Bayesian Modeling of Spatial Data

Peter Congdon, School of Geography, Queen Mary University of London, London, UK.
Email: p.congdon@qmul.ac.uk

Abstract. Bayesian inference and applications have been a central aspect in recent developments in spatial statistics. This influence has rested on advances in computer based estimations via Markov Chain Monte Carlo including recent improvements to random walk MCMC approaches (e.g. Hamiltonian Monte Carlo). Bayesian ideas have been particularly influential in spatial econometrics, disease mapping and analysis of point referenced spatial data. Starting with models for univariate spatial data at a single time point, models can be extended to multivariate outcomes or to a space-time framework. Spatial models vary in how spatial dependence or correlation is represented, including neighbourhood dependence in Markov fields models, explicit spatial decay in point referenced spatial process models, and spatial lags or residual correlation effects in spatial autoregressive models. Aims of spatial analysis also differ, from ensuring regression analysis to allow for spatial dependence (spatial autoregressive models), to detecting disease clusters or elevated relative risk (disease mapping), to spatial prediction and interpolation (spatial process models). Bayesian implementation has been facilitated by a much improved computational environment centred on the R package.

1 Introduction

Bayesian applications in spatial statistics have multiplied considerably in the last two decades, faciliated by improved estimation using Markov chain Monte Carlo (MCMC) methods, and by advances in relevant statistical theory. Application areas where Bayesian ideas have impacted include spatial epidemiology, spatial ecology, spatial econometrics, political science, and geostatistics. In spatial epidemiology, Bayesian studies include spatial smoothing of rare health outcomes, modelling spatial clustering in disease risks (e.g. Richardson et al., 2004) for administrative areas (spatial lattice data), and models for health impacts of environmental point sources (Wakefield and Morris, 2001), while in spatial ecology, applications include habitat and remote sensing models (Carroll et al., 2010). Applications in spatial econometrics concentrate on models for behaviour by economic actors (house purchasers, firms, etc.) involved in spatially defined behaviours (e.g. Le Sage and Pace, 2009), while spatial applications in political science focus on spatially defined electoral and legislative processes, and political outcomes. Another major application context is the continuous spatial framework of geostatistics (point data) with diverse applications including geology, infectious epidemiology and meteorology (e.g. Gelfand and Banerjee, 2017; Diggle and Ribeiro, 2007). Spatial process models and point pattern analysis are the two main methodological frameworks for point data.

Bayesian analysis in such applications is distinct from frequentist approaches in the need to consider the specification of prior densities for parameters θ. Such densities can potentially include prior evidence (e.g. from previous studies) where available, and possibly subject matter based constraints (e.g. confining a spatial correlation parameter to positive values). Prior densities may vary in their informativeness, meaning essentially the degree of concentration in the mass: a diffuse or flat prior will spread the prior density over a wide range of values (e.g. as in a uniform prior for a probability or rate), whereas an informative prior will concentrate potential values within a narrower range. The prior density $\pi(\theta)$ for a parameter is updated by the likelihood of the data $L(\theta|y) = p(y|\theta)$, and posterior inferences are based on the updated parameter density $\pi(\theta|y) = kp(y|\theta)\pi(\theta)$, and particularly the log posterior $log[p(y|\theta)] + \log[\pi(\theta)]$.

Modern spatial data analysis using Bayesian principles will usually use Monte Carlo Markov Chain (MCMC) sampling methods in the updating stage, for example Gibbs sampling (Casella and George, 1992), and Metropolis-Hastings methods in general. However, MCMC methods based on random walk samplers may involve protracted sampling, and delayed convergence. There is now improved implementation of alternative MCMC methods, such as Metropolis Adjusted Langevin Algorithm (MALA) and Hamiltonian Monte Carlo (HMC), which have improved search and convergence behaviour, especially for high-dimensional models (Septier and Peters, 2016; Durmus et al., 2017; Hoffman and Gelman, 2014). Protracted computing is also avoided in the Integrated Nested Laplace Approximation (INLA) approach, a computationally efficient alternative to MCMC (Blangiardo et al., 2013; Opitz, 2017; Illian et al., 2013).

It is clear then that applied spatial research using Bayesian inference therefore rests heavily on suitable computational tools, such as suitable MCMC procedures. In this regard, there have been substantial developments and improvements in the computing environment for Bayesian spatial statistics. These have been concentrated especially on the R platform, and include the development of new packages for Bayesian analysis using conditional autoregressive priors (such as CARBayes), packages for efficient covariance modelling for large point datasets such as spNNGP and spBayes, and greater facility to estimate certain spatial models using the package rstan, which is based on HMC.

2 Spatially Autoregessive Regression

Consider the normal linear regression with continuous outcomes y_i and errors ϵ_i, $i = 1, ..., n$. In applications to observations for discrete spatial units (also called lattice data), an *iid* assumption regarding the errors is likely to be invalid, and instead there will often be co-variation in errors for closely co-located areas i and j. For example, if the model is for crime rates then positive regression residuals may tend to be spatially clustered because crime rates themselves are spatially clustered (e.g. Ratcliffe, 2010). In spatial econometrics, an adaptation of linear regression tackles potential spatial correlation in terms of spatially

lagged dependence in observations or errors. This is analogous to similar forms of lagged dependence often applied in time series regression, namely using lagged values of the response as predictors, and assuming serially correlated errors in time.

Thus consider an $n \times n$ matrix W of contiguity dummies, with $w_{ij} = 1$ if areas i and j are adjacent, and $w_{ij} = 0$ otherwise (with $w_{ii} = 0$). Alternatively distance based interactions might be specified, for instance $w_{ij} = \exp(-\gamma d_{ij})$, where d_{ij} are distances, and $\gamma > 0$ reflects distance decay. From W may be obtained the row-standardised matrix $W^* = [w_{ij}^*] = [w_{ij}/\sum_j w_{ij}]$. The most general model, known as a spatial autoregressive regression, includes a spatial lag in both errors and observations, namely (for y of dimension $n \times 1$, predictors X of dimension $n \times K$, and β of dimension $K \times 1$).

$$y = \lambda W_1 y + X\beta + \epsilon, \qquad (1)$$
$$\epsilon = \rho W_2 \epsilon + u,$$
$$u \sim \mathcal{N}(0, \sigma^2 I),$$

where λ and ρ are unknown correlation parameters. The coefficients λ and ρ have bounds $\{\frac{1}{\omega_{1,\min}}, \frac{1}{\omega_{1,\max}}\}$ and $\{\frac{1}{\omega_{2,\min}}, \frac{1}{\omega_{2,\max}}\}$ respectively, where $\omega_{j,\min}$ and $\omega_{j,\max}$ are the minimum and maximum eigenvalues of W_j. For W standardized within rows, $\omega_{\max}$ is one, and since spatial correlation is usually positive, a prior on λ or ρ constrained to $[0,1]$ is often used, for example, a beta prior, $\pi(\lambda) = Beta(a_\lambda, b_\lambda)$.

