mitchell [26] based this model on the Leroux CAR prior, with ρ set to be 0.99 to ensure strong global spatial smoothing which could then be altered locally through estimating $\{w_{ij}|i \sim j\}$. Here, the elements in \mathbf{W} are modelled so the partial autocorrelations can be reduced between certain adjacent random effects. This approach can have binary or non-binary elements in \mathbf{W} .

The similarity between areas is determined by including non-negative dissimilarity metrics in the model, i.e. $z_{ij} = (z_{ij1}, \dots, z_{ijq})$ where $z_{ijk} = |z_{ik} - z_{jk}|/\sigma_k$ and σ_k is the standard deviation of $|z_{ik} - z_{jk}|$ over all pairs of contiguous areas.

The set of w_{ij} are determined using regression parameters $\alpha = (\alpha_1, \dots, \alpha_q)$. These can be based on social or physical factors. Physical boundaries (e.g. river/railway line, or the distance between centroids) can be used if the aim is to explain the spatial pattern in the response and include covariates in the model. Alternatively, covariate information can be used to construct the dissimilarity metrics if the aim is to identify the locations of any boundaries [25].

$$R_i = S_i$$

$$S_i | s_{\setminus i} \sim \mathcal{N}\left(\frac{0.99 \sum_j w_{ij}(\alpha) s_j + 0.01 \mu_0}{0.99 \sum_j w_{ij}(\alpha) + 0.01}, \frac{\sigma_s^2}{0.99 \sum_j w_{ij}(\alpha) + 0.01}\right).$$

The default binary formulation is:

$$w_{ij}(\alpha) = \begin{cases} 1 & \text{if } \exp\left(-\sum_{k=1}^q z_{ijk} \alpha_k\right) \geq 0.5 \text{ and } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha_k \sim \text{Uniform}(0, M_k) \text{ for } k = 1, \dots, q$$

where M_k is fixed so that a maximum of 50% of borders could be defined as boundaries [26]. The non-binary formulation (which does not allow identification of hard boundaries, but does allow for localised smoothing) is:

$$w_{ij}(\alpha) = \exp\left(-\sum_{k=1}^q z_{ijk} \alpha_k\right)$$

$$\alpha_k \sim \text{Uniform}(0, 50) \text{ for } k = 1, \dots, q.$$

10.2.2.2 Localised Autocorrelation

The spatially smooth random effects in this model are augmented with a piecewise constant intercept (cluster model). This allows for large jumps in the risk surface between adjacent areas if they are in different clusters. The approach by Lee and Sarran [28] partitions the I areas into a maximum of G clusters, each with their

own intercept term $(\lambda_1, \dots, \lambda_G)$. The model is thus given by:

$$\begin{aligned}
 R_i &= S_i + \lambda_{z_i} \\
 S_i | s_{\setminus i} &\sim \mathcal{N}\left(\frac{\sum_j w_{ij} s_j}{\sum_j w_{ij}}, \frac{\sigma_s^2}{\sum_j w_{ij}}\right) \\
 \lambda_g &\sim \text{Uniform}(\lambda_{g-1}, \lambda_{g+1}) \text{ for } g = 1, \dots, G \\
 f(Z_i) &= \frac{\exp(-\delta(Z_i - G^*)^2)}{\sum_{r=1}^G \exp(-\delta(r - G^*)^2)} \\
 \delta &\sim \text{Uniform}(1, M)
 \end{aligned}$$

where $f(Z_i)$ denotes a shrinkage prior on Z_i which shrinks extreme values towards the middle intercept value. Label switching is prevented by ordering the cluster means $(\lambda_1, \dots, \lambda_G)$ so that $\lambda_1 < \lambda_2 < \dots < \lambda_G$. The penalty term $\delta(Z_i - G^*)^2$ where $G^* = (G + 1)/2$ means that if G is odd then each data point will be shrunk towards a single intercept λ_{G^*} , but if G is even, there may be two different intercept terms used even if there is a spatially smooth residual structure. Lee and Sarran [28] thus recommend setting G to be a small odd number, such as 3 or 5. Area i is assigned to one of the G intercepts by $Z_i \in \{1, \dots, G\}$, and there is no spatial smoothing imposed on the indicator vector \mathbf{Z} . The clustering is purely non-spatial, and it is the CAR prior on the S_i term that accounts for spatial autocorrelation [28].

10.2.2.3 Locally Adaptive Model

The locally adaptive model takes a similar approach to the above dissimilarity model, except that here the boundaries are not identified by the use of additional information and the modelled w_{ij} are binary only. Lee and Mitchell [27] again based this on the Leroux CAR model:

$$S_i | s_{\setminus i} \sim \mathcal{N}\left(\frac{\rho \sum_j w_{ij} s_j}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\sigma_s^2}{\rho \sum_j w_{ij} + 1 - \rho}\right).$$

Here ρ can be estimated in the model, or fixed at a specified value. (Lee and Mitchell [27] recommend 0.99.)

The spatial weights matrix starts out as the binary, first-order, adjacency matrix given by Eq. (10.1) and is subsequently updated at each iteration which allows the weights corresponding to neighbours to be estimated as either 1 or 0 (with w_{ij} fixed at zero for non-neighbouring areas). Because only weights corresponding to neighbouring areas are estimated, this approach should be more computationally feasible than areal wombling [30] where all values in \mathbf{W} are estimated.

The matrix \mathbf{W} is estimated as follows. For adjacent areas i and j : if the marginal 95% credible intervals (CIs) of s_i and s_j overlap, then set $w_{ij} = 1$; else set $w_{ij} = 0$. It is therefore not a ‘fully’ Bayesian method of estimation for these terms, as they are

not considered to be random variates. For further details, refer to Lee and Mitchell [27], who implemented this using INLA.

10.2.2.4 Weighted Sum of Spatial Priors

The BYM model with its spatially structured component S_i and its unstructured spatial component U_i was extended by Lawson and Clark [23] to be able to incorporate discontinuities:

$$R_i = p_i S_i + (1 - p_i) Z_i + U_i. \quad (10.2)$$

The Z component models abrupt discontinuities between areas. Although a range of options is possible, Lawson and Clark [23] based the prior for this parameter on the total absolute difference in risk between neighbouring areas, i.e.

$$\pi(Z_1, \dots, Z_N) \propto \frac{1}{\sqrt{\lambda}} \exp\left(-\frac{1}{\lambda} \sum_{i \sim j} |Z_i - Z_j|\right)$$

where λ acts as a constraining term.

Note that if $p_i = 1$ in Eq. (10.2), then the model reverts to the BYM model. Conversely, if $p_i = 0$, then the model is entirely discontinuous.

10.2.2.5 Leroux Scale Mixture Model

Using a scale mixture model within a Leroux prior also enables detection of abrupt changes between areas, with the advantage over the above approaches of incorporating non-normality (heavy tailed distributions). This was proposed by Congdon [12] as follows:

$$S_i | s_{\setminus i} \sim \mathcal{N}\left(\frac{\rho \sum_j w_{ij} s_j}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\sigma_s^2}{\kappa_i [\rho \sum_j w_{ij} + 1 - \rho]}\right).$$

If $\rho = 0$, this reduces to an unstructured iid scale mixture Student-t density, which is a heavy-tailed distribution. Small values of $\kappa_j (< 1)$ will indicate areas differ from their neighbours and result in less smoothing between neighbouring areas. The scale mixture is implemented by $\kappa_i \sim \text{Gam}(0.5\nu, 0.5\nu)$, where ν is a hyperparameter.

