Spatial and Spatio-temporal Epidemiology 2 (2011) 79–89

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

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

A comparison of conditional autoregressive models used in Bayesian disease mapping

Duncan Lee

School of Mathematics and Statistics, University Gardens, University of Glasgow, Glasgow G12 8QW, United Kingdom

ARTICLE INFO

Article history:
Received 14 October 2010
Revised 14 February 2011
Accepted 7 March 2011
Available online 12 March 2011

Keywords:
Conditional autoregressive models
Disease mapping
Spatial correlation

ABSTRACT

Disease mapping is the area of epidemiology that estimates the spatial pattern in disease risk over an extended geographical region, so that areas with elevated risk levels can be identified. Bayesian hierarchical models are typically used in this context, which represent the risk surface using a combination of available covariate data and a set of spatial random effects. These random effects are included to model any overdispersion or spatial correlation in the disease data, that has not been accounted for by the available covariate information. The random effects are typically modelled by a conditional autoregressive (CAR) prior distribution, and a number of alternative specifications have been proposed. This paper critiques four of the most common models within the CAR class, and assesses their appropriateness via a simulation study. The four models are then applied to a new study mapping cancer incidence in Greater Glasgow, Scotland, between 2001 and 2005.

1. Introduction

Modelling data that relate to contiguous spatial units, such as electoral wards or pixels, is a common problem in a number of statistical applications, including disease mapping (MacNab et al., 2006), geographical association studies (Lee et al., 2009), image analysis (Molina et al., 1999) and agricultural field trials (Besag and Higdon, 1999). The response variables in these applications typically display spatial dependence, that is, observations from units close together are more similar than those relating to units further apart. A number of statistical approaches have been adopted for modelling spatial correlation in such data, including geostatistical models (Biggeri et al., 2006), simultaneously autoregressive models (Kissling and Carl, 2008) and conditional autoregressive models (MacNab, 2003).

In this paper we focus on disease mapping, the aim of which is to map the spatial pattern in disease risk over a predefined study region. Bayesian hierarchical models are typically used in such analyses, where any spatial correlation in the disease data is modelled at the second level of the hierarchy by a set of random effects. These effects are most commonly represented by a conditional autoregressive (CAR) prior distribution, which is a type of Markov random field. A number of models have been proposed within this general class of CAR priors, including the intrinsic and convolution models (both Besag et al., 1991), as well as alternatives proposed by Cressie (1993) and Leroux et al. (1999). However, to our knowledge, no formal comparison has been made of the appropriateness of each of these prior models. Therefore this paper presents such a critique, by comparing both their theoretical properties and practical performance.

The remainder of this paper is organised as follows. Section 2 provides a background to Bayesian disease mapping, while Section 3 presents the four commonly used CAR prior models. Section 4 compares the performance of the four models via simulation, while Section 5 applies them to a new disease mapping study of cancer incidence in Greater Glasgow, UK. Finally, Section 6 contains a concluding discussion and areas for future work.

E-mail address: Duncan.Lee@glasgow.ac.uk

1877-5845/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sste.2011.03.001

2. Disease mapping

In disease mapping studies the region of interest is split into n contiguous small-areas, such as census tracts or electoral wards, and the aim of the study is to detect which areas exhibit elevated disease risks. The observed numbers of disease cases in each small-area are collectively denoted by $\mathbf{y} = (y_1, \ldots, y_n)$, where y_k denotes the number of cases in area k. To fairly assess which areas exhibit elevated levels of disease risk, the numbers of cases expected to occur in each small-area are calculated. These expected numbers of cases are denoted by $\mathbf{E} = (E_1, \ldots, E_n)$, and are based on the size and demographic structure of the population living within each small-area. They are calculated by dividing the population living in each small-area into a number of strata, based on their age and sex. The number of people in each stratum is multiplied by the incidence rate for that stratum, and the results are summed over strata to produce the expected number of cases. From these data the simplest measure of disease risk is the standardised incidence ratio (SIR), which is calculated for area k as $\text{SIR}_k = y_k/E_k$. Values above one represent areas with elevated levels of disease risk, while values below one correspond to comparatively healthy areas. However, elevated risks (as measured by the SIR) are likely to happen by chance if E_k is small, which can occur if the disease in question is rare and/or the population at risk is small.

To overcome this problem a Bayesian model-based approach is typically adopted, which estimates the set of disease risks using covariate information and a set of random effects. The random effects borrow strength from values in neighbouring areas, which reduces the likelihood of excesses in risk occurring by chance. The class of Bayesian hierarchical models typically used in this context have been described in detail by Elliott et al. (2000), Banerjee et al. (2004) and Lawson (2008), and a general formulation is given by

$$\begin{aligned} &Y_k|E_k, R_k \sim \text{Poisson}(E_k R_k) \quad \text{for } k = 1, \ldots, n, \\ &\ln(R_k) = \mu + \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta} + \phi_k, \\ &\beta_i \sim N(0, 10) \quad \text{for } i = 1, \ldots, p, \\ &\mu \sim N(0, 10). \end{aligned} \tag{1}$$

In the above model R_k denotes the risk of disease in area k, which is modelled by an intercept term μ, a set of p covariates $\mathbf{x}_k^{\mathrm{T}} = (x_{k1}, \ldots, x_{kp})$ and a random effect ϕ_k. The regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ and the intercept term μ are assigned weakly informative Gaussian prior distributions, with mean zero and variance 10. The random effects are included to model any overdispersion and/or spatial correlation in the data, that persist after adjusting for the available covariate information. Overdispersion can occur because the Poisson likelihood enforces the restriction that $\text{Var}[Y_k] = \mathbb{E}[Y_k]$, whereas in most studies of this type, $\text{Var}[Y_k] > \mathbb{E}[Y_k]$. The existence of overdispersion and spatial correlation is likely due to the presence of unmeasured risk factors, which thus cannot be included as covariates in the model. Inference for this type of model is typically based on Markov chain Monte-Carlo (MCMC) simulation, utilising a combination of Gibbs sampling and Metropolis–Hastings steps.

