

Representing Spatial Dependence and Spatial Discontinuity in Ecological Epidemiology: a Scale Mixture Approach

Peter Congdon, Queen Mary University of London, p.congdon@qmul.ac.uk

Abstract Variation in disease risk underlying observed disease counts is increasingly a focus for Bayesian spatial modelling, including applications in spatial data mining. Bayesian analysis of spatial data, whether for disease or other types of event, often employs a conditionally autoregressive prior, which can express spatial dependence commonly present in underlying risks or rates. Such conditionally autoregressive priors typically assume a normal density and uniform local smoothing for underlying risks. However, normality assumptions may be affected or distorted by heteroscedasticity or spatial outliers. It is also desirable that spatial disease models represent variation that is not attributable to spatial dependence. A spatial prior representing spatial heteroscedasticity within a model accommodating both spatial and non-spatial variation is therefore proposed. Illustrative applications are to human TB incidence. A simulation example is based on mainland US states, while a real data application considers TB incidence in 326 English local authorities.

Keywords. Spatial. Bayesian. Conditional autoregressive. Heteroscedasticity. Scale mixture. Tuberculosis.

1 Introduction

Modelling variation in disease or other events underlying observed totals for geographic areas is important for detecting elevated rates (Beale et al., 2008). In disease mapping, the observations often consist of incidence totals for chronic or infectious disease. Such data are subject to stochastic variations, and the underlying area specific incidence risks are often the focus in data mining studies. In such studies, the objects include extraction of underlying spatial and spatiotemporal patterns, including detection of elevated risk (hotspots) and spatial outliers (Shekhar et al, 2015). The particular focus of this paper is on ecological epidemiology, in the sense of focusing on population aggregates (Morgenstern, 1995), namely geographic areas, and on environmental and socio-economic risk factors for infectious disease (Ploubidis et al, 2012). The applications are to human infectious disease, namely TB incidence.

Different forms of spatial correlation analysis or model have been proposed in disease applications (human and veterinary), environmental science, ecology, crime and other settings. For example, Wikle (2003) reviews hierarchical spatial models applied in environmental science, including irregular lattice data (such as geographic areas) and regular lattice data (such as air pollution grids). Beale et al (2010) consider how regression findings for spatial ecology data are affected by the method used (if at all) to reflect spatial dependence. To exemplify hierarchical models for veterinary data, Pioz et al (2012) apply simultaneous autoregressive

(SAR) models to investigate bluetongue spread in French municipalities, while Farnsworth and Ward (2009) apply Bayesian conditional autoregressive (CAR) models to avian influenza H5N1 outbreak data. In such applications, identifying elevated risk in particular areas, detecting elevated risk clusters, or assessing significant predictors of risk, is emphasized, in methods recognizing the explicitly spatial structure of the data. However, the underlying assumptions of such techniques should be assessed, and subject to modification when indicated.

Hierarchical models involving spatial random effects, both CAR and SAR forms, can be estimated by classical methods (Alam et al, 2015; Horabik and Nahorski, 2010) or Bayesian methods (Waller and Carlin, 2010; Lesage, 1997). CAR spatial priors imply local smoothing of outcome rates, that is smoothing towards the local rather than global average (Gelman, 1996; Waller and Carlin, 2010). Such local discontinuity is demonstrated in the England TB application considered below. Marked variability in risks has been detected in other area studies of infectious disease (Duarte-Cunha et al, 2015; Varga et al, 2015; Ploubidis et al, 2012), whereas spatial variability in relative risks of chronic diseases (cancer, diabetes, etc) is generally less pronounced. When there are spatial discontinuities in risk, it is preferable to allow differing strengths of association between neighbouring areas, as opposed to uniform local smoothing under CAR priors (Gelman, 1996; Smith et al, 2015).

Bayesian applications in disease mapping and ecological epidemiology commonly employ a CAR prior (Lee, 2011) to express spatial clustering in underlying risks (Besag et al., 1991; Best, 1999), including human TB incidence (Nunes, 2007; Maciel et al, 2010). Most applications of CAR priors assume a normal density for the underlying risks combined with uniform local smoothing. However, normality assumptions may be vitiated by heteroscedasticity linked to spatial outliers or to marked discrepancies in risk between neighbouring areas. It is also desirable that spatial disease models represent variation in area disease risks that is not attributable to spatial dependence (i.e. heterogeneity as against clustering). Some spatial priors may represent this feature by using more than one set of random effects, but at the cost of identifiability.

This paper considers modification of the local smoothing principle when there are spatial discontinuities, namely discrepant levels of outcome rates (e.g. disease or crime incidence) between neighbouring areas. In particular, we consider modifications of the normality assumption for area random effects based on a scale mixture version of the Leroux et al (1999) model, allowing for heterogeneity and clustering in a single set of random effects, but with the scale mixture providing adaptivity to local discontinuity and spatial outliers. The relevance of such an approach is illustrated with simulated data on TB incidence in 49 mainland US states, and an application to observed TB incidence in 326 English local authorities.

2 Defining Conditional Spatial Priors

As discussed by Besag and Kooperberg (1995), one may use properties of the multivari-

ate normal to obtain the univariate conditional autoregressive prior from a joint spatial prior and vice versa. Thus consider a joint multivariate normal density for the spatial risk effects $s = (s_1, \dots, s_n)$ for n areas, with mean zero and covariance Σ_s ,

$$p(s) = \frac{1}{(2\pi)^{n/2}} |\Sigma_s|^{-0.5} \exp(-0.5s' \Sigma_s^{-1} s). \quad (1)$$

Denote $Q = [q_{ij}] = \Sigma_s^{-1}$ as the precision matrix, and $s_{[i]} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$ as the totality of effects omitting the i^{th} effect. The conditional distributions for each s_i take a univariate normal form (Rue and Held, 2005, p. 22), namely

$$s_i | s_{[i]} \sim N\left(\sum_{j \neq i} \left[-\frac{q_{ij}}{q_{ii}}\right] s_j, \frac{1}{q_{ii}}\right). \quad (2)$$

Following Besag and Kooperberg (1995, p 734) define $h_{ii} = 0$, and set

$$h_{ij} = -q_{ij}/q_{ii} \quad (i \neq j).$$

Also set $q_{ii} = a_i/\delta$ with variance parameter δ , so that

$$h_{ij} = -q_{ij}\delta/a_i. \quad (3)$$

The density (2) is then in the conditional autoregressive form specified by Besag (1974), namely

$$s_i | s_{[i]} \sim N\left(\sum_{j \neq i} h_{ij} s_j, \delta/a_i\right).$$

To obtain the joint density from the conditional one, symmetry of Q means $-q_{ij} = -q_{ji}$, so that from (3) the constraint

$$h_{ij} a_i = h_{ji} a_j \quad (4)$$

applies.

3 Conditional Autoregressive Spatial Priors

Various schemes for defining the h_{ij} and a_i can be used. A measure of spatial dependence $0 \leq \omega \leq 1$ is included by setting

$$h_{ij} = \omega \frac{w_{ij}}{\sum_{k \neq i} w_{ik}}, \quad a_i = \sum_{k \neq i} w_{ik},$$

where w_{ij} represent spatial interactions between areas i and j . If the interactions are specified as symmetric with $w_{ij} = w_{ji}$, and also with $w_{ii} = 0$, the symmetry constraint (4) is ensured, with $h_{ij} a_i = \omega w_{ij} = h_{ji} a_j$.

A common approach sets $w_{ij} = 1$ for adjacent areas and $w_{ij} = 0$ otherwise, with $a_i = \sum_{k \neq i} w_{ik} = d_i$ then equal to the number, d_i , of areas adjacent to area i . Equivalently d_i is the number of areas in the locality N_i of area i (the areas surrounding area i , and excluding area i itself). This provides the conditionally autoregressive $CAR(\omega)$ prior, with

$$s_i | s_{[i]} \sim N\left(\omega \bar{A}_i, \frac{\delta}{d_i}\right), \quad (5)$$

where \bar{A}_i is the average of the s_j in locality N_i , i.e.

$$\bar{A}_i = \frac{\sum_{j \in N_i} s_j}{d_i}.$$

Lower values of ω imply lesser degrees of spatial dependence between the s_i , though the

limiting case when $\omega = 0$ has the disadvantage that the variance is not constant but depends on the number of neighbours d_i . The *CAR*(1) prior (Besag et al, 1991) specifies relative risks entirely determined by spatial dependence, with

$$s_i | s_{[i]} \sim N\left(\sum_{j \in N_i} s_j / d_i, \frac{\delta}{d_i}\right).$$

In any set of area disease rates, some spatial correlation is typically detected, and this motivates spatial priors which imply borrowing of strength from nearby areas. However, there may also be particular local variations in illness risks unrelated to those in surrounding areas, namely unstructured variation without spatial dependence. In principle, the *CAR*(ω) prior (also called the proper CAR prior) can represent various levels of spatial dependence through the ω parameter, but this parameter does not calibrate well with marginal measures of spatial correlation, such as Moran's I (Banerjee et al, 2004; Rodrigues and Assunção, 2012). Values of ω exceeding 0.99 are needed to achieve modest values of I .

