A Bayesian Localized Conditional Autoregressive Model for Estimating the Health Effects of Air Pollution

Authors
Duncan Lee, Alastair Rushworth, and Sujit K. Sahu

Presented by
Nayeon Kwon

December 7, 2022

Overview

Introduction

- The research goal is to estimate the long-term health effects of air pollution with spatial small-area disease incidence data.
- The challenge comes from the spatial autocorrelation structure in the data.
- The spatial autocorrelation structure is explained by random effects modeled by a globally smooth conditional autoregressive (CAR) model.
- The problem is these smooth random effects confound the effects of air pollution, which are also globally smooth.
- To avoid this collinearity, a Bayesian localized conditional autoregressive (LCAR) model is developed for the random effects.
- The LCAR model is not only able to model areas of spatial smoothness, but also it is able to capture step changes in the random effects surface unlike the currently available globally smooth models.

Motivating Study

- The study region: the city of Glasgow and the river Clyde estuary
- The region is partitioned into $n = 271$ administrative units (Intermediate Geographies; IG).
- The response variable: the number of admissions to hospitals in each IG in 2011 with a diagnosis of respiratory disease
  - Excluded admissions to psychiatric and obstetric hospitals
- Expected counts: external standardization was used based on age- and sex-specific respiratory disease rates.

Motivating Study (Cont'd)

Covariates

- Air pollutants: Concentrations of carbon monoxide (일산화탄소; CO), *nitrogen dioxide (이산화질소; NO2), sulfur dioxide (이산화황; SO2), *PM10 (미세먼지), and *PM2.5 (초미세먼지)
- Others: a measure of socio-economic deprivation, *the percentage of people living in each IG who receive job seekers allowance (JSA), measures of ethnicity, access to alternative forms of health care, and a measure of urbanicity
- Used covariates that are measured in 2010

*Covariates that showed a significant relationship with the response variable

Motivating Study (Cont'd)

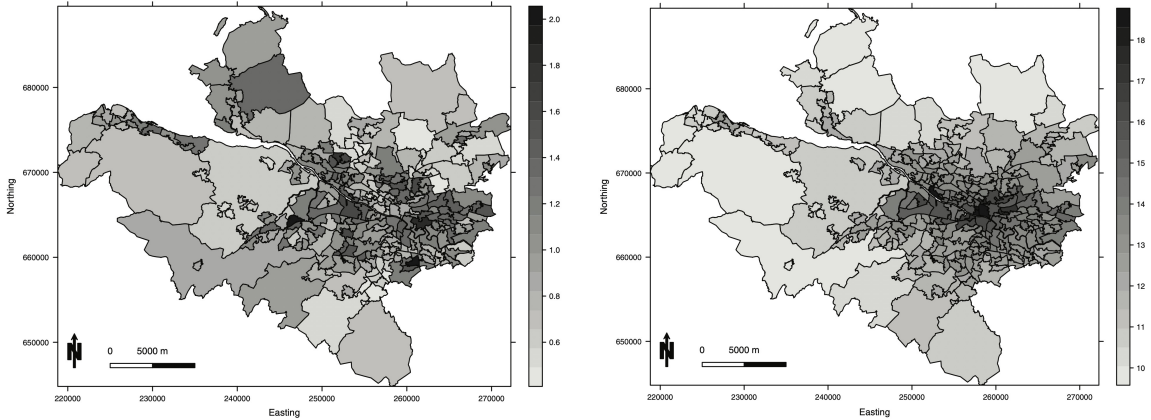

Figure: Maps displaying the spatial pattern in the standardized incidence ratio for respiratory disease in 2011 (left panel) and the modeled yearly average concentration of PM10 in 2010 (right panel).

