A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health

DUNCAN LEE[*]

School of Mathematics and Statistics, University of Glasgow, 15 University Gardens, Glasgow G12 8QW, UK

duncan.lee@glasgow.ac.uk

SABYASACHI MUKHOPADHYAY

School of Mathematics, University of Southampton, Building 54, Salisbury Road, Southampton SO17 1BJ, UK

ALASTAIR RUSHWORTH

Department of Mathematics and Statistics, University of Strathclyde, Livingston Tower, 26 Richmond Street, Glasgow G1 1XH, UK

SUJIT K. SAHU

School of Mathematics, University of Southampton, Building 54, Salisbury Road, Southampton SO17 1BJ, UK

SUMMARY

In the United Kingdom, air pollution is linked to around 40000 premature deaths each year, but estimating its health effects is challenging in a spatio-temporal study. The challenges include spatial misalignment between the pollution and disease data; uncertainty in the estimated pollution surface; and complex residual spatio-temporal autocorrelation in the disease data. This article develops a two-stage model that addresses these issues. The first stage is a spatio-temporal fusion model linking modeled and measured pollution data, while the second stage links these predictions to the disease data. The methodology is motivated by a new five-year study investigating the effects of multiple pollutants on respiratory hospitalizations in England between 2007 and 2011, using pollution and disease data relating to local and unitary authorities on a monthly time scale.

Keywords: Air pollution estimation; Bayesian spatio-temporal modeling; Health effects analysis.

1. INTRODUCTION

Air pollution remains a global public health problem in the United Kingdom, with an estimated 40000 premature deaths attributable each year (Royal College of Physicians, 2016). Nitrogen dioxide (NO_2) is

---

[*]To whom correspondence should be addressed.

predicted to exceed European Union (EU) limits until after 2030 in some urban areas of the United Kingdom (Department for the Environment Food and Rural Affairs, 2015), meaning that pollution will remain a key environmental problem for some time. The long-term effects of air pollution are predominantly estimated from cohort studies such as Miller *and others* (2007), but such studies are expensive and time consuming to implement. Therefore spatio-temporal areal unit studies such as Lee *and others* (2009) and Greven *and others* (2011) are also used, and as they make use of freely available population-level data they are quick and inexpensive to implement. Thus although they can be prone to ecological bias, they are used to independently corroborate the evidence from cohort studies.

Poisson log-linear models are used to analyse data from these studies, where the spatio-temporal pattern in disease risk is modeled by known covariates and a set of spatio-temporal random effects. The latter account for any residual spatio-temporal autocorrelation remaining in the disease data, which could be caused by unmeasured confounding, neighborhood effects and grouping effects. However, this study design presents a number of statistical challenges, and this article presents a rigorous statistical framework for addressing them.

The first challenge is the estimation of air pollution at fine spatio-temporal scales, and existing studies typically use either measured data (e.g. Elliott *and others*, 2007 and Greven *and others*, 2011) or estimated concentrations from an atmospheric pollution model (e.g. Lee *and others*, 2009 and Haining *and others*, 2010). Bayesian fusion models for combining modeled and measured pollution data have been proposed by Berrocal *and others* (2009, 2012) and Sahu *and others* (2010) for pollution prediction purposes, and a challenge to combining them with a health model is the *change of support problem*, as there will be spatial variation in the pollution surface within an areal unit. This problem is typically ignored by computing the average concentration in each unit (see Elliott *and others*, 2007, Lee *and others*, 2009), leading to the possibility of ecological bias (see Wakefield and Shaddick, 2006). A further source of pollution uncertainty is posterior uncertainty in the point-level predictions, which is typically ignored by computing a point estimate.

The other main modeling challenge is accounting for residual spatio-temporal autocorrelation via the random effects, and typically globally smooth Gaussian Markov random field (GMRF) priors are used (e.g. Rushworth *and others*, 2014). However, global smoothness is restrictive, as the residual autocorrelation will likely be localized and present between some pairs of adjacent areal units but absent between other pairs. Such localized structure is present in the England study motivating this article (see Figure 1 top right), and Lee and Sarran (2015) have shown that using globally smooth models produces poor-health effect estimates in a purely spatial domain. This article therefore makes two key contributions to the literature. First, Section 3 proposes a rigorous statistical framework for estimating the long-term health effects of air pollution, that is the first to simultaneously address the challenges outlined above. Second, we apply our methodology to a new comprehensive study of air pollution and health in England between 2007 and 2011, which is outlined in Section 2. The results of the study are presented in Section 4 after the methodological development, while Section 5 presents a concluding discussion.

2. STUDY DESIGN AND EXPLORATORY ANALYSIS

2.1. *Disease data*

The study region is mainland England, United Kingdom, partitioned into $k = 1, \ldots, K = 323$ Local and Unitary Authorities (LUA), and data are available for $t = 1, \ldots, T = 60$ months between 2007 and 2011. Counts of the number of respiratory hospitalizations for LUA k and month t are denoted by Y_{kt}, which have a median value of 111 and a range from 6 to 2485. The monthly time scale matches the study by Greven *and others* (2011), whereas the majority of studies such as Lee *and others* (2009) utilise yearly data. An advantage of the monthly scale is that it requires less aggregation of the data away from the
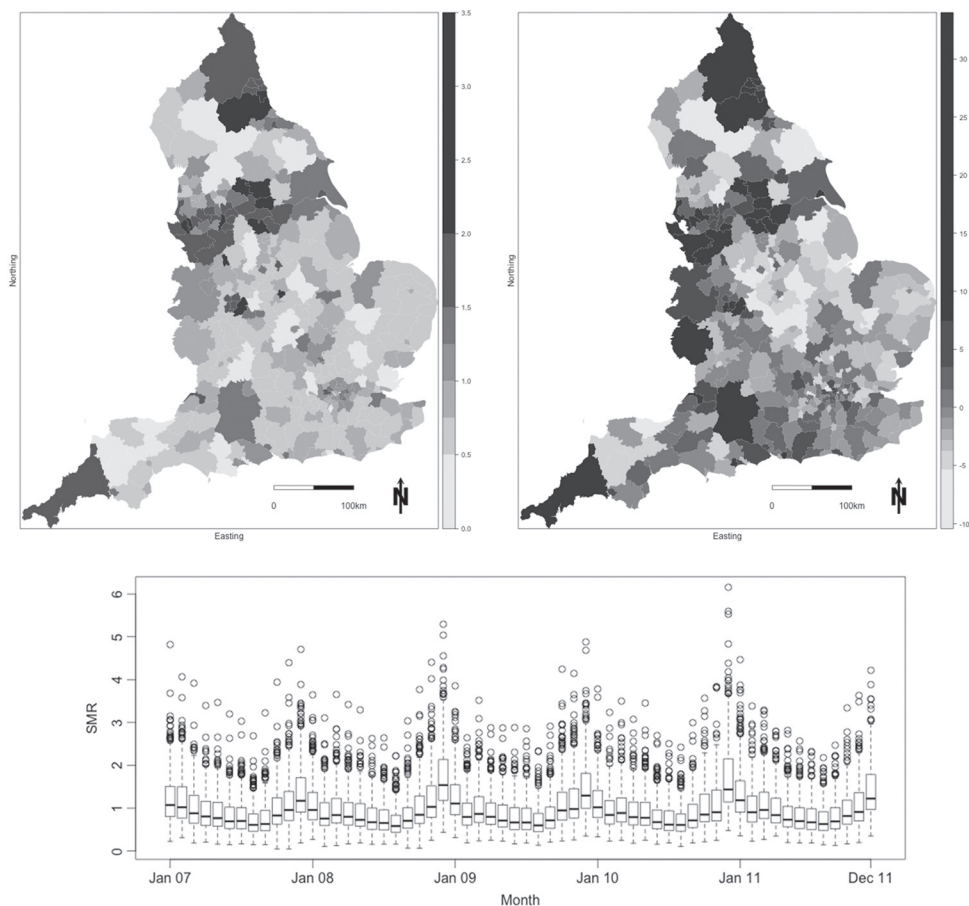
Fig. 1. Display of the spatial (top left) and temporal (bottom) patterns in the SMR data. In the boxplots, the whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range away from the box. The top right panel shows the spatial pattern in the residuals from fitting a Poisson generalised linear model. Both spatial patterns are averages (means) over all time periods.