A widely used scheme for error dependence is a reduced version of the above known commonly as the spatial error model, with $W_1 = 0$, namely

$$y = X\beta + \epsilon, \qquad (2)$$
$$\epsilon = \rho W \epsilon + u,$$

This model expresses spatial covariation in errors not accounted for by independent variables, measurement errors and mismatch between the spatial units used and the scale at which the process occurs (Anselin and Bera, 1998). The corresponding likelihood is

$$L(\beta, \sigma^2, \rho | y) = (2\pi)^{-n/2} \sigma^{-n} |I - \rho W| \exp(-\frac{1}{2\sigma^2} \varepsilon' \varepsilon), \qquad (3)$$

where $\varepsilon = (I - \rho W)(y - X\beta)$. The spatial errors model may be expressed as

$$y = X\beta + (I - \rho W)^{-1} u, \qquad (4)$$

or equivalently

$$y \sim \mathcal{N}(X\beta, \sigma^2 [(I - \rho W)'(I - \rho W)]^{-1}). \qquad (5)$$

The spatial error model may be estimated straightforwardly using Hamiltonian Monte Carlo in rstan. For K predictors including an intercept, a relatively compact code is as follows:

```
model="
data {int n;
real y[n]; //the response
matrix[n,K] X; // predictor matrix
matrix<lower=0>[n,n] W; // row standardised interaction matrix
matrix<lower=0,upper=1>[n,n] I; // Identity Matrix
```

```
}
transformed parameters {vector[n] linpred;
linpred <- X*beta;
}
parameters {real beta;
real<lower = 0> sigma;
real<lower=-1,upper=1> rho;
}
model {y ~multi_normal_prec(linpred, crossprod(I - rho * W)/(sigma*sigma));
}
```

Spatially lagged effects in the dependent variable rather than errors (i.e. $W_2 = 0$ in the general model above) lead to the spatial lag model

$$y = \lambda W y + X\beta + u, \qquad (6)$$

where the u are *iid* errors. This model is often used to represent neighborhood diffusion or spillover effects, as in applications to technical innovation and house prices respectively. The corresponding likelihood is

$$L(\beta, \sigma^2, \lambda | y) = (2\pi)^{-n/2} \sigma^{-n} |I - \lambda W| \exp(-\tfrac{1}{2\sigma^2} u'u), \qquad (7)$$

where $u = (I - \lambda W)y - X\beta$.

This model can be used to illustrate the principles of prior specfication and of MCMC sampling (via Metropolis-Hastings and Gibbs sampling). Thus one typically assumes prior independence between λ and other parameters, but adopts a normal prior for β that conditions on the sampled value of σ^2. Then with an inverse Gamma prior on σ^2, one has

$$\pi(\sigma^2) = IG(\sigma^2 | a_0, b_0) = \tfrac{b_0^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-(a_0+1)} \exp(-b_0/\sigma^2), \qquad (8)$$

$$\pi(\beta | \sigma^2, c_0, d_0) = \mathcal{N}(c_0, d_0\sigma^2).$$

The prior on λ is uniform between bounds determined by the eigenvalues ω of W,

$$\pi(\lambda | \omega, W) = U(\tfrac{1}{\omega_{\min}}, \tfrac{1}{\omega_{\max}}).$$

In general, choice of MCMC sampling technique is dependent on the form of the full conditional densities $p(\theta_r | \theta_{[r]}, y)$ for parameter θ_r conditional on all other parameters $\theta_{[r]}$, and the data y.

For example, to implement Gibbs sampling usually requires that the full conditional densities have a known form that permits direct sampling. Thus letting $A = (I - \lambda W)$, the full conditional density for β is normal, permitting Gibbs sampling, namely

$$p(\beta | \sigma^2, \rho, y) = \mathcal{N}(c_1, d_1\sigma^2), \qquad (9)$$

$$c_1 = (X'X + 1/d_0)^{-1}(X'Ay + c_0/d_0),$$

$$d_1 = (X'X + 1/d)^{-1}.$$

The full conditional density for σ^2 is inverse Gamma (again allowing Gibbs sampling), namely

$$p(\sigma^2 | \beta, \rho, y) = IG(a_1, b_1), \qquad (10)$$

$$a_1 = a_0 + 0.5n, \qquad b_1 = b_0 + 0.5(Ay - X\beta)'(Ay - X\beta).$$

However, the full conditional density for λ has the form

$$p(\lambda|\sigma^2, \beta, y) = k|I - \lambda W| \exp(\tfrac{1}{2\sigma^2} u'u) U(\tfrac{1}{\omega_{\min}}, \tfrac{1}{\omega_{\max}}), \qquad (11)$$

where k is an unknown constant. This is not a standard density and so more general Metropolis-Hastings sampling is needed. Let $p(\lambda^{(t)})$ denote the value of $p(\lambda|\sigma^2, \beta, y)$ at the current value $\lambda^{(t)}$ in an MCMC sampling sequence $t = 1, ..., T$. Let $p(\lambda_{new})$ be the value of the same conditional density at a candidate value generated by a proposal density. Let h be a random number between zero and one. Then in Metropolis sampling, the candidate value replaces the current value either if $p(\lambda_{new}) > p(\lambda^{(t)})$, or if $p(\lambda_{new}) < p(\lambda^{(t)})$, but $h < p(\lambda_{new})/p(\lambda^{(t)})$.