In practice, to represent a mix between spatial dependence and simple unstructured variation, called clustering and heterogeneity respectively by Clayton et al (1993), a common strategy is the so-called convolution prior (Ugarte et al, 2005; Waller and Carlin, 2010). This represents the unknown area relative risk as a sum of a pure spatial effect following a *CAR*(1) prior, combined with an iid (or unstructured) random effect. Thus denote observed disease counts as y_i , expected counts as E_i (expected disease counts in the demographic sense) and known area risk variables (predictors) as X_i . Then one might specify

$$y_i \sim Po(\rho_i E_i),$$

$$\log(\rho_i) = X_i \beta + s_i + h_i, \quad (6.1)$$

$$s_i | s_{[i]} \sim N\left(\sum_{j \in N_i} s_j / d_i, \frac{\delta}{d_i}\right), \quad (6.2)$$

$$h_i \sim N(0, \phi), \quad (6.3)$$

where ρ_i denotes an area specific relative risk, and ϕ is a variance term for iid unstructured effects h_i . A drawback with this scheme is that identifiability may be impeded by the presence of two sets for random effects representing one underlying aspect of the data, namely variation in area illness risks.

4 The Leroux et al Spatial Prior

A scheme for area effects, incorporating both clustering and heterogeneity, involves scale adjustments

$$a_i = (1 - \lambda) + \lambda \sum_{j \neq i} w_{ij},$$

with the parameter $0 \leq \lambda \leq 1$ providing a measure of spatial dependence (Leroux et al, 1999). This scheme, which may be represented as the LLB prior by virtue of its authors, has the benefit that only one set of random effects is involved in representing the pattern of area illness risks. This provides improved identifiability as compared to the convolution prior (Lee, 2011). The case $\lambda = 0$ corresponds to a lack of spatial interdependence (and i.i.d

or unstructured errors s_i), with the advantage that the conditional variance is then simply δ , independent of $\sum_{j \neq i} w_{ij}$. By contrast, $\lambda = 1$ leads to an $CAR(1)$ model, with purely spatial interdependence. In typical datasets λ will be intermediate between these extreme values.

The symmetry condition $h_{ij}a_i = h_{ji}a_j$ is maintained by setting

$$h_{ij} = \frac{\lambda w_{ij}}{(1-\lambda) + \lambda \sum_{j \neq i} w_{ij}},$$

since $h_{ij}a_i = \lambda w_{ij} = \lambda w_{ji} = h_{ji}a_j$. So the conditional prior is

$$s_i | s_{[i]} \sim N\left(\frac{\lambda}{1-\lambda+\lambda \sum_{j \neq i} w_{ij}} \sum_{j \neq i} w_{ij} s_j, \frac{\delta}{1-\lambda+\lambda \sum_{j \neq i} w_{ij}}\right), \quad (7)$$

with δ a scale parameter. When $\lambda = 0$ one obtains normal iid effects $s_i \sim N(0, \delta)$. If the w_{ij} are defined by contiguity one obtains (Leroux et al, 1999, p 181)

$$s_i | s_{[i]} \sim N\left(\frac{\lambda}{1-\lambda+\lambda d_i} \sum_{j \in N_i} s_j, \frac{\delta}{1-\lambda+\lambda d_i}\right). \quad (8)$$

5 Adaptiveness to non-normality and spatial discontinuities

Proposals to modify spatial priors to achieve greater robustness have been made, including the presence of heteroscedasticity and heavier tails (excess kurtosis) than under the normal. Thus Yan (2007), Brewer and Nolan (2007), and Reich and Hodges (2008) propose modified CAR priors to accommodate heteroscedasticity. Other forms of modified spatial prior are considered by Nathoo and Ghosh (2013) and Lawson and Clark (2002). These schemes are all modifications of the CAR prior, or of the convolution prior, as considered in section 3. Modifications of the pure spatial $CAR(1)$ prior, without allowance for spatially unstructured variation, may be appropriate for particular applications, such as dental decay as in Reich and Hodges (2008), but for area illness data an allowance for heterogeneity is generally needed. Modification of the proper $CAR(\omega)$ prior are left with the problem that its ω parameter does not calibrate well with marginal measures of spatial correlation. Studies such as Yan (2007) and Lawson and Clark (2002) modify the convolution prior, with potential identifiability problems due to multiple sets of random effects. Thus Yan (2007) allows for heteroscedasticity in spatial effects via a double implementation of the $CAR(1)$ prior, namely

$$\begin{aligned} y_i &\sim Po(\rho_i E_i), \\ \log(\rho_i) &= X_i \beta + s_i + h_i, \\ s_i | s_{[i]} &\sim N\left(\sum_{j \in N_i} s_j / d_i, \frac{\delta_s}{d_i}\right), \\ h_i &\sim N(0, \phi_i), \\ \log(\phi_i) &= \mu_h + r_i, \\ r_i | r_{[i]} &\sim N\left(\sum_{j \in N_i} r_j / d_i, \frac{\delta_r}{d_i}\right). \end{aligned}$$

Here we modify the constant scale assumption of the LLB prior in (7) and (8) using a scale mixture, with the benefit of providing an indicator of potential outlier status for each area. To implement a scale mixture, define $\kappa_i \sim Ga(0.5\nu, 0.5\nu)$ where ν is a hyperparameter. The proposed model reduces to the scale mixture version of the Student t when $\lambda = 0$ (Boris

Choy and Chan, 2008). The κ_i have average 1 with small values of κ_i (under 1) acting as indicators of outlier status (West, 1984). Under this scale mixture modification, the symmetry condition (4) is maintained by setting

$$a_i = \kappa_i[(1 - \lambda) + \lambda \sum_{j \neq i} w_{ij}],$$

$$h_{ij} = \frac{\lambda w_{ij} \kappa_j}{[1 - \lambda + \lambda \sum_{j \neq i} w_{ij}]},$$

since $h_{ij} a_i = \lambda w_{ij} \kappa_j \kappa_i = \lambda w_{ji} \kappa_i \kappa_j = h_{ji} a_j$.

Then the model for incidence counts becomes

$$y_i \sim Po(\mu_i),$$

$$\mu_i = \rho_i E_i,$$

$$\log(\rho_i) = X_i \beta + s_i,$$

where conditional prior when the w_{ij} are binary indicators of adjacency is

$$s_i | s_{[i]} \sim N\left(\frac{\lambda}{1 - \lambda + \lambda d_i} \sum_{j \in N_i} \kappa_j s_j, \frac{\delta}{\kappa_i [1 - \lambda + \lambda d_i]}\right). \quad (9)$$

This prior reduces to an unstructured i.i.d scale mixture Student-t density

$$s_i \sim N(0, \delta / \kappa_i),$$

when $\lambda = 0$.

From (9) it can be seen that small κ_j values indicate areas discrepant in risk from their neighbours (i.e. they indicate outliers in spatial terms), and reduce the amount of spatial borrowing of strength. Equivalently stated, a clustering of small κ_j values can be taken as indicators of spatial volatility, namely discrepant illness risks in a set of adjacent areas. In regression applications, small κ_j values will also indicate where the regression predictions in the neighbourhood of area i , and their implied neighbourhood relative risk $\sum_{j \in N_i} \mu_j / \sum_{j \in N_i} E_j$, are discrepant from the modelled relative risk in area i itself μ_i / E_i .