individual level, but it does mean that Y_{kt} could include admissions driven by both chronic and acute pollution exposure.

The magnitude of Y_{kt} depends on the size and demographics of the population at risk, which we adjust for by computing the expected number of hospital admissions E_{kt} using indirect standardization from national age and sex-specific hospitalization rates. The spatial (top left panel) and temporal (bottom panel) patterns in the standardized morbidity ratio, $\mathrm{SMR}_{kt} = Y_{kt}/E_{kt}$ are displayed in Figure 1, where a value of 1.2 corresponds to a 20% increased risk compared to E_{kt}. The figure shows the highest risks are in cities in the center and north of England, such as Birmingham, Leeds, and Manchester, while the temporal pattern is strongly seasonal, with higher risks of admission in the winter due to factors such as influenza epidemics and cold temperature. We partially adjust for this by seasonally adjusting E_{kt} by a monthly correction factor, resulting in a modified SMR that does not exhibit seasonal behavior.

In previous studies, socio-economic deprivation is the key confounder, because areas that are impoverished have worse health on average than more affluent areas. However, poverty is multi-dimensional, and we represent it by proxy measures of unemployment rate and property price. Specifically, we have

Fig. 2. A map of England and Wales including the locations of the 142 monitoring sites plotted as "*".

measures of the proportion of the working age population in receipt of Job Seekers Allowance (denoted *JSA*), a benefit paid to those without employment, and the average property price (denoted *Price*), both of which are available from the UK Data Archive. Additionally, we also control for the potential effect of ethnicity via the percentage of the population that were born in the United Kingdom as of the 2001 UK Census. Finally we also control for monthly mean temperature, which will account for any year-to-year variation not captured by the seasonally adjusted $\{E_{kt}\}$.

2.2. *Pollution data*

Daily mean concentrations of nitrogen dioxide (NO_2), ozone (O_3), and particles less than $10\mu m$ (PM_{10}) and $2.5\mu m$ ($PM_{2.5}$) were obtained at $n = 142$ locations from the Automatic Urban and Rural Network (AURN, *http://uk-air.defra.gov.uk/networks*) in England and Wales, the latter included to increase the

available information. Their locations are displayed in Figure 2, and six sites are in Wales but close to the English – Welsh border, which thus provides information on the study region. These data were averaged to a monthly temporal resolution to align with the disease data, but contain a large number of missing (not present) values due to discontinuation of sites, introduction of new sites, instrument malfunction, and because not all sites measure all pollutants. The percentages of missing observations are summarized in the supplementary material (Section 2) available at *Biostatistics* online, which shows that roughly 35%, 60%, 65%, and 70% are missing for NO_2, O_3, $PM_{2.5}$, and PM_{10}, respectively. The majority of the missingness occurs in long temporal blocks (e.g. before a monitor is introduced), and it is noticeable that there are many more sites measuring $PM_{2.5}$ after 2008.

Numerical and graphical summaries are provided in the supplementary material (Section 2) available at *Biostatistics* online, by site type, which shows that 16 sites are Rural, 80 are Urban (which includes suburban sites), and 46 are roadside or kerbside (RKS). These results show that rural sites are less polluted than urban and RKS sites for NO_2, PM_{10}, and $PM_{2.5}$, while the converse is true for O_3, suggesting a negative correlation between them. Additionally, RKS sites have a larger spread for NO_2, $PM_{2.5}$, and PM_{10} due to many extreme observations. Finally, one can also see that pollution concentrations vary little from year to year.

We also use estimated hourly concentrations from the Air Quality Unified Model (AQUM, Savage *and others*, 2013), available on the corners of a 12 km^2 grid covering England. AQUM is a 3D weather and chemistry transport model used by the Met Office to deliver air quality forecast for the Department for the Environment, Food, and Rural Affairs (DEFRA) and for scientific research. These hourly estimates were averaged to daily (for ozone daily maximum values were computed as is standard) and then to monthly summary values, to temporally align with the measured concentrations. Bilinear interpolation is then used to estimate monthly AQUM values at the 142 AURN locations (not on the 12km grid), and Figure 3 in the supplementary material available at *Biostatistics* online shows scatter plots of the measured against the modeled concentrations. These plots show moderate correlations of 0.45 for NO_2, 0.69 for O_3, 0.46 for PM_{10}, and 0.37 for $PM_{2.5}$ respectively, as well as inherent negative bias in the modeled concentrations. Our use of AQUM output frees us from needing emission or meteorological data, since these variables are not significant after including AQUM outputs.

In the analysis, we predict pollution concentrations at the corners of the 12 km grid across England, but this results in 49 city LUAs not containing a prediction location. We therefore insert an additional prediction point within each of these LUAs, yielding 1516 prediction sites within England. These prediction sites do not have an associated site type classifier (Rural, Urban or RKS), so for prediction purposes each site is assigned the same type as its corresponding LUA (Rural or Urban), using the classification developed by the Office for National Statistics.

2.3. *Exploratory analysis*

A simple Poisson generalized linear model was applied to the disease data with the covariates ($PM_{2.5}$ was the pollutant included), and the residuals exhibited significant moderate spatio-temporal autocorrelation, with a median (over month) Moran's I statistic of 0.363 (spatial) and a median (over LUA) lag 1 autocorrelation coefficient of 0.422. The average spatial pattern in the residuals is displayed in the top right panel of Figure 1, and visually shows localized spatial autocorrelation that is strong between some neighboring pairs but non-existent between others.

