A rigorous statistical framework for estimating the long-term health impact of air pollution

Duncan Lee[1], Sabyasachi Mukhopadhyay[2], Alastair Rushworth[3] and Sujit K. Sahu[2]

February 12, 2016

[1]

Abstract

The adverse health impact of air pollution is known worldwide, and in the UK it has been linked to over 23,000 premature deaths in 2010 alone. However, a statistically rigorous estimation of its health impacts is challenging in spatio-temporal ecological studies, since there is spatial misalignment between the pollution and disease data, uncertainty in the estimated pollution surface, and complex residual spatio-temporal autocorrelation in the disease data. If ignored, these issues may bias the estimated health effects. This article thus develops a rigorous statistical modelling framework for addressing these issues, and is motivated by a new five-year study investigating the effects of multiple pollutants on monthly respiratory hospitalisations in England between 2007 and 2011. Air pollution estimation; Bayesian spatio-temporal modelling; Health effects analysis.

1 Introduction

Air pollution remains a global public health problem, with the World Health Organisation recently estimating that it was responsible for the premature deaths of 3.7 million people under the age of 60 in 2012 (World Health Organisation (2014)). In the United Kingdom (UK) there were estimated to be over 23,000 premature deaths from air pollution in 2010 alone, with an associated cost to the economy of $83,000 million USD (World Health Organisation Regional Office for Europe (2015)). Nitrogen dioxide (NO_2) emissions are predicted to exceed European Union limits until after 2030 in some urban areas of the UK such as Greater London, West Midlands and West Yorkshire (Department for the Environment Food and Rural Affairs (2015)),

[1]Address for correspondence: *Duncan.Lee@glasgow.ac.uk* School of Mathematics and Statistics, University of Glasgow, UK, [2] Mathematical Sciences, University of Southampton, UK. [3] Department of Mathematics and Statistics, University of Strathclyde, UK.

meaning that pollution will remain a key environmental problem for some time. Health impact assessments of the long-term effects of air pollution are based predominantly on evidence from cohort studies such as Miller *et al.* (2007) and Hajat *et al.* (2013), but such studies are expensive and time consuming due to the long-term follow up period required for the cohort.

Therefore spatio-temporal areal unit studies are increasingly being used (see Best *et al.* (2000) and Greven *et al.* (2011)), which utilise geographical and temporal contrasts in air pollution and population-level disease risk. Poisson log-linear models are used in these studies, where the spatio-temporal pattern in disease risk is modelled 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 after removing covariate effects, which could be caused by unmeasured confounding, neighbourhood effects and grouping effects. However, this study design presents a number of statistical challenges, and this paper presents a rigorous statistical framework for addressing them.

The first is the estimation of air pollution at fine spatio-temporal scales, and existing studies have used either measured data (e.g. Elliott *et al.* (2007) and Greven *et al.* (2011)) or estimated concentrations from an atmospheric model (e.g. Lee *et al.* (2009) and Haining *et al.* (2010)). Bayesian statistical models for fusing modelled and measured pollution data have been proposed by Berrocal *et al.* (2009) and Sahu *et al.* (2010), and have been used in a cohort study by Sacks *et al.* (2014). However, fusion models are yet to be utilised in an areal unit study, and an additional challenge here is the *change of support problem* (Gelfand *et al.* (2001)), as there will be spatial variation in the estimated pollution surface within an areal unit. This problem is typically ignored by computing the average concentration in each unit (see Elliott *et al.* (2007), Lee *et al.* (2009)), which leads to the possibility of ecological bias (see Wakefield and Shaddick (2006)). A further source of pollution uncertainty that should be accounted for arises from posterior uncertainty in the point-level pollution predictions, which is typically ignored by the use of a point estimate such as the mean as in Sacks *et al.* (2014).

The other main challenge in areal unit studies is accounting for the residual spatio-temporal autocorrelation in the disease data via the random effects, and typically globally smooth Gaussian Markov random field (GMRF) priors are used (e.g. Rushworth *et al.* (2014)). However, global smoothness is too simplistic to accurately represent the residual autocorrelation, because it will likely exhibit localised structure that is present between some pairs of adjacent areal units but absent between other pairs. Such localised structure is present in the England study motivating this paper (see Figure 1 top right), and Lee *et al.* (2014) and Lee and Sarran (2015) have shown that utilising globally smooth random effect models in this context produces poor health effect estimates in a purely spatial domain. Thus this paper makes two key contributions to the lit-

erature. 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, UK, 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. Hospital admission records with a primary diagnosis of respiratory disease were aggregated to obtain hospitalisation counts Y_{kt} for LUA k and month t. These counts have a median value of 111 (range 6 to 2485), and monthly aggregation was used since the aim of the study was to estimate the chronic effects of pollution. Additionally, aggregation to a finer temporal scale would result in confidentiality issues resulting from small disclosive counts in some LUA.

The magnitude of Y_{kt} depends on the size and demographic structure of the population at risk, which is controlled for by computing the expected number of hospital admissions E_{kt} using indirect standardisation from national age and sex specific hospitalisation rates (as in Lawson *et al.* (2012)) . The Standardised Morbidity Ratio, $\text{SMR}_{kt} = Y_{kt}/E_{kt}$ is an exploratory measure of risk, and a value of 1.2 corresponds to a 20% increased risk compared to E_{kt}. The spatial (top left panel) and temporal (bottom panel) patterns in SMR are displayed in Figure 1, where the former shows the highest risks are observed in cities in the centre and north of England, such as Birmingham, Leeds and Manchester. The temporal pattern is strongly seasonal, with higher risks of admission in the winter due to factors such as influenza epidemics. 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 behaviour.