The precision matrix has the following diagonal terms [12]:

$$\{\mathbf{T}\}_{ii} = \frac{1}{\sigma_s^2} \kappa_i \left[(1 - \rho) + \rho \sum_{j \neq i} w_{ij} \right]$$

and off-diagonal terms:

$$\{\mathbf{T}\}_{ij} = \frac{1}{\sigma_s^2} \rho \kappa_i \kappa_j \mathbb{I}(i \sim j).$$

10.2.2.6 Skew-Elliptical Areal Spatial Model

Another approach that focused on incorporating skewness was introduced by Nathoo and Ghosh [36]. Here

$$R_i = \eta_i^{-0.5} (\delta |Z_i| + S_i)$$

where $\delta |Z_i|$ is the skewing component where Z_i is a set of skewing variables each independently drawn from a standard normal distribution, η provides the scale mixing and S_i is from the CAR model, i.e.

$$S_i | s_{\setminus i} \sim \mathcal{N} \left(\kappa \frac{\sum_j w_{ij} s_j}{\sum_j w_{ij}}, \frac{\sigma_s^2}{\sum_j w_{ij}} \right)$$

where κ is a spatial smoothing parameter (note that if κ is set to 0 then the distribution corresponds to uncorrelated skew-t random effects) and other terms are defined as before.

Two versions were proposed by Nathoo and Ghosh [36]. The first aims to ensure each R_i has a skew-elliptical distribution, with the marginal distribution for each spatial effect belonging to the skew-t family of distributions.

The second is a semiparametric version that uses an approximation to a Dirichlet process to allow for data-driven departures from the parametric version. This accommodates uncertainty in the mixing structure, and gives greater flexibility in the tail behaviour of marginal distributions [36].

10.2.2.7 Hidden Potts Model

In contrast to the above approaches, this model is based on a hidden Markov field, so spatial correlation occurs in an additional latent hierarchy of the model [47]. This approach was proposed by Green and Richardson [18] and assigns each area to one of several risk categories. The spatial random effect is modelled on the log scale,

as a K -component mixture model, where each component represents a different risk category, and the allocation of each area to a component follows a spatially correlated process. The number of components K is considered unknown and is estimated by the model.

$$R_i = \log(S_{z_i})$$

$$S_k \sim \text{Gamma}(a, b) \text{ for } k = 1, \dots, K$$

$$K \sim \text{Uniform}(1, K_{\max}).$$

The Potts model is proposed as the allocation model,

$$p(z|\psi, K) = \exp(\psi U(z) - \delta_k(\psi))$$

where $\psi > 0$ is the interaction parameter to be estimated and $U(z) = \sum_{i \sim j} \mathbb{I}(z_i = z_j)$ is the number of like labelled pairs of neighbouring areas. This model allows for discontinuities between areas in different risk categories and also for the amount of spatial correlation to vary by risk category. However, it does require careful MCMC implementation due to having an unknown number of risk categories and unknown area allocation to these categories. It is also more often implemented in high-dimensional data rather than disease mapping, as its greater flexibility generally has more advantages as data complexity increases [47].

10.2.2.8 Spatial Partition Model

Closely related to the above Hidden Potts model are the spatial partition models [14, 21]. These also have K non-overlapping clusters of areas, each with a constant relative risk, and K is unknown [10]. The key differences are in defining the clusters and the hyperprior specifications [10]. Specifically, the spatial partition model assigns up to K areas as cluster centres, which are allocated with a uniform prior probability, and the number of clusters is chosen according to the distribution $p(K = k) \propto (1 - c)^k$ where $c \in [0, 1)$ is fixed *a priori*. Smaller values of c makes this prior less informative, with the limiting case $c = 0$ yielding a uniform distribution. The remaining $N - K$ areas are then assigned to their nearest cluster, according to the minimal number of boundaries that have to be crossed. Both this model and the above hidden Potts model have been criticised for forcing discontinuities into a surface, and for assuming constant relative risk within a cluster [23].

10.2.2.9 Local Spline Model

An extension to the global spline models described in Sect. 10.2.1 that results in a less smooth surface is the incorporation of unstructured random effects as in the

penalised random individual dispersion effects (PRIDE) model, originally proposed by Perperoglou and Eilers [37]. Here

$$R_i = f(c_{1i}, c_{2i}) + \gamma_i$$

where γ_i is an area-specific random effect, whose vector follows a multivariate normal distribution [17]. This means that the covariance matrix captures the unstructured heterogeneity by containing an identity matrix multiplied by a variance component, in addition to the eigenvalues from the P-spline model component [17].

10.3 Case Study

10.3.1 Data

Since the dissemination of actual cancer data is restricted due to privacy and confidentiality requirements of the data custodians, simulated data that reflected the general distributions of actual data were generated to enable data sharing and reproduction of the presented results (see contact the authors for data and model code). Two datasets were generated that reflected the incidence of cancer types with a strong socioeconomic gradient: one with low total counts per geographical area over ten years (median of 2, range 0–19), considered a rare cancer, and one with higher counts over 10 years (median 25 cases, range 0–163), considered a common cancer. The focus on socioeconomic gradients meant we expected neighbouring areas having different socioeconomic levels would have different incidence rates.

The areas used were statistical areas 2 (SA2s) based on the 2011 Australian Statistical Geography Standard (ASGS) boundaries [4]. After excluding some areas with no/nominal resident populations, the number of areas was 2153. The median population of the included SA2s was 9055 (range: 3–50,251). Land area size varied from 0.8 to 520,000 km², with a median of 15.6 km².

10.3.2 Model Selection

Of the fourteen models introduced in Sect. 10.2 and described in Table 10.1, five were excluded from the application. Two of these were on theoretical grounds: the localised P-spline and the proper CAR models. The localised P-spline model was not investigated because implementing the P-spline had many challenges within the Australian context of vastly differing area sizes. The disadvantages of the proper CAR formulation such as the potentially limited breadth of estimates have limited appeal for spatial modelling [5]. We attempted to run a Hidden Potts model, spatial partition model and skew-elliptical areal spatial model, but were unable to

successfully achieve this due to the computational complexity of the models, so they are also excluded from this section. The skew-elliptical model was unable to compile in WinBUGS [31], while the multidimensionality required for the spatial partition model and Hidden Potts model became too unwieldy.

10.3.3 Model Variants

Of the nine models successfully implemented, multiple variants were considered for the global P-spline model, CAR dissimilarity models, localised autocorrelation models, and the locally adaptive models, and these are detailed below. Specifications for the geostatistical model are also documented. These resulted in a total of 13 versions of models applied to the simulated data (Table 10.1).

Table 10.1 Software used for models applied to simulated data

Models investigated	Authors	Software used
<i>Global spatial smoothing</i>		
BYM (Intrinsic CAR)	Besag et al. [8]	R (CARBayes)
Proper CAR	Besag [6]	–
Leroux	Leroux et al. [29]	R (CARBayes)
Geostatistical	Clements et al. [11]	JAGS
P-spline (tensor)	Lang and Brezger [22]	JAGS
P-spline (radial)	Ruppert et al. [42]	JAGS
<i>Local spatial smoothing</i>		
CAR dissimilarity model (binary)	Lee and Mitchell [26]	R (CARBayes)
CAR dissimilarity model (non-binary)	Lee and Mitchell [26]	R (CARBayes)
Localised autocorrelation ($G = 3$)	Lee and Sarran [28]	R (CARBayes)
Localised autocorrelation ($G = 5$)	Lee and Sarran [28]	R (CARBayes)
Locally adaptive model (ρ estimated)	Lee and Mitchell [27]	R (INLA)
Locally adaptive model ($\rho = 0.99$)	Lee and Mitchell [27]	R (INLA)
Weighted sum of spatial priors	Lawson and Clark [23]	WinBUGS
Leroux scale mixture	Congdon [12]	WinBUGS
Skew-elliptical areal spatial	Nathoo and Ghosh [36]	–
Hidden Potts	Green and Richardson [18]	–
Spatial partition	Denison and Holmes [14], Knorr-Held and Raßer [21]	–
Local spline	Goicoa et al. [17], Perperoglou and Eilers [37]	–

10.3.3.1 Global P-spline Model

Two formulations of the global P-spline model were implemented: the first uses a tensor product (refer to Sect. 10.2.1.5 for a definition) to define the basis, and the second uses a radial basis based on distances. No further modifications were made to the tensor product version.