While developed for continuous data, these techniques can be adapted to binary, multinomial or ordinal outcomes using latent outcome representations. For binary data defined over cases $i = 1, ..., n$

$$y_i \sim Bern(\pi_i) , \qquad (12)$$

the spatial autoregressive and spatial error models can be applied using a latent variable model, whereby

$$y_i = 1 \qquad if \qquad z_i > 0, \qquad (13)$$
$$y_i = 0 \qquad if \qquad z_i \leq 0,$$

where z_i can be interpreted as the utility difference $U_{1i} - U_{0i}$ between binary options, with $Pr(y_i = 1) = Pr(U_{1i} > U_{0i}) = Pr(z_i > 0)$ (Smith and LeSage, 2004). For example, the spatial autoregressive model for dichotomous outcomes based on the underlying latent variables is

$$z = \lambda W z + X\beta + u, \qquad u \sim \mathcal{N}(0, I), \qquad (14)$$

so that

$$z = (I - \lambda W)^{-1} X\beta + v, \qquad (15)$$
$$v = (I - \lambda W)^{-1} u \sim N_n(0, [(I - \lambda W)'(I - \lambda W)]^{-1}).$$

3 Disease Mapping: Conditional Autoregressive Priors in Spatial Epidemiology

Simultaneous autoregressive schemes are primarily designed for continuous univariate responses, whereas count variables (usually leading to Poisson or binomial likelihoods) are typically observed in spatial health applications. Although transformations of count variables may be applied, leading to approximate normality (e.g. the Anscombe transform), inverse transformation is sometimes subject to bias. By contrast, conditional autoregressive priors provide a straightforward approach for analyzing discrete data outcomes. Such priors are especially used in Bayesian analysis of disease patterns, often termed Bayesian disease mapping, with aims including detection of spatial disease clusters and the detection of aereas with elevated disease risk. Instead of focusing on the joint multivariate distribution of the entire vector s, the multivariate density can be equivalently expressed in terms of conditional priors for the univariate density of each area's error, s_i, conditioning on errors in all other areas $s_{[i]} = \{s_j, j \neq i\}$. MCMC sampling is facilitated by use of conditional priors. Certain restrictions on the form of the spatial weight matrix W and the conditional precision of the s_i need to be followed to ensure a valid joint density is obtained from the collection of conditional priors (Besag and Kooperberg, 1995).

Conditional priors can be used in all forms of generalized linear model, including linear regression with $y = X\beta + s$. The conditional autoregressive (CAR) prior (Bell and Broemeling, 2000) expresses each s_i as

$$s_i|s_{[i]} \sim \mathcal{N}(\rho\sum_{j\neq i}w_{ij}s_j, \sigma^2), \qquad (16)$$

where the conditional mean is a weighted average of errors in other areas, and ρ is bounded by the inverses of the minimum and maximum eigenvalues of W. Using a standardised weight matrix leads to what are often termed intrinsic conditional autoregressive or $ICAR(\rho)$ priors (Stern and Cressie, 2000), with

$$s_i|s_{[i]} \sim \mathcal{N}(\rho\sum_{j\neq i}w_{ij}s_j/\sum_{j\neq i}w_{ij}, \sigma^2/\sum_{j\neq i}w_{ij}). \qquad (17)$$

The upper bound for ρ is now one, and a uniform prior on ρ with values between zero and one is often reasonable.

A popular scheme, analogous to random walk priors in time series, in fact assumes $\rho = 1$. Additionally estimation of distance decay parameters can be avoided by taking $w_{ij} = 1$ for adjacent areas, and $w_{ij} = 0$ otherwise. Define $M_i = \sum_{j\neq i}w_{ij}$ as the number of areas adjacent to area i, and let L_i denote this collection of areas. Then the $ICAR(1)$ prior is

$$s_i|s_{[i]} \sim \mathcal{N}\left(\sum_{j\in L_i} s_j/M_i, \sigma^2/M_i\right). \qquad (18)$$

The joint prior version of this scheme is technically improper (Sun et al., 1999), but propriety is achieved in practice by re-centering the sampled s_i to sum to zero under MCMC sampling (Rodrigues and Assunçao, 2008).

For example, suppose y_i denotes small area disease counts, with expected events E_i obtained using region-wide incidence rates, and with $\sum_i y_i = \sum_i E_i$. The outcomes may be taken as Poisson, $y_i \sim Po(E_i\theta_i)$, where θ_i denotes relative disease risk in area i. Classical estimation may take the θ_i as fixed effects, with (implicit) flat priors, producing relative risk estimates $\hat{\theta}_i = y_i/E_i$. These are potentially misleading as indicators of disease variations, since the resulting maps may be distorted by low event counts or populations, and small changes in event totals may produce major shifts in estimates $\hat{\theta}_i$. Instead one possible plausible scheme for spatial borrowing of strength suggests two forms of underlying random variation: a smooth spatial signal s_i following an $ICAR(1)$ prior with variance σ_s^2, and a iid term u_i for representing idiosyncratic local effects (Besag et al., 1991; Mollie, 1996), leading to the so-called convolution prior with

$$log(\theta_i) = X_i\beta + u_i + s_i, \qquad (19)$$

where the iid errors are normal, $u_i \sim \mathcal{N}(0, \sigma_u^2)$.

This model has potential identifiability issues: only the total error $t_i = u_i + s_i$ is identi-

fied by the data and estimates of variances σ_s^2 and σ_u^2 may be sensitive to priors adopted (e.g. Yan, 2006). Lee (2011) shows limitations of the convolution prior in both weak and strong spatial correlation situations. Riebler et al. (2016) develop a modified convolution prior retaining the two random effects, but with a single scale parameter δ for the composite effects

$$t_i = u_i + s_i = \sqrt{\delta}[\sqrt{1-\rho}\theta_i + \sqrt{\rho}\phi_i^*]. \qquad (20)$$

Here $\theta_i \sim \mathcal{N}(0,1)$ are iid effects, the ϕ_i^* are scaled versions of spatial effects ϕ_i following a ICAR(1) prior, and $\rho \in [0,1]$ governs the proportion of residual variance due to spatial dependence.

Spatial effects subject to a sum-to-zero constraint are included in the BUGS program (https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/), specifically via the car.normal option. BUGS relies on Metropolis-Hastings MCMC, and can be called from R using the R2OpenBUGS package. The car.normal option is also available in the R nimble package. For example, a BUGS program for the model (19) is

```
model {for (i in 1 : n) { y[i] ~dpois(mu[i])
            mu[i] <- E[i]*theta[i]
             log(theta[i]) <- inprod(beta[],X[i,])+ s[i] + u[i]}
        # ICAR prior for spatial random effects:
             s[1:n] ~car.normal(adj[], weights[], M[], tau.s)
             for(k in 1:MM) {weights[k] <- 1}
             tau.u ~dgamma(0.5, 0.0005)
             tau.s ~dgamma(0.5, 0.0005)
    for (k in 1:K) {beta[k] ~dnorm(0,0.001)}}}
```

where M[1:n] is a vector containing the neighbour totals M_i, MM is the total of neighbours in the region, adj[MM] is an area adjacency list, and tau.s and tau.u are precision parameters. The ICAR(1) prior can also be estimated using rstan, but estimation is via Hamiltonian MC, and involves the joint likelihood — see http://mc-stan.org/users/documentation/case-studies/icar_stan.html.