3. Conditional autoregressive models

In disease mapping studies the random effects $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$ are commonly modelled by the class of conditional autoregressive (CAR) prior distributions, which are a type of Markov random field model. These models are specified by a set of n univariate full conditional distributions $f(\phi_k|\boldsymbol{\phi}_{-k})$ (where $\boldsymbol{\phi}_{-k} = (\phi_1, \ldots, \phi_{k-1}, \phi_{k+1}, \ldots, \phi_n)$), for $k = 1, \ldots, n$, rather than by a single multivariate distribution $f(\boldsymbol{\phi})$. Spatial correlation between the random effects is determined by a binary $n \times n$ neighbourhood matrix W, whose jkth element w_{jk} is equal to one if areas (j, k) are defined to be neighbours, and is zero otherwise. If two areas are defined to be neighbours their random effects are correlated, while random effects in non-neighbouring areas are modelled as being conditionally independent given the remaining elements of $\boldsymbol{\phi}$. The most common approach is to assume that areas (j, k) are neighbours (i.e., $w_{jk} = 1$) if and only if they share a common border, which is denoted in this paper by $j \sim k$. A number of different conditional autoregressive prior models have been proposed in a disease mapping context, and the remainder of this section describes the four that are most commonly used.

3.1. Intrinsic model

The simplest CAR prior is the intrinsic autoregressive (IAR) model, which was proposed by Besag et al. (1991) and has full conditional distributions given by

$$\phi_k|\boldsymbol{\phi}_{-k}, W, \tau_I^2 \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\phi_j, \frac{\tau_I^2}{n_k}\right). \tag{2}$$

The conditional expectation of ϕ_k is equal to the mean of the random effects in neighbouring areas, while the conditional variance is inversely proportional to the number of neighbours n_k. This variance structure recognises the fact that in the presence of strong spatial correlation, the more neighbours an area has the more information there is in the data about the value of its random effect. The variance parameter τ_I^2 controls the amount of variation between the random effects, and the choice of hyperprior is discussed in Section 3.5.

The above model is the simplest possible CAR prior, and as a result is rather restrictive. Its single parameter does not determine the strength of the spatial correlation between the random effects, because multiplying each ϕ_k by 10 will increase τ_I^2 but leave the spatial correlation structure unchanged. Therefore model (2) can only represent strong spatial correlation structures, and is hence not appropriate if the data are only weakly correlated. In addition, the joint distribution for $f(\boldsymbol{\phi})$ corresponding to (2) is improper, because it is possible to add a constant to each ϕ_k without changing the distribution. This impropriety can be remedied by enforcing a constraint such as, $\sum_{j=1}^{n}\phi_j = 0$, which can be implemented numerically at each iteration of an MCMC algorithm.

3.2. Convolution model

The convolution model was also proposed by Besag et al. (1991), and combines the intrinsic model with an

additional set of independent random effects. The model is given by

$$\begin{aligned}&\phi_k = \theta_k + \psi_k,\\ &\theta_k|\sigma^2 \sim N(0,\sigma^2),\\ &\boldsymbol{\psi} = (\psi_1,\ldots,\psi_n)|W,\tau_I^2 \sim \text{IAR}(W,\tau_I^2),\end{aligned} \tag{3}$$

where $\boldsymbol{\psi}$ is represented by the intrinsic CAR prior described in the previous subsection. The second set of random effects $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n)$ is independent between areas, and different strengths of spatial correlation can be represented by varying the relative sizes of the two components $(\boldsymbol{\theta},\boldsymbol{\psi})$. However, the disadvantage of this flexibility is that each data point is represented by two random effects, and hence only their sum $\theta_k + \psi_k$ is identifiable. Eberly and Carlin (2000) assess the extent of this problem, and find that MCMC convergence is slow for this model, and that the individual components (θ_k,ψ_k) are not reliably estimated.

3.3. Cressie model

An alternative approach for modelling varying strengths of spatial correlation was proposed by Cressie (1993) and Stern and Cressie (2000), who use a single set of random effects, but introduce an additional spatial correlation parameter. The implementation of their model we consider here is given by

$$\phi_k|\boldsymbol{\phi}_{-k},W,\tau^2,\rho,\mu \sim N\left(\rho \times \frac{1}{n_k}\sum_{j\sim k}\phi_j + (1-\rho)\mu, \frac{\tau^2}{n_k}\right). \tag{4}$$

This CAR prior has the same conditional variance as the intrinsic model, while the conditional expectation is a weighted average of the mean of the random effects in neighbouring areas and an overall mean μ. The existence of μ in model (4) means that the intercept term in Eq. (1) is not required. The weight parameter ρ controls the strength of the spatial correlation between the random effects, with $\rho = 0$ corresponding to independence, while increasing its value towards one corresponds to increasingly strong spatial correlation ($\rho = 1$ simplifies to the intrinsic model). This set of full conditional distributions correspond to a proper multivariate Gaussian distribution for $\boldsymbol{\phi}$ if $0 \leqslant \rho < 1$, which has a constant mean of μ. The covariance matrix is equal to $\tau^2 Q_C^{-1}$, where Q_C has jkth element equal to n_k if $j = k$, $-\rho$ if $j \sim k$, and zero otherwise. The major drawback with this model is the form of the conditional variance, which is unappealing when ρ is close to zero. This is because in the absence of spatial correlation (when $\rho = 0$) there is no reason for the conditional variance of ϕ_k to be inversely proportional to the number of neighbours, as they provide no information about ϕ_k.

3.4. Leroux model

The final model considered here was originally proposed by Leroux et al. (1999), and has been further explored by MacNab (2003). It is based on a single set of random effects $\boldsymbol{\phi} = (\phi_1,\ldots,\phi_n)$, which are represented by the multivariate Gaussian distribution

$$\boldsymbol{\phi}|W,\tau^2,\rho,\mu \sim N\left(\boldsymbol{\mu},\tau^2[\rho W^* + (1-\rho)I_n]^{-1}\right). \tag{5}$$

In common with model (4), this prior has a constant non-zero mean $\boldsymbol{\mu} = (\mu,\ldots,\mu)$, which is consequently not required in Eq. (1). The precision matrix is given by $Q_L = \rho W^* + (1-\rho)I_n$, where I_n is an $n \times n$ identity matrix and the elements of W^* are equal to

$$w_{jk}^* = \begin{cases} n_k & \text{if } j = k\\ -1 & \text{if } j \sim k\\ 0 & \text{otherwise.}\end{cases} \tag{6}$$

The precision matrix is hence a weighted average of spatially dependent (represented by W^*) and independent (represented by I_n) correlation structures, where the weight is equal to ρ. This model can represent a range of weak and strong spatial correlation structures, with the special case of $\rho = 0$ simplifying to a model with independent random effects. The joint distribution (5) is proper if $0 \leqslant \rho < 1$, while $\rho = 1$ corresponds to the improper intrinsic model given by (2). The univariate full conditional distributions corresponding to (5) are given by

$$\phi_k|\boldsymbol{\phi}_{-k},W,\tau^2,\rho,\mu \sim N\left(\frac{\rho\sum_{j\sim k}\phi_j + (1-\rho)\mu}{n_k\rho + 1 - \rho}, \frac{\tau^2}{n_k\rho + 1 - \rho}\right). \tag{7}$$

The conditional expectation is a weighted average of the random effects in neighbouring areas and the overall mean μ, while the conditional variance has a more attractive form than that in model (4). When there is strong spatial correlation in the data ρ will be close to one and the conditional variance is approximately τ^2/n_k, which is the same as in the intrinsic model (2). In contrast, if the random effects are independent $\rho = 0$, and the conditional variance of ϕ_k is a constant (equal to τ^2). This is at odds with model (4), but is more theoretically appealing because there is no longer any information about ϕ_k in the neighbouring random effects.