The radial P-spline model had the knots evenly spaced at intervals of 5 degrees of latitude and longitude, as shown in Fig. 10.1. Knots which were too distant from the centroids of SA2 areas were subsequently dropped. A total of 47 knots were retained for modelling. Based on these knots, Δ was set to 500.

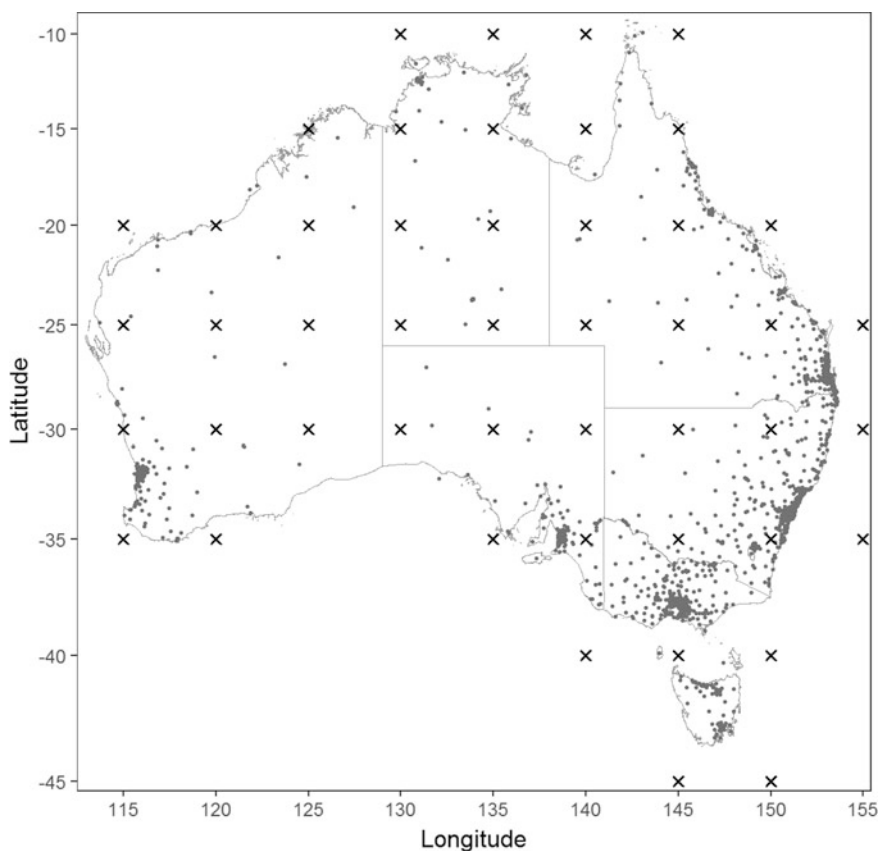


Fig. 10.1 Location of knots (crosses) in relation to SA2 centroids (dots) for the P-spline radial model

10.3.3.2 CAR Dissimilarity Model

The CAR dissimilarity model can also be applied in a variety of forms. As discussed in Sect. 10.2.2.1, the weighting matrix can be binary or non-binary, and the dissimilarity measure can be based on distance, geographical features (such as railways or mountains), or covariate information. Here we examine both binary and non-binary forms of this model based on the Socioeconomic Indexes for Areas (SEIFA) dissimilarity. This gives a continuous score for each area which is designated based on a range of socioeconomic measures, including house ownership, car ownership, employment and internet access. Several indices are available, and we used the Index of Relative Socioeconomic Disadvantage. Further details on SEIFA are available in Australian Bureau of Statistics [ABS] [3].

10.3.3.3 Localised Autocorrelation Models

Two variants of this model were assessed based on the value of G , the maximum number of clusters, being set to 3 or 5. See Sect. 10.2.2.2 for discussion of these choices.

10.3.3.4 Locally Adaptive Models

Two variants of this model were assessed based on the value of ρ , the spatial autocorrelation parameter, one being set to 0.99 (as recommended by Lee and Mitchell [27]) and the other allowed to vary between 0 and 1. The aim of fixing the value of ρ close to one is to ensure there is spatial smoothing occurring when $w_{ij} > 0$. Note that if $\rho = 0$ then w_{ij} vanishes from the model and cannot be used to determine if discontinuities are present. Setting ρ to 1 is not ideal, as the precision matrix would become singular.

10.3.3.5 Geostatistical Model

The geostatistical model had two adjustments made to provide a better fit. First, the priors for λ and k were changed according to the possible values of spatial correlation observed given different combinations of λ , k , and distances d_{ij} . This exploratory analysis suggested using

$$\lambda \sim \text{Uniform}(0.01, 1)$$

$$k \sim \text{Uniform}(0.1, 20).$$

To allow for further flexibility, λ and k were replaced by one of $\{\lambda_1, \dots, \lambda_5\}$ and $\{k_1, \dots, k_5\}$ respectively according to the remoteness of the area (major city, inner regional, outer regional, remote, and very remote) to allow the degree of smoothing to vary between the five levels of remoteness. Second, to make this model computationally feasible, the distance matrix $\{d\}_{ij}$ was modified by imposing a remoteness-specific radius of influence $\{r_1, \dots, r_5\}$ on each area, such that areas beyond this threshold are not considered neighbours. These radii were $\{50, 100, 200, 400, 800\}$ km respectively. This induces a Markov random field (MRF) structure which should have only a negligible effect on parameter estimation while greatly increasing computational efficiency. Some remote and very remote areas are relatively close to major city and inner regional areas, which can lead to some areas having more than 1000 neighbouring SA2s, thereby drastically reducing any computational gains. Therefore, the imposed MRF was further modified to exclude major city areas as neighbours of remote areas, and to exclude both major city and inner regional areas as neighbours of very remote areas. This is also sensible given the differences in cancer incidence and underlying influences between these areas [13]. This was achieved by setting the distances to these excluded areas to infinity. The result of these adjustments lead to

$$\{\mathbf{S}\}_{ij} = \begin{cases} \exp\left(-(\lambda_{z_i} d_{ij})^{k_{z_i}}\right) & \text{if } d_{ij} \leq r_{z_i} \\ 0 & \text{if } d_{ij} > r_{z_i} \end{cases}$$

$$S_i = f(\mathbf{S}) = \frac{1}{N_i} \sum_{j=1}^{N_r} \{\mathbf{S}\}_{ij}$$

where N_i is the number of areas within a radius of r_{z_i} units from the centroid of area i (including area i), $N_r = \max_i \{N_i\}$, and z_i represents the degree of remoteness for area i , where $z_i = 1$ corresponds to an area in a major city.

10.3.4 Statistical Software

Code for implementing the models in freely available software (Table 10.1) is available on request, as are the data sets.

The main software used to implement the statistical models were WinBUGS [31] and JAGS [38], which were run via R [39] using the packages R2WinBUGS [44] and R2jags [45] respectively, and also the R package CARBayes [25]. R-INLA [33] was also used for one model.

10.3.5 Model Comparison

Models were compared using several criteria, which are described below. The posterior SIR was calculated as $\exp(\mu_i) = \exp(\alpha + R_i)$, as no covariates were included in these models. The median, lower and upper bounds of the 80% CIs were calculated as the 50th, 10th and 90th percentiles of the posterior, respectively.