Modeling

Notations

- $A = \{A_1, \ldots, A_n\}$: the study region partitioned into n areal units
- $Y = (Y_1, \ldots, Y_n)$: the vector of observed number of disease cases
- $E = (E_1, \ldots, E_n)$: the vector of expected number of disease cases
- $X = (X_1^T, \ldots, X_n^T)^T$: the matrix of p covariates and a column of ones for the intercept term, where the values relating to areal unit A_k are denoted by $X_k^T = (1, x_{k1}, \ldots, x_{kp})$

A Bayesian hierarchical model

$$\begin{aligned} Y_k | E_k, R_k &\sim \text{Poisson}(E_k R_k) \text{ for } k = 1, \ldots, n, \\ \ln(R_k) &= x_k^T \beta + \phi_k. \end{aligned} \tag{1}$$

where the disease counts are assumed to be conditionally independent given the covariates and the randon effects.
Here, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ denotes the vector of covariate effects, and R_k represents diesease risk in areal unit A_k.

Modeling (Cont'd)

- The random effects $\boldsymbol{\phi} = (\phi_1, ..., \phi_n)$ capture any residual spatial autocorrelation present in the disease data, and are typically assigned a CAR prior.
- Models within the CAR class are typically specified as a set of n univariate full conditional distributions $f(\phi_k|\boldsymbol{\phi}_{-k})$ for $k = 1, ..., n$, where $\boldsymbol{\phi}_{-k} = (\phi_1, ..., \phi_{k-1}, \phi_{k+1}, ..., \phi_n)$.
- The adjacency information comes from a binary $n \times n$ neighborhood matrix W, where $w_{ki} = \begin{cases} 1 & \text{if } k \sim i \\ 0 & \text{if } k \not\sim i \end{cases}$.
- The intrinsic model (Besag et al., 1991; IAR) is the simplest prior in the CAR class, and its full conditional distributions are given by

$$\phi_k|\phi_{-k}, \tau^2, \mathrm{W} \sim N\left(\frac{\sum_{i=1}^n w_{ki}\phi_i}{\sum_{i=1}^n w_{ki}}, \frac{\tau^2}{\sum_{i=1}^n w_{ki}}\right). \tag{2}$$

Modeling (Cont'd)

- Modeling ϕ as $\phi \sim N(0, \mathrm{Q(W)}/\tau^2)$, where $\mathrm{Q(W)} = \mathrm{diag(W1)} - \mathrm{W}$, and 1 is an n dimensional vector of ones.
- This prior is appropriate if the residuals from the covariate component of the model, that is $\ln(\mathrm{Y/E}) - \mathrm{X}\boldsymbol{\beta}$, are spatially smooth across the entire region, because the partial autocorrelation between (ϕ_k, ϕ_j) conditional on the remaining random effects ϕ_{-kj} is

$$\mathrm{Corr}\left[\phi_k, \phi_j | \phi_{-kj}, \mathrm{W}\right] = \frac{w_{kj}}{\sqrt{\left(\sum_{i=1}^{n} w_{ki}\right)\left(\sum_{i=1}^{n} w_{ji}\right)}}. \tag{3}$$

- The above equation shows that all pairs of random effects relating to geographically adjacent areal units are partially autocorrelated ($w_{kj} = 1$), which smoothes the random effects across geographical borders.

Modeling (Cont'd)

- To account for localized spatial smoothing, we could treat $W = \left\{ w_{kj} | k \sim j, k > j \right\}$ as a set of binary random quantities.
- The neighborhood matrix W is always assumed to be symmetric so that changing w_{kj} also changes w_{jk}, while the other elements in W relating to non-neighboring areal units remain fixed at zero.
- Equation(3) shows that this allows (ϕ_k, ϕ_j) corresponding to adjacent areal units to be conditionally independent or autocorrelated.
- However, full estimation of W as a set of separate unknown parameters results in a highly overparameterized precision matrix for ϕ.

LCAR

- The LCAR prior treats the elements in W relating to contiguous areal units as a set of binary random quantities. This model comprises a joint distribution for an extended set of random effects $\tilde{\phi}$ and the set of edges W.
  - Hereafter, we refer to W as the set of *edges*, and define any edge $w_{kj} \in W$ that is estimated as zero as being removed.
  - We assume that W is NOT fixed.
- We decompose the joint prior distribution as $f(\tilde{\phi}, W) = f(\tilde{\phi}|W)f(W)$.