3.5. Hyperpriors

3.5.1. Hyperprior for the correlation parameter ρ

For the Cressie and Leroux models the correlation parameter is assigned a discrete hyperprior, because it is leads to faster MCMC inference than if a continuous distribution was specified. This is because for each value of ρ proposed by the MCMC algorithm the determinant of the precision matrix needs to be calculated, which will be computationally demanding if the number of data points n is large. If the hyperprior is continuous a new value of ρ is proposed at each iteration of the MCMC algorithm, where as if a discrete prior is adopted only a small number of ρ values are possible (and hence only a small number of determinant calculations are required). The discrete hyperprior adopted here is given by

$$\rho \sim \text{discrete uniform}(a_1,\ldots,a_r).$$

A uniform prior is adopted (i.e., the possible values of ρ, $a_1,\ldots,a_r$ have the same prior probability) because before the analysis it is unknown whether the data contain strong, moderate or weak spatial correlation. The discrete prior is

implemented on the ρ scale (rather than say a Fisher transformation of it) because it is interpretable, with 0 corresponding to independence of the random effects, while increasing its value towards 1 corresponds to increasingly strong spatial correlation. Negative spatial correlation is rarely seen in disease mapping data, so the minimum allowable value for ρ is set at $a_1 = 0$. For models (4) and (7) to correspond to proper multivariate Gaussian distributions, ρ must be less than 1. Therefore we set $a_r = 0.95$, because the simulation study in the next section shows that for both models this is large enough to represent strong spatial correlation. In the absence of any prior information the remaining possible values $a_2, \ldots, a_{r-1}$ are assumed to be equally spaced between 0 and 0.95. Finally, the number of possible values r needs to be specified, and a value of 20 is used in this paper, giving possible values of $0, 0.05, 0.1, \ldots, 0.9, 0.95$. A sensitivity analysis to this value is conducted in Section 5.

3.5.2. Hyperprior for the variance parameters $(\tau_I^2, \sigma^2, \tau^2)$

Following the work of Gelman (2006), all variance parameters are assigned uniform $(0, M)$ priors on the standard deviation scale. The commonly used class of inverse-gamma (ϵ, ϵ) priors are sensitive to the value of ϵ if the true variance is close to zero, and are therefore not used. A value of $M = 10$ is specified as the upper limit of the uniform prior throughout this paper, although a small sensitivity analysis is conducted to this value in Section 5.

3.6. Inference

Inference for all models is based on Markov Chain Monte-Carlo simulation, using a combination of Gibbs sampling and Metropolis-Hastings steps. All regression and intercept parameters are updated using Metropolis steps, utilising a random walk proposal distribution. The random effects are updated using a block Metropolis-Hastings algorithm, where the block size can be tuned to obtain the desired acceptance rates. Variance parameters are Gibbs sampled from their full conditional inverse-gamma distributions, while ρ is straightforward to update because it has a discrete sample space. A function to run the Leroux model in the statistical package R (R Development Core Team, 2009) as well as supporting documentation and an example data set, are available in the Supplementary material accompanying this paper.

4. Simulation study

In this section, we present a simulation study, that compares the performance of the four CAR priors described in the previous section.

4.1. Data generation and study design

Simulated disease data are generated for the 271 intermediate geographies in the Greater Glasgow health board, which is the region used in the cancer mapping study in Section 5. The data are generated from model (1), where for simplicity, the percentage of the population who are income deprived is the only covariate. Its regression parameter is equal to $\beta = 0.1$, while the intercept term is fixed at $\mu = -0.2$. The expected numbers of disease cases, **E**, are those used in Section 5 for all cancer cases. The purpose of this study is to determine how well each of the CAR priors can represent different types of spatial correlation, and we simulate disease data under each of the following scenarios:

- Scenario 1: *Independence* – The random effects are generated from independent normal distributions, with mean zero and standard deviation equal to 0.2.
- Scenario 2: *Moderate spatial dependence* – The random effects are a convolution of the independent and spatially correlated processes used in Scenarios 1 and 3.
- Scenario 3: *Strong spatial dependence* – The random effects are generated from a multivariate Gaussian distribution with mean zero, with a correlation matrix specified by the Matern class with smoothness parameter equal to 2.5. The spatial range is fixed at 5 km, which corresponds to the median correlation between pairs of areas in the study region being 0.5.

Two hundred sets of disease counts were generated under each of the three scenarios, and the four CAR priors outlined in Section 3 were applied in each case. Each simulated data set is generated from a different realisation of the random effects, because it prevents the results from being affected by the particular set of random effects drawn. The relative performances of the four models are assessed by the following metrics.

- *Regression parameter:* β – Bias and root mean square error (RMSE) for the estimated regression parameter, presented as a percentage of its true value (0.1).
- *Disease risks:* R_k – Bias and RMSE for the set of disease risks $R_k = \exp(\mu + \phi_k + x_k\beta)$, which are again presented as a percentage of their true values.
- *Residual spatial correlation:* A permutation test (at the 5% level) based on Moran's I statistic is applied to the residuals from each model, and the percentage of significant results is reported.

4.2. Results

The results from all metrics and models are presented in Table 1. Overall, all of the models produce close to unbiased estimates of β and R_k, with relative percentage biases ranging between −0.85% and 3.01% across all models and scenarios. In the independence scenario (Scenario 1) the intrinsic model performs the worst in terms of RMSE for both the regression parameter and the disease risks, while the other models produce similar results. This is not surprising, as the intrinsic model is the only one considered here that cannot represent independence or weak spatial correlation. In contrast, in the presence of strong spatial correlation (Scenario 3) the convolution model performs worst in terms of RMSE. In addition, it failed to remove the spatial correlation in 73.5% of the data sets, which suggests that it is not appropriate for modelling strong spatial correlation. The model proposed by Cressie also performed

Table 1
Summary of the simulation study results. The bias and RMSE are presented as a percentage of the true values, while the fifth section summarises the percentage of data sets for which the model did not remove the spatial correlation. The bottom part of the table summarises the estimates of the spatial correlation parameter ρ.

