
Copyright © 2014 The Authors

[http://eprints.gla.ac.uk/95138/](http://eprints.gla.ac.uk/95138/)

Deposited on: 15 October 2014
A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London

Alastair Rushworth a,⇑, Duncan Lee a, Richard Mitchell b

a School of Mathematics and Statistics, University Gardens, University of Glasgow, Glasgow G12 8QW, UK
b Public Health and Health Policy, University of Glasgow, UK

A R T I C L E   I N F O

Article history:
Received 23 October 2013
Revised 5 February 2014
Accepted 6 May 2014
Available online 15 July 2014

Keywords:
Air pollution
Gaussian Markov Random Fields
Respiratory disease
Spatio-temporal autocorrelation

A B S T R A C T

It has long been known that air pollution is harmful to human health, as many epidemiological studies have been conducted into its effects. Collectively, these studies have investigated both the acute and chronic effects of pollution, with the latter typically based on individual level cohort designs that can be expensive to implement. As a result of the increasing availability of small-area statistics, ecological spatio-temporal study designs are also being used, with which a key statistical problem is allowing for residual spatio-temporal autocorrelation that remains after the covariate effects have been removed. We present a new model for estimating the effects of air pollution on human health, which allows for residual spatio-temporal autocorrelation, and a study into the long-term effects of air pollution on human health in Greater London, England. The individual and joint effects of different pollutants are explored, via the use of single pollutant models and multiple pollutant indices.

1. Introduction

Air pollution is well known to be detrimental to human health, and can exacerbate many respiratory problems. It is a much greater issue in highly urbanised environments compared with rural locations, because of elevated pollutant concentrations due to emissions from traffic and industry, and high population densities resulting in large populations at risk. The health impact of air pollution is known to be different over different exposure periods, with epidemiological studies having been conducted into the effects of exposure in both the short and the long term. Studies investigating the effects of short-term (acute) exposure are the most common, and utilise time series of health and pollution data recorded at daily or weekly intervals. One of the first such studies was Schwartz and Marcus (1990), and more recent examples include Peters et al. (1997), Schwartz (2000), Lee and Shaddick (2010) and Chang et al. (2011). Much less research has been conducted into the health impact of long-term (chronic) exposure to pollution, and individual level cohort studies investigating this problem include Hoek et al. (2002), Laden et al. (2006) and Beverland et al. (2012). However, cohort studies are both expensive and time consuming to implement, due to the need to follow up a large cohort of people over an extended period of time.

Recently, small area health and social statistics from government run repositories have been made publicly available, with examples being the Surveillance Epidemiology and End Results (SEER, http://seer.cancer.gov/) database in the USA, and the Health and Social Care Information Centre (HSCIC, https://indicators.ic.nhs.uk/webview/) indicator portal in the UK. These databases contain population level annual aggregated summaries of disease incidence and socio-economic status for a set of irregularly shaped admin-
The methodology developed in this manuscript was motivated by a new epidemiological study of air pollution...
concentrations and respiratory hospital admissions in London, UK. London has a long history of air pollution problems, with very high levels of smoke and SO$_2$ being observed since the industrial revolution. Famous pollution events include the ‘Great Smog’ of 1952, in which London was shrouded in a thick layer of airborne pollutants, predominantly originating from coal smoke. Many studies of the health impacts of this smog event have been undertaken, including Bell and Davis (2001), who estimate that it resulted in 12,000 excess deaths between December 1952 and February 1953. As a result of this event, greater regulation on black smoke and coal burning was introduced in the Clean Air Act of 1956, and air pollution was vastly reduced over subsequent years as a result. However, air pollution in London remains a critical public health issue today, with an estimated 4000 deaths a year attributable to poor air quality alone (Miller, 2010). In addition to being the most populous city in the European Union, London is also one of the most economically diverse, often with localised clusters of very affluent neighbourhoods bordering some of the most deprived. It is for these reasons that a spatio-temporal autocorrelation model is likely to prove vital, so that the residual spatio-temporal autocorrelation driven by observed and unobserved confounding factors is accounted for and does not distort the estimated effects of pollution exposure on health.