LCAR: Prior Distribution - $f(\tilde{\phi}|W)$

- We consider an extended vector of random effects $\tilde{\phi} = (\phi, \phi_*)$, where ϕ_* is a global random effect that prevents any unit from having no edges (ref eqn.2).
- The extended $(n+1) \times (n+1)$ dimensional neighborhood matrix corresponding to $\tilde{\phi}$ is given by

$$\tilde{W} = \begin{bmatrix} W & w_* \\ w_*^\top & 0 \end{bmatrix}. \tag{4}$$

where $w_* = (w_{1*}, ..., w_{n*})$ and $w_{k*} = \mathbb{1}\left[\sum_{i \sim k}(1 - w_{ki} > 0)\right]$.

- Based on $\tilde{W}$, we propose modeling $\tilde{\phi}$ as $\tilde{\phi} \sim N(0, \tau^2 Q(\tilde{W}, \epsilon)^{-1})$, where the precision matrix is given by

$$Q\left(\tilde{W}, \epsilon\right) = \mathrm{diag}\left(\tilde{W}1\right) - \tilde{W} + \epsilon I. \tag{5}$$

The addition of ϵI ensures the precision matrix is diagonally dominant and hence invertible.

LCAR: Prior Distribution - $f(\tilde{\phi}|W)$ (Cont'd)

- The full conditional distributions corresponding to the LCAR model are given by

$$\begin{aligned} \phi_k|\tilde{\phi}_{-k}, \tau^2, \mathrm{W} &\sim N\left(\frac{\sum_{i=1}^n w_{ki}\phi_i + w_{k*}\phi_*}{\sum_{i=1}^n w_{ki} + w_{k*} + \epsilon}, \frac{\tau^2}{\sum_{i=1}^n w_{ki} + w_{k*} + \epsilon}\right) \; k = 1, ..., n, \\ \phi_*|\tilde{\phi}_{-*} &\sim N\left(\frac{\sum_{i=1}^n w_{i*}\phi_i}{\sum_{i=1}^n w_{i*} + \epsilon}, \frac{\tau^2}{\sum_{i=1}^n w_{i*} + \epsilon}\right). \end{aligned} \tag{6}$$

- The shows that ϕ_* influences the conditional expectation of any other random effect that corresponds to an areal unit with at least one edge removed.
- All edges being retained in the model: IAR model for global spatial smoothing
- If all edges are removed the random effects are independent with a constant mean and variance.

LCAR: Prior Distribution - $f(W)$

- The dimensionality of W is $Nw = 1^{\top}W1/2$, and as each edge is binary the sample size has size 2^{Nw}.
- The simplest approach would be to assign each edge an independent Bernoulli prior, but this is likely to result in W being weakly identifiable.
- Therefore we treat W as a single random quantity, and propose the following discrete uniform prior for its neighborhood matrix representation $\tilde{W}$.

$$\tilde{W} \sim \text{discrete uniform}\left(\tilde{W}^{(0)}, \tilde{W}^{(1)}, ..., \tilde{W}^{(Nw)}\right). \tag{7}$$

- $\tilde{W}^{(Nw)}$ retains all Nw edges in the model ($w_{kj} = 1 \; \forall \, w_{kj} \in W$).
- Moving from $\tilde{W}^{(j)}$ to $\tilde{W}^{(j-1)}$ removes an edge from W (one additional $w_{kj} = w_{jk} = 0$ $\because$ the neighborhood matrix is assumed to be symmetric).
- $\tilde{W}^{(0)}$ contains no edges.
- This restriction reduces the sample space of W to being one-dimensional.