Metric	Scenario	Model			
		Intrinsic	Convolution	Cressie	Leroux
Bias – β	1	−0.74	−0.41	−0.85	−0.52
	2	−0.08	0.84	1.65	1.41
	3	1.77	3.01	2.20	1.77
RMSE – β	1	18.0	15.6	14.8	14.5
	2	11.2	11.5	11.9	11.8
	3	12.8	16.3	14.5	13.1
Bias – R_k	1	−0.57	−0.56	−0.57	−0.57
	2	−0.26	−0.30	−0.30	−0.31
	3	−0.18	−0.37	−0.24	−0.20
RMSE – R_k	2	9.6	9.2	9.4	9.3
	3	6.5	9.2	7.2	6.7
% Spatial correlation	1	0	0	0	0
	2	0	0	0	0
	3	0	73.5	0	0
Mean value of ρ	1	–	–	0.195	0.042
	2	–	–	0.621	0.342
	3	–	–	0.950	0.948

less well in terms of RMSE than the intrinsic or Leroux models in the presence of strong spatial correlation, which suggests that one of the latter is the appropriate model in this scenario. In summary, the model proposed by Leroux appears to be the most appropriate across the three scenarios considered here, as it performs consistently well in the presence of independence and strong spatial correlation.

Finally, the mean value (over the 200 data sets) of the spatial correlation parameter ρ in the Cressie and Leroux models is displayed in the bottom part of Table 1. The table shows that on average both models produce estimates of approximately the appropriate size, with values close to zero under independence (Scenario 1) and close to one in the presence of strong spatial correlation (Scenario 3). However, under independence the true value of ρ should be zero, and the estimates from the Leroux model are much closer to this than the values from the Cressie model.

5. Application

This section presents a study mapping the spatial pattern of cancer risk in Greater Glasgow, Scotland, between 2001 and 2005.

5.1. Data description

The data for this study are publicly available, and come from the Scottish Neighbourhood Statistics database, which is available on-line at http://www.sns.gov.uk. The study region is the Greater Glasgow and Clyde health board, which contains the largest city in Scotland (Glasgow) as well as the surrounding area. The health board is split up into $n = 271$ administrative units called intermediate geographies (IG), which have a median area of 124 hectares and a median population of 4239.

5.1.1. Cancer data

In this study we map the spatial pattern of: (a) lung cancer cases; and (b) all cancer cases; which are classified by the International classification of disease – 10th revision (ICD-10) as (a) C33–C34 and (b) C00–C97. respectively. Our response variables are the numbers of new cases diagnosed between 2001 and 2005, for each of the 271 intermediate geographies that comprise our study region. The expected numbers of cases are calculated by external standardisation, using age and sex adjusted rates for the whole of Scotland, which were obtained from the information services division (ISD) of the National Health Service. Maps of the standardised incidence ratios for both cancer types are shown in Fig. 1, where the top panel relates to lung cancer while the bottom one displays all cancer cases. The figure shows that cancer incidence in Greater Glasgow is higher than in the rest of Scotland, with average SIRs across the study region of 1.186 (lung cancer) and 1.034 (all cancer), respectively. Within the study region the majority of the highest SIR values are in the east, which corresponds to the heavily deprived east end of Glasgow.

5.1.2. Covariates

A small number of covariates are available to describe the spatial variation in cancer risk across Greater Glasgow. The first is a modelled estimate of the percentage of the population in each IG who smoke, and further details about its construction are available from Whyte et al. (2007). A number of measures of socio-economic deprivation are also available, but the majority of these are highly correlated with the smoking covariate, and therefore should not be used as it would lead to collinearity problems. Therefore we represent deprivation by the natural log of the median house price in each area (correlation with smoking of −0.69), which is the measure of deprivation