To mitigate against potential confounding, a number of covariates were obtained. In spatial epidemiological studies such as Haining *et al.* (2010) the key spatial confounder is socio-economic deprivation, because areas that are impoverished have worse health on average than more affluent areas. However, poverty is multi-dimensional and difficult to measure, and we represent it by proxy measures of unemployment rate and property price. Firstly, we obtained the proportion of the working age population in receipt of Jobs Seekers Allowance (denoted *JSA*), a benefit

paid to those without employment. Secondly, we obtained average property price (denoted *Price*) data, and both variables were downloaded from the UK Data Archive (*http://www.data-archive.ac.uk*). The confounding effects of population demography have been partially accounted for in E_{kt} (population size and age-sex structure), an offset in the regression models, and we additionally control for the potential effect of ethnicity via a covariate measuring the percentage of the population of each LUA that were born in the UK as of the 2001 UK Census (downloaded from *www.hscic.gov.uk/*). Finally, in time series study designs the key confounder is temperature, and although the large-scale seasonal pattern is controlled for by seasonally adjusting $\{E_{kt}\}$, we adjust for any year-to-year variation by including monthly mean temperature as a covariate in the model.

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 $2.5\mu m$ ($PM_{2.5}$) in size 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), and are displayed in the left panel of Figure 2. These data were aggregated to a monthly temporal resolution by averaging to align with the disease data. These data 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 summarised in Table 1 in the supplementary material, 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 specifically noticeable that there is a large increase in the number of monitoring sites that measure $PM_{2.5}$ after 2008.

Numerical and graphical summaries of these data are provided in the supplementary material (Table 2 and Figure 1) by site type, which shows that 16 sites are classified as Rural, 80 as Urban (which includes suburban sites) and 46 as roadside or kerbside (RKS). The table and figure 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. The figure also shows that RKS sites have a larger spread for NO_2, $PM_{2.5}$ and PM_{10} due to many large extreme observations. Finally, the concentrations vary little from year to year, which is displayed in Figure 2 in the supplementary material.

Estimated hourly concentrations are also available from the Air Quality Unified Model (AQUM) developed by Savage *et al.* (2013), on the corners of a 12 kilometer square grid covering England (see the right panel of Figure 2). AQUM is a 3-dimensional weather and chemistry transport

model, and is used by the Met Office to deliver the UK national air quality forecast for 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 and the hospitalisation counts. 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 shows scatter plots of the measured against the modelled 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 modelled concentrations. To reduce these biases, and to interpolate at unobserved locations, appropriate statistical methods are required and our fusion modelling in Section 3.1 achieves this. Our use of AQUM output frees us from needing emission or meteorological data, since these variables are not significant after including AQUM outputs (Sahu and Bakar (2012)).

Finally, we predict pollution concentrations at the corners of the 12 kilometre grid across England, but this results in 49 LUAs (out of the 323) not containing a prediction location. We rectify this by inserting an additional prediction point within each of these 49 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 Poisson generalised linear model (assuming independence) was applied to the disease data, with $PM_{2.5}$ concentrations, JSA, Price (square root transformation as it improved the fit), temperature and ethnicity as covariates. The residuals from this model exhibit moderate spatio-temporal autocorrelation, with a median Moran's I statistic of 0.363 (spatial) and median 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 shows localised spatial autocorrelation which is strong between some neighbouring 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 modelled and measured data, incorporate both spatial variation and posterior uncertainty in the pollution predictions when estimating the health effects, and control for complex localised 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 two 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 for this model is based on Markov chain Monte Carlo (MCMC) simulation

3.1 Stage 1 - pollution fusion model

3.1.1 Model specification

Let $z(\mathbf{s}_i, t)$, and $x(\mathbf{s}_i, t)$ denote the square root of the measured and modelled AQUM pollution concentrations 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 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 $\epsilon(\mathbf{s}_i, t)$:

$$\begin{aligned} Z(\mathbf{s}_i, t) &= \mu(\mathbf{s}_i, t) + \eta(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \qquad \epsilon(\mathbf{s}_i, t) \sim \mathrm{N}(0, \sigma^2_\epsilon), \\ \mu(\mathbf{s}_i, t) &= \gamma_0 + \gamma_1 x(\mathbf{s}_i, t) + \sum_{j=2}^{r} \delta_j(\mathbf{s}_i) \left(\gamma_{0j} + \gamma_{1j} x(\mathbf{s}_i, t)\right). \end{aligned} \tag{1}$$

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, 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 additionally include a spatially varying coefficient process, $\gamma(\mathbf{s}_i)$, for the $x(\mathbf{s}_i, t)$ to account for spatially varying multiplicative bias. However, we do not consider this here because it is likely to mask the site-type specific effects of $x(\mathbf{s}_i, t)$, resulting in weak model identifiability and slow MCMC convergence.

We consider three modelling 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 zero mean and a Matérn covariance function:

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

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. We do not include an auto-regressive extension of model (2) Sahu *et al.* (2007) here, because this model can thus be compared to simple Kriging capturing only spatial autocorrelation. The third model has a non-stationary covariance structure following

Sahu and Mukhopadhyay (2015), which is based on predictive process methodology (Banerjee *et al.* (2008)). First, define m knot-locations, $\mathbf{S}_m^* = (\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*)$, where m is chosen to optimise out of sample predictive performance. 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 (2). Then our non-stationary model replaces $\eta(\mathbf{s}_i, t)$ in (1) by $\tilde{\eta}(\mathbf{s}_i, t) = \mathbb{E}[\eta(\mathbf{s}_i, t)|\boldsymbol{\eta}_t^*]$, 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 modelled jointly by the zero-mean GP given in (2). Thus by writing $\tilde{\boldsymbol{\eta}}_t = (\tilde{\eta}(\mathbf{s}_1, t), \ldots, \tilde{\eta}(\mathbf{s}_n, t))^\top$ we have

$$\tilde{\boldsymbol{\eta}}_t = \mathbf{C}^*(\phi, \nu)\mathbf{H}_{\eta^*}^{-1}(\phi, \nu)\boldsymbol{\eta}_t^*, \tag{3}$$