To demonstrate the MCMC sampling involved in this model, define $\bar{s}_i = \sum\limits_{j \in L_i} s_j/M_i$, and respecify the relative risk model as

$$log(\theta_i) = u_i + s_i, \qquad (21)$$
$$u_i \sim \mathcal{N}(X_i\beta, \sigma_u^2).$$

Then the full conditional for each spatial error is

$$p(s_i|s_{[i]}, \sigma_s^2, \beta, u_i) = k_1 \exp\{y_i s_i - E_i\theta_i - 0.5M_i(s_i - \bar{s}_i)/\sigma_s^2\} \qquad (22)$$

while the full conditional for each iid error is

$$p(u_i|\beta, \sigma_u^2, \beta, s_i) = k_2 \exp\{y_i u_i - E_i\theta_i - 0.5(u_i - X_i\beta)/\sigma_u^2\}. \qquad (23)$$

These conditionals can be sampled one at a time, or via block updating, using Metropolis-Hastings algorithms (Lee, 2011).

A widely used alternative to the convolution prior was proposed by Leroux et al. (1999) and involves a single set of random effects h_i and a parameter λ which varies from zero and one according as spatial pattern is missing or predominant. The conditional prior for this model is

$$h_i | h_{[i]} \sim \mathcal{N}\left(\frac{\lambda}{1-\lambda+\lambda\sum_{j\neq i} w_{ij}}\sum_{j\neq i} w_{ij}h_j, \frac{\sigma_h^2}{1-\lambda+\lambda\sum_{j\neq i} w_{ij}}\right), \qquad (24)$$

and when the w_{ij} are defined by adjacency (as is generally the practice in Bayesian disease mapping) this reduces to

$$h_i | h_{[i]} \sim \mathcal{N}\left(\frac{\lambda}{1-\lambda+\lambda M_i}\sum_{j\in L_i} h_j, \frac{\sigma_h^2}{1-\lambda+\lambda M_i}\right). \qquad (25)$$

Conventional spatial priors may not adequately model spatial discontinuities. Among ways to better represent discontinuities, and also avoid distorting the smooth spatial signal, the convolution prior may use a Student-t distribution for the *iid* effect. This may be implemented using scale mixing, namely

$$u_i \sim \mathcal{N}(0, \sigma_u^2/\kappa_i), \qquad (26)$$
$$\kappa_i \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right),$$

where ν is the Student t degrees of freedom parameter. Areas with significantly lower κ_i are potential spatial outliers. The spatial prior itself may be adapted to be heavier-tailed: more robust alternatives include the double exponential prior

$$p(s) \propto \chi exp[-0.5\chi|s_i - s_j|^2], \qquad (27)$$

where χ is a scaling parameter. Another option are mixture priors, such as

$$log(\theta_i) = \gamma + u_i + \eta_i s_{1i} + (1 - \eta_i)s_{2i}, \qquad (28)$$

where γ is an intercept, s_{1i} is $ICAR(1)$, and s_{2i} follows the double exponential form, and η_i has a beta distribution (Lawson and Clark, 2002).

Conditional autoregressive schemes for multivariate effects are relatively straightforward. Suppose there are J sets of spatial effects s_{ji} for each area i. These might be relevant when there are J outcome variables each with a spatially distributed regression residual, but can also be used in other ways: for example, in discrete mixtures over spatial effects, or when regression coefficients show a spatial patterning. The latter scenario is sometimes denoted a spatially varying coefficient (SVC) modelling, and has the same intention (in terms of representing spatial heterogeneity) as techniques such as geographically weighted regression (Wheeler and Waller, 2009); also see [31] on "Geographically Weighted Regression" in this Handbook.

Under the multivariate conditional autoregressive or $MCAR(\rho, J)$ prior (Mardia, 1988) the conditional prior for the ith area effect vector $s_i = (s_{1i}, s_{2i}, ..., s_{Ji})$, given such effects for other areas, $s_{[i]} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_n)$, is multivariate normal of dimension J with conditional outcome-specific means

$$\mu_{ji} = E(s_{ji}|s_{[i]}) = \rho\sum_{k\neq i} w_{ik}s_{jk}/\sum_{k\neq i} w_{ik}, \qquad (29)$$

where the same ρ applies across all J outcomes. When the w_{ik} are binary and based on contiguity, the outcome-specific conditional means are

$$\mu_{ji} = \rho \sum_{k \in L_i} s_{jk}/M_i, \qquad (30)$$

namely locality averages of spatial effects for outcome j, with corresponding within area conditional precision matrices

$$Prec(s_i|s_{[i]}) = M_i\Phi, \qquad (31)$$

where Φ is $J \times J$. Taking $\rho = 1$ in the $MCAR(\rho, J)$ leads to the multivariate version of the $ICAR$ prior.

4 Models for Point Referenced Data (Spatial Processes)

The preceding discussion and examples consider continuous and discrete outcomes for zones (also called 'lattice' data). For point referenced data, the response $y(s)$ varies continuously over locations s within some region R. Data are collected at a finite set of locations, and using a model for these locations, predictions are obtained for unobserved sites. For such data the influence of proximity of locations on covariation in the outcome or residuals needs to be explicitly considered.

Consider point data in two dimensional space, $s_i = (s_{1i}, s_{2i})$, with s_{1i} denoting longitude, and s_{2i} denoting latitude. A starting point for estimating the effect of proximity is provided by a distance metric such as Euclidean interpoint distances, $d_{ij} = |s_i - s_j|$. A baseline assumption is that the covariance is isotropic, namely independent of location and a function only of distance: so for points s and s', separated by distance $d = |s - s'|$, one has $\Sigma(s, s') = \Sigma(d)$. Let $y(s)$ be a univariate response, and let $y(s), w(s)$, and $u(s)$ be $n \times 1$, with a predictor matrix $X(s)$ of dimension $n \times K$.