LCAR: Prior Distribution - $f(W)$ (Cont'd)

- We propose eliciting the set of candidate values $\left(\tilde{\mathrm{W}}^{(0)}, \tilde{\mathrm{W}}^{(1)}, ..., \tilde{\mathrm{W}}^{(Nw)}\right)$ from disease data prior to the study period.
- Let $\left((\mathrm{Y}_1^p, \mathrm{E}_1^p), ..., (\mathrm{Y}_r^p, \mathrm{E}_r^p)\right)$ denote these vectors of observed and expected disease counts for the r time periods prior to the study period.
- The general likelihood model (1) gives the vector of expectations for the study data as $\mathbb{E}\left[\mathrm{Y}\right] = \mathrm{E}\,\exp(\mathrm{X}\boldsymbol{\beta} \times \boldsymbol{\phi})$, which is equivalent to $\ln(\mathbb{E}\left[\mathrm{Y}\right]/\mathrm{E}) = \mathrm{X}\boldsymbol{\beta} \times \boldsymbol{\phi}$.
- Then as $\boldsymbol{\phi} \sim N(0, \tau^2 \mathrm{Q}(\tilde{\mathrm{W}}, \epsilon)^{-1}_{1:n})$, we make the approximation

$$\phi_j^p = \ln\left[\frac{\mathrm{Y}_j^p}{\mathrm{E}_j^p}\right] \approx \ln\left[\frac{\mathrm{Y}}{\mathrm{E}}\right] \sim \text{approx } N(\mathrm{X}\boldsymbol{\beta}, \tau^2 \mathrm{Q}(\tilde{\mathrm{W}}, \epsilon)^{-1}_{1:n}) \quad \text{for } j = 1, ..., r. \tag{8}$$

- Based on this approximation, the prior elicitation takes the form of an iterative algorithm (begins at $\tilde{\mathrm{W}}^{(Nw)}$ and algorithm continues till it reaches $\tilde{\mathrm{W}}^{(0)}$).

LCAR: Prior Distribution - $f(W)$ (Cont'd)

- The algorithm moves from $\tilde{\mathrm{W}}^{(j)}$ to $\tilde{\mathrm{W}}^{(j-1)}$ by computing the joint approximate Gaussian log-likelihood for $(\phi_1^p, ..., \phi_r^p)$ based on (8).
- This is given by

$$\begin{aligned}\ln\left[f(\phi_1^p, ..., \phi_r^p|\tilde{\mathrm{W}}^{(*)})\right] &= \sum_{j=1}^{r} \ln\left[N(\phi_j^p|\mathrm{X}\hat{\boldsymbol{\beta}}, \hat{\tau}^2\mathrm{Q}(\tilde{\mathrm{W}}^{(*)}, \epsilon)_{1:n}^{-1})\right], \\ &\approx \frac{r}{2}\ln(\left|\mathrm{Q}(\tilde{\mathrm{W}}^{(*)}, \epsilon)_{1:n}\right|) - \frac{nr}{2}\ln(\hat{\tau}^2) \\ &\quad - \frac{1}{2\hat{\tau}^2}\sum_{j=1}^{r}(\phi_j^p - \mathrm{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\mathrm{Q}(\tilde{\mathrm{W}}^{(*)}, \epsilon)_{1:n} \\ &\quad \times (\phi_j^p - \mathrm{X}\hat{\boldsymbol{\beta}}),\end{aligned} \tag{9}$$

where the constant in the likelihood function has been removed.

LCAR: Prior Distribution - $f(W)$ (Cont'd)

- This prior elicitation approach removes edges from W in sequence conditional on the current value of W.
- However, this approach requires (9) to be evaluated $Nw(Nw+1)/2$ times, which makes the approach computationally intensive.
- This computational burden is reduced by estimating $(\hat{\boldsymbol{\beta}}, \hat{\tau}^2)$ by maximum likelihood based on $\tilde{\mathrm{W}}^{(j)}$, $\hat{\boldsymbol{\beta}} = (\mathrm{X}^{\mathrm{T}}\mathrm{Q}(\tilde{\mathrm{W}}^{(j)}, \epsilon)_{1:n}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Q}(\tilde{\mathrm{W}}^{(j)}, \epsilon)_{1:n}((1/n)\sum_{j=1}^{r}\boldsymbol{\phi}_j^p)$, and $\hat{\tau}^2 = (1/nr)\sum_{j=1}^{r}(\boldsymbol{\phi}_j^p - \mathrm{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\mathrm{Q}(\tilde{\mathrm{W}}^{(j)}, \epsilon)_{1:n}(\boldsymbol{\phi}_j^p - \mathrm{X}\hat{\boldsymbol{\beta}})$.