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:

$$\boldsymbol{\eta}_t^* \sim N\left(\varrho\boldsymbol{\eta}_{t-1}^*, \sigma_\eta^2\mathbf{H}_{\eta^*}(\phi, \nu)\right), \quad \text{for } t = 1, \ldots, T, \tag{4}$$

with $\boldsymbol{\eta}_0^* = \mathbf{0}$ and ϱ is the autoregressive parameter. The number of knots m is chosen by cross-validation, and 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 kilometre 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 normalised 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 gamma$(2, 1)$), 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 minimises out of sample prediction errors, 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 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 *et al.* (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

$$\hat{Z}_{kt} \;=\; \frac{1}{|A_k|}\int_{A_k} Z(\mathbf{s},t)d\mathbf{s} \;\approx\; \frac{1}{n_k}\sum_{j=1}^{n_k} Z(\mathbf{v}_{kj},t), \tag{5}$$

where $|A_k|$ is the area of the kth LUA and $(\mathbf{v}_{k1},\ldots,\mathbf{v}_{kn_k})$ form a grid of prediction locations within the kth LUA. However, $Z(\mathbf{v}_{kj},t)$ in (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 summarised by its posterior predictive distribution:

$$\pi(z(\mathbf{v}_{kj},t)|\mathbf{z}) = \int \pi(z(\mathbf{v}_{kj},t)|\mathbf{S}^*_m,\boldsymbol{\eta}^*,\boldsymbol{\theta},\mathbf{z})\pi(\mathbf{S}^*_m,\boldsymbol{\eta}^*,\boldsymbol{\theta}|\mathbf{z})d\mathbf{S}^*_m d\boldsymbol{\eta}^* d\boldsymbol{\theta}, \tag{6}$$

where $\boldsymbol{\theta}=(\boldsymbol{\gamma},\varrho,\sigma^2_\epsilon,\sigma^2_\eta,\phi,\nu)^T$, $\boldsymbol{\eta}^*=(\boldsymbol{\eta}^*_1,\ldots,\boldsymbol{\eta}^*_T)$, 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), 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 (6) by composition sampling, and a corresponding sample from the posterior predictive distribution of (5) is computed as $z^{(\ell)}_{kt} = \frac{1}{n_k}\sum_{j=1}^{n_k} z^{(\ell)}(\mathbf{v}_{kj},t)$. The uncertainty in the LUA level pollution predictions are summarised in these L samples, but a point estimate $\hat{z}_{kt}=\frac{1}{L}\sum_{\ell=1}^{L}\hat{z}^{(\ell)}_{kt}$ 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 in the kth LUA and tth month (Y_{kt},E_{kt}) are related to a vector of covariates $\mathbf{u}_{kt}$ including measures of poverty, ethnicity and temperature, and a representative measure of 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 (6) at location $\mathbf{v}_{kj}$. A common model for these data is given by

$$\begin{aligned} Y_{kt} &\sim \text{Poisson}\,(E_{kt}R_{kt}) \quad k=1,\ldots,K,\;\; t=1,\ldots,T, \\ R_{kt} &= \exp(\mathbf{u}^\top_{kt}\boldsymbol{\beta}+\hat{z}_{kt}\beta_z+\psi_{kt}), \end{aligned} \tag{7}$$

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 *et al.* (2014) uses the autoregressive decomposition:

$$\begin{aligned} \boldsymbol{\psi}_t|\boldsymbol{\psi}_{t-1} &\sim \text{N}\left(\alpha\boldsymbol{\psi}_{t-1}, \tau^2\mathbf{Q}(\mathbf{W},\rho)^{-1}\right) \qquad t = 2, \ldots, T, \\ \boldsymbol{\psi}_1 &\sim \text{N}\left(\mathbf{0}, \tau^2\mathbf{Q}(\mathbf{W},\rho)^{-1}\right), \end{aligned} \tag{8}$$

where $\alpha, \rho \sim \text{Uniform}(0,1)$ while $\tau^2 \sim \text{IG}(2,1)$. Temporal autocorrelation is induced by the mean function $(\alpha\boldsymbol{\psi}_{t-1})$ while spatial autocorrelation is induced 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 *et al.* (1999), 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 $K \times K$ neighbourhood 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:

$$\psi_{k1}|\boldsymbol{\psi}_{-k1} \sim \text{N}\left(\frac{\rho\sum_{i=1}^K w_{ki}\psi_{i1}}{\rho\sum_{i=1}^K w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{i=1}^K w_{ki} + 1 - \rho}\right). \tag{9}$$

Here the conditional expectation is a weighted average of the random effects in neighbouring areal units, with the amount of spatial smoothing controlled globally by ρ. However, (7) - (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 (7) when there is within-area variation in pollution can lead to (ecological) bias in its estimated health effects. Consider an idealised model for (Y_{ktj}, E_{ktj}), which relate to the proportion of the population who experience pollution exposure $\hat{z}(\mathbf{v}_{kj}, t) = \frac{1}{L}\sum_{\ell=1}^L z^{(\ell)}(\mathbf{v}_{kj}, t)$. 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 $(\mathbf{v}_{k1}, \ldots, \mathbf{v}_{kn_k})$, an appropriate aggregated model for $Y_{kt} = \sum_{j=1}^{n_k} Y_{ktj}$ is:

$$\begin{aligned} Y_{kt} &\sim \text{Poisson}\left(E_{kt}R_{kt}\right), \\ R_{kt} &= \exp(\mathbf{u}_{kt}^\top\boldsymbol{\beta} + \psi_{kt})\sum_{j=1}^{n_k} E^*_{ktj}\exp(\hat{z}(\mathbf{v}_{kj}, t)\beta_z), \end{aligned} \tag{10}$$

where $E^*_{ktj} = E_{ktj}/E_{kt}$ and $\sum_{j=1}^{n_k} E^*_{ktj} = 1$. The consequence is that (7) exponentiates the spatially averaged pollution concentrations $\hat{z}_{kt}$ while (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 show 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 (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:

$$\begin{aligned} Y_{kt} &\sim \text{Poisson}\left(E_{kt}R_{kt}\right) \quad k = 1, \ldots, K, \quad t = 1, \ldots, T, \\ R_{kt} &= \exp(\mathbf{u}_{kt}^{\top}\boldsymbol{\beta} + Z_{kt}\beta_z + \psi_{kt}), \\ Z_{kt} &\sim \pi(z_{kt}|\mathbf{z}). \end{aligned} \tag{11}$$

