Spatio-temporal modelling of respiratory health in London

Alastair Rushworth, Duncan Lee and Richard Mitchell

University of Glasgow

September 12, 2013

Acknowledgements

Work is joint with

- `Duncan Lee` University of Glasgow
- `Richard Mitchell` University of Glasgow

Funded by the EPSRC

EPSRC

Engineering and Physical Sciences
Research Council

Background

- **Importance:** air pollution is well known to have a negative impact on human health and is still an important public health issue.
- London has a particularly rich history of issues with air pollution and it's subsequent effects. eg. 'Pea-soupers' and the 'Great smog'.
- Estimated 4000 additional deaths due to poor air quality alone in London (Miller, 2010).
- Difficult to unpick the different contributors to ill health without detailed data for risk factors and pollution exposure.

Goals

- **Quantify the impacts of pollutants on respiratory health in Greater London using an ecological design and data at the small area level**
- Investigate the effects of different pollutants, and composite indicators.
- Adequately account for unmeasured confounding
- Achieve some level of computational ease - important if models are to be widely adopted.

Talk structure

- Introducing the London data
- Importance of spatio-temporal modelling
- Results
- Conclusions and discussion.

Data

- London has 624 (non-overlapping) electoral wards for which data are available for the period spanning 2002 to 2009 inclusive.
- Health data are available as annualised total of respiratory hospital admissions for each of the areal units.
- Modelled air pollution maps available from DEFRA (www.uk-air.defra.gov.uk)
- Covariate data Average house prices (Price) and proportion of people claiming jobseekers allowance (JSA)

Both JSA and Price are proxy measures for income and housing deprivation.

Data summaries

- Spatially averaged respiratory admission and pollutant concentrations ($\mu g m^{-3}$) between 2002 and 2008.

	2002	2003	2004	2005	2006	2007	2008
Resp.	117.00	126.00	132.00	143.00	141.00	144.00	151.00
PM_{10}	17.60	20.20	24.60	23.40	22.70	23.80	20.30
$PM_{2.5}$	11.50	17.10	16.70	14.90	15.10	13.50	14.10
CO	372.00	372.00	343.00	338.00	264.00	251.00	229.00
NO_2	34.80	36.70	31.90	33.70	32.40	34.40	30.30
NO_X	59.50	62.00	53.70	56.30	53.10	57.80	50.10
SO_2	3.75	6.02	3.10	3.02	3.08	3.16	2.37

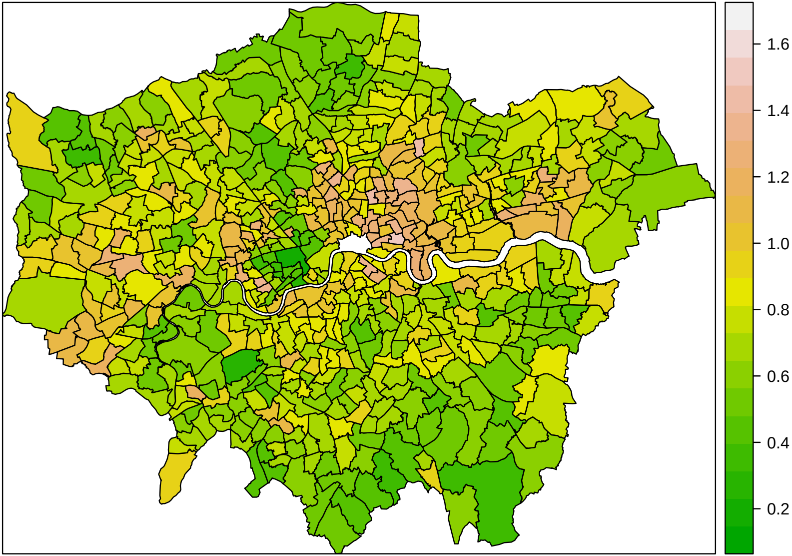
Data visualisation - SIR

As an exploratory measure we plot the Standardised Incidence Ratio (SIR) for each areal unit i, where

