Gaussian Markov Random Field models for capturing localized residual spatio-temporal confounding in air pollution and health studies

Alastair Rushworth

University of Glasgow

International Conference of the Royal Statistical Society, Sheffield

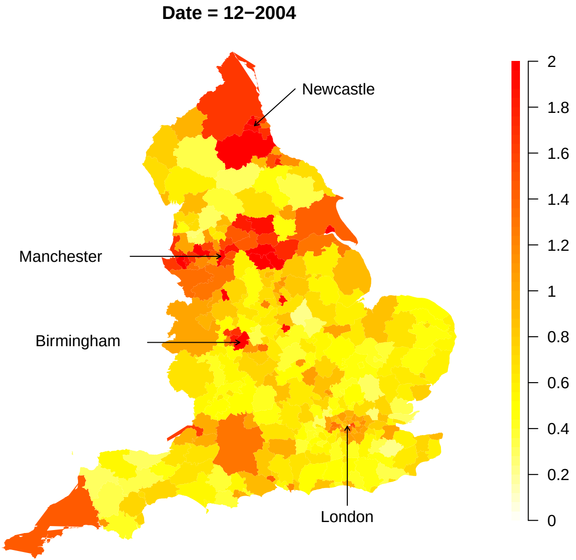
2[nd] September 2014

Acknowledgements

Work in this talk part of a larger collaboration:

University of Glasgow	Duncan Lee Richard Mitchell
University of Southampton	Sujit Sahu Sabyasachi Mukhopadhyay
UK Met Office	Paul Agnew Christophe Sarran
Funded by the EPSRC	EPSRC Engineering and Physical Sciences Research Council

Context: England LHAs - Respiratory admission SIR



- 123 months
- 323 Local Health Authorities (LHAs)
- January 2001 to December 2011
- 42636 data points in total

Modelling disease counts in space

Typical spatial model might look like

$$
\begin{aligned}
Y_i | E_i, R_i &\sim \mathrm{Poisson}(E_i R_i) \\
\ln(R_i) &= \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \phi_i, \qquad i = 1, \ldots, N
\end{aligned}
$$

Where

Y_i	disease counts	R_i	disease risk
E_i	expected cases	$\sum_{j=1}^{p} x_{ij}\beta_j$	covariates

$$
\pi(\phi) \propto \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i\sim j} w_{ij}(\phi_i - \phi_j)^2\right)\right]
$$

Localised smoothing

Different strategies achieving this:

Treat ϕ's prior variance, σ, as spatially varying

- Brewer and Nolan (2007); Reich and Hodges (2008)

Treat the weights w_{ij} as random variables

- Ma, Carlin and Banerjee (2010); Lee and Mitchell (2013)

Use a clustering or grouping prior for the random effects

- Richardson and Green (2002); Knorr-Held and Raßer (2000)

CAR2 model

Ma, Carlin and Banerjee (2010)

Starting with Poisson model and Intrinsic CAR for random effects:

$$
\begin{aligned}
Y_i|E_i, R_i &\sim Poisson(E_i R_i) \\
\log(R_i) &= \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \phi_i, \qquad i = 1, \ldots, N \\
p(\phi|\sigma, \{w_{ij}\}) &= C(\sigma, \{w_{ij}\}) \exp\left(-\frac{1}{2\sigma^2}\sum_{i\sim j} w_{ij}(\phi_i - \phi_j)^2\right)
\end{aligned}
$$

Normalising term $C(\sigma, \{w_{ij}\})$ included to emphasise that w_{ij} is now being treated as unknown.

CAR2 model

Ma, Carlin and Banerjee (2010)

Idea: Call the non-zero w_{ij}, w_{ij}^{+}, and treat as Bernoulli random variables with unknown probabilities, p_{ij}.