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

3.2.3 (C) - Localised spatio-temporal autocorrelation

The global smoothness assumption of (8) is unrealistic, because as evidenced by Figure 1 the residual spatial autocorrelation often exhibits a mixture of spatial smoothness and sharp discontinuities. We account for this by allowing spatially neighbouring random effects to be correlated (inducing smoothness) or conditionally independent (no smoothing), by modelling the non-zero elements of the neighbourhood 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 *et al.* (2015), because it is the only such model designed for a spatio-temporal setting. Each adjacency parameter in $\mathbf{w}^{+}$ is modelled on the interval $[0, 1]$, by placing a multivariate Gaussian prior on the transformation $\mathbf{g}^{+} = \log\left(\mathbf{w}^{+}/(\mathbf{1} - \mathbf{w}^{+})\right)$. We utilise a shrinkage prior for $\mathbf{g}^{+}$ with a constant mean and variance (μ, ζ^2), which is given by:

$$f(\mathbf{g}^{+}|\zeta^2, \mu) \propto \exp\left[-\frac{1}{2\zeta^2}\left(\sum_{g_{ik} \in \mathbf{g}^{+}} (g_{ik} - \mu)^2\right)\right], \quad \zeta^2 \sim \text{IG}(2, 1). \tag{12}$$

Here the random effects surface $\mathbf{g}^{+}$ is not smoothed spatially, for example by a second level GMRF prior, because work by Rushworth *et al.* (2015) 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 *et al.* (2015) 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 $\zeta^2 \to 0$ the prior becomes the global smoothing prior (8). Further discussion of these points can be found in Rushworth *et al.* (2015) .

4 Results

We now present the results of the England respiratory hospitalisation 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 modelling results

Table 1 summarises the predictive performance of a range of pollution models for all four pollutants (NO_2, O_3, PM_{10}, $PM_{2.5}$), based on a cross-validation exercise. Each model was fitted to a random subset of 127 monitoring sites, and was then used to predict the concentrations at the remaining 15 sites (see Figure 2). The first row of the table for each pollutant corresponds to simple Kriging performed separately for each month, which was implemented using the `fields` package (Furrer *et al.* (2013)) in `R`. This simple strategy provides a benchmark against which to compare the results of the proposed models. These include model (1) with $\eta(\mathbf{s}_i, t) = 0$ denoted (*Linear*), the same model with $\eta(\mathbf{s}_i, t)$ modelled by the GP (2) denoted (*GP*), and the full non-stationary model with $\tilde{\eta}(\mathbf{s}_i, t)$ given by (3) - (4). 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) kilometers 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 $\phi_1 \dots \phi_5$ 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.

All results are based on 5000 MCMC iterations, which were obtained after discarding the first 1000 iterations at which point convergence was assessed to have been reached. In all cases we take $m = 25$, which was chosen by an out of sample validation method among the possible values of 16, 25, 36, 49 and 100. A number of model performance criteria are presented in Table 1, including bias, root mean square prediction error (RMSPE), mean absolute prediction error (MAPE), coverage of the 95% prediction intervals and the Bayesian predictive model choice criterion (PMCC) proposed by Gelfand and Ghosh (1998) which one wishes to minimise. Using Kriging as a benchmark, the simple *Linear* and *GP* models generally perform poorer than Kriging with

regard to RMSPE and MAPE with the exception of NO_2. However, the full non-stationary model outperforms Kriging for all optimal values of the spatial range parameter ϕ, exhibiting lower RMSPE and either similar or slightly better coverage probabilities. Additionally, for NO_2 and O_3 Kriging exhibits bias unlike the models proposed here, which is positive for NO_2 and negative for O_3. The empirical coverage values are lower for all four models, including Kringing and linear, for PM than either of NO_2 and O_3. This may be due to the high percentage of missing data for both PM_{10} and $PM_{2.5}$ as evidenced by the lower validation sample sizes, 301 and 231, for these two pollutabnts compared to, 601 and 364, for the other two. In the next section we apply the best performing model in terms of RMSPE to data from all $n = 142$ monitoring locations to make 5000 predictions of pollution at each of the 1516 grid box corners across England for each of the 60 months in the study.

4.2 Disease modelling results

Estimated single-pollutant health effects from a range of disease models are displayed in Table 2, where the models are summarised in Table 3. Single-pollutant analyses were undertaken in common with the majority of the existing literature, and because it is single pollutant effect estimates that inform regulatory standards. These considerations also justfiy our use of univariate fusion modelling of the pollutants instead of multivariate ones. All results are presented as relative risks for a one standard deviation increase in each pollutants value, which is 16.07, 7.30, 4.90, and 4.11 (μgm^{-3}), respectively for NO_2, O_3, PM_{10} and $PM_{2.5}$. The covariates in each model include one of the four pollutants, two measures of socio-economic deprivation (JSA and the square root of property price), and the measures of ethnicity and temperature. The effects of these covariates are presented below for model M1 (see description below) as a posterior median and 95% credible interval: 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 ethnicity and increased disease risk, although this is likely confounded by inner-city deprivation, where ethnic diversity tends to be largest. House price is also a good proxy for overall deprivation, with increased house prices strongly associated with decreases in risk. Finally, colder temperatures are associated with increased respiratory admissions as expected. Inference for all models in this section is based on 500,000 MCMC iterations, with the first 100,000 being discarded as the burn-in period. These choices ensured that convergence was reached for each model and good mixing was achieved for all parameters. The remainder of this section quantifies the impact on health effects estimation of three aspects of the model, which coincide with labels (A) - (C) in Section 3.2.

4.2.1 (A) - Allowing for spatial variation in pollution within an areal unit

We first assess the impact of ignoring or allowing for spatial variation in pollution concentrations within an LUA, the former possibly resulting in ecological bias. This is achieved by comparing the ecological model (7) denoted by **M1** with the correct aggregate model (10) denoted by **M2**, and for the latter E_{ktj} were the estimated expected counts around each prediction location computed using electoral ward level (a much smaller unit than LUA) population data. In both cases uncertainty in the estimated pollution surface is ignored and the random effects are modelled by the localised GMRF model given by (8) and (12). A comparison of models **M1** and **M2** in Table 2 illustrates a number of key points. First, the estimated relative risks are almost identical between the two models for each pollutant, suggesting that ignoring spatial variation in pollution within an areal unit does not bias the estimated health effects. This is because the estimated effect sizes are small, and thus the bias term discussed in Wakefield and Shaddick (2006) is very small (of the order of β_z^2 where for NO_2 $\hat{\beta}_z = 0.00160$). This result was also observed by Lee and Sarran (2015), and they additionally showed by simulation that negligible bias would occur unless the effect size was much larger. As a result within-area variation in pollution is ignored in the remainder of this paper.