3. Methodology

We propose a two-stage Bayesian hierarchical model for estimating the long-term health effects of air pollution, that is the first to simultaneously predict pollution using both modeled and measured data,

incorporate both spatial variation and posterior uncertainty in the pollution predictions when estimating the health effects, and control for localized spatio-temporal autocorrelation in the disease data The first stage is a fusion model producing posterior predictive distributions of pollution at the 1516 prediction locations, while in stage 2 the corresponding health effects are estimated. We do not propose a single joint model because this would allow the disease counts to influence the predicted pollution concentrations, which is firstly implausible and secondly it is the relationship in the opposite direction we wish to make inference on. Inference is based on Markov chain Monte Carlo (MCMC) simulation, and further details are provided in the supplementary material (Sections 3 and 4) available at *Biostatistics* online and in the software section after the discussion.

3.1. *Stage 1 – pollution fusion model*

3.1.1. *Model specification*   Let $z(\mathbf{s}_i, t)$ and $x(\mathbf{s}_i, t)$ respectively denote the square roots of the measured AURN pollution concentration and the bilinearly interpolated modeled AQUM concentration at location $\mathbf{s}_i$, $i = 1, \ldots, n$ in month $t = 1, \ldots, T$ for a single pollutant. The square root scale is used as pollution is non-negative and skewed to the right, although all predictive accuracy measures are presented on the original scale. We model $z(\mathbf{s}_i, t)$ as Gaussian with a mean $\mu(\mathbf{s}_i, t)$, a spatio-temporal process $\eta(\mathbf{s}_i, t)$, and a white noise process $\epsilon(\mathbf{s}_i, t)$ as follows:

$$Z(\mathbf{s}_i, t) = \mu(\mathbf{s}_i, t) + \eta(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \qquad \epsilon(\mathbf{s}_i, t) \sim \mathrm{N}(0, \sigma_\epsilon^2), \tag{3.1}$$

$$\mu(\mathbf{s}_i, t) = \gamma_0 + \gamma_1 x(\mathbf{s}_i, t) + \sum_{j=2}^{r} \delta_j(\mathbf{s}_i) \left(\gamma_{0j} + \gamma_{1j} x(\mathbf{s}_i, t) \right).$$

Here $\mu(\mathbf{s}_i, t)$ comprises site-type specific regressions on $x(\mathbf{s}_i, t)$, where (γ_0, γ_1) are the global slope and intercept terms while $(\gamma_{0j}, \gamma_{1j})$ are the increments for site type j. Here $r = 3$ for Rural, Urban, and RKS site types, and rural corresponds to the baseline level $(j = 1)$. Finally, $\delta_j(\mathbf{s}_i)$ equals one if site $\mathbf{s}_i$ is of the jth site type and zero otherwise. Note, one could extend the above with a spatially varying coefficient process $\gamma(\mathbf{s}_i)$, however this is likely to mask the site-type specific effects of $x(\mathbf{s}_i, t)$, resulting in weak model identifiability and slow MCMC convergence.

Sahu *and others* (2010) and Berrocal *and others* (2012) have proposed alternatives to (3.1), but they are full fusion models in the sense that they acknowledge $x(\mathbf{s}_i, t)$ to be areal data indexed by grid cells rather than point locations $\mathbf{s}_i$. However, as mentioned above bilinear interpolation has been used to estimate $x(\mathbf{s}_i, t)$ at the $n = 142$ observation sites from the available data at grid box corners. Hence, the change of support issues considered by the above references are not necessary here.

We consider three modeling possibilities for $\eta(\mathbf{s}_i, t)$, which differ in their complexity. The first assumes $\eta(\mathbf{s}_i, t) = 0$ for all sites $\mathbf{s}_i$ and times t, which is used for comparison purposes with the other two models. The second is an independent over time Gaussian process (GP) with a zero mean and a Matérn covariance function:

$$\boldsymbol{\eta}_t = (\eta(\mathbf{s}_1, t), \ldots, \eta(\mathbf{s}_n, t))^\top \sim \mathrm{N}(\mathbf{0}, \sigma_\eta^2 \mathbf{H}_\eta(\phi, \nu)). \tag{3.2}$$

Here $\mathbf{H}_\eta(\phi, \nu)_{ij} = C(||\mathbf{s}_i - \mathbf{s}_j||; \phi, \nu)$, a Matérn correlation function with decay and smoothness parameters ϕ and ν, respectively. The third model has a non-stationary covariance structure following Sahu and Mukhopadhyay (2015), which is based on predictive process methodology (Banerjee *and others*, 2008). First, define m knot-locations, $\mathbf{S}_m^* = (\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*)$, where m is chosen to minimize out of sample root mean square prediction error. Given $\mathbf{S}_m^*$ let $\boldsymbol{\eta}_t^* = (\eta(\mathbf{s}_1^*, t), \ldots, \eta(\mathbf{s}_m^*, t))^\top$ be a zero mean GP with covariance function (3.2). Then our non-stationary model replaces $\eta(\mathbf{s}_i, t)$ in (3.1) by $\tilde{\eta}(\mathbf{s}_i, t) = \mathbb{E}\left[\eta(\mathbf{s}_i, t) | \boldsymbol{\eta}_t^*\right]$, the

expectation of $\eta(\mathbf{s}_i, t)$ given the predictive process $\boldsymbol{\eta}_t^*$. The $(n + m) \times 1$ vector $(\boldsymbol{\eta}_t, \boldsymbol{\eta}_t^*)$ is modeled jointly by the zero-mean GP given in (3.2). Thus by writing $\tilde{\boldsymbol{\eta}}_t = (\tilde{\eta}(\mathbf{s}_1, t), \dots, \tilde{\eta}(\mathbf{s}_n, t))^\top$, we have

$$\tilde{\boldsymbol{\eta}}_t = \mathbf{C}^*(\phi, \nu) \mathbf{H}_{\eta^*}^{-1}(\phi, \nu) \boldsymbol{\eta}_t^*, \tag{3.3}$$

where $\mathbf{C}^*(\phi, \nu)$ is the $n \times m$ cross-correlation matrix between $\boldsymbol{\eta}$ and $\boldsymbol{\eta}^*$, (i.e. $(\mathbf{C}^*)_{ij} = C(||\mathbf{s}_i - \mathbf{s}_j^*||; \phi, \nu)$) and $\mathbf{H}_{\eta^*}(\phi, \nu)$ is the $m \times m$ correlation matrix of $\boldsymbol{\eta}_t^*$ (i.e. $\mathbf{H}_{\eta^*}(\phi, \nu)_{ij} = C(||\mathbf{s}_i^* - \mathbf{s}_j^*||; \phi, \nu)$). Thus $\tilde{\boldsymbol{\eta}}_t$ is a linear function of the m-dimensional $\boldsymbol{\eta}_t^*$ instead of the n-dimensional $\boldsymbol{\eta}_t$, leading to computational savings when m is much smaller than n. However, for our data $n = 142$ is not large, hence this dimension reduction will be more beneficial to larger data sets. Finally, we introduce temporal dependence via the autoregressive model:

$$\boldsymbol{\eta}_t^* \sim N\left(\varrho \boldsymbol{\eta}_{t-1}^*, \sigma_\eta^2 \mathbf{H}_{\eta^*}(\phi, \nu)\right), \quad \text{for } t = 1, \dots, T, \tag{3.4}$$

with $\boldsymbol{\eta}_0^* = \mathbf{0}$ and ϱ is the autoregressive parameter. Sahu and Mukhopadhyay (2015) show that random locations $\mathbf{S}_m^*$ are preferable to a space filling design. We allow $\mathbf{S}_m^*$ to come from M potential locations, which are vertices of a 1 km grid covering England. Additionally, we consider a probability surface $p(\mathbf{s}_j^*)$ for these M locations, where $\sum_{j=1}^M p(\mathbf{s}_j^*) = 1$ and $p(\mathbf{s}_j^*) \geq 0$. Here $p(\mathbf{s}_j^*)$ is the normalized population density, encouraging knots to be placed at high density areas. The locations $\mathbf{S}_m^*$ are thus updated in the MCMC algorithm via a Metropolis-Hastings step, with proposals drawn from the prior $p(\mathbf{s}_j^*)$.