Then a spatial process model specifies

$$y(s) = X(s)\beta + w(s) + u(s), \qquad (32)$$
$$w(s) \sim \mathcal{N}(0, \Sigma_w(d)),$$

with $n \times n$ covariance matrix $\Sigma_w(d)$ representing spatial dependence, and where $u(s)$ are iid errors with variance τ^2. The prior $w(s) \sim \mathcal{N}(0, \Sigma_w(d))$ is known as a Gaussian process prior and may also be written $w(s) \sim GP(0, \Sigma_w(d))$. To define $\Sigma_w(d)$ appropriately, techniques such as variogram analysis can be used to explore covariation in regression residuals or investigate relevant assumptions such as isotropy (Irvine et al., 2007). Parametric functions can then be applied to represent $\Sigma_w(d)$.

Thus consider $\Sigma_w(d) = \sigma^2 R(d)$ in terms of a variance σ^2 (defined along the diagonal when $i = j$ and $d_{ii} = 0$), and $R(d) = [r_{ij}(d_{ij})]$ reflecting correlations between the errors $w(s_i)$ and $w(s_j)$, usually such that $r_{ii}(0) = 1$ and $R(d)$ is positive definite (Diggle and Ribeiro, 2007). Commonly used schemes include the exponential model

$$r_{ij} = exp(-d_{ij}/\phi), \qquad (33)$$

where ϕ is the range (distance at which spatial correlation ceases to be important), or the Gaussian function

$$r_{ij} = exp(-d_{ij}^2/\phi^2). \qquad (34)$$

Then the overall covariance is

$$\Sigma(d) = \sigma^2 R(d) + \tau^2 I, \qquad (35)$$

with limiting variance (as d_{ij} tends to zero) being $\tau^2 + \sigma^2$. Writing $V = \Sigma(d) = \sigma^2 R(d) + \tau^2$, $y = y(s)$, $X = X(s)$, the log-likelihood kernel is

$$-0.5 log|V| - 0.5(y - X\beta)'V^{-1}(y - X\beta). \qquad (36)$$

Prediction of y_{new} at a new site s_{new} under the linear model involves a vector of covariances $\lambda_i = Cov(s_{new}, s_i)$ between the new point and the sampled sites $s_i, i = 1, ..., n$, and the prediction is then a weighted combination of the existing point values with weights w_i determined by $w = \lambda V^{-1}$.

Bayesian inference and estimation for such models may provide additional scope for inferences not possible under classical estimation approaches. An example is provided by Irvine et al. (2007), regarding the "effective range" or distance beyond which the correlation between observations, $\rho(d) = \Sigma(d)/\Sigma(0)$, is less than or equal to 0.05. Precise estimation may be facilitated by informative priors, for example on distance-decay parameters such as ϕ, or on the nugget to sill ratio $\nu^2 = \tau^2/\sigma^2$ in a reparameterized covariance matrix $\Sigma = \sigma^2(R + \nu^2 I)$ (Diggle and Ribeiro, 2007, chapter 7).

Estimation may be slowed by computations for large dimension covariance matrices. Prior specification may improve estimation times: univariate or bivariate grid priors at selected points within a feasible range for ϕ and/or ν^2 allow prior calculation of Σ or Σ^{-1} at the grid points and thus reduced computation. Alternative strategies include lower dimension kernel based methods such as discrete convolution priors (Higdon, 2007), variational Bayes, and stochastic partial differentiation (SPDE) (Lindgren et al., 2011; Optiz, 2017). The latter is incorporated in the R-INLA package in R, whereby the continuous spatial domain $y(s)$ is approximated by a discrete Gaussian Markov random field process (GMRF). This approximation involves a triangulation (usually with $m \ll n$ nodes) of the spatial domain, and the density of the triangulation mesh determines how close the approximation is.

The computational burden is also reduced by using low-rank representations of the spatial field (Finley et al., 2015), obtained using a set of $r \ll n$ knots. Then denoting $y^* = \{y(s_1^*), y(s_2^*), ..., y(s_r^*)\}$ and inter-knot distances d^*, one has

$$y^*(s^*) \sim \mathcal{N}_r(X(s^*)\beta, \sigma^2 \Sigma^*(d^*, \theta) + \tau^2), \qquad (37)$$

with predictions or interpolations $\widetilde{y}(s)$ at generic locations s obtained as

$$\widetilde{y}(s) = c(s; \theta)[\Sigma^*(d^*, \theta)]^{-1} y^*, \qquad (38)$$

where $c(s; \theta)$ is an $r \times 1$ vector with i^{th} element $[c(s, s_i^*; \theta]$.

Under the nearest neighbour Gaussian process (NNGP) approach, incorporated in the R package spNNGP, a sparse precision matrix for the joint spatial density is obtained using neighbor sets $\mathcal{N}(s_i)$, for example, the m nearest neighbours of the point s_i. Then

$$y(s) \sim \mathcal{N}(X(s)\beta + w(s), \tau^2), \qquad (39)$$
$$w(s) \sim GP(0, C_\theta(s, s')),$$

and under the latent NNGP model, one has

$$w(s) \sim GP(0, C_*), \qquad (40)$$

where $C_*^{-1} = (I - A_*)^T D_*^{-1}(I - A_*)$ is the precision matrix of the latent process $w(s)$, A_* is a sparse, lower trangular matrix, and D_* is diagonal. Let $\mathcal{N}(s_i)$ be the m nearest points to s_i among points with index less than i. Following Finley et al. (2017), the ith row of A_* has nonzero entries for points with index $\mathcal{N}(s_i)$, with these entries obtained as

$$A_*[i, \mathcal{N}(s_i)] = C_\theta(s_i, \mathcal{N}(s_i))(C_\theta(\mathcal{N}(s_i), \mathcal{N}(s_i)) + \tau^2 I)^{-1}, \qquad (41)$$

and

$$D_*[i, i] = C_\theta(s_i, s_i) + \tau^2 - C_\theta(s_i, \mathcal{N}(s_i))(C_\theta(\mathcal{N}(s_i), \mathcal{N}(s_i)) + \tau^2 I)^{-1} C_\theta(\mathcal{N}(s_i), s_i). \qquad (42)$$

5 Space-Time Models

Longitudinal spatial observations raise similar issues to those for panel data generally, such as the modelling of temporal autocorrelation and permanent area effects. The main strands in spatial modelling (simultaneous autoregression, disease mapping and point referenced data) all have space-time representations, though Bayesian modelling in some applications has been relatively limited. Thus spatio-temporal variations on spatial lag and spatial error models have only recently been considered in Bayesian terms. For example, Debarsy et al. (2012) generalize the spatial lag model to incorporate time-lags in own area and neighbouring areas, as in

$$y_{it} = \phi y_{i,t-1} + \lambda W y_t + \theta W y_{t-1} + X_{it}\beta + \alpha_i + s_{it}. \qquad (43)$$