10.3.5.1 Convergence

Convergence was predominately based on calculating the Geweke convergence diagnostic [16] for each area's posterior SIR. A p-value for the test statistic below 0.01 was interpreted as suggestive evidence of non-convergence for that area. The trace and density plots for a subsample of areas were also examined for convergence.

10.3.5.2 Plausibility of Estimates

To determine how plausible the posterior SIR estimates were, the CI width was visually inspected, with unreasonably large CIs (with many of the 80% CIs spanning $\pm 5000\%$ or more of the median estimate) providing evidence the estimate was not well-defined; while very precise estimates (the majority within $\pm 4\%$) were evidence that uncertainty was not appropriately included. The magnitude of smoothing of the median posterior SIRs in comparison to the raw SIRs was also visually examined. A smoothed SIR which was very similar to the raw SIR was suggestive of under-smoothing, particularly in areas with small populations.

10.3.5.3 Model Goodness-of-Fit

Three model goodness-of-fit measures were considered: Deviance information criterion (DIC) [43], Watanabe-Akaike information criterion (WAIC) [48] and Moran's I on the residuals [35].

DIC and WAIC are both useful for comparing the predictive accuracy between models. Although DIC is a commonly used measure to compare Bayesian models, WAIC has several advantages over DIC, including that it closely approximates Bayesian cross-validation, it uses the entire posterior distribution and it is invariant to parameterisation [46]. For both these measures, smaller values indicate a better fitting model.

Moran's I was applied to the model residuals to determine if spatial autocorrelation was present after fitting the models. This measure can be quite sensitive to the spatial weights matrix used to define the spatial dependencies between areas, and while a range of spatial weights matrices (inverse-distance, third-order neighbours

etc) were considered, we used a matrix based on first-order neighbours. As values of Moran's I close to 0 indicate very low or no residual spatial autocorrelation, here we consider values above 0.2 to be suggestive of some positive spatial autocorrelation. The closer Moran's I is to zero, the better the model accounts for spatial autocorrelation [2].

10.3.5.4 Computational Time

The `microbenchmark` R package [34] was used to monitor computational time to run each model. The models were run on two different computers. However, the specifications of these computers were similar and any differences should have a negligible influence on computation time.

10.4 Results and Discussion

Substantive differences in the posterior estimates were observed between the 13 model variants applied, especially for the rare cancer (Table 10.2, Figs. 10.2, 10.3, 10.4, and 10.5). Depending on the model chosen, the modelled SIR estimates for the same geographical area could range from well below to well above the Australian average (Figs. 10.3 and 10.5).

While small numbers in geographical areas require smoothing, it remains possible that the neighbouring areas may have genuinely different incidence rates. These differences would be obscured during the smoothing process. Detecting these differences is problematic, and even many of the models designed to allow for local variation gave results similar to the BYM and Leroux models (Figs. 10.2, 10.3, 10.4, and 10.5), suggesting there was insufficient statistical power to adequately detect local differences. Of the models that obtained greater variation in the median SIR estimates between areas and less smoothing, there was often excessive uncertainty around these estimates, such as the localised autocorrelation model results (Figs. 10.2 and 10.4).

The number of area-specific SIR estimates that had evidence of non-convergence (based on Geweke p -value < 0.01) did vary between models and with the extent of data sparseness. In many cases, very wide CIs were symptomatic of non-convergence. For instance, the localised autocorrelation ($G = 3$) model for the rare cancer had 86% of area-specific SIRs with significant Geweke p -values, suggesting lack of convergence, and this model had among the widest CIs (Table 10.2). In contrast, models which had implausibly narrow CIs generally had very few/no areas with small Geweke p -values. However, overly narrow CIs are equally problematic as they over-exaggerate confidence in the plausibility of the estimates, which may actually be over- or under-smoothed.

In general, especially as data sparsity increased, our application of these models suggested that global models with more smoothing tended to have 'well-behaved',

Table 10.2 Quantitative summary of results across model criteria

	Posterior median smoothed SIR	N (%) Geweke $p < 0.01$	DIC	WAIC	Moran's I on residuals	Computation time (s)
<i>Rare cancer</i>						
BYM	0.95 (0.38, 1.34)	93 (4.3%)	14,511.3	7432.6	0.047	420.8
Leroux	0.95 (0.39, 1.34)	100 (4.6%)	14,523.9	7429.1	0.047	309.5
Geostatistical	0.94 (0.51, 1.84)	3 (0.1%)	14,646.9	7331.3	0.204	76,772.8
P-spline (tensor)	0.95 (0.13, 1.87)	0 (0.0%)	14,515.4	7279.9	0.205	8117.2
P-spline (radial)	0.93 (0.34, 1.97)	4 (0.2%)	14,760.3	7402.6	0.241	1960.1
CAR dissimilarity (binary)	0.94 (0.09, 4.27)	55 (2.6%)	14,360.3	7354.9	0.042	8462.3
CAR dissimilarity (non-binary)	0.94 (0.16, 3.62)	50 (2.3%)	14,315.4	7361.2	0.031	5316.0
Localised autocorrelation ($G = 3$)	0.96 (0.00, 1.71)	1858 (86.3%)	13,571.9	8107.0	0.054	1035.6
Localised autocorrelation ($G = 5$)	0.99 (0.00, 1.61)	151 (7.0%)	13,129.9	7435.9	0.068	1081.4
Locally adaptive (ρ estimated)	0.95 (0.38, 1.34)				0.067	1580.7
Locally adaptive ($\rho = 0.99$)	0.95 (0.38, 1.34)				0.064	264.9
Weighted sum of spatial priors	0.98 (0.09, 2.03)	64 (3.0%)	14,421.8	7414.2	0.094	6815.7
Leroux scale mixture	Would not run					
<i>Common cancer</i>						
BYM	1.00(0.74, 1.62)	25 (1.2%)	23,680.8	12,148.9	0.131	419.4
Leroux	1.00(0.74, 1.61)	33 (1.5%)	23,686.6	12,135.1	0.122	308.1
Geostatistical	0.98(0.69, 1.62)	14 (0.7%)	22,213.4	11,218.7	0.220	98,412.3
P-spline (tensor)	0.97(0.71, 1.57)	0 (0.0%)	22,173.0	11,110.1	0.228	7640.8

P-spline (radial)	0.97 (0.71,2.38)	0 (0.0%)	22,167.1	11,120.7	0.254	1831.6
CAR dissimilarity (binary)	1.00 (0.74,1.61)	26 (1.2%)	23,681.4	12,132.1	0.121	4617.5
CAR dissimilarity (non-binary)	0.98 (0.65,2.53)	51 (2.4%)	22,947.1	11,698.9	0.120	9188.9
Localised autocorrelation ($G = 3$)	1.00 (0.74,1.62)	27 (1.3%)	23,690.2	12,131.6	0.124	898.4
Localised autocorrelation ($G = 5$)	1.00 (0.74,1.61)	40 (1.9%)	23,690.9	12,132.2	0.124	907.2
Locally adaptive (ρ estimated)	1.00 (0.74,1.61)				0.148	556.0
Locally adaptive ($\rho = 0.99$)	1.00 (0.74,1.61)				0.147	138.2
Weighted sum of spatial priors	1.00 (0.75,1.62)	28 (1.3%)	23,761.2	12,218.1	0.182	6553.6
Leroux scale mixture	1.00 (0.78,1.52)	19 (0.9%)	23,983.4	12,311.3	0.198	9044.0

BYM Besag, York and Mollié, *DIC* Deviance Information Criterion, *SIR* Standardised incidence ratio, *WAIC* Watanabe-Akaike Information Criterion
Notes:

- The posterior median smoothed SIR is the median SIR estimate over all MCMC iterations for each area, then the median over all areas
- The lowest values in each column by type of cancer are shaded light grey. Any Moran's I above 0.2 are shaded a darker grey, as this is considered suggestive of some spatial correlation existing in the residuals
- The Leroux scale mixture model would not run for the rare cancer
- % Geweke, DIC and WAIC are unavailable for the locally adaptive results (which used INLA via custom functions), as it was not possible to obtain the necessary posterior samples to use in our calculations

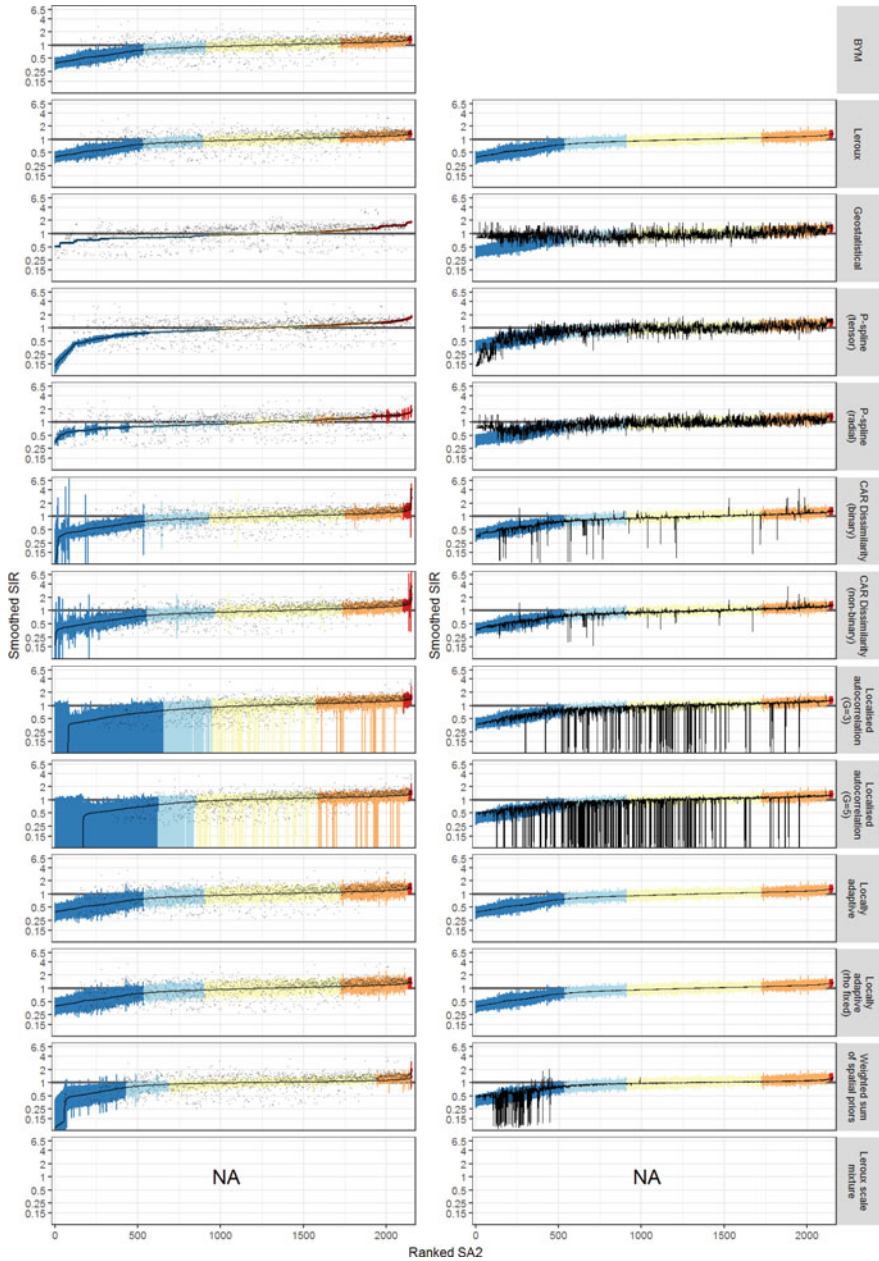


Fig. 10.2 Graphs of posterior SIR results by model, rare cancer. Note: Axes are consistent. Column 1 shows the 80% CI (shaded as per the tones on the maps in Fig. 10.3), the black line is the median SIR (in ascending order), the dots are the raw SIRs and the horizontal line at 1 represents the national average. For column 2, the 80% CIs are the BYM model, and the SA2s are ordered according to the BYM median SIR. The black line is the median estimate for the model named

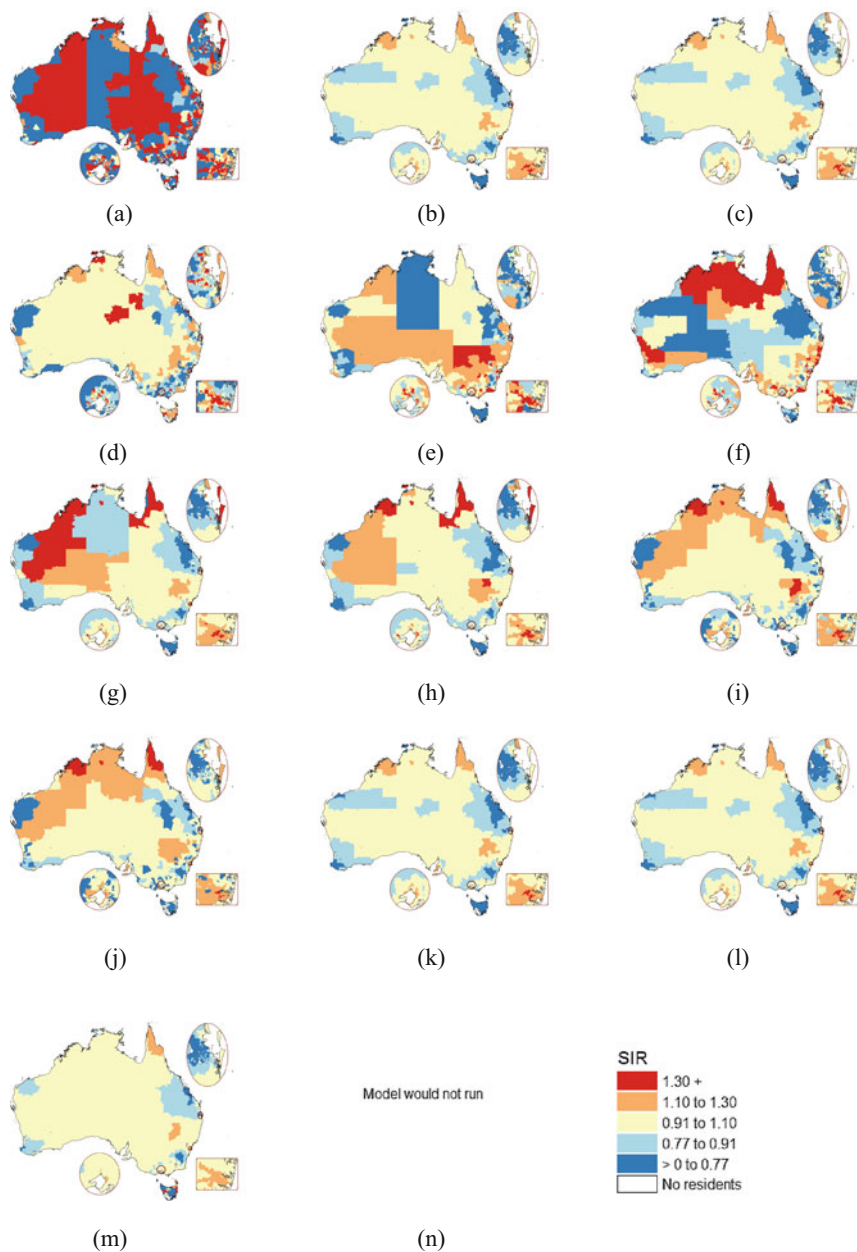


Fig. 10.3 Rare cancer median posterior SIR mapped by model. (a) Raw (observed/expected), (b) BYM, (c) Leroux, (d) Geostatistical, (e) P-spline (tensor), (f) P-spline (radial), (g) CAR dissimilarity (binary), (h) CAR dissimilarity (non-binary), (i) Localised autocorrelation ($G = 3$), (j) Localised autocorrelation ($G = 5$), (k) Locally adaptive (ρ estimated), (l) Locally adaptive ($\rho = 0.99$), (m) Weighted sum of spatial priors, (n) Leroux scale mixture