Let $\tau = 1/\delta$, and let $I(i \sim j) = I(j \sim i)$ denote that areas i and j are neighbours under binary adjacency. Then the precision matrix in the multivariate normal (1) has diagonal terms

$$Q_{ii} = \tau a_i = \tau \kappa_i [(1 - \lambda) + \lambda \sum_{j \neq i} w_{ij}],$$

and off diagonal terms

$$Q_{ij} = -\tau a_i h_{ij} = -\tau \lambda \kappa_i \kappa_j I(i \sim j).$$

A scale mixture approach to spatial dependence can be set within a broader literature on heavy tailed priors (e.g. student t, double exponential) that can be represented as two level hierarchical models (Yi and Xu, 2008). One application of such priors is to predictor selection in high dimensional regression, with a likelihood penalty function that is a normal scale mixture (e.g. Polson et al, 2014). Besag et al (1991) propose a double exponential prior for spatial effects as a robust alternative to the normal conditional autoregressive, with an application provided by Manda (2013).

Identification of random effects in spatial disease models is often problematic (e.g. MacNab, 2014; Nathoo and Ghosh, 2013), especially for models including multiple random effects, or when disease counts are relatively small. In the case of the model just discussed, identification of outliers (e.g. in terms of significantly low κ_i), as well as identification of elevated risks s_i , will be improved for larger disease counts and/or longer observation periods. Identification of hyperparameters may also be problematic, especially with small samples. For example, in student t binary regression with data augmentation, Gelman et al (2004, p 447) recommend a robust analysis with ν not estimated but preset at 4.

6 Simulation Example

A simulation example of the heteroscedastic LLB prior involves TB incidence with a spatial framework provided by the $n = 49$ mainland states (including the District of Columbia). Expected TB incidence counts E_i are obtained by applying actual US-wide age specific rates for TB in 2013 to state population estimates for 2013, taken from the US National Cancer Institute SEER site (<http://seer.cancer.gov/popdata/>). TB incidence rates are from the CDC National Tuberculosis Surveillance System, with just over 9500 incident cases in 2013, and an all ages rate of 3 per 100,000. Highest rates (over 6 per 100 thousand) are for the 75-84 and 85+ age groups.

We simulate TB incidence counts using these expected counts as offsets. The LLB hyperparameters (guide values) are set as $\lambda = 0.7$, $\tau = 3$, and with ν taking values 3, 10, and 25. Although the student t is defined for degrees of freedom of 2 or less, it has infinite variance, and Gelman et al (2003) mention that “t’s with one or two degrees of freedom have infinite variance and are not usually realistic in the far tails”. One hundred sets of random effects are generated from the multivariate normal $s_{1:n} \sim N(0, Q^{-1})$. Simulated TB incidence counts then obtained via a Poisson simulation $y_i \sim Po(E_i \rho_i)$, with $\log(\rho_i) = \beta_0 + s_i$, where $\beta_0 = -0.1$ and ρ_i is the simulated disease risk in state i . The R code used is set out in Appendix 1. Note that each of the 100 simulations involves a separate sample of $\kappa_i \sim Ga(0.5\nu, 0.5\nu)$.

Analyses to estimate the parameters from the 100 sets of simulated data $\{y, E\}$ (with E as in the simulations) are carried out using the WINBUGS package (Lunn et al, 2009). An exponential prior with mean 10 is adopted for ν (Fernandez and Steele, 1998; Geweke, 1993), a gamma prior with shape 1 and index 0.01 assumed for the inverse variance parameter τ , a normal prior with mean zero and precision 0.001 assumed for the fixed effect β_0 , and a uniform $U(0, 1)$ prior assumed on λ . Estimates are based on the last 5,000 iterations from two chain runs of 10,000 iterations, with convergence assessed using Brooks–Gelman–Rubin diagnostics (Brooks and Gelman, 1998).

The focus is on the posterior means for the main parameters of the LLB prior and risk

regression over the 100 samples, namely ν, λ, β_0 , and the variance of the spatial effects (which depends on both τ and the sampled κ_i). The posterior densities for ν tend to be positively skew, so Table 1 also includes results for the posterior summary of $\log(\nu)$. Because each simulation involves a distinct set of κ_i , the actual variance of the s_i will vary between simulations. This variance V_t of spatial effects for simulation t is recorded in the vector `var.s[]` in the code in Appendix 1. Table 1 sets out the percentiles (20th, 50th, 80th) of the 100 posterior means for $\nu, \log(\nu), \lambda$, and β_0 , and also the proportion of simulated datasets where the 95% credible interval for a parameter includes the guide value. Thus for the setting $\nu = 10$, 50 out of the 100 samples have posterior means for ν below 10.4, and 50 samples have posterior means above 10.4.

The expected E_i are relatively large, so the Poisson simulations may be subject to some excess dispersion, which to some extent attenuates the spatial structure present in the simulated data. Nevertheless, the recovered parameters effectively reproduce those used in generating the data. This feature is also apparent in a correlation between the actual and estimated V_t over the 100 samples of 0.97. Figure 1 plots the two series of V_t for the $\nu = 10$ option, including 95% credible intervals for the estimated V_t . Of substantive relevance in interpreting the parameters of the LLB model, there is a 0.72 correlation between the 100 posterior means for λ , and the corresponding posterior means for Moran's I, which are estimated from the s_i in each dataset. To further illustrate variation over the simulations, Figure 2 shows, for each simulated dataset, the posterior mean (and 95% interval) of $\log(\nu)$ under the $\nu = 3$ option.

One also wishes to reproduce the patterns of outlier status (areas with significantly low κ_i). This involves, for the setting $\nu = 10$ (and other hyperparameters as above), simulating 100 sets of y based on a single set of κ_i values (the "actual" κ_i), sampled from a gamma density, $\kappa_i \sim Ga(5, 5)$. The expected incidence counts are multiplied by 10 to increase the amount of information provided by the data. Re-estimation of κ_i from the simulated datasets shows a shrinkage effect, with posterior mean re-estimated κ_i closer to 1 than the actual κ_i (see Figure 3). However, the re-estimation does identify as outliers the states with unusually low actual κ_i . For the five states with the lowest actual κ_i , four have 95% credible intervals on the re-estimated κ_i that are entirely below 1, and no other states have re-estimated κ_i with credible intervals entirely below 1.

7 Application: TB Incidence for England Local Authorities

An application involves TB incidence data y for 326 English local authorities between 2011 and 2013. Two analyses are undertaken, one without predictors and one with two predictors: an index of multiple socio-economic deprivation (X_1) and population density (X_2). The impact of poverty on TB incidence is well documented (Lopez de Fede et al, 2008) and population density is associated with infectious disease risk as "the likelihood that a susceptible person will be exposed to an infectious tuberculosis patient increases with population

density" (Rieder, 1999). The two predictors are centred and divided by 100. Thus with predictors $X_i = (1, X_{1i}, X_{2i})$, under the scale mixture model we have

$$\begin{aligned} y_i &\sim Po(\mu_i), \\ \mu_i &= \rho_i E_i, \\ \log(\rho_i) &= X_i \beta + s_i, \\ s_i | s_{[i]} &\sim N\left(\frac{\lambda}{[1-\lambda+\lambda d_i]} \sum_{j \in N_i} \kappa_j s_j, \frac{\delta}{\kappa_i [1-\lambda+\lambda d_i]}\right), \\ \kappa_i &\sim Ga(0.5\nu, 0.5\nu). \end{aligned}$$

For the original Leroux et al (1999) scheme, the conditional prior for s_i is as in (8).

Prior settings are as in section 6, and inferences are from the last 5,000 iterations from two chain runs of 10,000 iterations, with convergence assessed using Brooks–Gelman–Rubin diagnostics. Table 2 contains parameter summaries and comparison of measures of fit between the original LLB model (section 4) and the heteroscedastic Leroux (section 5). Fit is assessed using the posterior predictive loss (PPL) criterion (Gelfand and Ghosh, 1998). Consider replicate observations y_{rep} sampled from the posterior predictive density $p(y_{rep}|y)$. The PPL involves defining $t(z) = z \log z - z$, and $\xi_i = t(y_{i,rep})$. Letting η_i and φ_i denote posterior means for $y_{i,rep}$ and ξ_i , the PPL is

$$2 \sum_i \{\varphi_i - t(\eta_i)\} + 2(k+1) \sum_i \left\{ \frac{t(\eta_i) + kt(y_i)}{k+1} - t\left(\frac{\eta_i + ky_i}{k+1}\right) \right\}.$$

where the left term is a penalty complexity, and different k values put different stress on fit and parsimony. In Table 2, two values of k are used, $k = 0.5$ and $k = 5$, with the latter putting more stress on goodness of fit.