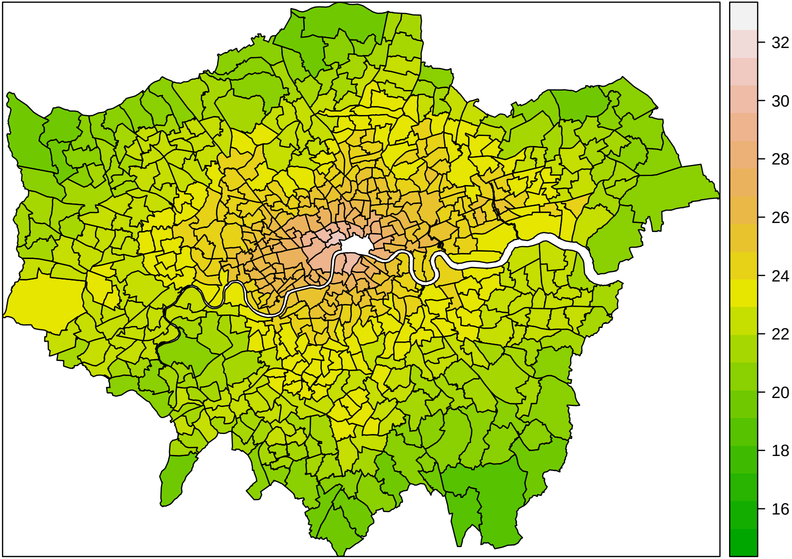
$$\mathrm{SIR}_i = \frac{\mathrm{Observed\ Incidence}_i}{\mathrm{Expected\ Incidence}_i}$$

Where the expected number of cases are calculated using external standardisation based on the population, gender and age distribution of each area.

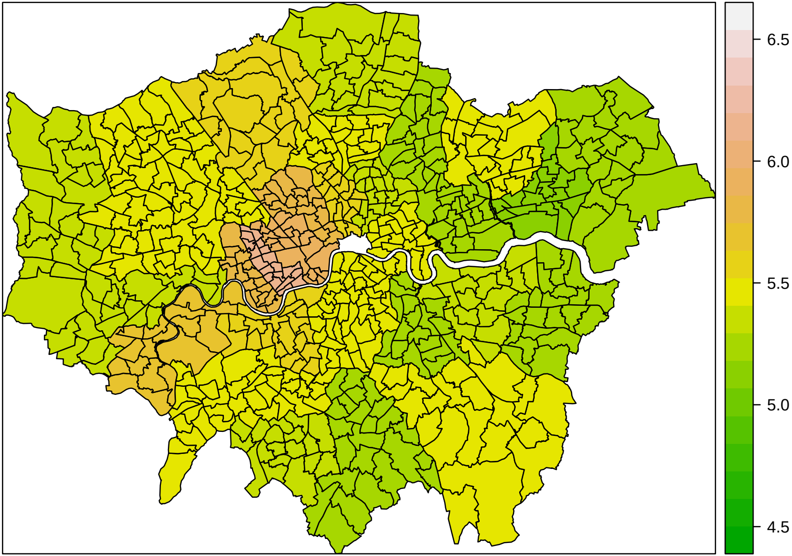
SIR - London respiratory admissions data (2001)



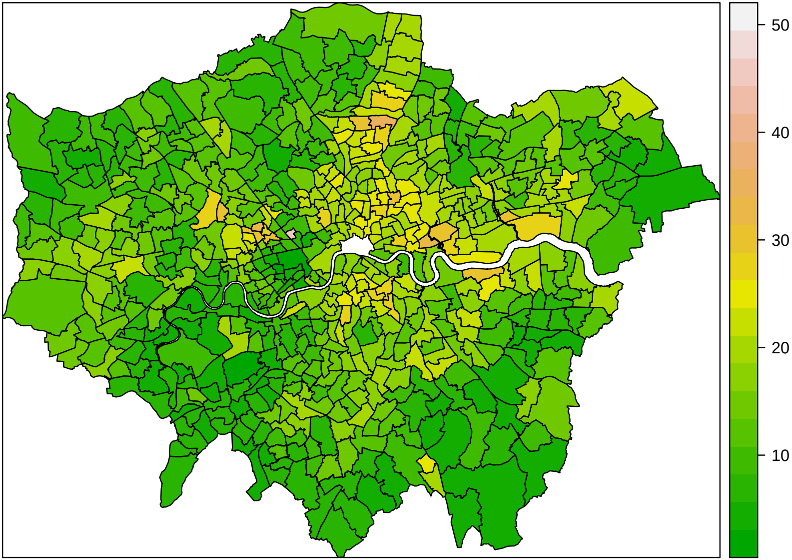
PM_{10} - London air pollution data (2001)



London house prices (2001)



Job seekers allowance (2001)

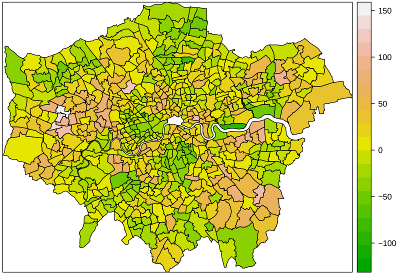


Residual structure - GLM

Fitting the simple model for counts Y_{ij} using GLM in R....

$$\begin{aligned} Y_{ij}|E_{ij}, R_{ij} &\sim \mathrm{Poisson}(E_{ij}R_{ij}) \\ \ln(R_{ij}) &= \alpha + \mathrm{JSM}_{ij}\beta_1 + \mathrm{Price}\beta_2 + \mathrm{Poll}_{ij}\beta_3 \end{aligned}$$

...but residuals are spatially correlated...