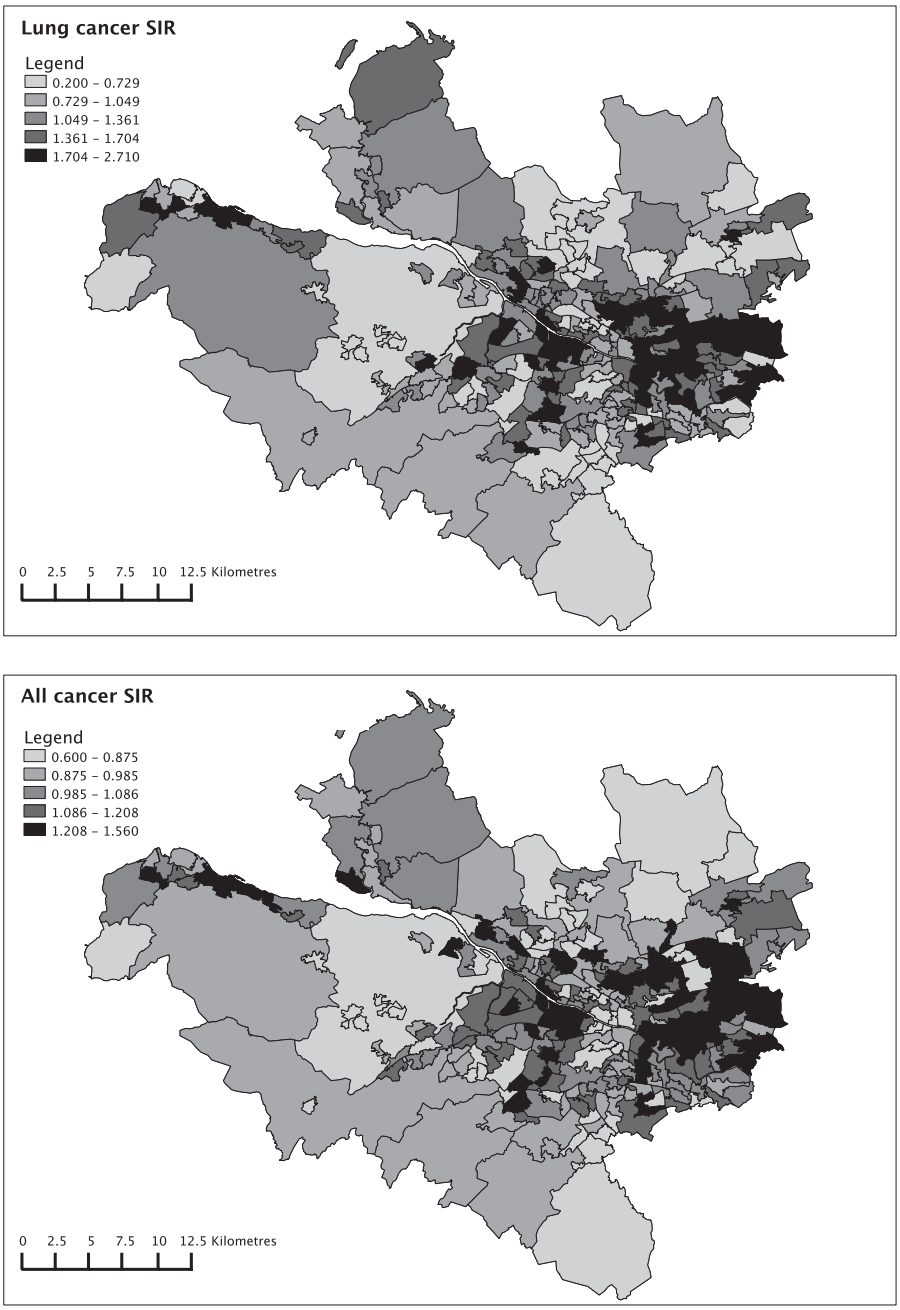

Fig. 1. Standardised incidence ratios (SIR) for lung cancer (top) and all cancer (bottom) in Greater Glasgow between 2001 and 2005.

available that is least correlated with smoking. In addition, the percentage of school children from ethnic minorities (i.e., non-white) is also available, which is used here as a proxy measure of the ethnic make-up of each area.

The final covariate is the estimated annual mean concentration of particulate matter air pollution in 2001, which is measured as PM_{10}. Particulate matter concentrations have been shown to be significantly or borderline significantly associated with lung cancer in previous studies (see, for example, Pope et al., 2002; Vineis et al., 2004; Jerrett et al., 2005), which is the reason for its inclusion here. However, modelled estimates are only available for each 1 km grid square across the UK, and not at the intermediate geography resolution. These gridded estimates can be obtained from the Department for Environment Food and Rural Affairs (DEFRA, http://laqm1.defra.gov.uk), and further details about their construction are available from Steadman et al. (2002). Here, we use the median concentration over the grid squares lying within each intermediate geography as our exposure measure.

5.2. Modelling

The cancer data sets are represented using the general model (1), where the covariates included are those described above. Both data sets are modelled using the four prior distributions for ϕ described in Section 3. Inference for all models is based on 50,000 MCMC samples generated from five Markov chains, that were initialised at dispersed locations in the sample space. Each chain is burnt in until convergence (40,000 iterations), and the next 10,000 samples are used for the analysis.

The residuals from each model were tested for the presence of spatial correlation, using a permutation test based on Moran's I statistic. For both cancer types all models adequately remove the spatial correlation present in the data, as all the corresponding p-values (not shown) were greater than 0.05. The deviance information criterion (DIC, Spiegelhalter et al., 2002) was also calculated in each case, and is a measure of how well a model fits a set of data. Low values of the DIC indicate a better fitting model, and the results across the four models are very similar. For both data sets the intrinsic model appears to be the worst fit, with DIC values of 1717 and 2205, respectively, for lung and all cancer cases. The remaining three models have similar DIC values, ranging between 1705 and 1708 for lung cancer and 2200 and 2201 for all cancer.

5.3. Results

The results of our study are presented below. The first part displays the level of spatial correlation estimated by the models, the second describes the covariate effects, while the third presents the fitted risk surfaces.

5.3.1. Spatial correlation

The posterior distributions of the spatial correlation parameter (ρ) are shown in Fig. 2, for both the Cressie and Leroux models and both cancer types. The figure shows that in all cases the data are informative about the value of ρ, as the posterior distributions are not similar to the uniform priors that were adopted. Secondly, the spatial correlation parameters from the two models do not have the same calibration, as their posterior distributions are different for both cancer types. Finally, the lung cancer data appear to contain weak spatial correlation after adjusting for the covariates (posterior median of 0.2 from the Leroux model), while the all cancer data contain mod-

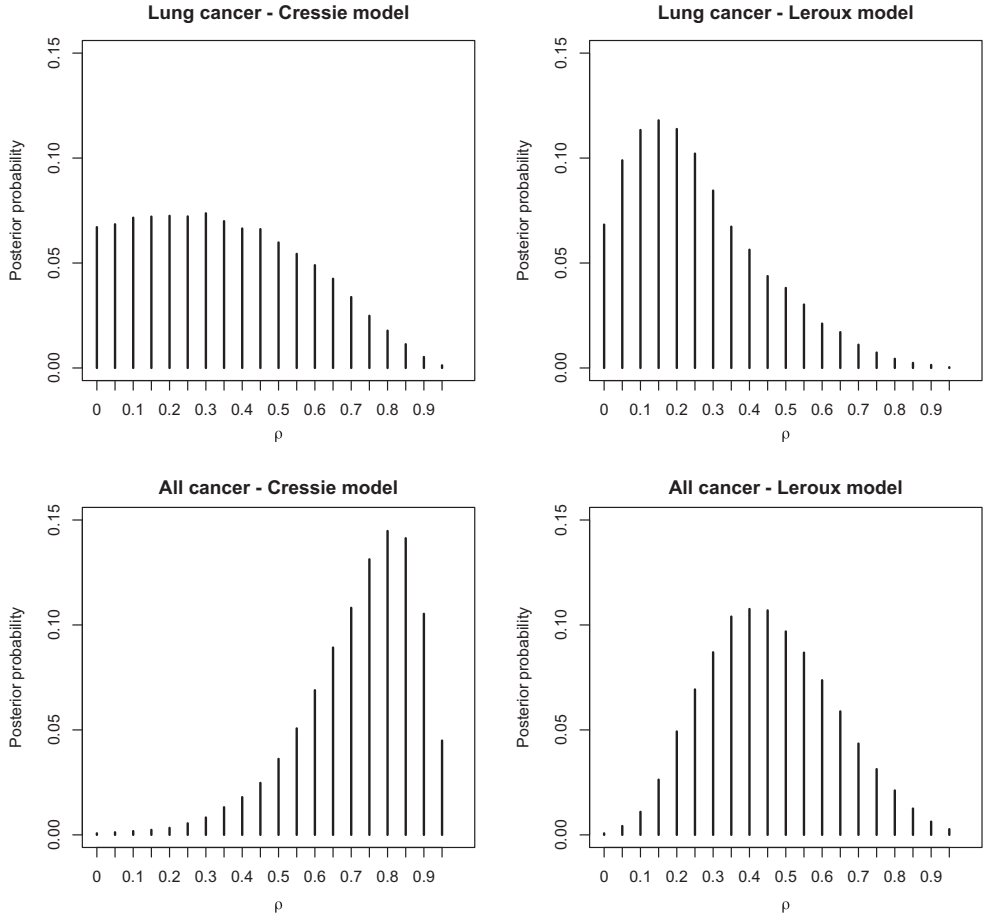

Fig. 2. Posterior distributions for the spatial correlation parameter from the models proposed by Cressie and Leroux.