2.1. Health data

The data analysed in this paper consist of a set of annualised counts ($Y_{ij}$) of the numbers of hospital admissions for respiratory disease (International Classification of Diseases codes J00-J99) for each of the 624 electoral wards that make up Greater London (indexed by $i$), and for each of the years spanning 2003–2009 (indexed by $j$). Although each of the electoral wards have approximately similar population sizes they are not identical, and the observed numbers of admissions will depend on these differences as well as the demographic structure therein. Therefore, the expected numbers ($E_{ij}$) of admissions were calculated by external standardisation, using age and gender specific respiratory admissions rates for the UK. Insight into the spatial distribution of risk can be obtained by looking at maps of the Standardised Incidence Ratio (SIR) defined as $\text{SIR}_{ij} = \frac{Y_{ij}}{E_{ij}}$, which is shown for 2005 in the top panel of Fig. 1. Spatially, the highest values of the SIR appear to be concentrated around the east of Central London, and on the western periphery around Heathrow Airport and the M1 motorway. These are persistent features during the time period for which data are available, and largely correspond to socio-economic deprivation across the city (not shown). Overall, there appears to be little change in the SIR over the 7 year period as the median values for each year vary between 0.75 and 0.8.

2.2. Pollutant and covariate data

One of the most important contributors to respiratory disease and ill-health in general is socio-economic depriva-
tion, which is multi-factorial and cannot be measured directly. One approach is to use a deprivation score such as the English indices of deprivation (https://www.gov.uk/government/organisations/department-for-communities-and-local-government/series/english-indices-of-deprivation) provided by the Department for Communities and Local Government. However, such indices are typically unavailable at the electoral ward level for the temporal extent for which the respiratory health data are available. As a result, proxy measures of deprivation and socio-economic status are available, such as median house price ($\text{Price}$) in each area and the proportion of the population in each electoral ward that are in receipt of Job Seekers Allowance ($\text{JSA}$). The $\text{JSA}$ data are available for each time period and electoral ward, while $\text{Price}$ is only available at Local Authority level (32 Local Authority areas make up Greater London). Although the impact on respiratory disease of smoking prevalence at local area level will dwarf that of air pollution, smoking prevalence data are unavailable at the ward level and for the 7 year period of this study and are not included in the analysis. However, London borough-level smoking prevalence data are available for 2009 (London Knowledge and Intelligence Team at Public Health England, 2013), and this exhibited a linear relationship with $\text{JSA}$ (Pearson’s correlation coefficient of 0.67), which suggests that socio-economic proxy variables can control for the majority of the health effects due to smoking.

Ideally, pollution data from a network of ground monitoring stations would be used to represent population level exposure. However, the network available is too sparse to give a full spatial picture of air pollution concentrations in each of the 624 electoral wards across Greater London. Therefore background pollution maps based on dispersion models and provided by DEFRA (http://www.uk-air.defra.gov.uk) were used, which contain modelled annual mean concentrations in $\mu\text{g} \cdot \text{m}^{-3}$ for each of carbon monoxide ($\text{CO}$); nitrogen dioxide ($\text{NO}_2$); the total of nitrogen monoxide and nitrogen dioxide ($\text{NO}_x$); particulate matter less than 2.5 $\mu\text{m}$ in diameter ($\text{PM}_{2.5}$); particulate matter less than 10 $\mu\text{m}$ in diameter ($\text{PM}_{10}$) and sulphur dioxide ($\text{SO}_2$) each on a 1 km $\times$ 1 km grid. Each pollution variable was lagged by one year relative to the respiratory admission data, so that the pollution exposure occurs before the health events. Ideally, lags extending beyond a single year would be investigated in order to understand the impact of historic air pollution exposures, and ’distributed lag’ models could be used to investigate how these effects can accumulate with time. However, each additional lag introduced requires a reduction in the number of years of data available for the study, and a one year lag was used to limit this data loss. In order to align the pollution grids to the electoral ward scale, the median was calculated for each pollutant in each electoral ward of the modelled pollution tiles that were in that area. These concentrations are shown in Fig. 1 for $\text{PM}_{10}$ in 2005, as the concentration of each pollution variable are summarised in Table 1. While the median levels of $\text{CO}$, $\text{NO}_2$, $\text{NO}$, and $\text{SO}_2$ appear to be on an overall downwards trajectory, it is less clear whether $\text{PM}_{10}$ and $\text{PM}_{2.5}$ are rising or falling during this period. Median house price ($\text{Price}$) has been transformed by dividing by 1000 and taking the natural log, while the proportion of a ward claiming job seekers allowance is on the percentage
Fig. 1. (Top panel) Standard incidence ratio (SIR) for respiratory hospital admissions across Greater London for 2005. (Bottom panel) Average PM$_{10}$ pollution concentrations (µg m$^{-1}$) across Greater London for 2005.
scale – both of these conventions will be maintained for the remainder of the paper.