Also presented are predictive checks based on replicate observations. Posterior predictive probabilities $\Pr(y_{i,rep} \geq y_i | y)$ in extreme tails (e.g. values under 0.1 or over 0.9) indicate poorly fitted cases. The mixed predictive scheme (Marshall and Spiegelhalter, 2003), providing checks that are close to leave-one-out cross validation (Green et al, 2009), was also applied. This involves sampling new random effects $s_{i,rep}$, and then sampling replicate data $y_{i,rep,mixed}$ conditional on these new effects.

Table 2 shows that fit is generally improved under the heteroscedastic option, and predictive checks are also improved. The estimates for λ suggest that spatial dependence is not overly pronounced, and hence illustrate the broader principle that a spatial prior represent unstructured as well as structured variation: estimates of λ are all under 0.8. Figure 4 demonstrates disjunction between high risk and adjacent low risk areas. Table 2 also shows positive effects for both predictors but less precise effects under the scale mixture approach, in line with a general principle that neglecting heteroscedasticity may lead to mis-stated regression coefficient standard errors.

Table 3 contains a more detailed assessment of predictive discrepancies between the two approaches for the regression without predictors. As mentioned above, the κ_i effects will act to identify spatial outliers, with illness levels discrepant from their neighbours, and so

Table 3 contains the 20 areas with the lowest posterior mean κ_i under the scale mixture approach. One may assess spatial outlier status to some extent from the observed data. The first two columns of Table 3 contain maximum likelihood (ML) relative risks in each area $R_i = y_i/E_i$, and relative risks in the neighbourhoods N_i of each area, with ML estimates $L_i = \sum_{j \in N_i} y_j / \sum_{j \in N_i} E_j$.

Table 3 shows two types of outlier. One consists of major urban centres with high risk themselves, but a low risk hinterland (e.g areas 1,2,3, and 8 in the Table). For example, Figure 5 shows estimated relative risk patterns around area 2 (Leicester). These areas are underpredicted under the constant scale model, with mixed predictive $\Pr(y_{i,rep,mixed} \geq y_i | y)$ p-values under 0.025. Under the scale mixture model they have higher means μ_i , closer to the observed y_i , as there is less local borrowing of strength. The other type of outlier (e.g. areas 5 and 6 in the Table) are low risk areas with much higher risk neighbourhoods. These are overpredicted under the constant scale model, with $Pr(y_{rep} > y | y) = 0.91$ and $\Pr(y_{i,rep,mixed} \geq y_i | y) = 1$ for area 6. Under the scale mixture model, modelled means are reduced closer to the observed y_i . For all 20 areas, 19 have mixed predictive p-values under 0.05 or over 0.95 under a constant scale model, whereas under the scale mixture, this is reduced to 12 out of 20.

Table 4 contains the 10 areas with the lowest posterior mean κ_i under the scale mixture approach when the two covariates are included. These areas illustrate when modelled relative risk in area i itself μ_i/E_i are discrepant from implied relative risk $\sum_{j \in N_i} \mu_j / \sum_{j \in N_i} E_j$ in the locality of area i . These discrepancies may be related to covariate patterns. Under a scale mixture approach, local borrowing of strength is lessened, and Table 4 shows that the predicted TB counts μ_i are closer to the actual counts than under the constant scale LLB.

8 Conclusion

Different forms of spatial correlation analysis or modelling have been proposed in disease applications, ecological epidemiology, environmental science and other settings. Both Bayesian and frequentist estimation have been used. Common themes include identifying elevated risk areas or clusters of areas, and finding predictors of risk, while recognizing the explicitly spatial structure of the observations. For example, in a review of regression findings from spatial species abundance data, Dorfmann (2007) shows that ignoring spatial dependence (e.g. in regression residuals) leads to possible bias in parameter estimates and optimistic standard errors. However, while it is important to incorporate spatial dependence in models for area data, the assumptions of such techniques should be assessed, and subject to modification when the data so indicate. In particular, spatial discontinuities suggest a modification to the principle of uniform local smoothing.

In particular, Bayesian analyses of spatially arranged data often employ a conditionally

autoregressive prior, which can express spatial clustering commonly present in the underlying risks, but typically assume a normal density for risks and uniform conditional association. However, a more sensitive parameterisation with utility in detecting outliers and locally irregular risk patterns may be obtained by allowing for non-normality. Commonly applied conditionally autoregressive priors such as the proper CAR prior and the convolution prior also have potential deficits when the observations contain a mixture of spatial dependence and unstructured heterogeneity. The present paper has proposed a scale mixture version of the Leroux et al (1999) spatial prior, combining the benefit of adaptability when risks are discrepant in adjacent areas, and also a less problematic approach to representing a mixture of clustering and heterogeneity.

The analyses here show improved fit to infectious disease data, which may often show pronounced risk variability between areas. In England, high risk areas are often major urban centres, whereas the neighbouring suburban or rural hinterlands of such centres may be low risk. In such situations some modification of the uniform local borrowing of strength principle may be beneficial.

Appendix 1

The R code for simulating data for 49 mainland US states is as follows:

```
# 49 by 49 binary adjacency matrix
W <- read.table("adj_state.txt")
# numbers of neighbours
d=c(4,5,6,3,7,3,3,2,2,5,6,5,4,6,4,7,3,1,5,5,3,4,4,8,4,6,5,3,3,5,5,4,3,5,6,4,6,2,2,6,8,4,6,3,6,2,5,4,6)
# expected events (TB incidence)
E <- c(147,199.5,89.6,1144.7,157.6,111.2,28.5,20.2,621.5,293.7,47.4,389.3,
197.3,94.5,86.5,133.4,138.4,42.4,180.1,207.7,302.4,164.4,89.1,184.3,31.4,56,
83.8,41.3,272.8,62.7,606.1,297.6,22,354.4,115.3,121.5,399.6,
32.8,145.3,25.5,197.4,765.3,79.8,19.7,250.2,211.6,58.4,175.8,17.5)
# parameter and data definitions
N <- 49; Tau <- 3; lam=0.7; T <- 100; nu <- 10; nu.2 <- nu/2
kap <- Qdiag <- numeric(N); var.s <- numeric(T)
y <- matrix(,N,T); Q <- C <- matrix(,N,N)
library(mvtnorm)
# simulation
for (t in 1:T) {for(i in 1:N) { # scale mixture effects
kap[i] <- rgamma(1,nu.2,nu.2);
Qdiag[i] <- Tau*kap[i]*(1-lam+lam*d[i])}
for(i in 1:N) { for (j in 1:N) {
Q[i,j] <- (i==j)*Qdiag[i]-(1-(i==j))*Tau*lam*W[i,j]*kap[i]*kap[j]} }
C <- solve(Q)
s <- rmvnorm(1, mean = rep(0, nrow(C)), sigma=C, method=c("svd"))
```

```

eta <- log(E)-0.1+s
mu <- exp(eta)
var.s[t] <- var(s[1:N])
for (i in 1:N){y[i,t] <- rpois(1,mu[i])}}