Table 2
Estimates and 95% credible intervals for the regression parameters. The results are presented on the relative risk scale, for a standard deviation increase in each covariates value.

Covariate	Model			
	Intrinsic	Convolution	Cressie	Leroux
(a) Lung cancer				
Ethnicity	0.92 (0.88, 0.96)	0.93 (0.89, 0.97)	0.94 (0.90, 0.98)	0.94 (0.90, 0.98)
House price	0.92 (0.88, 0.96)	0.92 (0.87, 0.96)	0.91 (0.87, 0.95)	0.91 (0.87, 0.95)
PM_{10}	1.12 (1.02, 1.20)	1.12 (1.06, 1.19)	1.10 (1.06, 1.15)	1.10 (1.05, 1.15)
Smoking	1.27 (1.22, 1.34)	1.28 (1.22, 1.35)	1.29 (1.22, 1.35)	1.28 (1.22, 1.34)
(b) All cancer				
Ethnicity	0.95 (0.92, 0.07)	0.95 (0.93, 0.97)	0.95 (0.93, 0.97)	0.95 (0.93, 0.97)
House price	0.95 (0.93, 0.97)	0.95 (0.92, 0.97)	0.95 (0.92, 0.97)	0.95 (0.93, 0.97)
PM_{10}	1.05 (1.00, 1.09)	1.06 (1.02, 1.10)	1.06 (1.03, 1.09)	1.05 (1.02, 1.08)
Smoking	1.03 (1.00, 1.06)	1.03 (1.00, 1.06)	1.03 (1.01, 1.06)	1.03 (1.01, 1.06)

erate correlation (posterior median of 0.45 from the Leroux model).

5.3.2. Covariate effects

The effects of the covariates are presented in Table 2, which displays the estimates (posterior medians) as well as 95% credible intervals. All results are presented on the relative risk scale, for a one standard deviation increase in each covariates value. The table shows that the choice of CAR prior has little effect on the results, as both the relative risks and credible intervals are consistent across the four models. There is convincing evidence that areas with higher proportions of ethnic minorities are at less risk of developing cancer, with relative risks around 0.94 (lung) and 0.95 (all cancer) for a 12% increase in the non-white school population. Populations living in areas that are less deprived (as measured by the log of average house price) are also less at risk of cancer, with relative risks of 0.92 (lung) and 0.95 (all cancer), respectively.

Exposure to higher concentrations of particulate matter air pollution appears to be associated with increased cancer incidence, with relative risks of 1.10 (lung) and 1.05 (all cancer) for a 1.8 $\mu\,g^{-3}$ increase in the yearly average concentration. The association observed for lung cancer is consistent with the existing research described above, while the association for all cancer is much smaller. Finally, there is strong evidence that increasing the percentage of the population in each IG that smoke is related to increased lung cancer risk, with a relative risk around 1.28 for a 10% increase in the smoking prevalence. The results for all cancer are less strong, with a relative risk around 1.03. However, the smoking covariate is moderately correlated with house price (correlation of -0.69), so the models were re-run without the latter to observe whether there was any change in the estimated smoking effects. The observed relative risks increased to 1.36 (lung cancer) and 1.07 (all cancer), respectively, suggesting that adjusting for deprivation (as measured by house price) impacts on the magnitude of both associations.

5.3.3. Risk maps

Fig. 3 shows the estimated disease risks (i.e., posterior medians of R_k) from the Leroux model, where the scales are the same as those used for the SIRs in Fig. 1. The estimated risk maps are smoother than the raw SIR values, and are also less extreme. For example, the SIR for all cancer ranges between 0.61 and 1.55, while the corresponding model estimates range between 0.75 and 1.43. However, the estimated risk surface exhibits a similar spatial pattern to the SIR map, with the highest risks for both cancer types being observed in the heavily deprived east end of Glasgow in the east of the study region.

Fig. 4 summarises the uncertainty in the disease risks R_k, and splits the areas into three categories. Areas shaded in black exhibit elevated risks of cancer compared with the whole of Scotland, having 95% credible intervals for R_k that do not include the null risk of one. Areas shaded in mid-grey have credible intervals for R_k that include the null risk of one, while those coloured light grey have decreased risks (credible intervals that are less than one). The figure shows there are more areas that exhibit elevated risks than decreased risks, suggesting that Greater Glasgow has a higher risk of cancer compared with the Scottish average. The areas of elevated risk lie mainly in the east end of Glasgow and along the south bank of the main river (the Clyde), the latter of which is illustrated by the thin white line moving south east across the study region.

5.3.4. Sensitivity to hyperpriors

To assess the effects of the hyperpriors on posterior inference, a small sensitivity analysis was undertaken using the Leroux model. Firstly, the model was implemented with a Uniform$(0, M)$ prior distribution for τ (the standard deviation), where $M = 5, 10, 20$. The estimates of τ^2 were invariant to this change, with, for example, posterior medians ranging between 0.0214 and 0.0231 for all cancer cases. Secondly, the model was implemented with four different discrete prior distributions for the spatial correlation parameter ρ. As described in Section 2, a uniform prior with values ranging between 0 and 0.95 is adopted in this paper, where the possible values are equally spaced between the two endpoints. Here, we assess the sensitivity of the posterior distribution to having 10, 20, 30 and 40 possible values (i.e., $r = 10, 20, 30, 40$) between these endpoints. The posterior distributions of ρ exhibited almost identical shapes and centres for each value of r, with posterior medians for all cancer ranging between 0.422 and 0.463.