The second key point 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 uncertainty in the estimated pollution data to be propagated into the disease model

We assess the impact of ignoring or propagating uncertainty in the estimated pollution surface into the disease model by comparing the three approaches discussed in Section 3.2.2. Model **M1** ignores this uncertainty by using the posterior predictive median $\hat{z}_{kt}$, model **M3** re-samples a new $\{\hat{z}_{kt}^{(\ell)}\}$ value at each step of the MCMC algorithm when implementing the disease model (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). Firstly, 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 substantial NO_2 result 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%, while the $PM_{2.5}$ estimate is unchanged.

4.2.3 (C) - Allowing for residual spatio-temporal autocorrelation

Finally, we compare the impact on health effects estimation of changing the mechanism for capturing unmeasured residual spatio-temporal autocorrelation. Model **M5** ignores this autocorrelation, that is $\psi_{kt} = 0$, model **M6** uses the globally smooth model proposed by Rushworth *et al.* (2014), while model **M1** uses the localised model described here. 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 pollution effect sizes between the globally smooth and locally smooth random effects 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 real relationship, then the results are much more consistent across the different random effects specifications considered here.

A comparison of the fit to the data of the global (**M6**) and local (**M1**) random effects models shows Deviance Information Criterion values of: **M6** - 157,131, **M1** - 155,538, indicating a moderate improvement in overall fit under the localised approach. Additionally, the effective number of parameters (pd) underlying this improvement decreases substantially for the localised model (6205 compared to 8198), despite it being a more complex model in terms of algebraic formulation. This occurs because the posterior random effects variance, τ^2 is much lower for the localised 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 *et al.* (2015). The localised nature of the spatial autocorrelation in the random effects is determined by $\mathbf{w}^+$, and for these data $\mathbf{w}^+$ comprises 861 adjacency elements. Posterior inference for $\mathbf{w}^+$ is summarised in Section 5 of the supplementary material via $\varrho_{kj} = \mathbb{P}(w_{kj} < 0.5|\mathbf{Y})$, the posterior probability of the adjacency element w_{kj} being less than 0.5 which is the mid-point of the allowable unit interval.

5 Discussion

This article develops a statistically rigorous framework for predicting pollution concentrations and then estimating their effects on health for an ecological spatio-temporal study design. The disease model is the first of its kind to simultaneously account for spatial variation in the pollution concentrations within an areal unit, correctly propagate the uncertainty in the estimated pollution surface into the disease model, and control for unmeasured confounding via a flexible localised GMRF model that has been shown to outperform a commonly used global smoothing alternative. Thus it is the most statistically rigorous model ever developed for the spatio-temporal areal unit study design considered here. Software in the form of `R` code to implement the full Bayesian two-stage model is available on request, but the disease model with the localised GMRF prior can be implemented in the `R` package `CARBayesST`.

The methodological development here was motivated by a large study of air pollution and respiratory hospitalisations in England, which is the largest and most comprehensive study of its type ever to be conducted in England, in terms of its spatio-temporal coverage and sheer volume of data. The main conclusion of interest from a public health perspective is that increased exposure to concentrations of NO_2 is associated with an increased risk of respiratory ill health, with a $16.07\mu gm^{-3}$ increase in concentrations being associated with around a 2.8% increased risk of hospitalisation (assuming the localised GMRF model). This equates to around 17,000 more admissions across England per year on average over the study period, as there are around 613,000 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. The estimated effects for particulates were mostly much smaller (except for PM_{10} under **M4**) and typically the uncertainty intervals included the null risk of one. This suggests that it is NO_2 that poses the greatest ongoing health risk. We note in passing that it was not possible to study the health effect of ultra fine particulates as they are not routinely measured in the UK.

From a statistical perspective, the similarity in the estimated effects of allowing for or ignoring within area variation in exposure corroborate those found by Lee and Sarran (2015), and suggest that ecological bias is not a big problem in these types of 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 the localised model proposed here proved a better fit to the data, as measured by DIC, than the globally smooth GMRF model. However, if no association exists then the three autocorrelation models considered here show similar results. Finally, the approach to allowing for uncertainty in the estimated pollution surface when estimating its health effects has a substantial impact on the results, and treating

the pollution concentrations as a random variable to be updated in the model with an informative prior distribution appears to have worked well.

The modelling developments and case study presented here suggest several further avenues for research. The first is the estimation of the cost of pollution, which can be obtained by combining the health effects estimated here with data on government health spending to estimate the financial as well as public health cost of environmental pollution. A second important application of the developed methodologies will be to try these for 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 so doing, will be to gather reliable data on air pollution exposure as well as health outcomes. Further methodological developments, such as multivariate space-time modelling for air pollution and providing an index of overall air quality will likely bear fruitful research, as will using the model developed here to estimate its impact on human health. Additionally, a multivariate disease model could be developed, which allows the joint and separate effects of air pollution on multiple diseases to be investigated. Lastly, development of software packages and their documentation to automate the procedures will also be worthwhile investigating.

6 Software

Software in the form of `R` code to implement our model is available upon request and will be published along with the paper. A simplified version of the disease model (without the uncertainty propagation or ecological bias correction) is available via the `R` package `CARBayesST`.

Acknowledgments

The authors gratefully acknowledge the Health and Social Care Information Centre for providing the disease data, and to the UK Met Office for compiling them to LUA level and for providing the AQUM data. *Conflict of Interest*: None declared.

References

Bakar, K. S. and Sahu, S. K. (2015). `spTimer`: Spatio-temporal bayesian modelling using `r`. *Journal of Statistical Software*, **63**(15), in press.

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.

Best, N., Ickstadt, K., and Wolpert, R. (2000). Spatial Poisson Regression for Health and Exposure Data Measured at Disparate Resolutions. *Journal of the American Statistical Association*, **95**, 1076–1088.

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.