Transform and smooth p_{ij} using a further CAR prior.

$$
\begin{aligned}
w_{ij}^{+}|p_{ij} &\sim \mathit{Bernoulli}(p_{ij}) \text{ and } \mathrm{logit}(p_{ij}) = \mathbf{z}_{ij}^{'}\gamma + \theta_{ij} \\
p(\boldsymbol{\theta}|\zeta) &\propto \exp\left(-\frac{1}{2\zeta^2}\sum_{ij\sim kl}(\theta_{ij}-\theta_{kl})^2\right)
\end{aligned}
$$

Pros: Maintains binary nature of w_{ij}

Cons: Assumes relatively strong smoothness over the w_{ij}; a lot of parameters to estimate $\phi, w_{ij}, \boldsymbol{\theta}$ that can be hard to identify.

A localised model in space and time

Poisson model for counts in space and time (indexed by i and j, respectively)

$$
\begin{aligned}
Y_{ij}|E_{ij}, R_{ij} &\sim \text{Poisson}(E_{ij}R_{ij}) \\
\log(R_{ij}) &= \beta_0 + \phi_{ij}, \quad i = 1, \ldots, N \text{ and } j = 1, \ldots, T
\end{aligned}
$$

Letting $\tilde{\phi}_t = (\phi_{1t}, \ldots, \phi_{Nt})$, where $t = 1, \ldots, T$.

$$
f(\tilde{\phi}_1, \ldots, \tilde{\phi}_T) = f(\tilde{\phi}_1) \prod_{t=2}^{T} f(\tilde{\phi}_t|\tilde{\phi}_{t-1}),
$$

where

$$
\begin{aligned}
\tilde{\phi}_1 &\sim \text{N}\left(\mathbf{0}, \sigma^2 Q(W, \epsilon)\right)^{-1} \\
\tilde{\phi}_t|\tilde{\phi}_{t-1} &\sim \text{N}\left(\alpha\tilde{\phi}_{t-1}, \sigma^2 Q(W, \epsilon)^{-1}\right).
\end{aligned}
$$

A localised model in space and time

Idea: Treat the non-zero elements of w_{ij}, w_{ij}^{+}, as random variables on $[0, 1]$.

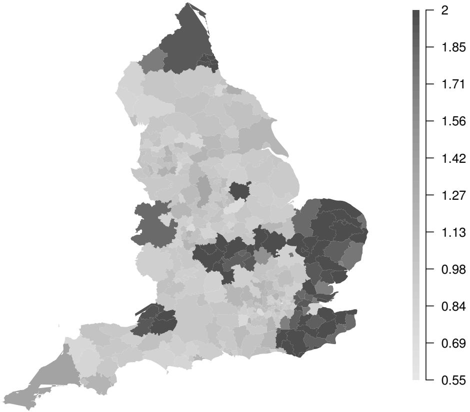
Then transform using $logit(w_{ij}^{+}) = z_{ij}$ and smooth these using a Leroux prior.

$$p(z_{ij}) \propto \exp\left[-\frac{1}{2\tau^2}\left(\rho\sum_{ij\sim kl}(z_{ij}-z_{kl})^2 + (1-\rho)\sum_{ij}(z_{ij})^2\right)\right]$$

Pros: Treating w_{ij} as non-binary makes things slightly easier; Leroux prior is a bit more flexible than the Intrinsic CAR model.

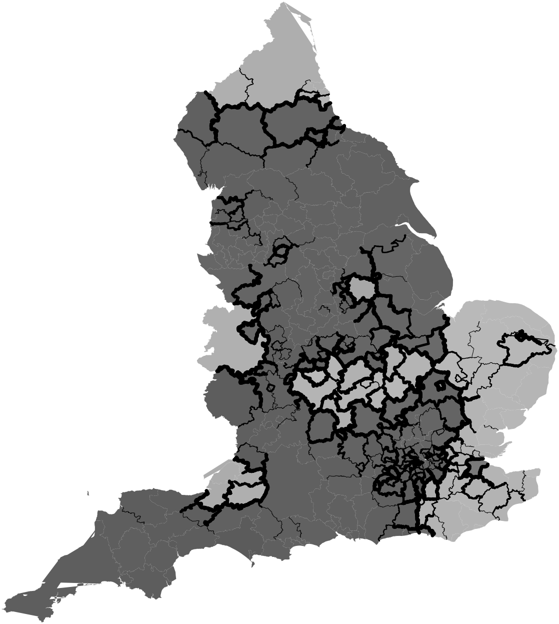
Cons: Still a lot of parameters to estimate