	Morans I	p-value
GLM	0.44	0.00

Models for spatial confounding

Introduce a set of spatially smooth random effects ϕ following a Gaussian Markov Random Field (GMRF) prior, options include:

- Intrinsic Autoregressive prior (Besag et al. (1991)); assumes strong smoothness of random effects.
- Besag-York-Mollie (Besag et al. (1991)); 2 sets of random effects (indep + spatial).
- Leroux prior (Leroux et al. (1999)); both indep and intrinsic models are special cases.

...and others. Also possible to use a geostatistical model, splines or other smoothers.

Models for spatial confounding

Leroux prior in space

Overall model is now

$$\begin{aligned} Y_{ij}|E_{ij}, R_{ij} &\sim \text{Poisson}(E_{ij}R_{ij}) \\ \ln(R_{ij}) &= \alpha + \texttt{JSM}_{ij}\beta_1 + \texttt{Price}_{ij}\beta_2 + \texttt{PM}_{10}\beta_3 + \phi_i \end{aligned}$$

Where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and

$$\boldsymbol{\phi} \sim N\left(\mathbf{0}, \sigma^2 \left[\rho W^* + (1-\rho)I\right]^{-1}\right)$$

Where W^* is $diag(W\mathbf{1}^T) - W$ and W is the adjacency matrix.

Need some additional machinery to allow the random effects dependence in time as well as space.

Other models for residual spatio-temporal structure

Many exist, although mainly used in settings where no covariates available.

- Bernardinelli et al. (1995) - spatially varying slopes and intercepts describing linear temporal changes
- MacNab and Dean (2001) - spatially varying spline components to describe non-linear temporal changes.
- Knorr-Held (2000) - studies different forms of space time interaction
- Ugarte. et al (2012) - P-spline ANOVA approach
- Lawson. et al (2012) - Bayesian mixture, some emphasis on clustering of temporal patterns

Models for spatio-temporal confounding

Leroux prior in space and time

Let $\tilde{\phi}_t = (\phi_{t1}, \ldots, \phi_{tN})$. Assume a model of the form

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

Desirable to allowing the prior to allow temporal independence and strong dependance as special cases. One approach would be

$$\begin{aligned} f(\tilde{\phi}_1) &\sim N\left(\mathbf{0}, \sigma^2 \left[\rho W^* + (1-\rho) I\right]^{-1}\right) \\ f(\tilde{\phi}_t | \tilde{\phi}_{t-1}) &\sim N\left(\alpha \tilde{\phi}_{t-1}, \sigma^2 \left[\rho W^* + (1-\rho) I\right]^{-1}\right) \end{aligned}$$

where $\alpha \in [0, 1]$ captures the strength of temporal dependence.

Prior distributions and inference

We assume

$$\begin{aligned}\rho, \alpha &\sim U[0,1] \\ \sigma^2 &\sim U[0,1000] \\ \alpha, \beta_1, \ldots, \beta_p &\sim N(0,100)\end{aligned}$$

- Samples from the marginal posterior of α can be drawn using Gibbs sampling
- Metropolis-Hastings is used for $\boldsymbol{\beta}, \rho, \sigma^2, \boldsymbol{\phi}$.
- For computational speed, the 624×7 vector $\boldsymbol{\phi}$ is updated using `C++`

Results - is a temporal component needed for London data?

To do this compare 3 scenarios under the proposed spatio-temporal model:

$$\alpha = \begin{cases} 0 & \text{no temporal dependence} \\ 1 & \text{strong temporal dependence} \\ \in [0,1] & \text{something in between} \end{cases}$$

	DIC	Morans I	p-value
$\alpha = 1$	36125.6	0.010	0.1490
$\alpha = 0$	36986.9	-0.043	0.0000
$\alpha = 0.85$	36074.1	0.019	0.0056

Table : DIC and residual correlation under each type of model in space and time

Composite indicator

It might be expected that the overall composition of the air is what increases ill health of a population.

As a simple measure take the mean of each pollutant variable as a composite measure at each site and time.
[Each pollutant standardised first so that they have mean 0 and variance 1]

$$\mathrm{Composite}_{ij} = \frac{\sum_{k=1}^{m} \mathrm{Pollutant}_{ijk}}{k}$$

Effects of individual pollutants and composite measure

	RR	95% CI
PM_{10}	1.022	(1.002,1.04)
$PM_{2.5}$	1.032	(1.015,1.051)
CO	1.023	(1,1.039)
NO_2	1.016	(0.998,1.033)
NO_x	1.012	(0.99,1.028)
SO_2	1.009	(0.993,1.024)
Composite	1.026	(1.004,1.046)

Table : Estimated effects of each pollutant and composite pollutant variable with corresponding DIC