```

References

- Alam M, Rönnegård L, Shen X (2015) Fitting conditional and simultaneous autoregressive spatial models in hglm. *The R Journal*, 7(2): 5-18
- Banerjee S, Carlin B, Gelfand A (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC
- Beale L, Abellan J, Hodgson S, Jarup L (2008) Methodologic issues and approaches to spatial epidemiology. *Environ Health Perspect.* 116(8):1105-10.
- Beale C, Lennon J, Yearsley J, et al (2010) Regression analysis of spatial data. *Ecol Lett.* 13(2):246-64.
- Berkhof J, van Mechelen I, Hoijtink H (2000) Posterior predictive checks: principles and discussion. *Computational Statistics* 15(3): 337–354.
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *J. R. Statist. Soc. B*, 36, 192-236
- Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-21.
- Besag J, Kooperberg C. (1995) On conditional and intrinsic autoregressions. *Biometrika*, 82, 733-746.
- Best N (1999) Bayesian ecological modelling. In: Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J.-F., Bertollini, R., (Eds.) *Disease Mapping and Risk Assessment for Public Health*. Wiley, New York, pp. 194–201
- Boris Choy S, Chan, J (2008) Scale mixtures distributions in statistical modelling. *Australian & New Zealand Journal of Statistics*, 50(2), 135-146.
- Brewer M, Nolan A (2007) Variable smoothing in Bayesian intrinsic autoregressions. *Environmetrics* 18, 841–857.
- Brooks S, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J. Computational and Graphical Stat*, 7: 434-45
- Carlin B, Louis T (2008) *Bayesian Methods for Data Analysis*, Third Edition. CRC
- Clayton D, Bernardinelli L, Montomoli C (1993) Spatial correlation in ecological analysis. *International Journal of Epidemiology*, 22(6): 1193-1202.
- Cressie N, Read T (1988) *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York.
- Dormann C (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16(2): 129-138.
- Duarte-Cunha M, Marcelo da Cunha G, Souza-Santos R (2015) Geographical heterogeneity in the analysis of factors associated with leprosy in an endemic area of Brazil: are we eliminating the disease? *BMC Infect Dis.* 2015, 15:196.
- Farnsworth M, Ward M (2009) Identifying spatio-temporal patterns of transboundary dis-

- ease spread: examples using avian influenza H5N1 outbreaks. *Veterinary Research*, 40(3): 1-14.
- Fernandez C, Steel M (1998) On Bayesian modeling of fat tails and skewness. *J. Am. Statist. Assoc.* 93, 359–71.
- Gelfand A, Ghosh S (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85(1): 1-11.
- Gelman A (1996) Bayesian model-building by pure thought: some principles and examples. *Statistica Sinica*, 6(1): 215-232.
- Gelman A, Carlin J, Stern H, Rubin D (2003) *Bayesian Data Analysis*, 2nd edition. Chapman & Hall/CRC Press, Boca Raton
- Geweke K (1993) Bayesian treatment of the independent student linear model. *Journal of Applied Econometrics* 8(S):19-40
- Green M, Medley G, Browne W (2009) Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression. *Veterinary Research*, 40(4), 1-10.
- Horabik J, Nahorski Z (2010) A statistical model for spatial inventory data: a case study of N₂O emissions in municipalities of southern Norway. *Climatic Change*, 103 (1-2): 263-276.
- Lawson A, Clark A (2002) Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine*, 21:359–370.
- Lee D (2011) A Comparison of Conditional Autoregressive Models Used in Bayesian Disease Mapping. *Spatial and Spatio-temporal Epidemiology*, 2, 79-89
- Leroux B, Lei X, Breslow N (1999) Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence. In *Statistical Models in Epidemiology: the Environment and Clinical Trials*, Halloran, M and Berry, D (eds), pp. 135-178. Springer-Verlag, New York.
- LeSage J (1997) Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1-2): 113-129.
- Lopez De Fede A, Stewart J et al (2008) Tuberculosis in socio-economically deprived neighborhoods: missed opportunities for prevention. *Int J Tuberc Lung Dis.* 12(12):1425-30.
- Lunn D, Spiegelhalter D, Thomas A, Best, N (2009) The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25), 3049-3067
- Maciel E, Pan W, Dietze R et al (2010) Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. *Int J Tuberc Lung Dis.* 14(11):1395-402.
- MacNab Y (2014) On identification in Bayesian disease mapping and ecological-spatial regression models. *Stat Methods Med Res.* 23(2):134-55.
- Manda S (2013) Macro Determinants of Geographical Variation in Childhood Survival in South Africa Using Flexible Spatial Mixture Models. In: Kandala N-B, Ghilagaber G (eds.) *Advanced Techniques for Modelling Maternal and Child Health in Africa*, Springer, Dordrecht, pp. 147-168.
- Marshall E, Spiegelhalter D (2003) Approximate cross validatory predictive checks in disease mapping. *Stat. Med.* 22:1649–1660
- Marshall E, Spiegelhalter D (2007) Identifying outliers in Bayesian hierarchical models: a

simulation-based approach. *Bayesian Analysis*, 2(2), 409-444.

Morgenstern H (1995) Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health*.16:61-81.

Nathoo F, Ghosh P (2013) Skew-elliptical spatial random effect modeling for areal data with application to mapping health utilization rates. *Statistics in Medicine*, 32(2): 290-306.

Nunes C (2007) Tuberculosis incidence in Portugal: spatiotemporal clustering. *Int J Health Geogr*. 2007 Jul 11;6:30.

Ploubidis G, Palmer M, Blackmore C et al (2012) Social determinants of tuberculosis in Europe: a prospective ecological study. *European Respiratory Journal*, 40(4): 925-930.

Polson N, Scott J, Windle J (2014) The Bayesian Bridge. *Journal of the Royal Statistical Society: Ser B*, 76(4): 713-733.

Reich B, Hodges J (2008) Modeling longitudinal spatial periodontal data: a spatially adaptive model with tools for specifying priors and checking fit. *Biometrics*. 64(3):790-9.

Rieder H (1999) *Epidemiologic Basis of Tuberculosis Control*. International Union Against Tuberculosis and Lung Disease, Paris.

Rodrigues E, Assunção R (2012) Bayesian spatial models with a mixture neighborhood structure. *Journal of Multivariate Analysis*, 109: 88-102.

Rue H, Held L (2005) *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

Shekhar S, Jiang Z, Ali R et al (2015) Spatiotemporal data mining: a computational perspective. *ISPRS International Journal of Geo-Information*, 4(4), 2306-2338.

Smith T, Wakefield J, Dobra A (2015) Restricted Covariance Priors with Applications in Spatial Statistics. *Bayesian Analysis*, 10(4): 965-990

Spiegelhalter D, Best N, Carlin B, van der Linde A (2002) Bayesian measures of model complexity and fit. *J Royal Statistical Society, Ser B*, 64, 583-641

Ugarte M, Ibáñez B, Militino A (2005) Detection of spatial variation in risk when using CAR models for smoothing relative risks. *Stochastic Environmental Research and Risk Assessment*, 19(1): 33-40.

Varga C, Pearl D, McEwen S et al (2015) Area-level global and local clustering of human *Salmonella* Enteritidis infection rates in the city of Toronto, Canada, 2007-2009. *BMC Infect Dis*. 2015, 15:359.

Waller L, Carlin B (2010) Disease Mapping. In: Gelfand A, Diggle P, Guttorp P, Fuentes M (eds) *Handbook of Spatial Statistics*, Chapman & Hall/CRC Press, Boca Raton, pp 217-243.

West M (1984) Outlier models and prior distributions in Bayesian linear regression. *J Royal Statistical Society, Ser B*, 46, 431-9

Wikle C (2003) Hierarchical models in environmental science, *International Statistical Review*, 71(2): 181-199

Yan J (2007) Spatial stochastic volatility for lattice data. *Journal of Agricultural, Biological, and Environmental Statistics*, 12(1): 25-40.

Yi N, Xu S (2008) Bayesian LASSO for quantitative trait loci mapping. *Genetics*, 179(2): 1045-1055.

Table 1 Recovered Parameter Estimates from 100 Simulated Datasets

(a) ν set to 3

Parameter	Percentiles of posterior means			Samples with 95% credible interval containing guide value
	20th	50th	80th	Percent
λ	0.44	0.59	0.67	99
ν	2.4	4.5	8.4	95
$\log(\nu)$	0.76	1.29	1.88	95
β_0	-0.26	-0.12	0.04	88

(b) ν set to 10

Parameter	Percentiles of posterior means			Samples with 95% credible interval containing guide value
	20th	50th	80th	Percent
λ	0.52	0.65	0.72	99
ν	7.8	10.4	12.5	100
$\log(\nu)$	1.68	2.09	2.36	97
β_0	-0.26	-0.14	0.01	91

(c) ν set to 25

Parameter	Percentiles of posterior means			Samples with 95% credible interval containing guide value
	20th	50th	80th	Percent
λ	0.53	0.65	0.73	100
ν	16.1	23.4	31.0	99
$\log(\nu)$	2.22	2.65	3.03	99
β_0	-0.28	-0.15	-0.02	84

Table 2 Goodness of fit and Parameter Summaries, Models without and including predictors