2.3. Exploratory analysis using a generalised linear model

Initially, a Poisson generalised linear model of the form
\[
Y_{ij} \sim \text{Poisson}(E_{ij}R_{ij}),
\]
\[
\log(R_{ij}) = \beta_0 + \beta_1 Y_i + \beta_2 Y_{ij-1} + \beta_3 Y_{ij-2}
\]
was fitted to the data with just the covariates, to assess the presence of residual spatio-temporal correlation. Here, \(R_{ij}\) represents disease risk in area \(i\) at time \(j\). However, the residuals that result from this analysis exhibit strong spatio-temporal autocorrelation, with an associated Moran’s I statistic \(I = 0.3434\) and an associated p-value of 0.0099, suggesting that some source of variation in respiratory disease risk has not been adequately accounted for. This autocorrelation is particularly evident from the plot of the spatial residuals shown in the top panel of Fig. 2, in which high levels of spatial smoothness are visible. Furthermore, there is evidence of non-separable space–time smoothness in the spatial residuals, which can be seen in the bottom panel of Fig. 2. The figure displays the difference between two successive years residuals, and under the assumption of a separable structure a ‘flat’ overall surface would be expected. However, the surface exhibits substantial spatial variation, suggesting the presence of non-separable space–time structure. Therefore in the next section we propose a model that allows for non-separable space–time residual autocorrelation in an intuitive manner.

3. Modelling

The study region is partitioned into a set of \(N\) non-overlapping areal units indexed by \(i \in \{1, \ldots, N\}\), and data are observed for each of these units for \(j \in \{1, \ldots, T\}\) consecutive time periods. A Bayesian hierarchical model is proposed for these data, with inference based on MCMC simulation. The first level of the hierarchical model is given by
\[
\begin{align*}
Y_{ij} | E_{ij}, R_{ij} & \sim \text{Poisson}(E_{ij}R_{ij}), \\
\log(R_{ij}) & = \mathbf{x}_i^T \boldsymbol{\beta} + \phi_j, \\
\beta_k & \sim N(0, 1000), \quad k \in \{1, \ldots, p\},
\end{align*}
\]
where \(Y_{ij}\) and \(E_{ij}\) are the observed and expected numbers of disease cases in areal unit \(i\) during time period \(j\), and are described in Section 2.1. Here \(\mathbf{x}_i\) is a \(p \times 1\) vector of covariates relating to areal unit \(i\) during time period \(j\), while \(\beta\) is the associated \(p \times 1\) vector of regression parameters. For the epidemiological study discussed in Section 2, the covariate component is given by \(\beta_0 + \beta_1 Y_i + \beta_2 Y_{ij-1} + \beta_3 Y_{ij-2}\), where \(Y_{ij}\) is generic notation for one of the pollutants summarised in Section 2.2.