We complete our model by specifying vague but proper prior distributions for the regression parameters ($N(0, 10^4)$), variance parameters (Inverse Ggamma(2, 1)), and autoregressive parameter ($N(0, 10^4)$ truncated to the interval $(-1, 1)$ to ensure stationarity). The Matérn covariance parameters (ϕ, ν) are estimated by an empirical Bayes approach that minimizes out of sample root mean square prediction error, due to the issues regarding inconsistent estimation as outlined by Zhang (2004). The estimates for ϕ are described in the results section while $\nu = 0.2$ was chosen as the optimal value in an initial validation exercise with hold out data for 15 sites among the possible values 0.1, 0.2, and 0.5 (exponential). This indicates a sub-exponential smoothness in the underlying GP is optimal, which is perhaps due to the sparsity and the missingness in the data. Previous studies have shown that the predictions are not greatly sensitive to the choice of these correlation parameters (Sahu *and others*, 2007).

3.1.2. *Prediction from the model* For the kth LUA (denoted A_k) and tth month, the average pollution concentration is estimated by Monte Carlo integration as

$$\hat{Z}_{kt} = \frac{1}{|A_k|} \int_{A_k} Z(\mathbf{s}, t) d\mathbf{s} \approx \frac{1}{n_k} \sum_{j=1}^{n_k} Z(\mathbf{v}_{kj}, t), \tag{3.5}$$

where $|A_k|$ is the area of the kth LUA and $(\mathbf{v}_{k1}, \dots, \mathbf{v}_{kn_k})$ form a grid of prediction locations within the kth LUA. However, $Z(\mathbf{v}_{kj}, t)$ in (3.5) is unknown and as a result $\hat{Z}_{kt}$ is a random variable whose uncertainty should be propagated into the disease model. First, the uncertainty in $Z(\mathbf{v}_{kj}, t)$ is summarized by its posterior predictive distribution:

$$\pi(z(\mathbf{v}_{kj}, t)|\mathbf{z}) = \int \pi(z(\mathbf{v}_{kj}, t)|\mathbf{S}_m^*, \boldsymbol{\eta}^*, \boldsymbol{\theta}, \mathbf{z}) \pi(\mathbf{S}_m^*, \boldsymbol{\eta}^*, \boldsymbol{\theta}|\mathbf{z}) d\mathbf{S}_m^* d\boldsymbol{\eta}^* d\boldsymbol{\theta}, \tag{3.6}$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \varrho, \sigma_\epsilon^2, \sigma_\eta^2, \phi, \nu)^T$, $\boldsymbol{\eta}^* = (\boldsymbol{\eta}_1^*, \dots, \boldsymbol{\eta}_T^*)$, and $\mathbf{z}$ denotes the complete set of pollution data. Here $\pi(z(\mathbf{v}_{kj}, t)|\mathbf{S}_m^*, \boldsymbol{\eta}^*, \boldsymbol{\theta})$ requires $\tilde{\eta}(\mathbf{v}_{kj}, t) = \mathbf{c}^*(\phi, \nu) \mathbf{H}_{\eta^*}(\phi, \nu)^{-1} \boldsymbol{\eta}_t^*$, analogous to (3.3), where $\mathbf{c}^*(\phi, \nu)_{1 \times m}$ has elements $\mathbf{c}_j^* = C(||\mathbf{v}_{kj} - \mathbf{s}_j^*||; \phi, \nu)$.

Samples $z^{(\ell)}(\mathbf{v}_{kj}, t)$, for $\ell = 1, \ldots, L$ are drawn from (3.6) by composition sampling, and a corresponding sample from the posterior predictive distribution of (3.5) is computed as $z_{kt}^{(\ell)} = \frac{1}{n_k} \sum_{j=1}^{n_k} z^{(\ell)}(\mathbf{v}_{kj}, t)$. The uncertainty in the LUA level pollution predictions is summarized in these L samples, but a point estimate $\hat{z}_{kt} = \frac{1}{L} \sum_{\ell=1}^{L} \hat{z}_{kt}^{(\ell)}$ can also be computed. However, such a point estimate ignores two sources of uncertainty, spatial variation in pollution within an LUA, and posterior uncertainty in $\hat{Z}_{kt}$. In the next section, we outline a range of disease models that either ignore or account for these uncertainties.

3.2. *Stage 2 — disease model*

The observed and expected numbers of disease cases (Y_{kt}, E_{kt}) are related to a vector of covariates $\mathbf{u}_{kt}$ measuring poverty, ethnicity, temperature, and a single pollutant. The latter is obtained from posterior predictive samples $\mathcal{Z}_{kt L \times n_k}$ from stage 1, which contains elements $z^{(\ell)}(\mathbf{v}_{kj}, t)$ denoting the ℓth sample from (3.6) at location $\mathbf{v}_{kj}$. A common model for these data is given by

$$Y_{kt} \sim \text{Poisson}\,(E_{kt} R_{kt}) \quad k = 1, \ldots, K, \ t = 1, \ldots, T, \qquad (3.7)$$
$$R_{kt} = \exp(\mathbf{u}_{kt}^\top \boldsymbol{\beta} + \hat{z}_{kt} \beta_z + \psi_{kt}),$$

where $\hat{z}_{kt}$ is the point estimate of pollution defined above, β_z is the effect of air pollution on health and ψ_{kt} is a spatio-temporal random effect. Globally smooth GMRF priors are typically used to model $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_T)$, where $\boldsymbol{\psi}_t = (\psi_{1t}, \ldots, \psi_{Kt})$. The model proposed by Rushworth *and others* (2014) uses the autoregressive decomposition:

$$\boldsymbol{\psi}_t | \boldsymbol{\psi}_{t-1} \sim \text{N}\left(\alpha \boldsymbol{\psi}_{t-1}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho)^{-1}\right) \qquad t = 2, \ldots, T, \qquad (3.8)$$
$$\boldsymbol{\psi}_1 \sim \text{N}\left(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho)^{-1}\right),$$