Illustrative simulations

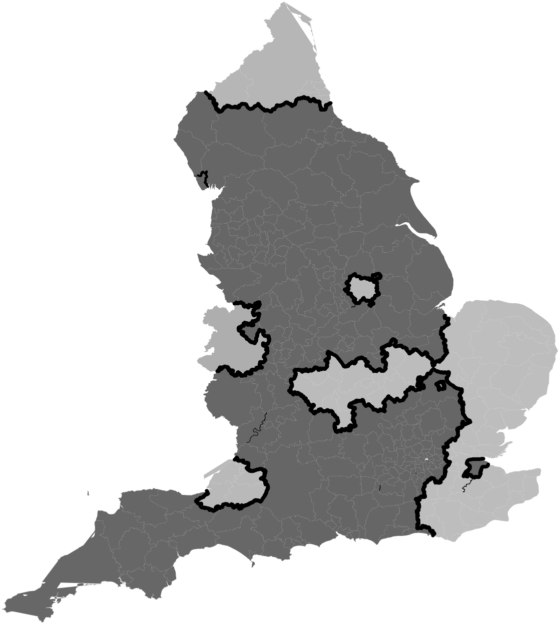


- English local authorities used as a spatial template
- What happens to boundary identification with increased temporal replication?
- What effect does ‘smoothing’ boundaries have? Can ρ be fixed?

Correlated boundaries: T = 10; $\rho = 0.993$



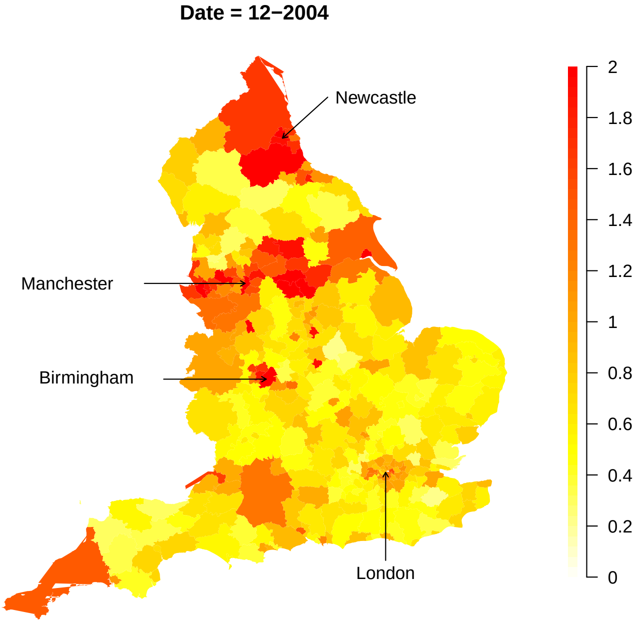
Uncorrelated boundaries: T $= 10$; $\rho = 0$



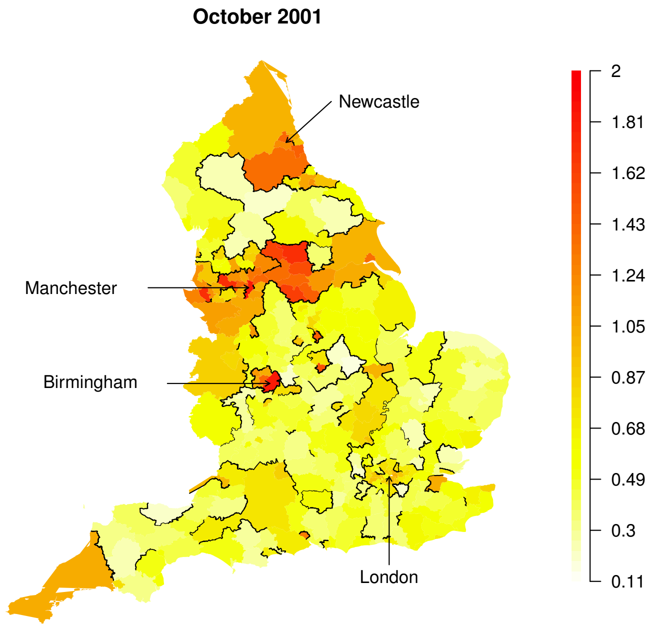
Validation of adaptive model by simulation