The random effects \(\phi_j\) are included in (2) to allow for any residual spatio-temporal autocorrelation in the data after the covariate effects have been removed, and are represented by a GMRF prior distribution. Numerous GMRF priors have been proposed for spatio-temporal random effects relating to areal unit data in the related field of disease mapping, although their application in long-term air pollution and health studies is rare (Lawson et al. (2012) being one such example). Both separable Knorr-Held and Besag (1998) and non-separable Knorr-Held (1999) spatio-temporal structures have been proposed, and as the former makes the restrictive assumption that the residual spatial structure is the same for all time periods which from Section 2.3 is unlikely to be realistic, we consider a non-separable model here. The non-separable model of Knorr-Held (1999) contains both spatial and temporal main effects and a non-separable interaction term, and is thus appropriate when the aim of the analysis is to identify these constituent parts of the spatio-temporal structure in the data. However, in the ecological regression context considered here the random effects are nuisance parameters included to account for any residual spatio-temporal autocorrelation in the data, and are not of direct interest. Therefore here we follow the less highly parameterised model proposed by Ugarte et al. (2012), and decompose the single set of random effects \(\phi = (\phi_1, \ldots, \phi_T)\) as
\[
\phi_1, \ldots, \phi_T \sim f(\phi_1) \prod_{j=2}^T f(\phi_j | \phi_{j-1}),
\]
where \(\phi_1 = (\phi_{1j}, \ldots, \phi_{1k})\) denotes the vector of random effects for time period \(j\). This decomposition induces temporal autocorrelation by explicitly allowing \(\phi_j\) to depend on \(\phi_{j-1}\), while \(\phi_1\) is specified marginally as \(\phi_1\) does not exist. The GMRF prior specified for \(f(\phi_1)\) induces spatial autocorrelation into the random effects at time period 1 by means of a binary \(N \times N\) adjacency matrix \(W = (w_{ik})\), which is based on the contiguity structure of the \(N\) areal units. Element \(w_{ik} = 1 \iff\) areal unit \(i\) shares a border with areal unit \(k\), otherwise \(w_{ik} = 0\), and also \(w_{ii} = 0\) \(\forall i\). The joint prior distribution for \(\phi_1\) is given by \(\phi_1 \sim N(0, \Sigma \cdot Q(\rho, W)^{-1})\), where spatial autocorrelation is induced by the precision matrix \(Q(\rho, W)\). A number of GMRF specifications have been used in the spatial

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>372.00</td>
<td>372.00</td>
<td>343.00</td>
<td>338.00</td>
<td>264.00</td>
<td>251.00</td>
<td>229.00</td>
</tr>
<tr>
<td>NO₂</td>
<td>34.80</td>
<td>36.70</td>
<td>31.90</td>
<td>33.70</td>
<td>32.40</td>
<td>34.40</td>
<td>30.30</td>
</tr>
<tr>
<td>NO₃</td>
<td>59.50</td>
<td>62.00</td>
<td>53.70</td>
<td>56.30</td>
<td>53.10</td>
<td>57.80</td>
<td>50.10</td>
</tr>
<tr>
<td>PM₂.₅</td>
<td>11.50</td>
<td>17.10</td>
<td>16.70</td>
<td>14.90</td>
<td>15.10</td>
<td>13.50</td>
<td>14.10</td>
</tr>
<tr>
<td>PM₁₀</td>
<td>17.60</td>
<td>20.20</td>
<td>24.60</td>
<td>23.40</td>
<td>22.70</td>
<td>23.80</td>
<td>20.30</td>
</tr>
<tr>
<td>SO₂</td>
<td>3.75</td>
<td>6.02</td>
<td>3.10</td>
<td>3.02</td>
<td>3.08</td>
<td>3.16</td>
<td>2.37</td>
</tr>
</tbody>
</table>

### Table 1

Median annual air pollution concentrations of London wards between 2002 and 2008 inclusive.
Fig. 2. (Top panel) Standardised residuals plot for 2005 after fitting a generalised linear model, where clear spatial autocorrelation is evident. (Bottom panel) Plot showing the difference between the residuals for the years 2004 and 2005, which suggests that the spatiotemporal structure is non-separable.
modelling literature for $Q(\rho, W)$, most of which are special cases of conditional autoregressive (CAR) models. The one we adopt here was proposed by Leroux et al. (2000) and is given by $Q(\rho, W) = \rho(\text{diag}(W) - W) + (1 - \rho)I$, where $I$ is the $N \times N$ identity matrix and $1$ is an $N \times 1$ vector of ones. This matrix is proper if $\rho \in (0, 1)$. The spatial structure amongst $\phi_i$ can be observed more clearly from the univariate full conditional distributions which are given by

$$
\phi_i | \phi_{-i} \sim N \left( \frac{\rho \sum_{k=0}^{N-1} W_{ik} \phi_k}{\rho \sum_{k=0}^{N-1} W_{ik} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{k=0}^{N-1} W_{ik} + 1 - \rho} \right).
$$