Their model did not include permanent area effects α_i but the earlier model of Kakamu and Wago (2008) does. Their model has the specification

$$y_{it} = \lambda \sum_j w_{ij} y_{jt} + X_{it}\beta + \alpha_i + u_{it}, \qquad (44)$$

where the u_{it} are iid, $u_{it} \sim \mathcal{N}(0, \sigma^2)$. Assuming the α_i are random, for example $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$, the stage 1 likelihood for period t is

$$p(y_t | \lambda, \beta, \alpha, \sigma^2) = (2\pi\sigma^2)^{-n/2} |I - \lambda W| \exp(-0.5 u_t' u_t / \sigma^2), \qquad (45)$$

where $u_t = y_t - \lambda W y_t - X_t\beta - \alpha$. As for cross-sectional spatial lag or error models, computational savings may be achieved by taking uniform grid priors on λ, allowing pre-calculation of the log-determinants, $log|I - \lambda W|$, at each grid point.

Conditional hierarchical space-time priors may have benefits in MCMC applications and are applicable straightforwardly in area-time analysis involving binomial or Poisson count data. For example, in Poisson modelling of area health risks ρ_{it}, one may, by analogy to the random intercept-random slope model of conventional panel models, assume spatially structured area-specific random variation for both the level and the growth effect, so that neighbouring areas have similar trends in relative risk (Bernardinelli et al., 1995). For equally spaced time points and expected events E_{it}, one has (for $i = 1, ...n;\ t = 1, ..., T$)

$$y_{it} \sim Po(E_{it}\rho_{it}), \qquad (46)$$

$$log(\rho_{it}) = \alpha + \delta_t + \lambda_{1i} + \lambda_{2i}t,$$

where the level effects λ_{1i} describe the stable relative risk pattern, while trend parameters λ_{2i} describe incremental changes in relative risk. The broad scale trend is represented by parameters δ_t, which for T small may be modelled as fixed effects with a corner constraint

(e.g. $\delta_1 = 0$). The two sets of spatial effects $\{\lambda_{1i}, \lambda_{2i}\}$ can be assigned a bivariate conditional autoregressive prior, $MCAR(\rho, 2)$, as discussed above.

To allow for local heterogeneity, space-time priors can incorporate the convolution principle, combining a pure spatial signal with an *iid* term, as in
$$log(\rho_{it}) = \alpha + \delta_t + s_{1i} + u_{1i} + (s_{2i} + u_{2i})t, \qquad (47)$$
where u_{1i} and u_{2i} are *iid*, and (s_{1i}, s_{1i}) are separately $ICAR$, or jointly $MCAR$ with $\rho = 1$. Setting $c_{ji} = s_{ji} + u_{ji}$ one has
$$log(\rho_{it}) = \alpha + \delta_t + c_{1i} + c_{2i}t. \qquad (48)$$
While some realignment of spatial risks is likely over time, one may, however, seek to model persistent differentials. With $c_{it} = s_{it} + u_{it}$ denoting a convolution scheme, correlation through time can be represented by a simple autoregressive order one (AR1) process, with
$$log(\nu_{it}) = \alpha + \delta_t + c_{it} + \lambda c_{i,t-1}, \qquad \lambda \in (-1, 1), \qquad t > 1 \qquad (49)$$
with initial time model (at $t = 1$) being
$$log(\nu_{i1}) = \alpha + \delta_1 + \frac{c_{i1}}{(1-\lambda^2)^{0.5}}. \qquad (50)$$

Space and time dependence in area-time interactions c_{it} can also be represented using a Kronecker product of the relevant structure matrices (which define the inverse covariance matrices in the joint prior) (Lagazio et al., 2001). Assume an $ICAR(1)$ scheme for spatial errors, with spatial interaction matrix W based on adjacency, has a joint multivariate normal prior with inverse covariance $\tau_s K_s$, where τ_s is a precision parameter, and the off-diagonal terms $K_{s[ij]}$ are -1 for neighbouring areas i and j, and $K_{s[ij]} = 0$ otherwise. Diagonal terms in $K_{s,i,i}$ are given by M_i, the number of neighbours of area i. If a first order random walk (RW) prior in time is assumed with K_t as the structure matrix in the joint prior, then the off-diagonal elements are $K_{t[ab]} = -1$ for adjacent times a and b, and $K_{t[ab]} = 0$ otherwise. Diagonal terms equal one when $a = b = 1$ or $a = b = T$, and equal two for other diagonal terms. Then an area-time interaction effect d_{it} formed by crossing an $RW1$ time prior with a $ICAR(1)$ spatial effect has a joint prior with precision specified by the Kronecker product
$$\tau_d K_s \otimes K_t.$$
The corresponding conditional priors (for d_{it} conditioning on all other interactions) have precisions $\tau_d M_i$ when $t = 1$ or $t = T$, and $2\tau_d M_i$ otherwise. With L_i denoting the neighbourhood of area i, the prior conditional means $\overline{d}_{it}$ for d_{it} are
$$\overline{d}_{i1} = d_{i2} + \sum_{j \in L_i} d_{j1}/M_i - \sum_{j \in L_i} d_{j2}/M_i, \qquad (51)$$
$$\overline{d}_{it} = 0.5(d_{i,t-1} + d_{i,t+1}) + \sum_{j \in L_i} d_{jt}/M_i - \sum_{j \in L_i} (d_{j,t+1} + d_{j,t-1})/(2M_i), \qquad 1 < t < T$$
$$\overline{d}_{iT} = d_{i,T-1} + \sum_{j \in L_i} d_{jT}/M_i - \sum_{j \in L_i} d_{j,T-1}/M_i.$$
For identification, the d_{it} should be doubly centred at each iteration (over areas for a given t, and over times for a given area i).