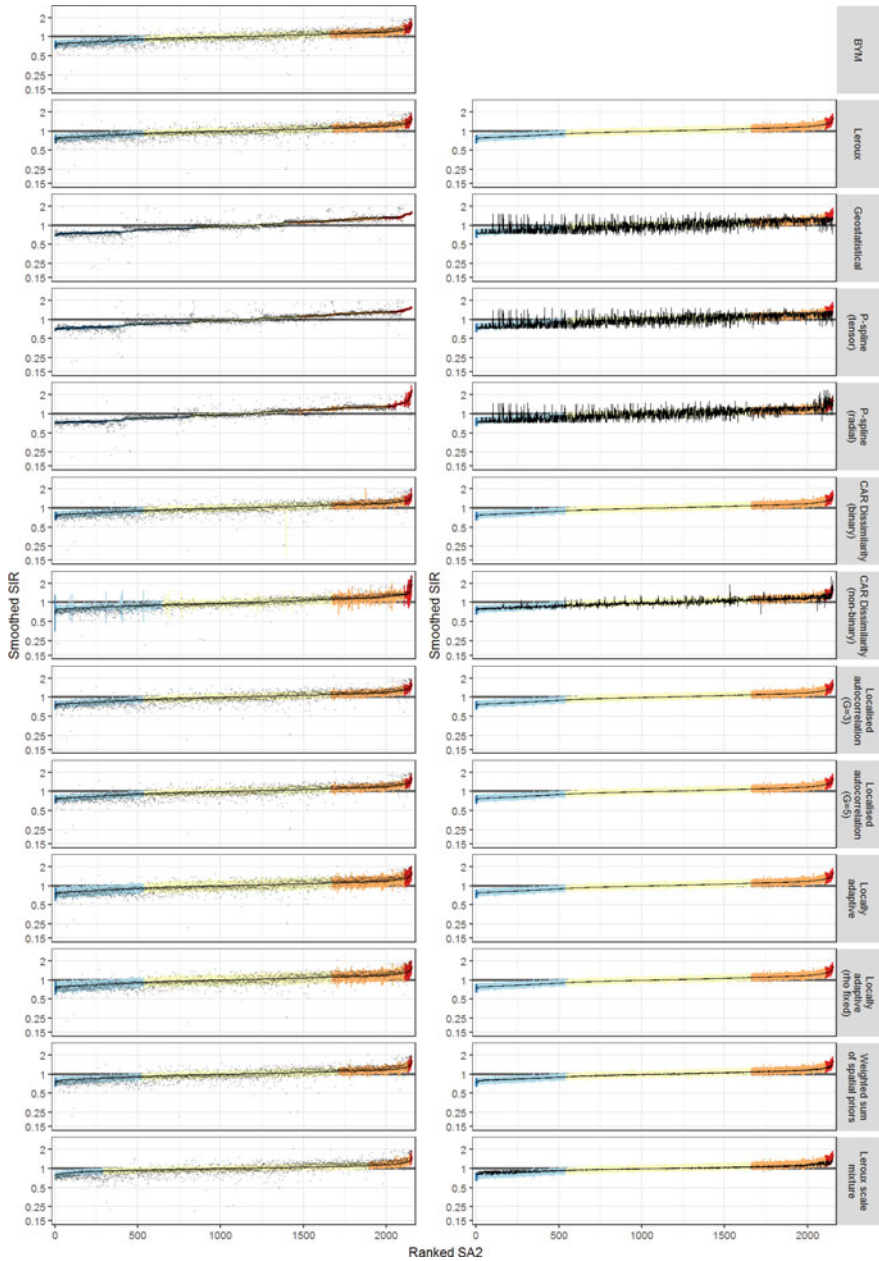


Fig. 10.4 Graphs of posterior SIR results by model, common cancer. Note: Axes are consistent. Column 1 shows the 80% CI (shaded as per the tones on the maps in Fig. 10.5), the black line is the median SIR (in ascending order), the dots are the raw SIRs and the horizontal line at 1 represents the national average. For column 2, the 80% CIs are the BYM model, and the SA2s are ordered according to the BYM median SIR. The black line is the median estimate for the model named

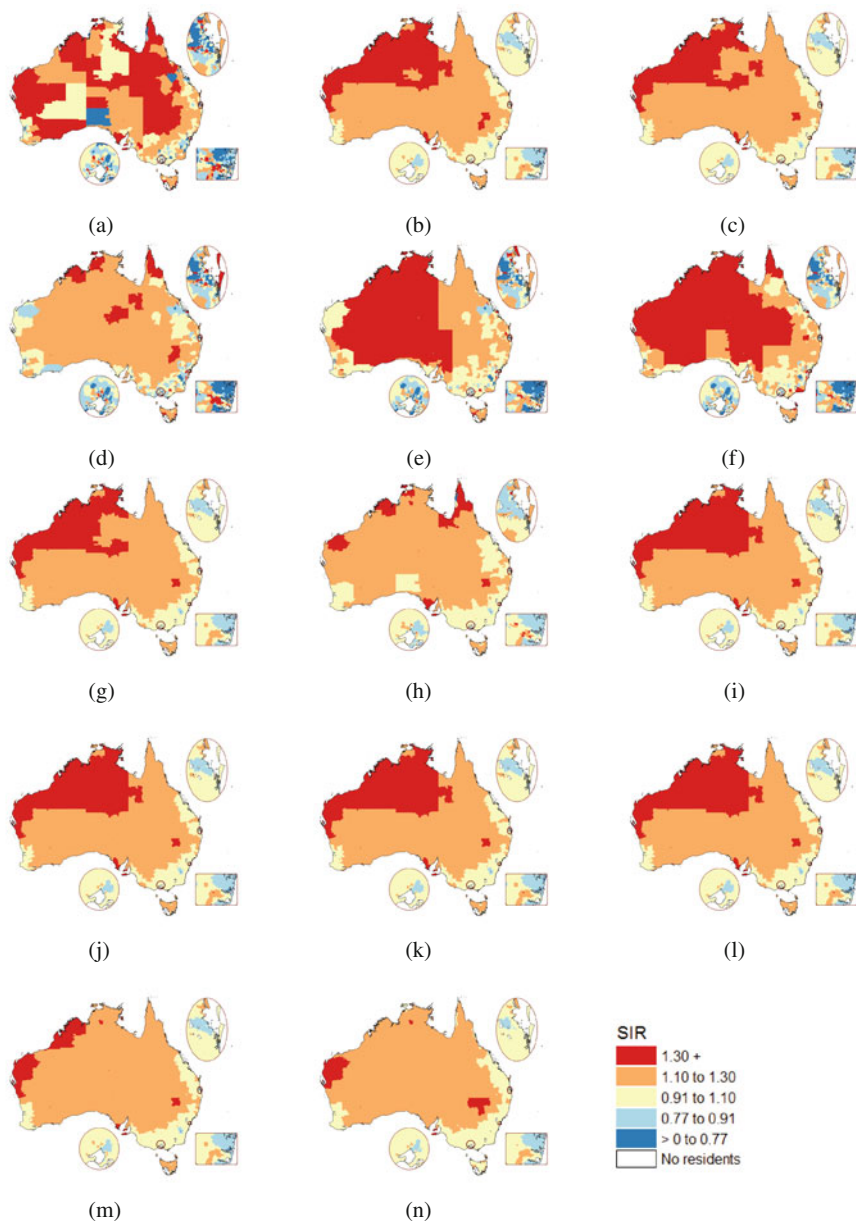


Fig. 10.5 Common cancer median posterior SIR mapped by model. (a) Raw (observed/expected), (b) BYM, (c) Leroux, (d) Geostatistical, (e) P-spline (tensor), (f) P-spline (radial), (g) CAR dissimilarity (binary), (h) CAR dissimilarity (non-binary), (i) Localised autocorrelation ($G = 3$), (j) Localised autocorrelation ($G = 5$), (k) Locally adaptive (ρ estimated), (l) Locally adaptive ($\rho = 0.99$), (m) Weighted sum of spatial priors, (n) Leroux scale mixture