Summary of covariate and parameter estimates

Summary of covariates

	RR	95% CI
JSA	1.206	(1.195,1.219)
Houseprice	0.945	(0.924,0.969)

Summary of model parameters

	Median	2.5%	97.5%
σ^2	0.0349	0.032	0.0379
ρ	0.9552	0.9242	0.9766
α	0.8456	0.8179	0.8721

Conclusions

- New model for air pollution and health that is effective in capturing residual structure in space and time
- The model was applied to a large data set for London, an analysis that is the first of its kind.
- For the London data, tangible improvement in fit and estimation if space-time structure acknowledged
- The new results show that pollution is still a significant factor in respiratory ill health, despite long-term improvements in urban air quality.
- Model is fast to fit.

Further work

Further development of composite pollution measures.

- Interpretation presently unclear
- Relax assumption that each pollutant contributes equally

Spatio-temporal residual structure not likely to be really smooth.

- Much work in the spatial domain eg. Lu and Carlin (2007); Lee and Mitchell, (2012) but little or none in the space-time setting.
- Approaches involve treating the adjacency structure as not fixed - computationally hard in space, more so in space-time.

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*

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*.

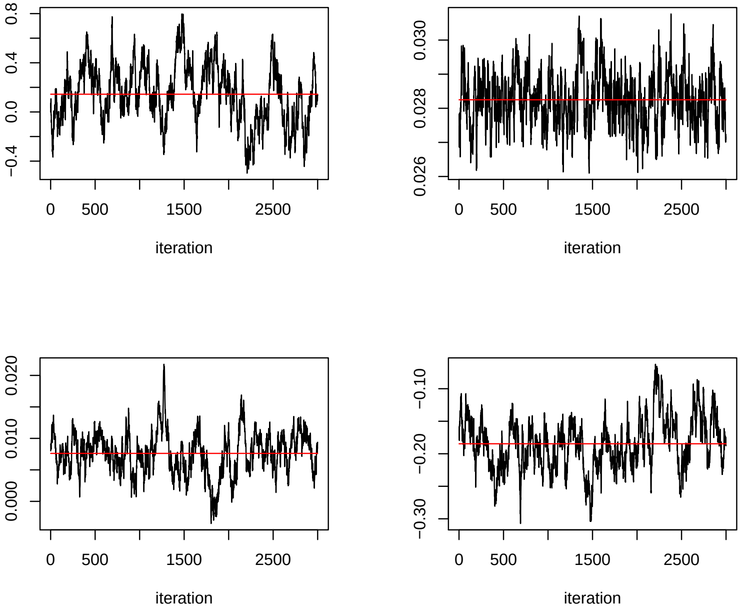
Lu, H. and B. Carlin (2005). Bayesian Areal Wombling for Geographical Boundary Analysis. *Geographical Analysis*.

Miller., B. G. (2010). Report on estimation of mortality impacts of particulate air pollution in london. *Institute of Occupational Medicine*.

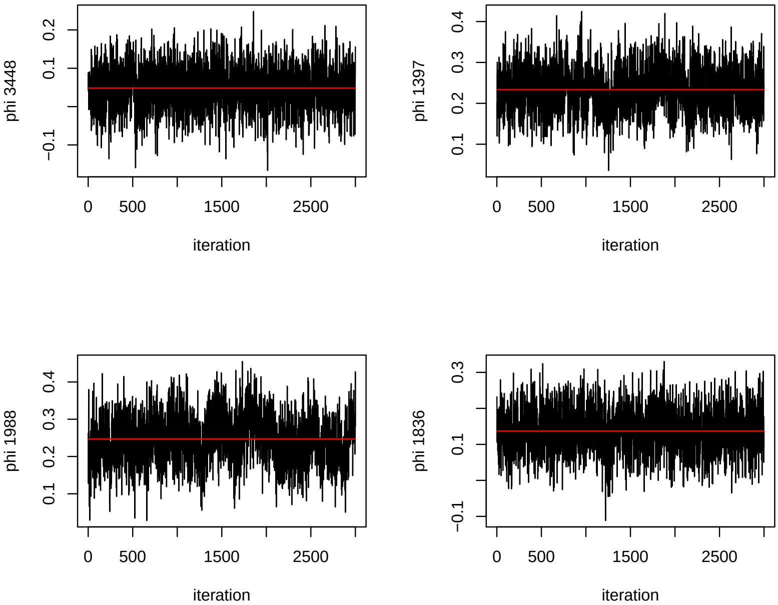
Lee, D. and R. Mitchell (2012). Boundary detection in disease mapping studies. *Biostatistics*.

Thanks for your attention

Trace plots: Fixed effects



Trace plots: ϕ



Trace plots: $\alpha, \rho, \frac{1}{\sigma^2}$