Model Fit and Checks

<i>Model without predictors</i>		LLB constant scale	Scale Mixture LLB
Fit Measures	PPL (k=0.5)	645.8	630.6
	PPL (k=5)	677.4	661.9
Predictive Checks	Total observations overpredicted, with $\Pr(y_{i,rep} > y_i \mathbf{y}) > 0.9$	5	0
	Total observations underpredicted, with $\Pr(y_{i,rep} > y_i \mathbf{y}) < 0.1$	0	0
	Total observations overpredicted, with $\Pr(y_{i,rep,mixed} > y_i \mathbf{y}) > 0.9$	30	25
	Total observations underpredicted, with $\Pr(y_{i,rep,mixed} > y_i \mathbf{y}) < 0.1$	38	33

<i>Model with predictors</i>		LLB constant scale	Scale Mixture LLB
Fit Measures	PPL (k=0.5)	622.2	616.3
	PPL (k=5)	666.6	660.2
Predictive Checks	Total observations overpredicted, with $\Pr(y_{i,rep} > y_i \mathbf{y}) > 0.9$	5	3
	Total observations underpredicted, with $\Pr(y_{i,rep} > y_i \mathbf{y}) < 0.1$	0	0
	Total observations overpredicted, with $\Pr(y_{i,rep,mixed} > y_i \mathbf{y}) > 0.9$	31	27
	Total observations underpredicted, with $\Pr(y_{i,rep,mixed} > y_i \mathbf{y}) < 0.1$	31	25

Parameter Summaries (posterior mean, 95% credible intervals)

<i>Model without predictors</i>		LLB constant scale	Scale Mixture LLB
λ	Spatial dependence	0.59 (0.37,0.89)	0.57 (0.37,0.86)
ν	Scale mixing parameter		8.1 (4.2,15.9)
<i>Model with predictors</i>		LLB constant scale	Scale Mixture LLB
λ	Spatial dependence	0.76 (0.49,0.98)	0.75 (0.51,0.98)
ν	Scale mixing parameter		10.42 (4.19,27.31)
β_1	Effect of deprivation	4.98 (4.01,6.11)	4.96 (3.71,6.23)
β_2	Effect of population density	1.34 (0.98,1.7)	1.25 (0.75,1.81)

Table 3 Areas Ranked by Outlier Status, No Predictors

Number	Name	Events	Heteroscedastic LLB						Constant scale LLB		
			Relative Risk (MLE)	Spatial lag relative risk (MLE)	κ	μ	$\Pr(y_{rep} > y y)$	$\Pr(y_{rep, mixed} > y y)$	μ	$\Pr(y_{rep} > y y)$	$\Pr(y_{rep, mixed} > y y)$
1	Peterborough	170	2.12	0.33	0.48	167.6	0.48	0.03	166.3	0.40	0.01
2	Leicester	528	3.57	0.44	0.48	526.2	0.49	0.03	524.7	0.46	0.00
3	Blackburn-Darwen	157	2.55	0.79	0.49	154.1	0.39	0.03	152.7	0.39	0.00
4	Preston	106	1.66	0.30	0.55	103.9	0.40	0.05	103.4	0.41	0.02
5	Tandridge	3	0.09	1.12	0.56	5.7	0.78	0.98	6.6	0.87	1.00
6	Bromsgrove	2	0.05	1.59	0.57	5.0	0.79	0.98	5.9	0.91	1.00
7	Rushmoor	81	1.86	0.35	0.58	78.9	0.42	0.04	78.2	0.40	0.02
8	Luton	257	2.74	0.29	0.60	255.3	0.44	0.05	255.0	0.45	0.02
9	Crawley	75	1.47	0.40	0.60	72.9	0.42	0.06	72.2	0.39	0.03
10	Reading	159	2.02	0.37	0.63	156.9	0.44	0.06	156.7	0.44	0.02
11	Slough	248	3.91	1.71	0.64	246.2	0.44	0.07	245.6	0.45	0.02
12	Sheffield	279	1.04	0.32	0.64	276.6	0.45	0.07	276.4	0.45	0.04
13	Birmingham	1238	2.54	1.20	0.64	1234.0	0.43	0.04	1233.0	0.45	0.00
14	Southampton	130	1.07	0.17	0.65	128.3	0.42	0.06	127.7	0.43	0.04
15	South Staffordshire	4	0.09	0.85	0.65	7.1	0.78	0.97	7.4	0.83	0.98
16	Newcastle upon Tyne	122	0.84	0.19	0.66	119.9	0.44	0.07	119.8	0.43	0.04
17	Redditch	46	1.31	0.24	0.69	44.2	0.40	0.07	44.0	0.39	0.04
18	Woking	57	1.34	0.37	0.70	55.2	0.43	0.08	55.1	0.41	0.06
19	Rossendale	6	0.20	1.13	0.71	8.5	0.71	0.95	8.8	0.74	0.98
20	Halton	2	0.04	0.36	0.71	4.2	0.77	0.96	4.5	0.82	0.98

Table 4 Areas Ranked by Outlier Status, Regression with Predictors

Name	Events	Scale mixture LLB		Constant scale LLB		
		κ	μ_i	Model RR (μ_i/E_i)	Model neighbour- hood RR	μ_i
Brent	896	0.57	892.9	7.22	2.41	889.7
Peterborough	170	0.60	166.3	2.06	0.35	165.1
Barnsley	20	0.60	23.7	0.24	0.85	24.8
Swale	8	0.62	11.1	0.20	0.50	11.7
Woking	57	0.67	54.2	1.26	0.39	53.7
North Lincolnshire	47	0.68	43.5	0.61	0.27	43.0
Kirklees	287	0.69	283.1	1.53	1.00	281.9
Newham	1072	0.71	1068.0	9.34	2.99	1068.0
Tandridge	3	0.71	6.8	0.21	1.10	7.5
Rushmoor	81	0.72	78.0	1.79	0.36	77.6

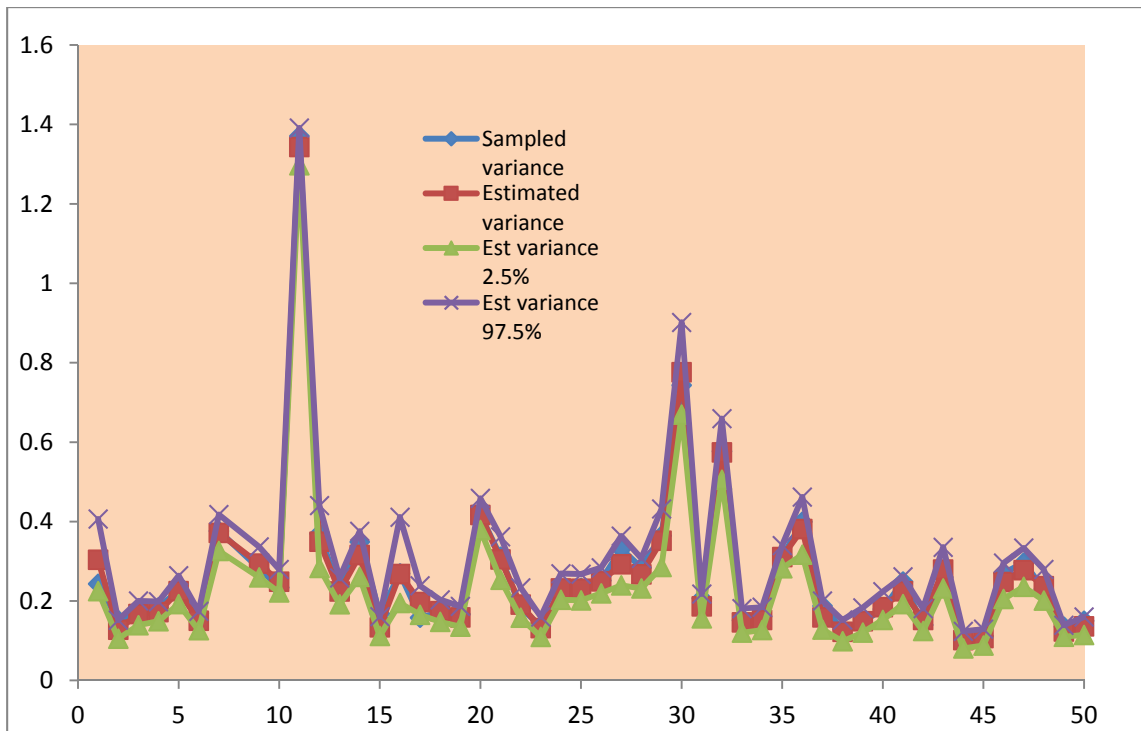


Figure 1a Simulated and Estimated Spatial Effect Variances: US Mainland States (1st 50 Samples)