reliable estimates, while local models tended to struggle in producing plausible estimates (Figs. 10.2 and 10.4). The estimates for the binary CAR dissimilarity model (based on socioeconomic differences) in our study were often unreliable, and this is likely due to its tendency to remove too many neighbours. This is expected to also apply to other formulations, such as distance-based models.

The DIC and WAIC (Table 10.2) measures of goodness of model fit were generally in consensus for a given cancer type, apart from the localised autocorrelation models which had among the lowest DIC, but highest WAIC. Some models fit the data well for one type of simulated data, but not the other. For example, the geostatistical and P-spline models fit the common cancer quite well, but resulted in poor to average model fit for the rare cancer.

Moran's I statistic (Table 10.2) generally indicated that the residual spatial autocorrelation is quite small. The only models with noticeable remaining correlation were the centroid based geostatistical and P-spline models, and this apparent correlation may result from using a weights matrix based on first-order neighbours when calculating Moran's I.

Computational time varied substantially across the models, with times for the rare cancer ranging from 5 minutes (Leroux model in CARBayes) to over 20 hours (geostatistical model in JAGS). Models able to be run in CARBayes were generally very fast, while models run in JAGS or WinBUGS took longer (approximately between 0.5 and 2.5 hours, excluding the geostatistical model). Of note though, are the implications these varying computing times may have when many models need to be run, such as considering multiple cancer types, or repeating models to test different hyperprior specifications. While increasing computing specifications may reduce these times, it is still an important consideration when choosing between two (or more) otherwise well performing models.

It is a tenet of statistical research that the choice of model depends on the data characteristics and the aims of the analysis. However, when data are sparse and there is extreme variation in area size, such as are consistent with our simulated data, we found that the geostatistical or P-spline models generally had poor performance. The geostatistical model is prohibitively slow for these type of data, and when combined with the unpredictable model fit, this model is not recommended.

A previous comparison by Adin et al. [1] of the global P-spline model against the moving average and CAR models found the P-spline performed well for sparse disease mapping, although Goicoa et al. [17] found it to be more prone to detecting more false high-risk areas than either the CAR or a local P-spline model. This model is also rather complex to implement, requiring a penalty matrix and the number of knots to be specified, both of which are subjective and can have a large impact on model fit. The main concern with the P-spline model, however, was the specification of the basis matrix using the tensor product, which does not adequately address the fact that the SA2s are irregular in shape and the distances between their centroids can be vastly different. The radial basis version of the P-spline model was designed to address this, but aside from being computationally faster, it provided similar levels of smoothing and a worse model fit.

The BYM and Leroux models may be prone to over-smoothing when neighbouring areas have abrupt differences [23, 27], but they generally converged, provided a reasonable model fit with plausible estimates and were computationally efficient to implement. The Leroux model may be preferred over the BYM model to avoid the inability of the BYM model to identify both the structured and unstructured spatial random effects separately, but we found that in some cases it struggled to achieve convergence for its mixing parameter.

The locally adaptive models provided results similar to that of the BYM model, with slightly wider credible intervals. The main disadvantage was the difficulty in obtaining samples from the posterior due to the script calling up INLA from within another function.

A non-binary dissimilarity model may also provide an adequate fit, as this smooths more than a P-spline but less than BYM or Leroux. The non-binary dissimilarity formulation using the SEIFA covariate worked quite well for both cancer types, with noticeably less constraining of modelled SIR estimates than under BYM or Leroux. Whether these SIR estimates are appropriate or are under-smoothed will depend on data characteristics and the aims of the analysis.

Note that the final specification of each model requires additional sensitivity analyses to determine the influence of the priors and hyperpriors, the topic of which was outside the scope of this chapter.

10.5 Conclusion

The number of Bayesian spatial models available continues to increase, along with the capacity of the computing software and hardware. Determining the optimal amount of smoothing in spatial analyses remains difficult, but our study demonstrates the benefits of running a range of model types and provides insights into the relative merits of the different models for the study dataset. Comparing estimates from several different model types is important to assess consistency of results when conducting a spatial analysis

In summary, in sparse data contexts, the BYM, Leroux, locally adaptive, non-binary CAR dissimilarity models or some versions of localised autocorrelation models may outperform the other models examined. We suggest considering using centroid-based smoothing models only when areas are of similar size and shape.

References

1. A. Adin, M.A. Martinez-Beneito, P. Botella-Rocamora, T. Goicoa, M.D. Ugarte. Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive, and moving average models. *Stoch. Environ. Res. Risk Assess.* **31**(2), 403–415 (2017). <https://doi.org/10.1007/s00477-016-1269-8>