For point referenced data the cross-sectional spatial process model generalizes to a point model with continuous space and time

$$y(s,t) = X(s,t)\beta + w(s,t) + u(s,t), \qquad (52)$$
$$w(s,t) \sim \mathcal{N}(0, \Sigma_w(s,t)),$$
$$u(s,t) \sim \mathcal{N}(0, \tau^2 I).$$

A stationary, isotropic covariance assumption sets
$$\Sigma_w(s,t) = cov[w(s,t), w(s',t')] = \Sigma_w(|s-s'|, |t-t'|), \qquad (53)$$
and one option is a separable form with no space-time interaction,
$$cov[w(s,t), w(s',t')] = \sigma^2 R_1(s-s'; \phi_1) R_2(t-t'; \phi_2), \qquad (54)$$
where $R_1 = (r_{1ij}), R_2 = (r_{2ij})$ are $n \times n$ and $T \times T$ respectively. For example, if the times are discrete and equally spaced, then $r_{1ij} = \exp(-d_{ij}/\phi_1)$, and $r_{2ij} = \exp(-|i-j|/\phi_2)$ define exponential decay in space and time. Space-time interactions require non-separable covariance specifications (Gelfand and Banerjee, 2017).

6 Point Pattern Data

Point pattern data are generated when random events are observed over some region R, and both the number of events n and their locations s_i are random. Examples are location of persons with a disease, locations of riots, or locations of fast food outlets. Among research questions that may arise are measuring the intensity of events, and whether events (points) are more likely in certain sub-regions than others, i.e. is the process clustered. Variations in intensity may be investigated using nonhomogenous Poisson processes, or log Gaussian Cox processes, or (approximately) by subdividing R into discrete cells, and analysing totals of events in each cell (Illian et al., 2013). The inhomogeneous Poisson process with intensity function $\lambda(.)$ is defined by two properties, namely (a) the number of events $\mathcal{N}(R) = n$ is Poisson with mean $\int_R \lambda(u)du$, and (b) conditional on $\mathcal{N}(R) = n$, the actual locations are independent and identically distributed in R with density proportional to $\lambda(.)$. Taking $log[\lambda(u)] = \log(\lambda_0) + S(u)$, the intensity function will can be expressed in terms of a background intensity λ_0 and a regression term $S(u)$ (e.g. a trend function in latitude and longitude). In a Bayesian analysis the background intensity will typically be assigned a gamma prior, with prior mean expected to approximately equal the number of events divided by the area of R (Leininger and Gelfand, 2017).

7 Focussed Clustering Models

In environmental epidemiology, disease risk may be related to proximity to one or more known or unknown hazard sites (e.g. Ismaila et al., 2007; Maule et al., 2007). Models for such effects may draw on point pattern or disease mapping model frameworks. A benchmark scheme in such situations includes both background risk and focused risk, with relative risk at location s at distance $d = |s - s_0|$ from a point source at s_0 represented as
$$\lambda(s) = \rho g(s) h(d), \qquad (55)$$
where ρ is the regional incidence rate, $g(s)$ is the background population at risk at location s, and $h(d)$ expresses disease risk postulated to reflect distance from the source. For example, one may take

$$h(d) = 1 + \eta f(d, \phi), \qquad (56)$$

where $f(d, \phi)$ is a distance decay function expressing lessened risk at greater distance, such as an exponential function, $f(d) = exp(-d/\phi)$, and where $1 + \eta$ defines relative risk at or near the source (where $d \approx 0$, and $f(d) \approx 1$). Provided $\phi > 0$, $h(d)$ tends to one as d tends to infinity (and $f(d)$ tends to zero). A simpler 'hot spot' clustering model specifies uniformly elevated risk $1 + \eta$ in a neighbourhood (defined by distances $d < \delta$) around the focus, but background risk elsewhere. If there are multiple foci one may generalize to

$$\lambda(s) = \rho g(s) h(d) = \rho g(s)[1 + \sum_k \eta_k f(d, \phi_k)]. \qquad (58)$$

Observed data for focussed clustering models may involve individual level disease status or small area disease totals. For the former type of outcome, the population density may be modelled via kernel methods, for example using small area population estimates. An alternative is to proxy the background population distribution using a control disease unrelated to exposure from the point source. Cases and controls have binary outcomes $y_i = 1$ and $y_i = 0$ respectively, and if there are individual risk factors X_i, the odds of being a case may be represented as

$$\pi_i/(1 - \pi_i) = \rho^*[1 + \eta f(d)]exp(X_i\beta), \qquad (59)$$

where $\rho^* = (a/b)\rho$, a and b are sampling proportions of cases and of controls respectively, and ρ is the population odds of disease

Focused clustering may be relevant to small area studies, for example in modelling area counts of cancer incidence according to distance from a source, or in modelling human flow behaviours (to hospitals or supermarkets) (e.g. Ismaila et al., 2007). For observations consisting of disease counts y_i in areas i, the background risk in an area might be approximated by the expected disease total E_i based on population totals or age structure. To account for spatial correlation effects distinct from the effect of distance from the focus (or foci), the Poisson mean μ_i might include spatial effects s_i as discussed above, area predictors (e.g. deprivation) and iid effects u_i also, as in

$$\mu_i = \rho E_i exp(X_i\beta)[1 + f(d_i)]exp(u_i + s_i). \qquad (60)$$

For example, setting $\alpha = \log(\rho)$, an exponential decay model would lead (Ma et al., 2007) to

$$log(\mu_i) = \alpha + \log(E_i) + X_i\beta + \log(1 + e^{-\phi d}) + u_i + s_i. \qquad (61)$$

8 Conclusion

The rapidly growing area of spatial statistics and cognate disciplines such as spatial econometrics and spatial epidemiology have been considerably influenced by Bayesian ideas. Feasibility of Bayesian analysis has much improved due to improved estimation techniques based on Markov Chain Monte Carlo and on the improved availability of software, as in numerous options for Bayesian spatial analysis in R. Much recent development has focused on space-time models (Fischer and LeSage, 2015), analysis of multivariate outcomes (Kavanagh et al., 2016), and on new estimation approaches (Morris, 2019; Finley et al., 2017),

References

Anselin L, Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah A, Giles D eds, Handbook of Applied Economic Statistics, pp 237-289. Marcel Dekker, New York

Bernardinelli, L, Clayton, D, Pascutto, C., Montomoli, C, Ghislandi, M (1995) Bayesian Analysis of space-time variations in disease risk, Stat. Med., 14 (21-22): 2433-2443

Besag J, Kooperberg, C (1995) On conditional and intrinsic autoregressions. Biometrika 82(4), 733-746

Besag, J, York, J, Mollie, A (1991) Bayesian image restoration, with two applications in spatial statistics, Annals of the Institute of Statistical Mathematics, 43(1),1-20

Blangiardo, M, Cameletti, M, Baio, G, Rue, H. (2013) Spatial and spatio-temporal models with R-INLA. Spatial and Spatio-temporal Epidemiology, 4(1), 33-49.