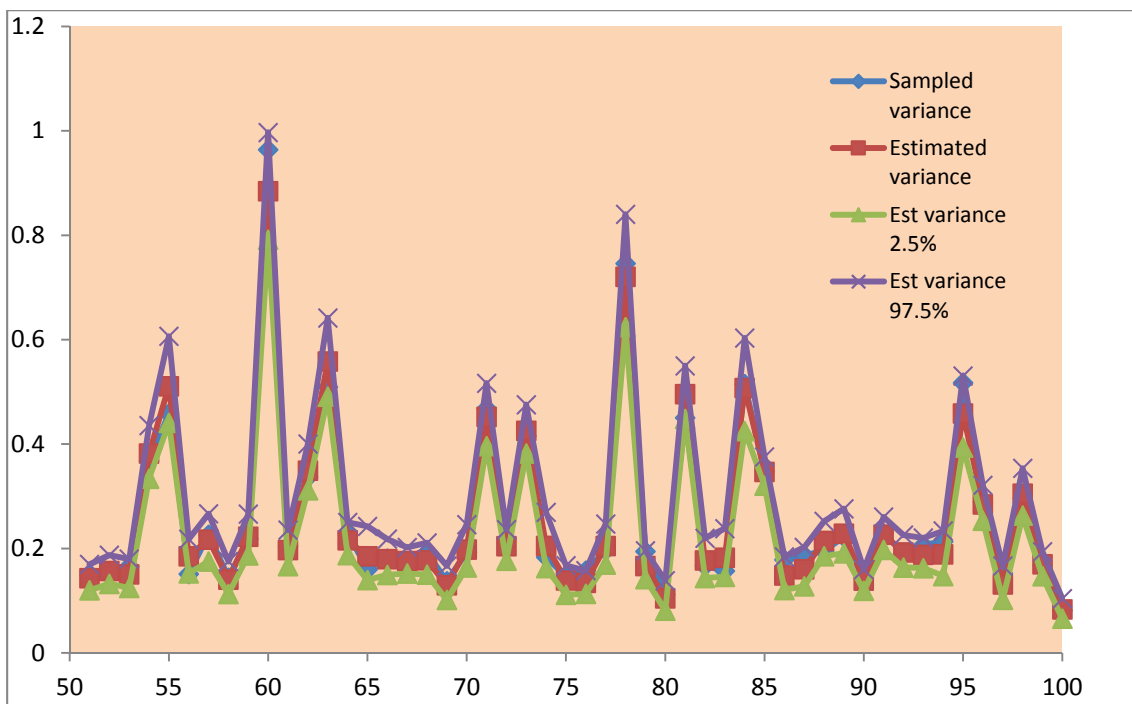


Figure 1b Simulated and Estimated Spatial Effect Variances: US Mainland States (2nd 50 Samples)

Figure 2. Posterior Intervals, re-estimated $\log(v)$, 100 Simulated Datasets with Setting $v=3$

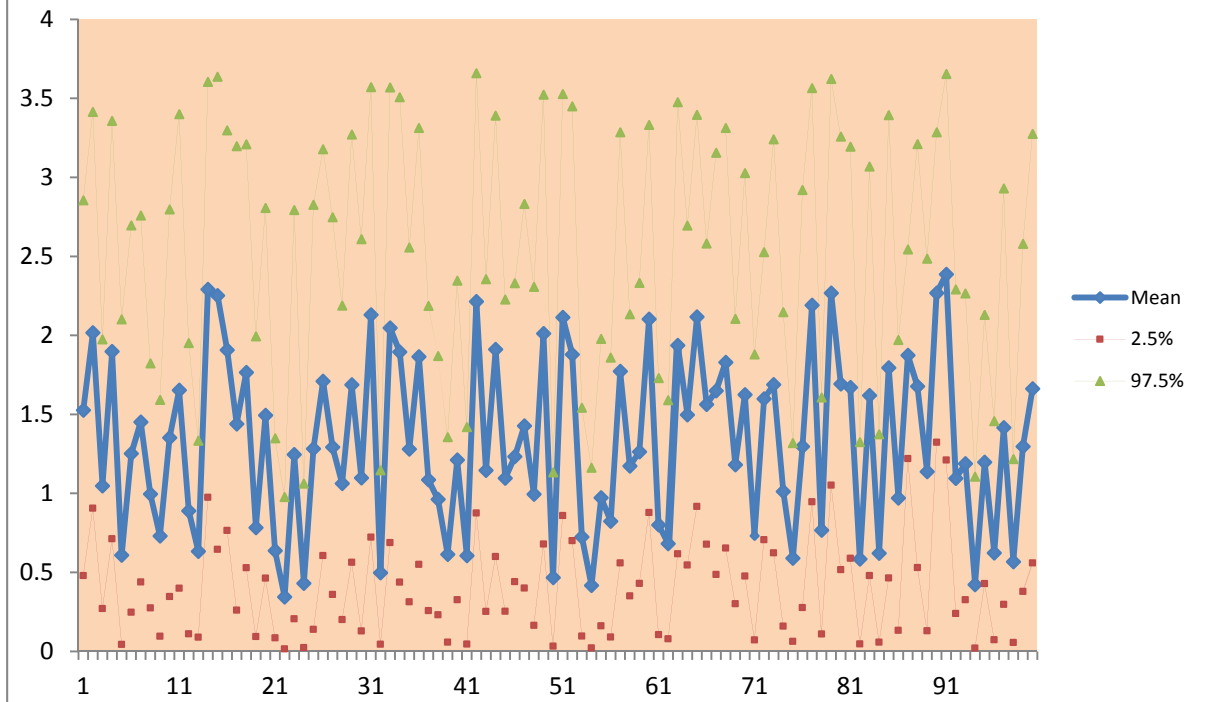
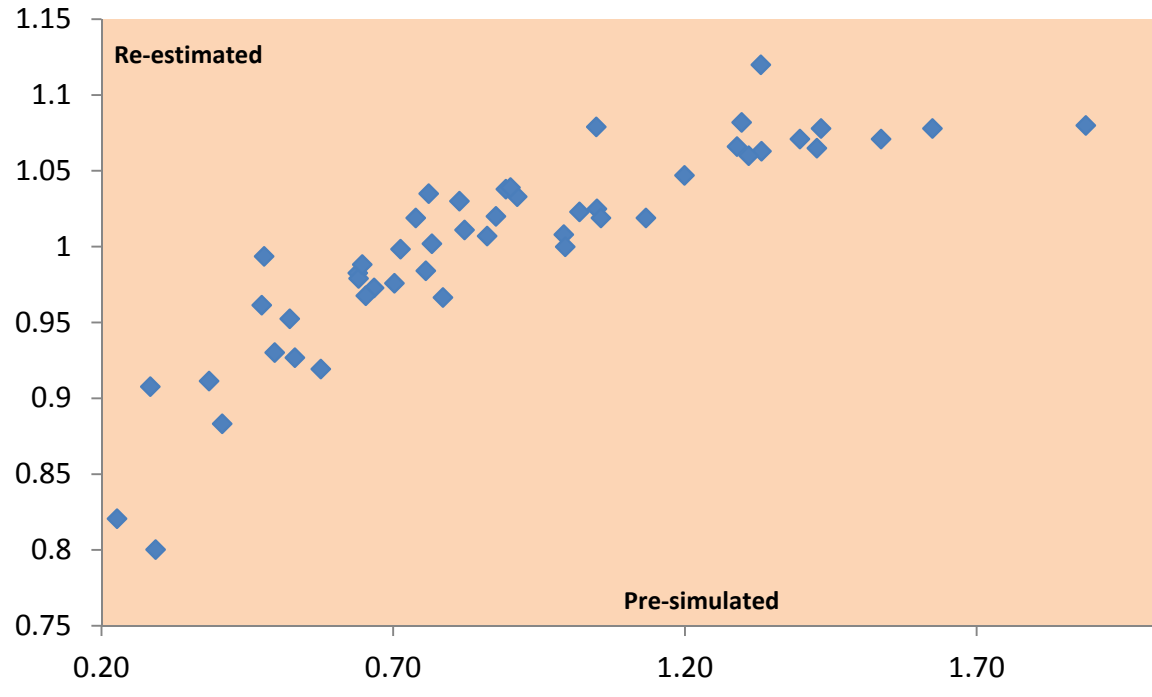


Figure 3 Pre-simulated and Posterior Mean Re-Estimated κ_i . Simulated Data (100 Datasets) with Preset κ_i .