LCAR: Overal Model

- The Bayesian hierarchical model proposed here combines the likelihood (1) with the priors (6) and (7) and is given by

$$
\begin{aligned}
Y_k|E_k, R_k &\sim \text{Poisson}(E_k R_k) \quad \text{for } k = 1, \ldots, n, \\
\ln(R_k) &= x_k^\top \beta + \phi_k, \\
\tilde{\phi} &\sim N(0, \tau^2 Q(\tilde{W}, \epsilon = 0.001)^{-1}), \\
\tilde{W} &\sim \text{discrete uniform}\left(\tilde{W}^{(0)}, \tilde{W}^{(1)}, ..., \tilde{W}^{(Nw)}\right), \\
\beta_j &\sim N(0, 1000) \quad \text{for } j = 1, ..., p, \\
\tau^2 &\sim \text{uniform}(0, 1000).
\end{aligned}
\tag{10}
$$

- Diffuse prior are specified for the regression parameters β and the variance parameter τ^2, while $\epsilon = 0.001$.

LCAR: Overal Model (Cont'd)

- Inference for this model is based on MCMC simulation, using a combination of Metropolis–Hastings and Gibbs sampling steps.
- The spatial structure matrix $\tilde{W}$ is updated using a Metropolis–Hastings step, where if the current value in the Markov chain is $\tilde{W}^{(j)}$, then a new value is proposed uniformly from the set $(\tilde{W}^{(j-q)}, ..., \tilde{W}^{(j-1)}, \tilde{W}^{(j+1)}, ..., \tilde{W}^{(j+q)})$.
  - Here q is a tuning parameter, which controls the mixing and acceptance rates of the update.

Summary of the Models

- BYM (Besag et al, 1991): global spatial smoothing
- LCAR & LM (Lee and Mitchell, 2013): local spatial smoothing
- HH (Hughes and Haran, 2013): smoothing orthogonal to the covariates

Simulation Study: Data Generation and Study Design

- Simulated data are generated for the 271 IGs.
- Disease counts are generated from model (1), where the size of the expected numbers E is varied to access its impact on model performance.
- The pollution covariate is generated as the average of two Gaussian spatial processes with different ranges, one of which has the same range and is confounded with the localized spatial autocorrelation.
- Both spatial processes are generated using the Matérn family of correlation functions, where the smoothness parameter equals 2.5.
- The regression coefficient for the covariate is fixed at $\beta = 0.1$
- The residual autocorrelation is also generated from a Gaussian process with a Matérn family of correlation function, where localized spatial structure is induced via a piecewise constant mean.
  - The piecewise constant mean has three distinct values $\{-1, 0, 1\}$.
  - These values are multiplied by a constant M to obtain the expectation.

Simulation Study: Data Generation and Study Design (Cont'd)

- The study is split into nine different scenarios comprising pairwise combinations of $M = 0.5, 1, 1.5$ and $E_k \in [10, 25], [50, 100], [150, 200]$.
  - The size of E quantifies disease prevalence.
  - M determines the extent of local rather than global residual autocorrelation.
- Each simulated data set consists of study data and three years of prior data, which is the number of prior data sets used in the Glasgow motivating study.
- The residual spatial autocorrelation for the prior data is generated by adding uniform random noise in the range $\left[-0.1, 0.1\right]$ to the realization generated for the real data.