Furrer, R., Nychka, D., and Sain, S. (2013). `fields`*: Tools for Spatial Data*. National Center for Atmospheric Research, Boulder, Colorado. `R` package version 6.9.1.

Gelfand, A., Zhu, L., and Carlin, B. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, **2**, 31–45.

Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, **85**, 1–11.

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.

Guhaniyogi, R., Finley, A. O., Banerjee, S., and Gelfand, A. E. (2011). Adaptive gaussian predictive process models for large spatial datasets. *Environmetrics*, **22**(8), 997–1007.

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 Spatio-temporal Epidemiology*, **1**, 123–131.

Hajat, A., Diez-Roux, A., Adar, S., Auchincloss, A., Lovasi, G., ONeill, N., Sheppard, L., and Kaufman, J. (2013). Air Pollution and Individual and Neighborhood Socioeconomic Status: Evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environmental Health Perspectives*, **121**, 1325–1333.

Lawson, A., Choi, J., Cai, B., Hossain, M., Kirby, R., and Liu, J. (2012). Bayesian 2-Stage Space-Time Mixture Modeling With Spatial Misalignment of the Exposure in Small Area Health Data. *Journal of Agricultural, Biological and Environmental Statistics*, **17**, 417–441.

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

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

Lee, D., Rushworth, A., and Sahu, S. (2014). A Bayesian Localised Conditional Autoregressive Model for Estimating the Health Effecs of Air Pollution. *Biometrics*, **70**, 419–429.

Leroux, B., Lei, X., and Breslow, N. (1999). *Estimation of disease rates in small areas: A new mixed model for spatial dependence. Statistical Models in Epidemiology, the Environment and Clinical Trials, Halloran, M and Berry, D (eds)*, pages 135–178. Springer-Verlag, New York.

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.

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 Spatio-temporal Epidemiology*, **10**, 29–38.

Rushworth, A., Lee, D., and Sarran, C. (2015). An adaptive spatio-temporal smoothing model for estimating trends and step changes in disease risk. *arXiv:1411.0924*.

Sacks, J., Rappold, A., Davis Jr, J., Richardson, D., Waller, A., and Luben, T. (2014). Influence of Urbanicity and County Characteristics on the Association between Ozone and Asthma Emergency Department Visits in North Carolina. *Environmental Health Perspectives*, **122**, 506–512.

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. and Bakar, K. S. (2012). A comparison of bayesian models for daily ozone concentration levels. *Statistical Methodology*, **9**(1), 144–157.

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*.

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.

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 modelling 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 modelling and the ecological fallacy. *Biostatistics*, **7**, 438–455.

World Health Organisation (2014). Ambient (outdoor) air quality and health: Fact Sheet 313. http://www.who.int/mediacentre/factsheets/fs313/en/.

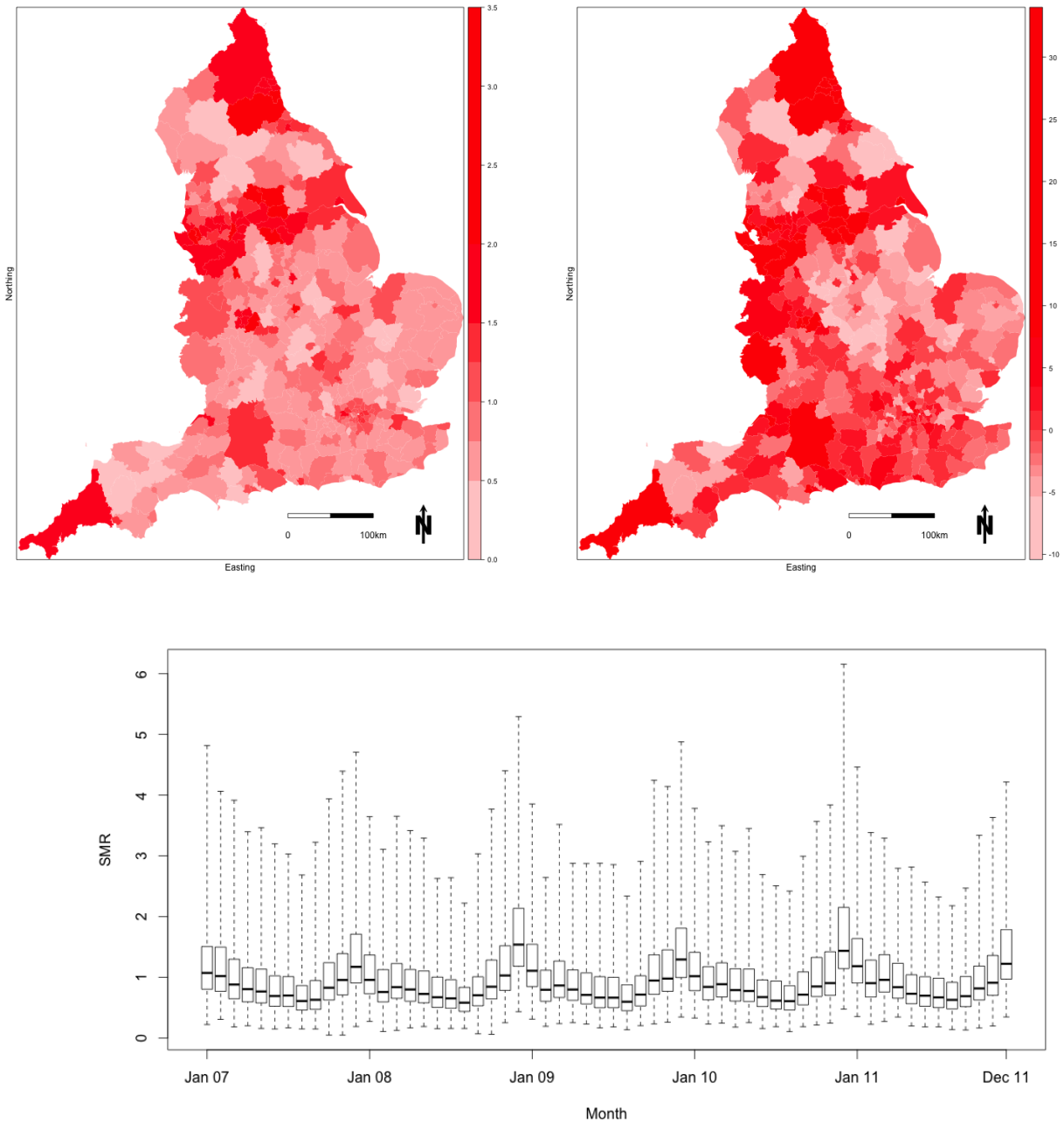
World Health Organisation Regional Office for Europe (2015). Economic cost of the health impact of air pollution in Europe: Clean air, health and wealth. WHO Regional Office for Europe.