		RMSE		
		non adaptive	$\rho = 0$	ρ varies
Temporal replication	**Time = 1**	0.1143	0.0971	0.0728
	Time = 5	0.0865	0.0499	0.0525
	Time = 20	0.0646	0.0390	0.0405
Relative risk	**a = 1**	0.0363	0.0386	0.0397
	a = 2	0.0868	0.0497	0.0524
	a = 3	0.1085	0.0587	0.0611
Expected cases	**E = 25**	0.1253	0.0686	0.0726
	E = 75	0.0866	0.0495	0.0521
	E = 200	0.0613	0.0375	0.0391

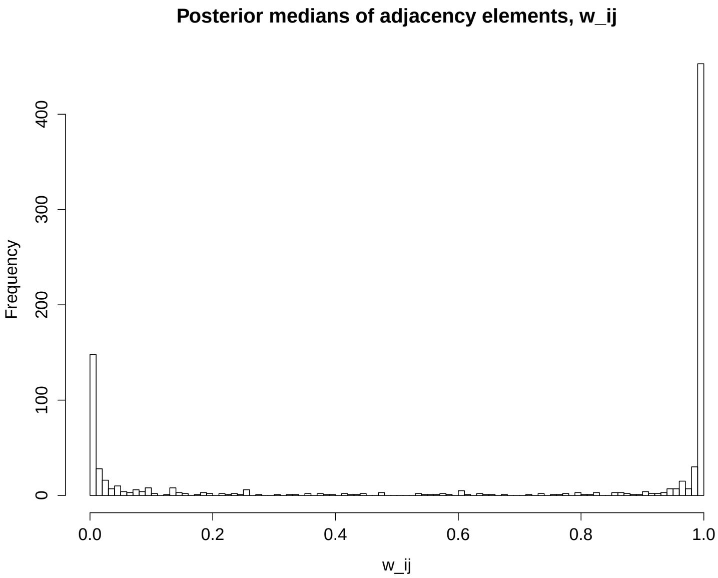
Application to England LHAs - Data



Application to England LHAs - Results



Application to England LHAs - Results



Comments

Adaptive model is succesful at identifying boundaries

- Simulation shows that performance is generally much better than for a standard model
- The prior distribution for boundaries is attractive as it encourages $w_{ij} = 1$ or $w_{ij} = 0$.
- Computation is important, and this model exploits matrix sparsity for speed.

In general, difficult to incorporate information about likely discontinuity structure into prior - compromise sometimes is required due to computational constraints:

- This is why CAR priors are nice: (matrix) sparsity; well established tricks for model-fitting; easy to write very efficient code.
- Example of alternative: Knorr-Held and Raßer (2000) use clustering prior that can induce smoothness; but requires RJMCMC.

Thanks for listening!

Some references

Besag, J., J. York, and A. Mollie (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics*

Lee, D. and R. Mitchell (2013). Locally adaptive spatial smoothing using conditional auto-regressive models. *Journal of the Royal Statistical Society: Series C*

Leroux, B., X. Lei, and N. Breslow. (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence, *Chapter Statistical Models in Epidemiology, the Environment and Clinical Trials, Springer-Verlag, New York*.

Ma, H. and Carlin, B. (2010) Hierarchical and Joint Site-Edge Methods for Medicare Hospice Service Region Boundary Analysis. *Biometrics*