Simulation Study: Data Generation and Study Design (Cont'd)

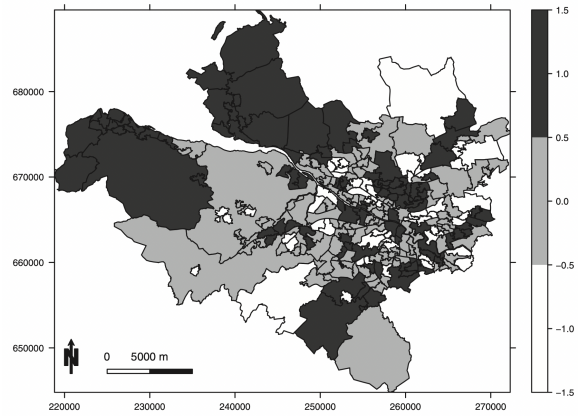

Figure: A map showing the piecewise constant mean function (with possible values $\{-1, 0, 1\}$) for the random effects that generate localized spatial correlation in the simulation study.

Simulation Results

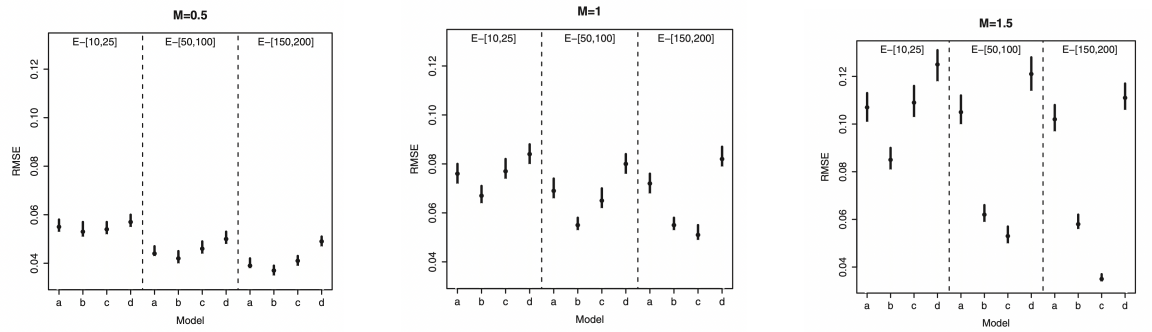

Figure: RMSE for the estimated regression parameter β. In each case, the dot represents the estimated RMSE while the black bars are bootstrapped 95% uncertainty intervals. The models are: (a) BYM, (b) LCAR, (c) LM, and (d) HH.

Simulation Results (Cont'd)

Table 1

Percentage coverages and average widths (in brackets) for the 95% credible intervals for the estimated regression parameter β. Here LM and HH refer to the models proposed by Lee and Mitchell (2013) and Hughes and Haran (2013)

		Model			
E	M	**BYM**	**LCAR**	**LM**	**HH**
[10, 25]	0.5	94.2 (0.204)	92.2 (0.193)	92.8 (0.197)	73.8 (0.131)
	1	94.2 (0.290)	92.8 (0.248)	91.0 (0.266)	53.0 (0.128)
	1.5	94.4 (0.392)	93.0 (0.298)	80.0 (0.284)	32.8 (0.122)
[50, 100]	0.5	92.6 (0.158)	90.2 (0.134)	86.6 (0.139)	46.2 (0.065)
	1	94.0 (0.257)	89.8 (0.184)	73.8 (0.148)	28.0 (0.063)
	1.5	90.8 (0.365)	92.8 (0.236)	79.0 (0.134)	20.4 (0.060)
[150, 200]	0.5	94.2 (0.147)	89.6 (0.113)	78.2 (0.099)	31.4 (0.042)
	1	90.2 (0.248)	85.8 (0.165)	67.0 (0.098)	18.0 (0.041)
	1.5	92.4 (0.353)	93.0 (0.218)	81.4 (0.087)	12.6 (0.040)

The Glasgow Study: Modeling

1. A simple Poisson log-linear model including the four non-pollution covariates was fitted to the data, and only JSA exhibited a significant relationship with respiratory disease risk.
2. The remaining three covariates were removed from the model.
3. Each of the pollutants was included in separate models due to their collinearity.
4. Random effects were added to the model, and we implement the four models (BYM, LCAR, LM, and HH).
5. For the LCAR model, the prior elicitation was based on respiratory disease data from 2008 to 2010.