In the above equation, $\phi_{-i}$ denotes the vector of random effects for time period $1$ except for $\phi_i$. From (4) it is clear that $\rho$ controls the spatial autocorrelation structure, with $\rho = 1$ corresponding to the intrinsic CAR prior (Besag et al., 1991) for strong spatial autocorrelation, where the conditional expectation is the mean of the random effects in geographically adjacency areal units. In contrast, $\rho = 0$ corresponds to independent random effects with constant mean and variance. Temporal autocorrelation is induced into the random effects by the conditional specifications $f(\phi | \phi_{-i})$, which are given by

$$
\phi_j | \phi_{-j} \sim N \left( \alpha \phi_{j-1}, \tau^2 Q(\rho, W)^{-1} \right) j \in \{2, \ldots, T\},
$$

where the precision matrix $Q(\rho, W)$ is as defined above. This model thus induces temporal autocorrelation through the conditional expectation, while spatial autocorrelation is induced via the precision matrix. The level of temporal autocorrelation is controlled by $\alpha$, with $\alpha = 0$ corresponding to temporal independence, while $\alpha = 1$ corresponds to strong temporal autocorrelation and is a first order random walk model. We specify weakly informative hyperparameters for the parameters $(\tau^2, \rho, \alpha)$ as

$$
\tau \sim U[0, 1000],
\alpha \sim U[0, 1],
\rho \sim U[0, 1],
$$

which allows their values to be informed by the data. Values of $(\rho, \alpha)$ equal to one correspond to non-stationary processes in space and time, while values in the interval (0,1) lead to stationary specifications. The random effects as modelled above are non-separable in space and time, as the spatial structure at time $j$ is equal to a proportion of the spatial structure at time $j-1$ plus error. The spatial structure thus evolves through time, with the magnitude and strength of this evolution being controlled by the hyperparameters $(\tau^2, \rho, \alpha)$. Our model is a straightforward extension of that proposed by Ugarte et al. (2012), which makes the restriction of strong temporal dependence by setting $\alpha = 1$.

To obtain posterior summaries for the model parameters $\Theta = (\beta, \phi, \tau^2, \rho, \alpha)$, samples were drawn from the posterior distributions using MCMC simulation, based on a mixture of Gibbs sampling and Metropolis-Hastings steps. The analyses presented in Section 4 are based on running the model for a burn in period of 10,000 iterations, after which visual diagnostics suggested that the Markov chains had converged. Inference in all cases was then based on a further 50,000 samples. Spatio-temporal models of this type are very computationally intensive to simulate form, due to the large number of random effects and their complex spatio-temporal autocorrelation structure. As a result, we have implemented our MCMC algorithm using an efficient C++ script written using the R package Rcpp (Eddelbuettel and François, 2011; Eddelbuettel, 2013). All of the software developed was implemented within the R (R Core Team, 2013) statistical programming language, and is available for download with this paper along with the hospital admissions, PM$_{2.5}$, JSA and Price data.

4. Results

4.1. Investigating $\alpha$

We first investigate whether the inclusion of the temporal smoothness parameter $\alpha$ was required, by fitting the model described in Section 3 to the data with $\alpha$ fixed at the values 0 and 1, as well as allowing it to be selected by the data. Each of the three different models were fitted to the London data, and for each the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) was calculated to give an indication of how well each model fits the data. The lowest DIC value of 36,073 is associated with the most flexible scenario in which $\alpha$ is estimated from the data, and is substantially lower than those associated with the scenarios in which $\alpha$ is held fixed (36,986 for $\alpha = 0$ and 36,125 for $\alpha = 1$ respectively). This illustrates that if the strength of temporal dependence is assumed fixed in advance, the model does not fit the data as well. The estimated posterior median was $\alpha = 0.85$, which suggests that while the residual temporal autocorrelation in the London data is strong, the temporal evolution of the random effects is stationary.