2. C. Anderson, L. Ryan, A Comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *Int. J. Environ. Res. Public Health* **14**(2), 146 (2017). <https://doi.org/10.3390/ijerph14020146>
3. Australian Bureau of Statistics [ABS], Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 'Statistical Area Level 2, Indexes, SEIFA 2011', data cube: Excel spreadsheet, cat. no. 2033.0.55.001. (2013). www.abs.gov.au/AUSSTATS/abs.nsf/DetailsPage/2033.0.55.0012011
4. Australian Bureau of Statistics [ABS], Australian Statistical Geography Standard (ASGS): Volume 1 - Main structure and greater capital city statistical areas, July 2011, cat. no. 1270.0.55.001 (2013). www.abs.gov.au/AUSSTATS/abs.nsf/DetailsPage/1270.0.55.001July%202011
5. S. Banerjee, B.P. Carlin, A.E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*. Monographs on Statistics and Applied Probability, 2nd edn., vol. 135 (CRC Press/Chapman & Hall, Boca Raton, 2014)
6. J. Besag, Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **36**(2), 192–236 (1974)
7. J. Besag, C. Kooperberg, On conditional and intrinsic autoregressions. *Biometrika* **82**(4), 733–746 (1995). <https://doi.org/10.1093/biomet/82.4.733>
8. J. Besag, J. York, A. Mollié, Bayesian image restoration with application in spatial statistics. *Ann. Inst. Stat. Math.* **43**(1), 1–20 (1991). <https://doi.org/10.1007/BF00116466>
9. N. Best, S. Cockings, J. Bennett, P. Wakefield, Ecological regression analysis of environmental benzene exposure and childhood leukaemia: sensitivity to data inaccuracies, geographical scale and ecological bias. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* **164** (1), 155–174 (2001). <https://doi.org/10.1111/1467-985x.00194>
10. N. Best, S. Richardson, A. Thomson, A comparison of Bayesian spatial models for disease mapping. *Stat. Methods Med. Res.* **14**(1), 35–59 (2005). <https://doi.org/10.1191/0962280205sm388oa>
11. A.C.A. Clements, N.J.S. Lwambo, L. Blair, U. Nyandindi, G. Kaatano, S. Kinung'hi, J.P. Webster, A. Fenwick, S. Brooker, Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop. Med. Int. Health* **11** (4), 490–503 (2006). <https://doi.org/10.1111/j.1365-3156.2006.01594.x>
12. P. Congdon, Representing spatial dependence and spatial discontinuity in ecological epidemiology: a scale mixture approach. *Stoch. Environ. Res. Risk Assess.* **31**(2), 291–304 (2017). <https://doi.org/10.1007/s00477-016-1292-9>
13. S.M. Cramb, K.L. Mengersen, P.D. Baade, Identification of area-level influences on regions of high cancer incidence in Queensland, Australia: a classification tree approach. *BMC Cancer* **11**, 311 (2011). <https://doi.org/10.1186/1471-2407-11-311>
14. D.G.T. Denison, C.C. Holmes, Bayesian partitioning for estimating disease risk. *Biometrics*, **57**(1), 143–149 (2001). <https://doi.org/10.1111/j.0006-341x.2001.00143.x>
15. L.E. Eberly, B.P. Carlin, Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Stat. Med.* **19**(17), 2279–2294 (2000). [https://doi.org/10.1002/1097-0258\(20000915/30\)19:17/18<2279::aid-sim569>3.0.co;2-r](https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2279::aid-sim569>3.0.co;2-r)
16. J. Geweke, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in *Bayesian Statistics*, vol. 4, ed. by J.M. Bernardo, J. Berger, A.P. Dawid, A.F.M. Smith (Oxford University Press, Oxford, 1992), pp. 169–193
17. T. Goicoa, M.D. Ugarte, J. Etxeberria, A.F. Militino, Comparing CAR and P-spline models in spatial disease mapping. *Environ. Ecol. Stat.* **19**(4), 573–599 (2012). <https://doi.org/10.1007/s10651-012-0201-8>
18. P.J. Green, S. Richardson, Hidden Markov models and disease mapping. *J. Am. Stat. Assoc.* **97**(460), 1055–1070 (2002). <https://doi.org/10.1198/016214502388618870>
19. C. Kandhasamy, K. Ghosh, Relative risk for HIV in India—an estimate using conditional autoregressive models with Bayesian approach. *Spatial Spatio-temporal Epidemiol.* **20**, 13–21 (2017). <https://doi.org/10.1016/j.sste.2017.01.001>

20. S.Y. Kang, S.M. Cramb, N.M. White, S.J. Ball, K.L. Mengersen, Making the most of spatial information in health: a tutorial in Bayesian disease mapping for areal data. *Geospat. Health* **11**(2), 428 (2016). <https://doi.org/10.4081/gh.2016.428>
21. L. Knorr-Held, G. Raßer, Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56**(1), 13–21 (2000). <https://doi.org/10.1111/j.0006-341x.2000.00013.x>
22. S. Lang, A. Brezger, Bayesian P-Splines. *J. Comput. Graph. Stat.* **13**(1), 183–212 (2004). <https://doi.org/10.1198/1061860043010>
23. A.B. Lawson, A. Clark, Spatial mixture relative risk models applied to disease mapping. *Stat. Med.* **21**(3), 359–370 (2002). <https://doi.org/10.1002/sim.1022>
24. D. Lee, A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial Spatio-temporal Epidemiol.* **2**(2), 79–89 (2011). <https://doi.org/10.1016/j.sste.2011.03.001>
25. D. Lee, *CARBayes Version 4.7: An R package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors* (University of Glasgow, Glasgow, 2017). <https://CRAN.R-project.org/package=CARBayes>
26. D. Lee, R. Mitchell, Boundary detection in disease mapping studies. *Biostatistics* **13** (3), 415–426 (2012). <https://doi.org/10.1093/biostatistics/kxr036>
27. D. Lee, R. Mitchell, Locally adaptive spatial smoothing using conditional auto-regressive models. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **62**(4), 593–608 (2013). <https://doi.org/10.1111/rssc.12009>
28. D. Lee, C. Sarran, Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics* **26**(7), 447–487 (2015). <https://doi.org/10.1002/env.2348>
29. B.G. Leroux, X. Lei, N. Breslow, Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Stat. Models Epidemiol. Environ. Clin. Trials* **116**, 179–191 (2000). https://doi.org/10.1007/978-1-4612-1284-3_4
30. H. Lu, C. Reilly, S. Banerjee, B. Carlin, Bayesian areal wombling via adjacency modelling. *Environ. Ecol. Stat.* **14**(4), 433–452 (2007). <https://doi.org/10.1007/s10651-007-0029-9>
31. D.J. Lunn, A. Thomas, N. Best, D. Spiegelhalter, WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* **10**(4), 325–337 (2000). <https://doi.org/10.1023/a:1008929526011>
32. Y.C. MacNab, Spline smoothing in Bayesian disease mapping. *Environmetrics* **18**(7), 727–744 (2007). <https://doi.org/10.1002/env.876>
33. T.G. Martins, D. Simpson, F. Lindgren, H. Rue, Bayesian computing with INLA: new features. *Comput. Stat. Data Anal.* **67**, 68–83 (2013). <https://doi.org/10.1016/j.csda.2013.04.014>
34. O. Mersmann, microbenchmark: accurate timing functions. R package version 1.4-6 (2018). <http://CRAN.R-project.org/package=microbenchmark>
35. P.A.P. Moran, Notes on continuous stochastic phenomena. *Biometrika* **37**(1/2), 17–23 (1950). <https://doi.org/10.2307/2332142>
36. F.S. Nathoo, P. Ghosh, Skew-elliptical spatial random effect modeling for areal data with application to mapping health utilization rates. *Stat. Med.* **32**(2), 290–306 (2013). <https://doi.org/10.1002/sim.5504>
37. A. Perperoglou, P.H.C. Eilers, Penalized regression with individual deviance effects. *Comput. Stat.* **25**(2), 341–361 (2010). <https://doi.org/10.1007/s00180-009-0180-x>
38. M. Plummer, JAGS version 4.3.0 user manual (2017). <https://CRAN.R-project.org/package=CARBayes>
39. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing (2018). <http://www.R-project.org/>
40. S. Richardson, Statistical methods for geographical correlation studies, in *Geographical and Environmental Epidemiology*, ed. by P. Elliot, J. Cuzick, D. English, R. Stern (Oxford University Press, Oxford, 1996), pp. 181–204
41. A. Riebler, S.H. Sørbye, D. Simpson, H. Rue, An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* **25**(4), 1145–1165 (2016). <https://doi.org/10.1177/0962280216660421>

42. D. Ruppert, M.P. Wand, R.J. Carroll, *Semiparametric Regression* (Cambridge University Press, Cambridge, 2003)
43. D.J. Spiegelhalter, N.G. Best, B.P. Carlin, V. der Linde, Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **64** (4):583–640 (2002). <https://doi.org/10.1111/1467-9868.00353>
44. S. Sturtz, U. Ligges, A. Gelman, R2WinBUGS: a package for running WinBUGS from R. *J. Stat. Softw.* **12**(3), 1–16 (2005). <https://doi.org/10.18637/jss.v012.i03>
45. Y.-S. Su, M. Yajima, R2jags: using R to run ‘JAGS’. R package version 0.5-7 (2015). <https://CRAN.R-project.org/package=R2jags>
46. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**(5), 1413–1432 (2017). <https://doi.org/10.1007/s11222-016-9709-3>
47. L.A. Waller, B.P. Carlin, Disease mapping, in *Handbook of Spatial Statistics*, ed. by A.E. Gelfand, P.J. Diggle, P. Guttorp, M. Fuentes (CRC Press, Boca Raton, 2010)
48. S. Watanabe, Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**, 3571–3594 (2010)