The Glasgow Study: Model Goodness of Fit

- Deviance information criterion (DIC)
  - DIC (Spiegelhalter et al., 2002) is useful for comparing the predictive accuracy between models.
  - Smaller values indicate a better fitting model.
- Moran's I
  - As values of Moran's I close to 0 indicate very low or no spatial autocorrelation.
  - The closer Moran's I is to zero, the better the model accounts for spatial autocorrelation (Anderson & Ryan, 2017).

Results from the Glasgow Study

Table 2

A summary of the overall fit of each model (top panel) and the estimated covariate effects (bottom panel)

	Model			
	BYM	**LCAR**	**LM**	**HH**
DIC	2124.0 (178.5)	2112.4 (173.4)	2115.8 (167.7)	2467.6 (157.1)
Moran's I	−0.025 (0.7078)	−0.082 (0.9834)	−0.121 (0.9997)	−0.089 (0.9909)
JSA	1.304 (1.268, 1.342)	1.283 (1.247, 1.320)	1.306 (1.272, 1.341)	1.318 (1.300, 1.336)
CO	0.997 (0.954, 1.038)	1.011 (0.973, 1.045)	0.998 (0.959, 1.036)	1.021 (1.006, 1.035)
NO_2	1.036 (0.998, 1.072)	1.040 (1.012, 1.067)	1.033 (1.003, 1.065)	1.043 (1.028, 1.059)
$PM_{2.5}$	1.029 (0.991, 1.067)	1.039 (1.007, 1.071)	1.026 (0.989, 1.063)	1.035 (1.021, 1.050)
PM_{10}	1.032 (0.994, 1.071)	1.040 (1.007, 1.073)	1.028 (0.993, 1.064)	1.034 (1.021, 1.048)
SO_2	1.009 (0.980, 1.040)	1.016 (0.989, 1.044)	1.010 (0.983, 1.037)	1.010 (0.998, 1.024)

Results from the Glasgow Study (Cont'd)

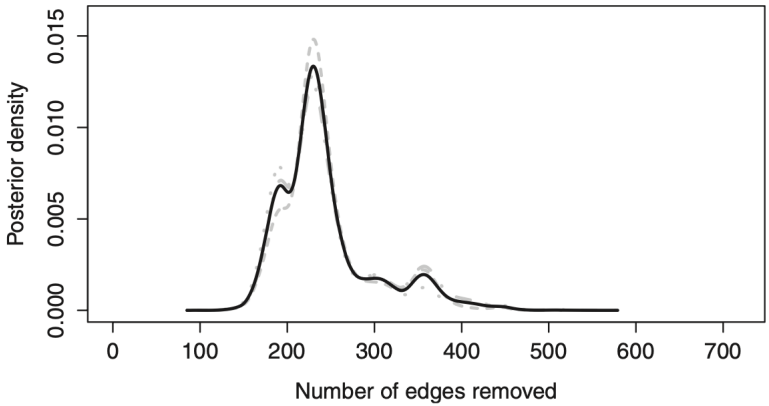

Figure: Posterior density for the number of edges removed from the model.

Conclusion

- The LCAR prior is flexible enough to capture either spatial smoothness or a distinct step change in the data between adjacent areal units.
- Inappropriate control for residual spatial autocorrelation can greatly retard fixed effects estimation, then its careful modeling is vital.

References

- Lee, D., Rushworth, A., & Sahu, S. K. (2014). A Bayesian localized conditional autoregressive model for estimating the health effects of air pollution. *Biometrics*, 70(2), 419-429.
- Lee, D. (2017). Carbayes version 4.6: An r package for spatial areal unit modelling with conditional autoregressive priors. Glasgow: University of Glasgow.