4.2. Estimation of the health impact of air pollution and other covariates

The main aim of this study is to estimate the human health impacts of different types of air pollution, and to investigate techniques for estimating the joint effects of numerous air pollutants simultaneously. For the former, each pollutant can be included separately in the spatio-temporal model described in Section 3, in order to identify its relative impact on respiratory health. The principle underlying the latter is that the air we breathe is a complex mixture of different pollutants, and thus its health effects may be different to those of individual pollutants. However, air pollution measurements exhibit strong linear correlations, because they may be generated by common processes or be driven by similar factors such as meteorology. This means that it is inappropriate to include a number of different pollutants in a single model as they are collinear, and thus their individual effects would not be well estimated. Therefore, we propose constructing an air quality indicator (AQI) composed of an appropriately chosen linear combination of the air pollution measures available.

A first approach might be to construct an AQI based on the average of the 6 air pollutants in time and space. This approach has been taken by a number of studies, and was further developed by Lee et al. (2011) who recognise
that the uncertainty present in the individual pollutant measures should be adequately accounted for in the final AQI. Even so, the restrictive assumption that each of the pollutants contributes equally to the resulting AQI is made, and does not take account of the correlation between the contributing pollutants. The UK Met Office calculate a Daily Air Quality Index (DAQI) by assigning each measured air pollutant a score between 1 and 10, and the overall DAQI is defined as the maximum of these scores (www.metoffice.gov.uk/guide/weather/air-quality). However, for the DAQI the assignment of scores and the thresholds that define them are based on daily pollution levels and it is unclear whether this type of AQI is appropriate for the annual data considered here. An alternative approach is to implement a dimension reduction analysis such as Principal Components Analysis (PCA), in order to identify a small number of orthogonal indicators based on different linear combinations of the pollutants that can account for the maximum level of variability in the multivariate pollution data. The principal component loadings that result from this analysis are shown in Table 2, which are the weights each pollutant is multiplied by to create the composite index.

The first principal component (PC) shown in Table 2 shows similar positive loadings across all pollutants, indicating that 56% of the variation in air pollution is due to the overall amount of air pollution, and this PC has a similar interpretation to the average pollution AQIs discussed earlier. The second PC describes a contrast between the effects of large particulate matter as measured by PM10 and SO2. The third PC contrasts the effects of Nitrogen Oxides with small particulate matter as measured by PM2.5 and SO2. Since the first three principal components account for 91% of the total variation in the air pollution data, only these three will be used as air quality indicators.

For each pollutant, Table 3 presents the relative risk associated with a 1 standard deviation increase with it’s associated 95% credible interval (CI). All of the estimated relative risks associated with air pollutants are greater than 1, indicating median increases in respiratory admissions of between 0.7% (SO2) and 2.7% (PM2.5) should be expected for 1 standard deviation increases in pollution. The adverse health effects associated with PM10, PM2.5 and CO are substantial at the 95% level, although all of the other pollutant’s relative risks had lower levels of the credible intervals that were less than one. The strongest overall effect was associated with PM2.5, with annual increases in respiratory hospital admissions of between 0.9% and 4.4% expected for a 2.03 μg m⁻³ increase in small particulate matter concentrations.

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>0.43</td>
<td>–0.31</td>
<td>–0.19</td>
<td>–0.75</td>
<td>0.34</td>
<td>–0.02</td>
</tr>
<tr>
<td>NO2</td>
<td>0.51</td>
<td>–0.04</td>
<td>–0.33</td>
<td>0.30</td>
<td>–0.23</td>
<td>–0.70</td>
</tr>
<tr>
<td>NOx</td>
<td>0.51</td>
<td>–0.04</td>
<td>–0.34</td>
<td>0.28</td>
<td>–0.20</td>
<td>0.71</td>
</tr>
<tr>
<td>PM2.5</td>
<td>0.38</td>
<td>0.27</td>
<td>0.64</td>
<td>–0.30</td>
<td>0.54</td>
<td>0.01</td>
</tr>
<tr>
<td>PM10</td>
<td>0.31</td>
<td>0.68</td>
<td>0.13</td>
<td>0.15</td>
<td>0.63</td>
<td>–0.01</td>
</tr>
<tr>
<td>SO2</td>
<td>0.24</td>
<td>–0.61</td>
<td>0.57</td>
<td>0.39</td>
<td>0.32</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2

Loadings corresponding to a PCA performed on the air pollution data and the cumulative proportion of the variance explained.

Table 3

A summary of the parameter estimates from the spatio-temporal model. The estimated covariate effects are relative risks (RR) for a one-standard deviation (Std. dev) increase in each variable’s value, shown in the final column.