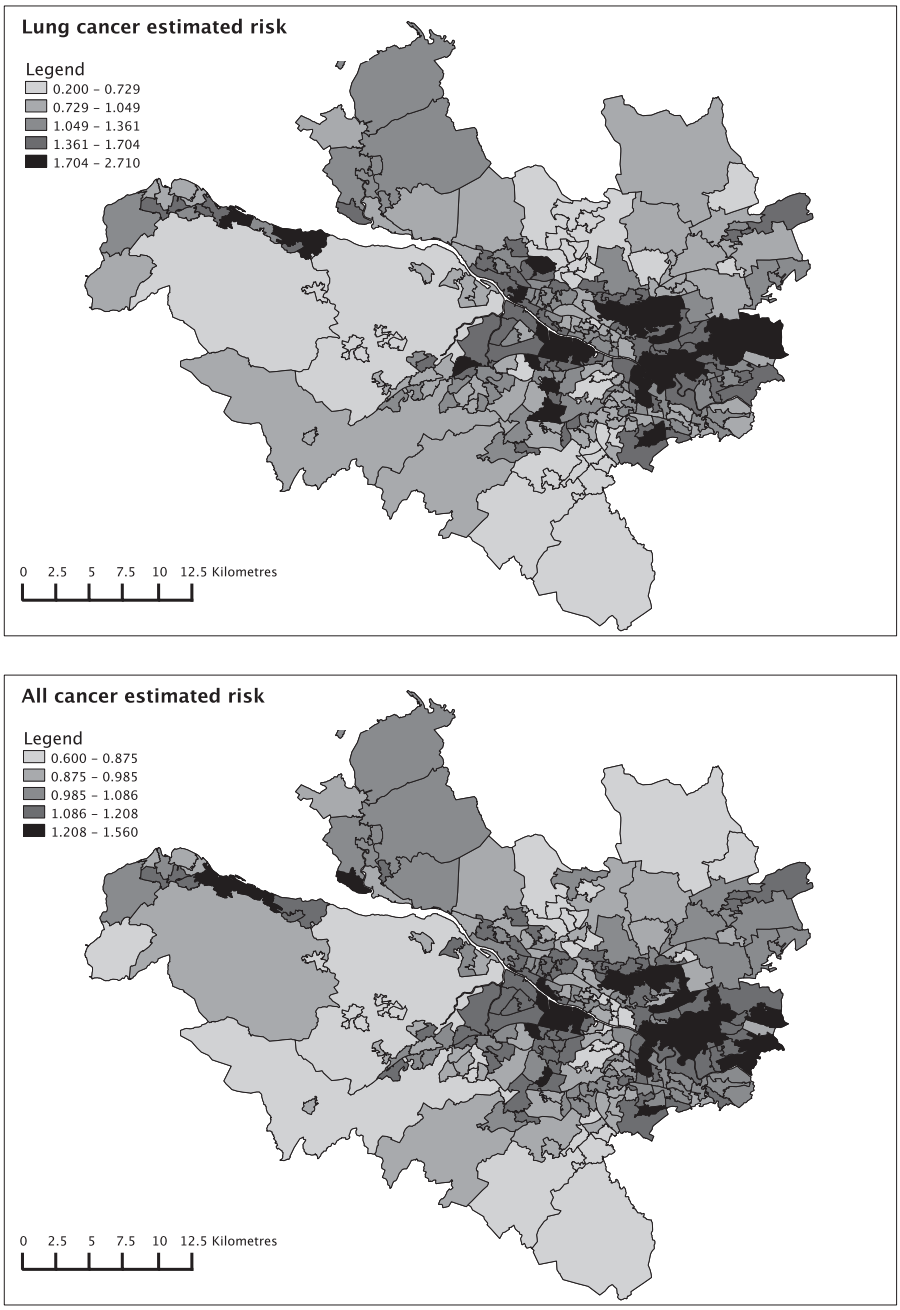

Fig. 3. Estimated disease risks from the Leroux model for lung cancer (top) and all cancer (bottom).

6. Discussion

This paper has critiqued four of the most commonly used conditional autoregressive prior distributions in Bayesian disease mapping, which include the intrinsic and convolution models (both Besag et al., 1991), as well as the alternatives proposed by Cressie (1993) and Leroux et al. (1999). The performance of these models has been quantified by simulation, specifically assessing the accuracy with which they can estimate regression parameters and disease risk surfaces. The paper then applies each of these models to a new study mapping cancer incidence in Greater Glasgow, Scotland, between 2001 and 2005.

The simulation study shows that all four prior models produce close to unbiased estimates of the regression parameters and the set of disease risks, regardless of the amount of spatial correlation in the data. However, there are differences in the corresponding root mean square errors of β and R_k. If the data do not contain spatial correla-