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

Model	RMSPE	MAPE	Bias	Coverage (%)	G	P	G+P
NO_2: Fitting SS =4822, validation SS=601							
Kriging	20.12	15.26	3.48	96.13	–	–	–
Linear	13.66	10.45	–1.35	99.83	105733	8002	113735
GP	15.14	12.39	2.48	98.32	2918	18594	21511
ϕ_3	10.78	8.17	1.12	99.16	4897	62710	67607
O_3: Fitting SS =3269, validation SS=364							
Kriging	8.95	7.08	–3.44	93.31	–	–	–
Linear	9.01	7.29	–0.60	99.45	76384	2010	78394
GP	9.60	7.90	2.36	100.0	1149	5992	7141
ϕ_3	6.53	5.12	0.72	96.70	1371	18716	20087
PM_{10}: Fitting SS =2463, validation SS=301							
Kriging	3.82	2.99	0.09	88.60	–	–	–
Linear	5.65	4.69	0.32	89.23	91873	91973	183846
GP	5.71	4.72	1.10	85.34	721	3928	4649
ϕ_2	3.29	2.55	–0.03	89.70	595	7617	8212
$PM_{2.5}$: Fitting SS =1820, validation SS=231							
Kriging	2.81	1.92	–0.76	82.53	–	–	–
Linear	5.17	4.24	–0.43	81.45	46590	46679	93268
GP	5.18	4.35	1.51	81.45	595	5466	6061
ϕ_2	2.72	1.93	–0.52	83.11	330	2765	3095

Table 1: Assessment of predictive performance for a range of models for all four pollutants. SS stands for sample size which is the number of monthly observations. Here **G** denotes goodness-of-fit while **P** denotes the predictive penalty.

Figure 1: Display of the spatial (top left) and temporal (bottom) patterns in the SMR data. 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.



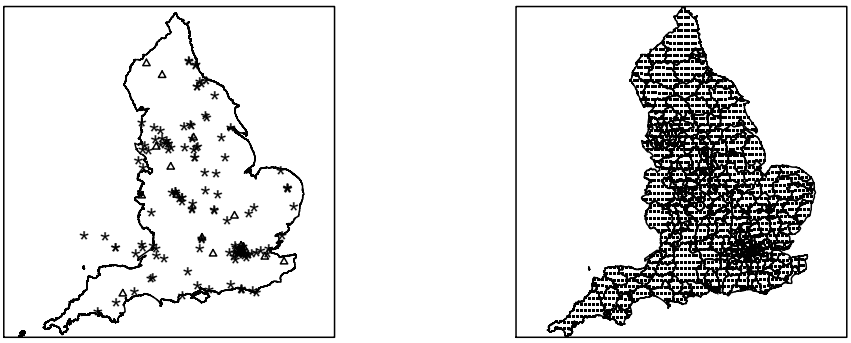


Figure 2: The left panel shows the 127 monitoring sites as stars and and 15 validation sites as triangles. The right panel shows the boundaries of the 323 local authorities in England and the corners of the 1516 predictive points. These grid points are corners of a 12 km^2 grid and some further locations added so that there is at least one prediction point within each of the 323 local authorities.

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)

Table 2: 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 16.07μgm^{-3} for NO_2, 7.30μgm^{-3} for O_3, 4.90μgm^{-3} for PM_{10}, and 4.11μgm^{-3} for $PM_{2.5}$. Note, the uncertainty intervals are credible intervals for all but model M5 where they are confidence intervals.

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

Table 3: A summary of the 6 models fitted to the disease data.

Supplementary materials for 'A rigorous statistical framework for estimating the long-term health impact of air pollution, with application to respiratory hospitalisation risk in England between 2007 and 2011'

S.1 Introduction

This supplementary material accompanies the main paper and contains the following content. Section S.2 provides additional exploratory data analysis for the pollution data used in the England study. Section S.3 provides brief details of the implementation of the pollution model, while Section S.4 provides similar details for the disease model.

S.2 Additional exploratory analysis of the pollution data

As discussed in the main paper pollution data from a monitoring network are often prone to large numbers of missing values, due to factors such as instrument malfunction, discontinuation of some sites, introduction of new sites during the study period, or the fact that not all sites monitor all pollutants. The missingness , broadly defined, for the AURN network is summarised in Table S.4.1 below, which shows constant missingness over time except for $PM_{2.5}$ which decreases after 2008 due to more sites measuring this pollutant. Numerical and graphical summaries of the observed pollution data are given by Table S.4.2 and Figures S.4.2 and S.4.3, which respectively present the data by either site type (16 Rural, 80 Urban and 46 RKS) or year. The site type figure shows greater average concentrations and greater levels of variation for RKS sites for NO_2, $PM_{2.5}$ and PM_{10}, where as the converse is true for O_3 in terms of mean concentrations. The concentrations of all four pollutants show little variation by year, with no discernible changes in average concentrations or the levels of variation.