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>RR</th>
<th>95% CI</th>
<th>St.Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>1.023</td>
<td>(1.000, 1.039)</td>
<td>75.83</td>
</tr>
<tr>
<td>NO2</td>
<td>1.013</td>
<td>(0.995, 1.030)</td>
<td>6.99</td>
</tr>
<tr>
<td>NOx</td>
<td>1.009</td>
<td>(0.987, 1.031)</td>
<td>14.67</td>
</tr>
<tr>
<td>PM2.5</td>
<td>1.027</td>
<td>(1.009, 1.044)</td>
<td>2.03</td>
</tr>
<tr>
<td>PM10</td>
<td>1.018</td>
<td>(1.001, 1.038)</td>
<td>2.85</td>
</tr>
<tr>
<td>SO2</td>
<td>1.007</td>
<td>(0.996, 1.019)</td>
<td>1.38</td>
</tr>
<tr>
<td>PC1</td>
<td>1.026</td>
<td>(1.006, 1.044)</td>
<td>1.82</td>
</tr>
<tr>
<td>PC2</td>
<td>1.003</td>
<td>(0.987, 1.025)</td>
<td>1.12</td>
</tr>
<tr>
<td>PC3</td>
<td>1.017</td>
<td>(1.001, 1.033)</td>
<td>1.12</td>
</tr>
<tr>
<td>JSA</td>
<td>1.207</td>
<td>(1.193, 1.220)</td>
<td>6.64</td>
</tr>
<tr>
<td>Price</td>
<td>0.950</td>
<td>(0.925, 0.973)</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Of the three PCs fitted to account for the joint impact of air pollution, the first and the third were estimated as having substantial impacts on health. These risks were 2.6% for PC1 and 1.7% for PC3. Since these variables are orthogonal it is appropriate to include them in the same model simultaneously, and is a more efficient use of the air pollution data. PC1 is a roughly even weighted average of all pollutants, and thus increasing overall pollution levels leads to a substantial increase in respiratory disease risk. PC2 exhibits no relationship to respiratory ill health, which means that changing the composition of air pollution to include relatively more coarse particles (PM10) at the expense of SO2, is not harmful to health. In contrast, PC3 does exhibit a substantial health impact, which means that a relative reduction in CO, NO2 and NOx and increase in fine particulate matter (PM2.5) results in substantial health effects.

It can also be seen from Table 3 that for a 6.6% increase in the proportion of a ward population claiming JSA, an increase in admissions of 20.7% would be expected, indicating that JSA is a very informative covariate and that increasing JSA, as a proxy for smoking has a much stronger impact than air pollution. The House price variable was also found to be strongly associated with decreased hospital admissions, with a decrease in risk of 5% associated with a 0.31 increase in the log House price. For example, this is the difference in admissions that would be expected between two electoral wards in which the average house prices are £200,000 and £270,000 on the original scale.

5. Discussion

In this paper we have proposed a novel spatio-temporal modelling approach for estimating the long-term
health impact of air pollution while allowing for non-separable residual spatio-temporal autocorrelation, and have applied this methodology to a new epidemiological study of the effects of air pollution on respiratory ill health in Greater London. While London has been the scene for many short-term time series studies in the past, this is the first study of the longer-term impact of pollution using a small-area spatio-temporal design. In addition, we are one of very few studies to take a multifactorial view of the health impact of pollution, as we have considered both single pollutant analysis, as well as composite indicators of air quality generated using Principal Components Analysis.

The main results of the study are that after accounting for socio-economic deprivation, air pollution concentrations are associated with an increase in the incidence of respiratory hospital admissions. In particular, one standard deviation increases in PM$_{2.5}$ and CO were found to significantly increase the rate of respiratory hospital admissions, by around 1.8% and 2.7% respectively. Furthermore, an increase in PC$_3$, which represents a roughly equal weighted average of all pollutants, is associated with a 2.6% increase in respiratory ill health, while an increase in PC$_3$ is associated with a 1.7% increase. The latter represents a change in the composition of pollution, obtained by increasing the amounts of PM$_{2.5}$ and SO$_2$ relative to CO, NO$_2$ and NO$_x$. This result thus re-enforces the significant association observed between PM$_{2.5}$ and ill health in the single pollutant analysis, and suggests that it is fine particles that may pose the greatest risk to respiratory ill health. Therefore air pollution still has a substantial population level health impact in London, even at the relatively low concentrations observed in recent years.