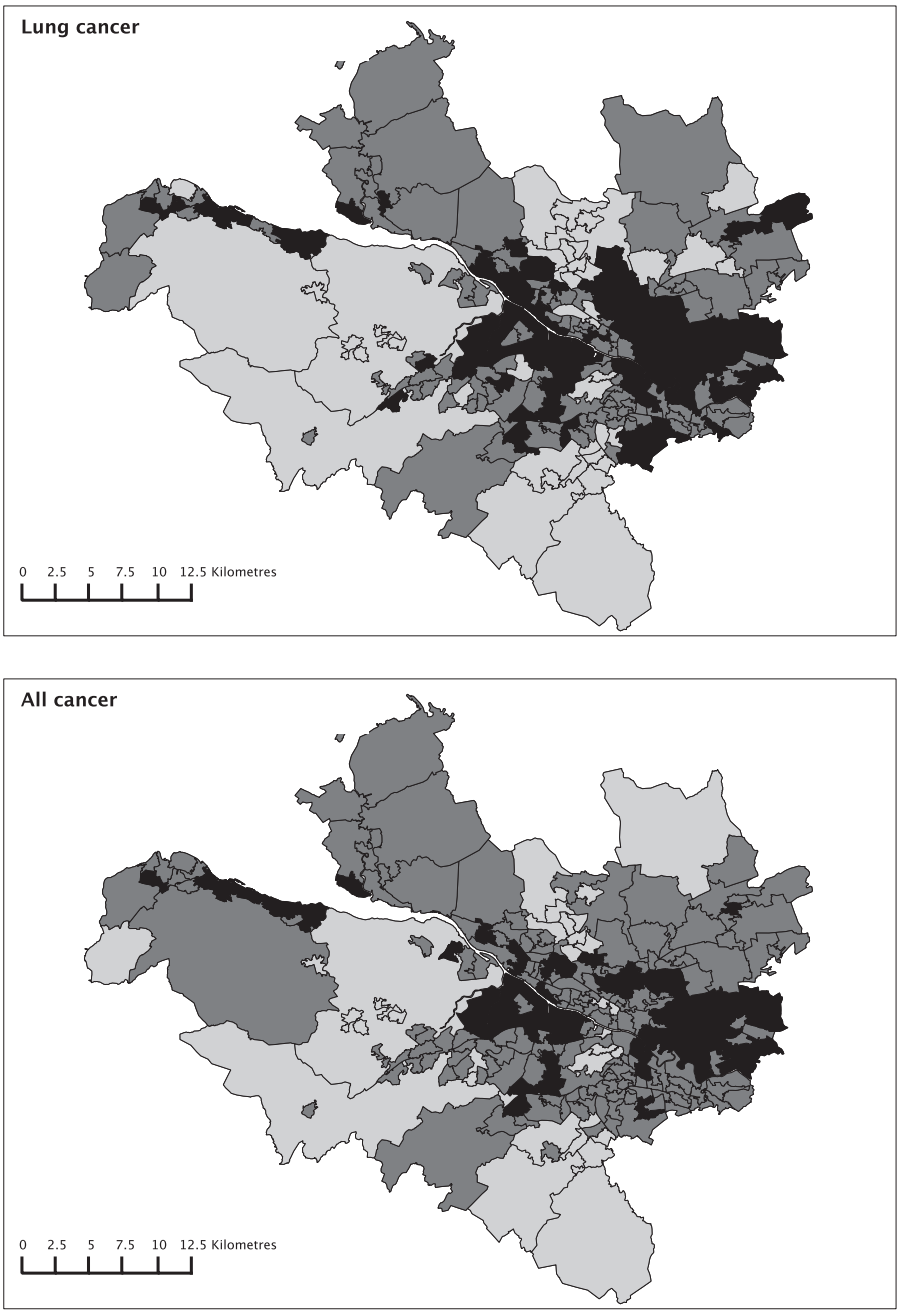

Fig. 4. Uncertainty in the estimated disease risks from the Leroux model for lung cancer (top) and all cancer (bottom). Areas shaded in black have elevated risks of disease (credible intervals for R_k that are greater than 1), areas shaded in mid grey have credible intervals that contain one, while areas in light grey have substantially decreased risks.

tion the intrinsic model has the largest RMSE, while in the presence of strong spatial correlation the convolution and Cressie models have the largest values. These results suggest that the model proposed by Leroux et al. (1999) is the best overall, because it produces consistently good results across the range of spatial correlation scenarios considered here. It is also the most theoretically appealing of the four prior models for a number of reasons. Firstly, it can represent a range of strong and weak spatial correlation structures with a single set of random effects, which is beyond the capability of the convolution and intrinsic models. Secondly, it corresponds to a proper joint distribution for the random effects, which is not true for the intrinsic model. Thirdly, its full conditional distributions (given by 7) have an appropriate mean and variance structure, regardless of whether the data are independent or contain

strong spatial correlation. This is not the case for the model proposed by Cressie (1993), which assumes the conditional variance is inversely proportional to the number of neighbouring areas, regardless of whether there is any spatial correlation in the data.

The study presented in Section 5 is based on ecological data, which means the results relate to the health of the overall population rather than applying directly to individuals. Overall, Greater Glasgow appears to have a higher risk of cancer than the Scottish average, with mean risks (from the Leroux model) of 1.24 (lung) and 1.05 (all) across the set of 271 intermediate geographies considered in this study. The south and east of the city of Glasgow have the highest risks of cancer, where as the more rural surrounding areas have much lower risks. These differences in risk appear to be partly due to the covariates. Increased levels of smoking, socio-economic deprivation and air pollution appear to inflate an areas risk of cancer, while increasing the proportion of ethnic minorities appears to decrease the risk.

All of the CAR priors considered here assume that if two areas share a common border their random effects will be correlated, which is unlikely to be true in all cases. This is especially true in urban areas, where neighbourhoods of rich and poor people are often geographically adjacent. Therefore future work will involve relaxing this assumption, and only forcing pairs of contiguous areas to have correlated random effects if those areas are in some way similar. Such similarity could be measured in numerous ways, including geographical distance or differences in their relative levels of socio-economic deprivation.

Acknowledgements

The author gratefully acknowledges the valuable comments and suggestions made by two referees, all of which have greatly improved the focus and presentation of this paper. The data and shapefiles used in this paper were provided by the Scottish Government.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.sste.2011.03.001.

References

Banerjee S, Carlin B, Gelfand A. Hierarchical modelling and analysis for spatial data. 1st ed. Chapman and Hall; 2004.
Besag J, Higdon D. Bayesian analysis of agricultural field experiments. J R Stat Soc Ser B 1999;61:691–746.
Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. Ann Inst Stat Math 1991;43:1–59.
Biggeri A, Dreassi E, Catelan D, Rinaldi L, Laglazio C, Cringoli G. Disease mapping in veterinary epidemiology: a Bayesian geostatistical approach. Stat Methods Med Res 2006;15:337–52.
Cressie N. Statistics for spatial data, revised ed.. New York: Wiley; 1993.
Eberly L, Carlin B. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Stat Med 2000;19:2279–94.
Elliott P, Wakefield J, Best N, Briggs D. Spatial epidemiology: methods and applications. 1st ed. Oxford University Press; 2000.
Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Anal 2006;1:515–33.
Jerrett M, Burnett R, Ma R, Pope C, Krewski D, Newbold K, et al. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 2005;16:727–36.
Kissling W, Carl G. Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecol Biogeogr 2008;17:59–71.
Lawson A. Bayesian disease mapping: hierarchical modelling in spatial epidemiology. 1st ed. Chapman and Hall; 2008.
Lee D, Ferguson C, Mitchell R. Air pollution and health in Scotland: a multi-city study. Biostatistics 2009;10:409–23.
Leroux B, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Halloran M, Berry D, editors. Statistical models in epidemiology, the environment and clinical trials. New York: Springer-Verlag; 1999. p. 135–78.
MacNab Y. Hierarchical Bayesian modelling of spatially correlated health service outcome and utilization rates. Biometrics 2003;59:305–16.
MacNab Y, Kmetic A, Gustafson P, Sheps S. An innovative application of Bayesian disease mapping methods to patient safety research: a Canadian adverse medical event study. Stat Med 2006;25:3960–80.
Molina R, Katsaggelos A, Mateos J. Bayesian regularization methods for hyperparameter estimation in image restoration. IEEE Trans Image Process 1999;8:231–46.
Pope C, Burnett R, Thun M, Calle E, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. J Am Med Assoc 2002;287:1132–41.
R Development Core Team, 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0. Availalbe from: <http://www.R-project.org>
Spiegelhalter D, Best N, Carlin B, Van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Ser B 2002;64:583–639.
Steadman J., Bush T., Vincent K. UK air quality modelling for annual reporting 2001 on ambient air quality assessment under Council Directives 96/62/EC and 1999/30/EC. Department for Environment Food and Rural Affairs; 2002.
Stern H, Cressie N. Posterior predictive model checks for disease mapping models. Stat Med 2000;19:2377–97.
Vineis P, Forastiere F, Hoek G, Lipsett M. Outdoor air pollution and lung cancer: recent epidemiologic evidence. Int J Cancer 2004;111:647–52.
Whyte B., Gordon D., Haw S., Fischbacher C., Harrison R. An atlas of tobacco smoking in Scotland: a report presenting estimated smoking prevalence and smoking attributable deaths within Scotland. NHS Health Scotland; 2007.