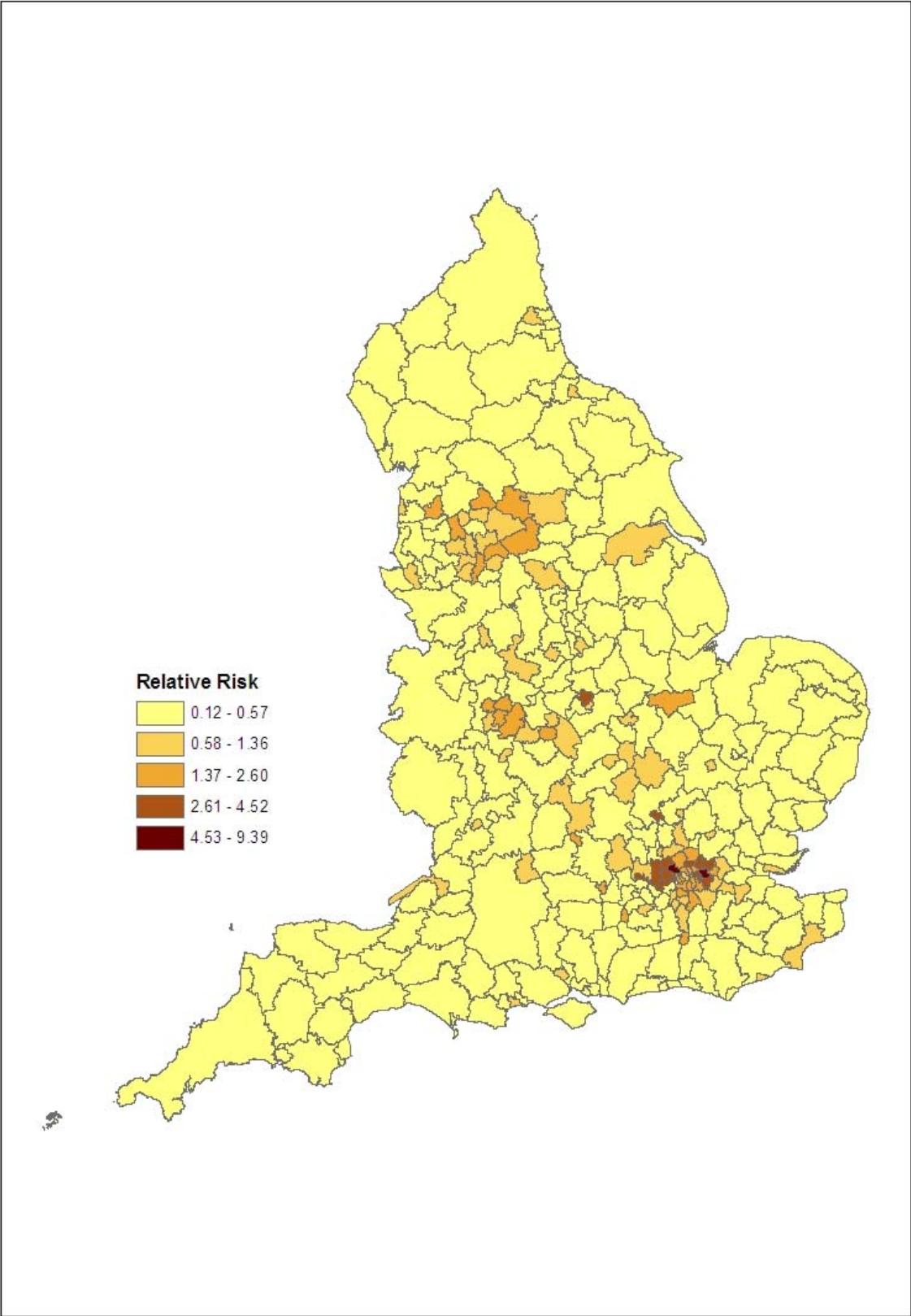


Figure 4 Modelled Relative Risks of TB Incidence, Scale Mixture Model

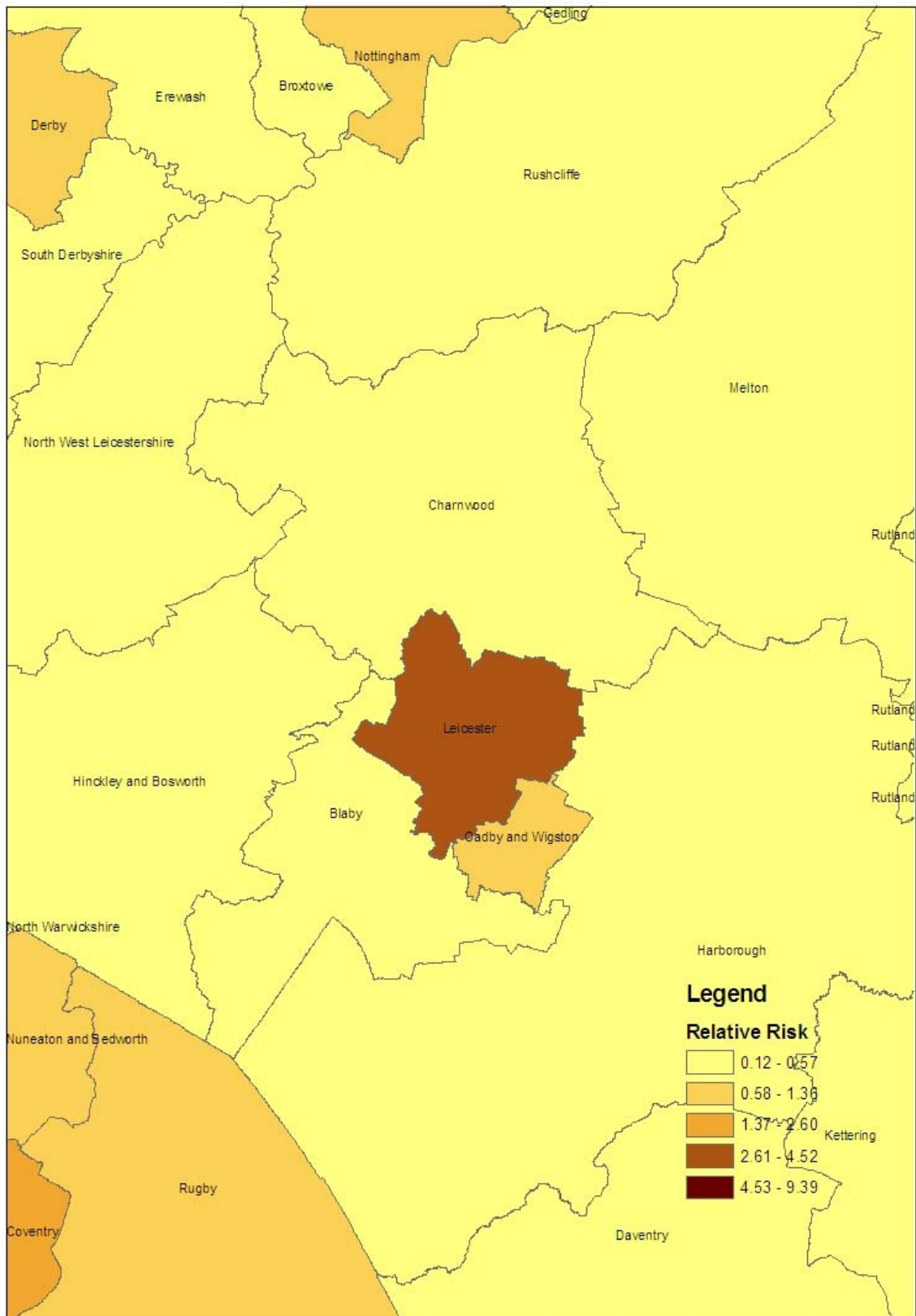


Figure 5 Modelled Relative Risks of TB Incidence around Leicester, Scale Mixture Model