where $\alpha, \rho \sim \text{Uniform}(0, 1)$ while $\tau^2 \sim \text{Inverse Gamma}(2, 1)$. Temporal autocorrelation is captured by the mean function $(\alpha \boldsymbol{\psi}_{t-1})$ while spatial autocorrelation is captured by the precision matrix $\mathbf{Q}(\mathbf{W}, \rho) = \rho[\text{diag}(\mathbf{W1}) - \mathbf{W}] + (1 - \rho)\mathbf{I}$. The latter was proposed by Leroux *and others* (2000), where $\mathbf{1}$ is a $K \times 1$ vector of ones while $\mathbf{I}$ is a $K \times K$ identity matrix. Spatial autocorrelation is induced by a binary $K \times K$ neighborhood matrix $\mathbf{W}$, where $w_{ki} = 1$ if areal units (k, i) share a common border and $w_{ki} = 0$ otherwise $(w_{kk} = 0$ for all $k)$. The spatial smoothing from this model is evident from its full conditional form, $f(\psi_{kt} | \boldsymbol{\psi}_{-kt})$ (where $\boldsymbol{\psi}_{-kt} = \boldsymbol{\psi}_t \setminus \psi_{kt}$) which for $t = 1$ is:

$$\psi_{k1} | \boldsymbol{\psi}_{-k1} \sim \text{N}\left(\frac{\rho \sum_{i=1}^{K} w_{ki} \psi_{i1}}{\rho \sum_{i=1}^{K} w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{i=1}^{K} w_{ki} + 1 - \rho}\right). \qquad (3.9)$$

Here the conditional expectation is a weighted average of the random effects in neighboring areal units, with the amount of spatial smoothing controlled globally by ρ. However, (3.7) – (3.9) make a number of restrictive assumptions, which we relax in (A) to (C) below.

3.2.1. *(A) - Spatial variation in pollution within an LUA* Wakefield and Shaddick (2006) show that using a representative measure $\hat{z}_{kt}$ in (3.7) when there is within-area variation in pollution can lead to (ecological) bias in its estimated health effect. Consider an idealized model for (Y_{ktj}, E_{ktj}), which relate to the proportion of the population who experience pollution exposure $\hat{z}(\mathbf{v}_{kj}, t) = \frac{1}{L} \sum_{\ell=1}^{L} z^{(\ell)}(\mathbf{v}_{kj}, t)$ at location $\mathbf{v}_{kj}$. Then an appropriate model is $Y_{ktj} \sim \text{Poisson}\left(E_{ktj} R_{ktj}\right)$, where $R_{ktj} = \exp(\mathbf{u}_{kt}^\top \boldsymbol{\beta} + \hat{z}(\mathbf{v}_{kj}, t)\beta_z + \psi_{kt})$. Assuming

conditional independence of $Y_{ktj}|E_{ktj}R_{ktj}$ across the n_k grid squares $(\mathbf{v}_{k1}, \ldots, \mathbf{v}_{kn_k})$ in area k, an appropriate aggregated model for $Y_{kt} = \sum_{j=1}^{n_k} Y_{ktj}$ is:

$$Y_{kt} \sim \text{Poisson}(E_{kt}R_{kt}), \tag{3.10}$$

$$R_{kt} = \exp(\mathbf{u}_{kt}^{\top}\boldsymbol{\beta} + \psi_{kt}) \sum_{j=1}^{n_k} E_{ktj}^* \exp(\hat{z}(\mathbf{v}_{kj}, t)\beta_z),$$

where $E_{ktj}^* = E_{ktj}/E_{kt}$ and $\sum_{j=1}^{n_k} E_{ktj}^* = 1$. The consequence is that (3.7) exponentiates the spatially averaged pollution concentrations $\hat{z}_{kt}$ while (3.10) averages the exponentiated risks $\exp(\hat{z}(\mathbf{v}_{kj}, t)\beta_z)$, and the resulting bias has been studied analytically by Wakefield and Shaddick (2006) and empirically by Lee and Sarran (2015). The latter shows that when β_z is small, as is the case here, then the bias is likely to be negligible.

3.2.2. (B) - Posterior uncertainty in pollution

Model (3.7) uses the posterior predictive mean $\hat{z}_{kt}$, which ignores the posterior uncertainty in the L samples $(z_{kt}^{(1)}, \ldots, z_{kt}^{(L)})$. We consider two distinct approaches to allow for this uncertainty, the first of which treats $(z_{kt}^{(1)}, \ldots, z_{kt}^{(L)})$ as the complete knowledge about the unknown Z_{kt}, and simply samples a new $z_{kt}^{(\ell)}$ value at each iteration of the MCMC algorithm when implementing the disease model independently of the other parameters. The second approach treats the L samples $(\hat{z}_{kt}^{(1)}, \ldots, \hat{z}_{kt}^{(L)})$ as the prior distribution for the unknown Z_{kt} in the disease model, leading to the extended model:

$$Y_{kt} \sim \text{Poisson}(E_{kt}R_{kt}) \quad k = 1, \ldots, K, \quad t = 1, \ldots, T, \tag{3.11}$$

$$R_{kt} = \exp(\mathbf{u}_{kt}^{\top}\boldsymbol{\beta} + Z_{kt}\beta_z + \psi_{kt}),$$

$$Z_{kt} \sim \pi(z_{kt}|\mathbf{z}).$$

A multivariate Gaussian approximation is made to the prior distribution $\pi(Z_{1t}, \ldots, Z_{Kt}|\mathbf{z})$ for all spatial units for each month t for ease in implementing the MCMC algorithm, and details are given in the supplementary material available at *Biostatistics* online.

3.2.3. (C) - Localized spatio-temporal autocorrelation

The global smoothness assumption of (3.8) is unrealistic, because as evidenced by Figure 1, the residual spatial autocorrelation exhibits a mixture of spatial smoothness and sharp discontinuities. Therefore, we allow spatially neighboring random effects to be correlated (inducing smoothness) or conditionally independent (no smoothing), by modeling the non-zero elements of the neighborhood matrix $\mathbf{W}$ as unknown parameters. These adjacency parameters are collectively denoted by $\mathbf{w}^+ = \{w_{ki}|k \sim i\}$, where $k \sim i$ means areas (k, i) share a common border. Estimating $w_{ki} \in \mathbf{w}^+$ equal to zero means (ψ_{kt}, ψ_{it}) are conditionally independent for all t given the remaining random effects, while estimating it close to one means they are correlated. Here we use the model proposed by Rushworth *and others* (2016), where each adjacency parameter in $\mathbf{w}^+$ is modeled on the interval [0, 1], by placing a multivariate Gaussian prior on the transformation $\mathbf{g}^+ = \log(\mathbf{w}^+/(\mathbf{1} - \mathbf{w}^+))$. We utilise a shrinkage prior for $\mathbf{g}^+$ with a constant mean and variance (μ, ζ^2), which is given by:

$$f(\mathbf{g}^+|\zeta^2, \mu) \propto \exp\left[-\frac{1}{2\zeta^2}\left(\sum_{g_{ik} \in \mathbf{g}^+}(g_{ik} - \mu)^2\right)\right], \quad \zeta^2 \sim \text{IG}(2, 1). \tag{3.12}$$

Here the random effects surface $\mathbf{g}^+$ is not smoothed spatially, for example by a second level GMRF prior, because work by Rushworth *and others* (2016) showed this results in poor estimation performance. Under small values of ζ^2, the elements of $\mathbf{g}^+$ are shrunk to μ, and here we follow the work of Rushworth *and others* (2016) and fix $\mu = 15$ because it avoids numerical issues when transforming between $\mathbf{g}^+$ and $\mathbf{w}^+$ and implies a prior preference for values of w_{ik} close to 1. That is as $\zeta^2 \rightarrow 0$ the prior becomes the global smoothing prior (3.8). Further discussion of these points can be found in Rushworth *and others* (2016) .