The limitations of data availability mean that the results of this study are subject to the following caveats. Firstly, this study assumes that impact of smoking can be accounted for by socio-economic variables, and that this relationship is linear. In Section 2, JSA was found to increase linearly with smoking prevalence at Borough level in 2009, which is consistent with these assumptions. However, interaction with other factors such as ethnic composition could mean that this specification is too simplistic. For example, Bhopal et al. (2004) find that ethnicity affects smoking rates even after controlling for wealth, and so incorporating ethnic composition data should be considered in future studies. In addition, population transience has the potential to dilute the area-specific impact of an exposure, if a large enough proportion of the population in each area have re-located in a short time period. For London boroughs, these movements can be particularly large, for example Hollis (2010) indicates that in 2008, both Hammersmith and Fulham, and Islington experienced population turnovers of over 27%. However, the impact these might have on air pollution and health studies is unclear, since Dennett and Stillwell (2008) show that a large proportion of UK internal population movement is due to residents in their late teens or early twenties, who in turn are at low risk of respiratory problems and are unlikely to make up the high-risk sub-population from which the respiratory admissions are drawn. It is therefore not expected that the effects of population transience would diminish the results presented, however we note that this is an issue for all small-area studies.

One of the key motivators of this work was to develop a model that captured the residual spatio-temporal autocorrelation in the respiratory disease data after the covariate effects have been removed, which if ignored can bias the estimated health effects of air pollution. A class of models based on GMRF priors was used for this purpose, which was a simple extension of that proposed by Ugarte et al. (2012) in the related field of disease mapping. Where our model differs from theirs is that we consider varying degrees of temporal autocorrelation in the random effects structure, and this flexibility was found to be necessary for the Greater London data. However, in a purely spatial setting (Reich et al., 2006) has shown that there is potential for collinearity between spatially smooth covariates and the globally spatially smooth random effects, which may lead to unstable fixed effects estimation. A related issue is that the residual spatio-temporal autocorrelation that the random effects are designed to model might not be globally smooth, as two bordering areal units might contain communities that have very different characteristics. This feature calls into question the use of border sharing as a proxy measure of similarity, with which a smooth residual structure is defined. Some recent developments try to address this issue in a purely spatial setting by treating the neighbourhood structure as a quantity that must be estimated as part of the modelling process. Approaches include Ma et al. (2010), Lee and Mitchell (2012), Lee and Mitchell (2013), and for each, attempts are made to avoid overparameterisation and reduce computational complexity. Therefore, an avenue of future work will be to investigate these phenomena in the spatio-temporal context considered here, to see what impact they may have on the estimated pollution-health relationships.

Part of the focus of this study was to investigate the impacts of different pollutants both individually and jointly, to determine if the air we breathe has a larger health impact than individual pollutants. The PCA undertaken for this purpose successfully yielded two transformed variables that were both found to be strongly related to respiratory admissions, indicating that additional information can be unlocked by taking a multifactorial view of pollution. However, this analysis has ignored two types of uncertainty, which should be allowed to propagate through the PCA. First a measure of variability should be associated with the modelled air pollutant concentrations themselves, and secondly the factor loadings obtained from the PCA have also been estimated and are thus subject to error. In future work, approaches to account more fully for these sources of uncertainty in the pollution data will be pursued.

**Funding**

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC), via grant EP/j017442/1.
Acknowledgements

The authors gratefully acknowledge the editor and a reviewer whose thoughtful comments have improved the content and presentation of this paper.

Appendix A. Supplementary data

The supplementary materials include annual PM$_{2.5}$ concentrations for each London ward between 2002 and 2008, and respiratory hospital admissions, JSA and Price for each ward between 2003 and 2009. The materials also include the C++ functions and an R script describing how to fit the spatio-temporal random effects model described in Section 3. Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.sste.2014.05.001.

References

Schwartz J. The distributed lag between air pollution and daily deaths. Environ Health Perspect 2001;109(Suppl. 3):389.