Carroll C, Johnson D, Dunk J, Zielinski W (2010) Hierarchical Bayesian spatial models for multispecies conservation planning and monitoring. Conservation Biology, 24 (6): 1538-1548

Casella G, George E (1992) Explaining the Gibbs sampler. The American Statistician, 46(3):167–174

Debarsy N, Ertur C, Sage J (2012) Interpreting dynamic space–time panel data models. Statistical Methodology, 9 (1-2): 158-171

Diggle P, Ribeiro, P (2007) Model-based Geostatistics. New York: Springer

Durmus, A., Roberts, G. O., Vilmart, G., Zygalakis, K (2017). Fast Langevin based algorithm for MCMC in high dimensions. The Annals of Applied Probability, 27(4), 2195-2237.

Finley A, Banerjee S, Gelfand A (2015) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Journal of Statistical Software, 63(13), 1-28.

Finley, A, Datta A, Cook B, Morton D, Andersen H, Banerjee S (2017) Applying Nearest Neighbor Gaussian Processes to Massive Spatial Data Sets: Forest Canopy Height Prediction Across Tanana Valley Alaska. https://arxiv.org/abs/1702.00434

Fischer, M, LeSage, J (2015). A Bayesian space-time approach to identifying and interpreting regional convergence clubs in Europe. Papers in Regional Science, 94(4), 677-702.

Gelfand, A, Banerjee, S (2017) Bayesian Modeling and Analysis of Geostatistical Data. Annual Review of Statistics and its Application, 4, 245-266.

Higdon D (2007) A primer on space-time modelling from a Bayesian perspective. Chapter 6 (pp 218-280) in Finkelstadt, Held and Isham (eds) Statistical Methods for Spatio-Temporal Systems. CRC Press.

Hoffman M, Gelman A (2014) The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15(1), 1593–1623

Illian, J, Martino, S., Sørbye, S, Gallego-Fernández, J, Zunzunegui, M., Esquivias, M,Travis, J (2013) Fitting complex ecological point process models with integrated nested Laplace approximation. Methods in Ecology and Evolution, 4(4), 305-315.

Irvine K, Gitelman A, Hoeting J (2007) Spatial design and properties of spatial correlation: effects on covariance estimation, Journal of Agricultural, Biological and Environmental Statistics, 12(4), pp. 1-20.

Ismaila A, Canty A, Thabane L (2007) Comparison of Bayesian and frequentist approaches

in modelling risk of preterm birth near the Sydney Tar Ponds, Nova Scotia, Canada. BMC Med Res Methodol. 2007 Sep 10;7:39.

Kakamu K, Wago H (2008) Small-sample Properties of Panel Spatial Autoregressive Models: Comparison of the Bayesian and Maximum Likelihood Methods. Spatial Economic Analysis, 3(3): 305-319

Kavanagh, L., Lee, D, Pryce, G. (2016) Is poverty decentralizing? Quantifying uncertainty in the decentralization of urban poverty. Annals of the American Association of Geographers, 106(6), 1286-1298.

Lagazio C, Dreassi, E, Bernardinelli, A (2001) A hierarchical Bayesian model for space-time variation of disease risk, Statistical Modelling, 1(1), 17:29

Lee D (2011) A Comparison of Conditional Autoregressive Models Used in Bayesian Disease Mapping. Spatial and Spatio-Temporal Epidemiology, 2(2), 79-89.

Leininger, T, Gelfand, A (2017) Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. Bayesian Analysis, 12(1), 1-30.

Ma B, Lawson A, Liu Y (2007) Evaluation of Bayesian models for focused clustering in health data. Environmetrics, 18(8): 871-887

Mardia, K (1988) Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. J. Multiv. Anal. 24(2), 265-284

Maule M, Magnani C, Dalmasso P, Mirabelli D, Merletti F, Biggeri A (2007) Modeling mesothelioma risk associated with environmental asbestos exposure. Environ Health Perspect. 115(7):1066-71.

Mollie, A. (1996) Bayesian mapping of disease. In: Gilks, W, Richardson, S., and Spiegelhalter, D. J. (eds.). Markov Chain Monte Carlo in Practice. London: Chapman & Hall: 359–379.

Opitz, T. (2017) Latent Gaussian modeling and INLA: A review with focus on space-time applications. J. French Statistical Society 158(3), 62-85

Morris, M (2019) Spatial Models in Stan: Intrinsic Auto-Regressive Models for Areal Data. http://mc-stan.org/users/documentation/case-studies/icar_stan.html

Pace, R, LeSage J (2010) Spatial Econometrics. In Handbook of Spatial Statistics, eds A Gelfand, P Diggle, M Fuentes, P Guttorp (eds.), Boca Raton, Florida: CRC (pp 245-262)

Ratcliffe, J (2010) Crime mapping: Spatial and temporal challenges, in Piquero, A. and Wiesburd, D. (Eds) Quantitative Criminology , Springer, pages 5-24.

Riebler, A, Sørbye S, Simpson D, Rue H (2016) An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Statistical Methods in Medical Research, 25(4), 145-1165

Richardson S, Thomson A, Best N, Elliott P (2004) Interpreting posterior relative risk estimates in disease-mapping studies.Environ Health Perspect. 112(9):1016-25.

Rodrigues A, Assuncao R (2008) Propriety of posterior in Bayesian space varying parameter models with normal data, Statistics & Probability Letters 78(15):2408-2411

Septier, F, Peters, G (2016) Langevin and Hamiltonian Based Sequential MCMC for Efficient Bayesian Filtering in High-Dimensional Spaces. J. Sel. Topics Signal Processing, 10(2), 312-327

Stern H, Cressie N (2000) Posterior predictive model checks for disease mapping models. Stat Med, 19 (17-18):2377–2397

Wakefield J, Morris S (2001) The Bayesian Modeling of Disease Risk in Relation to a Point Source. J Amer Stat Assoc, 96 (453): 77-91

Wheeler D, Waller L (2009) Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests, Journal of Geographical Systems, 11(1): 1-22

Wikle, C, Berliner, L, Cressie, N. (1998) Hierarchical Bayesian space-time models. Environmental and Ecological Statistics, 5(2), 117-154.

Yan J (2007) Spatial Stochastic Volatility for Lattice Data. J. Agric Biol. Env. Stat, 12(1): 25–40