4. Results

We now present the results of the England respiratory hospitalization study described in Section 2, where the first subsection quantifies the predictive performance of a range of pollution models, while the second subsection presents the health effect estimates.

4.1. *Pollution modeling results*

We assessed the relative performances of three variants of the pollution model (3.1) of increasing complexity, which are defined by: (a) $\eta(\mathbf{s}_i, t) = 0$ and denoted *Linear*; (b) $\eta(\mathbf{s}_i, t)$ modeled by the Gaussian process (3.2) and denoted *GP*; and (c) the full non-stationary Gaussian predictive process model $\tilde{\eta}(\mathbf{s}_i, t)$ given by (3.3) – (3.4) and denoted *GPP*. In the latter, the spatial range parameter ϕ is fixed at effective ranges of 3500 (ϕ_1), 3000 (ϕ_2), 600 (ϕ_3), 300 (ϕ_4), and 100 (ϕ_5) km respectively, and these choices are guided by the need to include moderate to large amounts of spatial correlation into the model. Only the results for the best value of ($\phi_1 \ldots \phi_5$) are shown in the table, although all values gave similar results as expected since the predictions are not very sensitive to the choice of the value of the decay parameter.

Table 1 summarises the results of a 10-fold cross-validation exercise conducted separately for each pollutant, where the $n = 142$ sites are split into 10 equally sized groups (8 groups of 14 and 2 of 15). Each model is then applied to data from 9 of the 10 groups, and the resulting fitted model is used to predict the concentrations in the 10th group. This procedure is conducted 10 times leaving each group out once. All results are based on $L = 5000$ MCMC iterations, which were obtained following a burn-in period of 5000 iterations at which point convergence was assessed to have been reached. In all cases, we take $m = 25$, which was chosen among the possible values of 16, 25, 36, 49, and 100 using a separate validation method, where data from 15 randomly chosen sites were set aside. Table 1 presents the bias, root mean square prediction error (RMSPE), mean absolute prediction error (MAPE), and coverage percentage of the 95% prediction intervals, from which a number of key themes are apparent.

First, all models show relatively little amounts of bias on the scale of the pollutants, with biases of less than 1.42 in all but one case. Second, the *GPP* model has better point prediction for all pollutants than the other models, with noticeably lower RMSPE and MAPE values. In contrast, the *GP* model exhibits poorest point prediction in all cases. As these results are averages over all $n = 142$ observation locations, we present scatter plots of the individual predictions from the *GPP* model against the observations in Section 5 of the supplementary material available at *Biostatistics* online. These results show that the predictions are well calibrated against the observations, except that the model is unable to capture the very high NO_2 values or the very low O_3 values.

Finally, none of the coverage percentages are that close to their nominal 95% levels, and are too low for the *Linear* and *GPP* models and too high for the *GP* models. These problems are caused by the irregular and sparse nature of the monitored data, which is particularly evident for $PM_{2.5}$ where few observations are available for the first two years (see the supplementary material Section 2). The *GP* model has almost 100% coverage, which is due to its 95% uncertainty intervals being very wide compared to the other models (see the supplementary material Section 5). This is because its spatio-temporal process

Table 1. *Summary of the 10-fold cross-validation for the* Linear, GP *and* GPP *models for all four pollutants. Here RMSPE and MAPE respectively denote root mean square prediction error and mean absolute prediction error*

Model	Bias	RMSPE	MAPE	Coverage (%)
		NO_2		
Linear	−1.15	19.92	14.52	63.13
GP	1.29	20.73	15.78	98.79
GPP	−1.19	18.94	14.03	88.63
		O_3		
Linear	−0.57	11.21	8.98	60.18
GP	2.80	11.77	9.38	99.28
GPP	1.41	8.86	6.73	87.84
		PM_{10}		
Linear	−0.28	5.91	4.70	65.51
GP	1.11	6.04	4.89	99.97
GPP	−0.23	4.29	3.14	80.85
		$PM_{2.5}$		
Linear	−0.34	5.02	4.02	78.61
GP	1.41	5.20	4.37	100
GPP	−0.06	3.10	2.29	74.83

Table 2. *A summary of the six models fitted to the disease data*

Model	Spatial Variation	Uncertainty Propagation	Spatio-temporal autocorrelation
M1	Ignored	Ignored	Localized
M2	Allowed	Ignored	Localized
M3	Ignored	Posterior	Localized
M4	Ignored	Prior	Localized
M5	Ignored	Ignored	None
M6	Ignored	Ignored	Global

$\eta(\mathbf{s}_i, t)$ is overly flexible in space and time with no temporal autocorrelation and a separate random effect for each spatial location. In contrast, the *GPP* model is autocorrelated in time and uses a reduced rank spatial predictive process, resulting in more borrowing of strength in the estimation and reduced uncertainty.

4.2. *Disease modeling results*

Following the cross-validation exercise described above, we applied the non-stationary *GPP* model to the pollution data from all $n = 142$ locations, and made $L = 5000$ predictions of each pollutant at each of the 1516 prediction locations $\mathbf{v}_{kj}$ across England for each of the 60 months in the study. These predictions were then utilized in a number of different disease models, whose specifications are summarized in Table 2. The results from fitting these single pollutant models are displayed in Table 3, where single-pollutant analyses were undertaken in common with the existing literature and because it is single pollutant effect estimates

Table 3. *Estimated health effects from each pollutant for a range of models. All results are presented as relative risks for a one standard deviation increase in pollution, which is $16.07 \mu gm^{-3}$ for NO_2, $7.30 \mu gm^{-3}$ for O_3, $4.90 \mu gm^{-3}$ for PM_{10}, and $4.11 \mu gm^{-3}$ for $PM_{2.5}$. Note, the uncertainty intervals are credible intervals for all but model M5 where they are confidence intervals*

Model	NO_2	O_3	PM_{10}	$PM_{2.5}$
M1	1.026 (1.017, 1.036)	0.983 (0.971, 0.995)	1.004 (0.994, 1.015)	1.003 (0.991, 1.016)
M2	1.026 (1.012, 1.038)	0.981 (0.966, 1.000)	1.004 (0.994, 1.016)	1.003 (0.993, 1.015)
M3	1.001 (0.999, 1.002)	1.000 (0.999, 1.001)	1.000 (0.998, 1.003)	1.001 (0.997, 1.004)
M4	1.028 (1.021, 1.033)	0.997 (0.994, 0.999)	1.026 (1.011, 1.039)	1.006 (0.993, 1.020)
M5	1.107 (1.099, 1.116)	0.998 (0.992, 1.004)	1.003 (0.997, 1.010)	0.998 (0.991, 1.004)
M6	1.031 (1.019, 1.043)	0.989 (0.978, 1.001)	1.007 (0.996, 1.017)	1.006 (0.997, 1.014)

that inform regulatory standards. All results are presented as relative risks for a one standard deviation (sd) increase in each pollutant, which is 16.07, 7.30, 4.90, and 4.11 (μgm^{-3}), respectively for NO_2, O_3, PM_{10}, and $PM_{2.5}$. Inference for all models in this section is based on 500000 MCMC iterations, with the first 100000 being discarded as the burn-in period. The covariate effects are described below, while the remainder of this section quantifies the impact of three aspects of the model, which coincide with labels (A)–(C) in Section 3.2.

The covariates in each model include a pollutant, socio-economic deprivation (measured by JSA and the natural log of property price), ethnicity, and temperature, and the results presented below relate to model M1 and are relative risks for a one sd increase in each covariate (JSA - Sd = 0.334; property price - Sd = 0.359; ethnicity - Sd = 7.98; and temperature - Sd - 4.68). The posterior medians and 95% credible intervals are: JSA - 0.998 (0.994, 1.002), Price - 0.834 (0.819, 0.850), Ethnicity 0.783 (0.769, 0.795), Temperature 0.979 (0.964, 0.994). These results indicate a strong association between increasing non-British ethnicity and increased disease risk, although this is likely confounded by inner-city deprivation, where ethnic diversity tends to be largest. Property price is a good proxy for overall deprivation, with increased prices strongly associated with decreases in risk. Finally, colder temperatures are associated with increased respiratory admissions as expected.

4.2.1. *(A) - Allowing for spatial variation in pollution within an areal unit*   We compare the ecological model (3.7) denoted by **M1** that ignores within area pollution variation with the correct aggregate model (3.10) denoted by **M2** that allows for this variation. For the latter, E_{ktj} are the estimated expected counts around each of the 1516 prediction locations $\mathbf{v}_{kj}$, computed using electoral ward level (a much smaller unit than LUA) population data. In this comparison, uncertainty in the estimated pollution surface is ignored and the random effects are modeled by the localized GMRF model given by (3.8) and (3.12). The main finding is that the estimated relative risks from each model are almost identical, suggesting that ignoring spatial variation in pollution within an areal unit does not bias the estimated health effects. This is likely to be because the effect sizes are small, making the bias term discussed in Wakefield and Shaddick (2006) negligible (of the order of β_z^2 where for NO_2 $\hat{\beta}_z = 0.00160$). This result is corroborated by Lee and Sarran (2015), as they showed that negligible bias would occur unless the effect size was much larger. Another possible reason for the similarity of the results is that 49 of the 323 LUAs only had a single prediction location due to their small geographical size. As a result, within-area variation in pollution is ignored in the remainder of this article.

The second main finding is that NO_2 exhibits substantial health effects, with a $16.07 \mu gm^{-3}$ increase being associated with nearly a 2.6% increased risk of disease. The estimated effects for the two particulate

matter metrics are borderline significant at the 5% level, with increased risks around 0.4%. In contrast, O_3 shows a borderline negative effect, which is clearly erroneous, and due to the negative correlation between this pollutant and all the others, resulting in an estimated effect of the opposite sign.

4.2.2. *(B) - Allowing for posterior uncertainty in the pollution predictions*   We compare three approaches to dealing with posterior uncertainty, the simplest being model **M1** that ignores it by using the posterior predictive median $\hat{z}_{kt}$. In contrast, model **M3** resamples a new $\{\hat{z}_{kt}^{(\ell)}\}$ value at each step of the MCMC algorithm (denoted Posterior in the table), while model **M4** treats the samples $(\hat{z}_{kt}^{(1)}, \ldots, \hat{z}_{kt}^{(L)})$ (based on a multivariate Gaussian approximation) as a prior distribution for the unknown Z_{kt} in the disease model (denoted Prior in the table). The results show that re-sampling a new $\{\hat{z}_{kt}^{(\ell)}\}$ value at each step of the MCMC algorithm results in all the estimated effects being greatly attenuated to the null risk of one, which occurs because the uncertainty in the posterior predictive pollution distributions outweigh the spatio-temporal variation in the pollution data, resulting in the estimated effects being washed away by large uncertainty. In contrast, the results for the prior scheme are mixed, with the NO_2 and $PM_{2.5}$ results remaining unchanged, the spurious O_3 result being attenuated towards one, while the PM_{10} estimate is actually increased to a substantial increased risk of 2.6%.

4.2.3. *(C) - Allowing for residual spatio-temporal autocorrelation*   Finally, we compare three approaches for controlling for unmeasured spatio-temporal autocorrelation. Model **M5** ignores this autocorrelation, that is $\psi_{kt} = 0$, model **M6** uses the globally smooth model given by (3.8), while model **M1** uses the localized model given by (3.12). The results show a much larger effect size for NO_2 if the residual spatial autocorrelation is ignored (**M5**), while the other pollutants show negligible differences. The differences in the estimated effect sizes between the globally smooth and locally smooth random effects models are also largest for NO_2, with a difference in estimated relative risk of around 0.5%. This suggests that if there is a substantial relationship between a pollutant and disease risk, then the choice of random effects model can impact the results. However, if there is no relationship then the results appear to be consistent.

The Deviance Information Criterion (DIC) statistics are **M6** - 157131 and **M1** - 155538, indicating a moderate improvement in overall fit under the localized approach. Additionally, the effective number of parameters decreases substantially for the localized model (6205 compared to 8198), despite it being more complex algebraically. This occurs because the random effects variance, τ^2 is much lower for the localized model (posterior medians of 0.0033 compared to 0.0248), resulting in stronger smoothing of the random effects and hence a simpler model with fewer effective parameters. This result is consistent with a similar observation made by Rushworth *and others* (2016), although as a caveat to these results we note that Plummer (2008) has shown that the DIC has a tendency to underpenalize complex models. The localized nature of the spatial autocorrelation is determined by $\mathbf{w}^+$, and posterior inference for $\mathbf{w}^+$ is summarized in Section 6 of the supplementary material available at *Biostatistics* online.

5. Discussion

This article develops the first pollution-health model for simultaneously predicting pollution concentrations using fusion methodology, and estimating its health effects accounting for within-area variation and posterior predictive uncertainty in pollution, as well as localized residual spatio-temporal autocorrelation. Software to implement the full model is available on `figshare` (see software section for the url), while the disease model with either a global or a localized GMRF prior is available via the R package `CARBayesST`. The methodology was motivated by a new study of air pollution and respiratory hospitalizations in England, and the main finding is that areas with an increased monthly exposure to NO_2 of $16.07 \mu gm^{-3}$ exhibit an estimated 2.8% increased risk of hospitalization (assuming model **M4**). If this

increase in pollution was applied uniformly to England it would equate to around 17000 extra admissions per year, as there are around 613000 admissions per year in England. This result, against the backdrop of many urban areas of England likely to exceed EU NO_2 emission targets until after 2030, suggests that NO_2 concentrations will be a major public health threat for the foreseeable future. These findings are in line with existing research, as Lee and Sarran (2015) and Haining *and others* (2010) (they considered the related pollutant NO_x) also found similar sized associations in areal unit studies, albeit in different study regions (Haining *and others*, 2010) or time scales (Lee and Sarran, 2015) making the results not directly comparable. Our estimated effects for particulates were mostly much smaller (except for PM_{10} under **M4**), and are generally similar or lower than what has been found in the existing literature such as Elliott *and others* (2007) (using the similar black smoke metric), Greven *and others* (2011), and Rushworth *and others* (2014).

The similarity in the estimated health effects of allowing for or ignoring within area variation in exposure corroborate the findings of Haining *and others* (2010) and Lee and Sarran (2015), and suggest that ecological bias is not a big problem in these studies where effect sizes are small. However, the choice of spatio-temporal autocorrelation model does impact on the estimated pollution-health effects if substantial associations are found, and Lee and Sarran (2015) also find this when comparing models with differing levels of complexity. In common with the results presented here, they find that localized autocorrelation models fit the data better, as measured by DIC, than simpler global smoothing alternatives. However, if no association exists then the three autocorrelation models considered here show similar results. Finally, the approach taken to allowing for uncertainty in the estimated pollution surface when estimating its health effects has a substantial impact on the results, and feeding the entire posterior predictive distribution through the disease model (model **M3**) has washed away the pollution-health effects due to the magnitude of the posterior uncertainty. In contrast, treating the posterior predictive pollution distribution as a prior in the disease model has produced similar results to ignoring the uncertainty except for PM_{10}, where the effect size is greatly increased. Blangiardo *and others* (2016) also found that incorporating uncertainty in pollution did not change the substantive conclusions, although they adopted a different approach and ran the disease model separately for each posterior predictive sample from the pollution model before combining the results. Quantifying the best approach to allowing for pollution uncertainty in a disease model is in need of further research, as a number of different approaches have been proposed to date with no systematic evaluation of their relative performances.

There are many other areas of future work in this field, including estimating the financial as well as health cost of pollution, which additionally requires data on government health spending. A second important application of the developed methodologies will be to apply them to data from mega-cities in the developing world where air pollution levels are high, such as New Delhi in India and Beijing in China. The main problem in doing so will be to gather reliable data on air pollution exposure as well as health outcomes. Finally, a key methodological development will be extending the methodology into a multivariate domain, so that the pollutants and their impacts on multiple diseases can be modeled jointly, which allows one to estimate the overall impact of air pollution on population health.

Software

The data and software used in this article are available on `figshare` at https://figshare.com/articles/Code_and_data_for_A_rigorous_statistical_framework_for_estimating_the_long-term_health_impact_of_air_pollution_/3581235, and a simplified version of the disease model without the uncertainty propagation or ecological bias correction is available via the `R` package CARBayesST.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

References

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of Royal Statistical Society, Series B* **70**, 825–848.

Berrocal, V., Gelfand, A. and Holland, D. (2009). A Spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological and Environmental Statistics* **15**, 176–197.

Berrocal, V., Gelfand, A. and Holland, D. (2012). Space-time data fusion under error in computer model output: An application to modeling air quality. *Biometrics* **68**, 837–848.

Blangiardo, M., Finazzi, F. and Cameletti, M. (2016). Two-stage Bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions. *Spatial and Spatio-temporal Epidemiology* **18**, 1–12.

Department for the Environment Food and Rural Affairs. (2015). Updated projections for Nitrogen Dioxide (NO2) compliance.

Elliott, P., Shaddick, G., Wakefield, J., Hoogh, C. and Briggs, D. (2007). Long-term associations of outdoor air pollution with mortality in Great Britain. *Thorax* **62**, 1088–1094.

Greven, S., Dominici, F. and Zeger, S. (2011). An Approach to the estimation of chronic air pollution effects using spatio-temporal information. *Journal of the American Statistical Association* **106**, 396–406.

Haining, R., Li, G., Maheswaran, R., Blangiardo, M., Law, J., Best, N. and Richardson, S. (2010). Inference from ecological models: estimating the relative risk of stroke from air pollution exposure using small area data. *Spatial and Spatio-temporal Epidemiology* **1**, 123–131.

Lee, D., Ferguson, C. and Mitchell, R. (2009). Air pollution and health in Scotland: a multicity study. *Biostatistics* **10**, 409–423.

Lee, D. and Sarran, C. (2015). Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics* **26**, 477–487.

Leroux, B., Lei, X. and Breslow, N. (2000). *Estimation of disease rates in small areas: A new mixed model for spatial dependence. Statistical Models in Epidemiology, the Environment and Clinical Trials, In Halloran, M and Berry, D (editors)*, New York: Springer-Verlag, pp. 135–178.

Miller, K., Sicovick, D., Sheppard, L., Shepherd, K., Sullivan, J., Anderson, G. and Kaufman, J. (2007). Long-Term Exposure to Air Pollution and Incidence of Cardiovascular Events in Women. *New England Journal of Medicine* **356**, 447–458.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* **9**, 523–539.

Royal College of Physicians. (2016). Every breath we take: The lifelong impact of air pollution. https://www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution.

RUSHWORTH, A., LEE, D. AND MITCHELL, R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater london. *Spatial and Spatio-temporal Epidemiology* **10**, 29–38.

RUSHWORTH, A., LEE, D. AND SARRAN, C. (2016). An adaptive spatio-temporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society Series C*, DOI: 10.1111/rssc.12155.

SAHU, S., GELFAND, A. AND HOLLAND, D. (2010). Fusing point and areal level space-time data with application to wet deposition. *Journal of the Royal Statistical Society: Series C* **59**, 77–103.

SAHU, S. K., GELFAND, A. E. AND HOLLAND, D. M. (2007). High-resolution space-time ozone modeling for assessing trends. *Journal of the American Statistical Association* **102**, 1221–1234.

SAHU, S. K. AND MUKHOPADHYAY, S. (2015). On generating a flexible class of anisotropic spatial models using gaussian predictive processes. *Technical Report, University of Southampton*.

SAVAGE, N. H., AGNEW, P., DAVIS, L. S., ORDÓÑEZ, C., THORPE, R., JOHNSON, C. E., O'CONNOR, F. M. AND DALVI, M. (2013). Air quality modeling using the met office unified model (aqum os24-26): model description and initial evaluation. *Geoscientific Model Development* **6**(2), 353–372.

WAKEFIELD, J. AND SHADDICK, G. (2006). Health exposure modeling and the ecological fallacy. *Biostatistics* **7**, 438–455.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99**, 250–261.