S.3 Implementation of the pollution model

The details for fitting the independent GP models are provided in Bakar and Sahu (2015). Following are the details for implementing the space-time non-stationary model (1) and (4) discussed in the main paper. The logarithm of the full posterior distribution is given by:

$$
\begin{aligned}
\log\left(\pi\left(\mathbf{S}_m^*, \boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{z}\right)\right) &\propto -\tfrac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(z(\mathbf{s}_i,t)-\mu(\mathbf{s}_i,t)-\tilde{\eta}(\mathbf{s}_i,t)\right)^2 \\
&- \tfrac{nT}{2}\log(\sigma_\epsilon^2)+\sum_{j=1}^{m}\log(\pi(\mathbf{s}_j^*))+\log\left(\pi(\boldsymbol{\eta}_0^*)\right) \\
&- \tfrac{mT}{2}\log(\sigma_\eta^2)-\tfrac{T}{2}\log|\mathbf{H}_{\eta^*}(\phi,\nu)| \\
&- \tfrac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\left(\boldsymbol{\eta}_t^*-\varrho\boldsymbol{\eta}_{t-1}^*\right)^\top\mathbf{H}_{\eta^*}(\phi,\nu)^{-1}\left(\boldsymbol{\eta}_t^*-\varrho\boldsymbol{\eta}_{t-1}^*\right) \\
&+ \log(\pi(\boldsymbol{\theta})),
\end{aligned}
$$

where $\boldsymbol{\theta}=(\boldsymbol{\gamma},\varrho,\sigma_\epsilon^2,\sigma_\eta^2,\phi,\nu)^T$ denotes the model parameters and $\pi(\boldsymbol{\theta})$ thus denotes its joint prior distribution. Inference uses Markov chain Monte Carlo (MCMC) simulation, via a combination of Gibbs sampling and Metropolis-Hastings steps. Our implementation follows that described in Section 3.2 of Guhaniyogi *et al.* (2011), except that we take a different approach to updating the knot-locations $\mathbf{S}_m^*$. We adopt the method developed by Sahu and Mukhopadhyay (2015), where we simulate m proposed knots from the prior outlined in the main paper without replacement, and then use a Metropolis-Hastings step to accept the proposed knots.The starting configuration of the knots is taken to be according to a space filling design and we tune the algorithm to have 15-40% acceptance rate as is common in practice.

S.4 Implementation of the disease model

The major challenge in implementing the disease model defined by (11), (8), (12) is computational, because there are 19,380 spatio-temporal observations (e.g the study by Elliott *et al.*, 2007 used around 800 data points, while the largest study region used in Lee *et al.*, 2009 was around 1500). The random effects $\boldsymbol{\psi}$ with the GMRF prior are updated in the McMC routine using computationally efficient algorithms written in C++, which utilised the sparseness of the neighbourhood matrix $\mathbf{W}$ by using its triplet form representation. The elements of W were proposed in a computationally efficient manner by updating $\boldsymbol{g}^+$ in blocks of size 10. In the acceptance ratio for each block proposal, a new determinant, $\det(\boldsymbol{Q}(\boldsymbol{W},\rho))$, is required, which was calculated efficiently by updating the (sparse) Cholesky decomposition, $\boldsymbol{L}$, so that $\boldsymbol{L}\boldsymbol{L}^\top=\boldsymbol{Q}(\boldsymbol{W},\rho)$, and as a result $\det(\boldsymbol{Q}(\boldsymbol{W},\rho))=\det(\boldsymbol{L})^2=\prod_{i=1}^{K}L_{ii}^2$, since L_{ii} is triangular. The health model, without the ecological bias correction or the uncertainty propagation, can be implemented in the `R` package `CARBayesST`, which can also implement similar models with binomial and Gaussian likelihoods.

The other major computational challenge is updating the set of spatio-temporal pollution concentrations Z_{kt} from (11). These concentrations should not be updated independently for each areal unit and time period, because the pollution model from stage one produces posterior pre-

Table S.4.1: Percentage of missing monthly observations at the 142 sites in each year for each of the four pollutants.

	2007	**2008**	**2009**	**2010**	**2011**
NO_2	38.38	38.08	36.97	37.21	34.62
O_3	52.26	59.14	60.09	58.68	58.33
$PM_{2.5}$	96.53	91.43	65.96	64.32	64.03
PM_{10}	64.96	66.26	69.37	71.48	69.13

dictive realisations of the spatio-temporal pollution surface that are correlated in space and time. However, randomly selecting one of the L posterior predictive samples (for all 19,380 data points) at random as a proposal, and then accepting or rejecting it via a Metropolis-Hastings step is not feasible due to the curse of dimensionality which results in very poor acceptance rates. Therefore we propose a multivariate Gaussian approximation to the posterior predictive distribution for each time period separately, that is:

$$\pi(Z_{1t}, \ldots, Z_{Kt}|\mathbf{z})^\top = \mathrm{N}\left(\hat{\mathbf{z}}_t, \hat{\mathbf{\Sigma}}_t\right) \qquad \text{for } \; t = 1, \ldots, T,$$

where the kth element of $\hat{\mathbf{z}}_t$ is given by $\hat{z}_{kt} = \frac{1}{Ln_k} \sum_{j=1}^{n_k} \sum_{\ell=1}^{L} z^{(\ell)}(\mathbf{v}_{kj}, t)$, while the k and k'th element of $\hat{\mathbf{\Sigma}}_t$ is given by:

$$\left(\hat{\mathbf{\Sigma}}_t\right)_{kk'} = \frac{1}{L-1} \sum_{\ell=1}^{L} (\hat{z}_{kt}^{(\ell)} - \hat{z}_{kt})(\hat{z}_{k't}^{(\ell)} - \hat{z}_{k't}).$$

While this specification disregards the temporal dependence between exposures that occur for a single region, the spatial dependence structure is largely preserved. The Gaussian approximation allows the univariate prior conditional distributions, $Z_{kt}|\mathbf{Z}_{-kt}$ for $k = 1, \ldots, K$ (where $\mathbf{Z}_{-kt}$ denotes the vector with the kth element removed) to be easily computed, making Metropolis-Hastings updating one element at a time straightforward. We note that a full spatio-temporal approximation could also be used, but that this would require a $KT \times KT$ covariance matrix to be constructed, from which it would be computationally demanding to calculate conditional distributions.

Type	Min	Median	Mean	Max	Sd	*N*
			NO_2			
Rural (16)	3.21	17.30	19.53	52.79	9.68	696
Urban (80)	8.67	48.18	49.81	135.90	17.92	2952
RKS (46)	17.77	71.55	76.54	227.30	33.27	1775
All (142)	3.21	50.46	54.67	227.28	29.61	5423
			O_3			
Rural (16)	30.03	67.06	68.60	113.40	13.29	788
Urban (80)	13.38	54.62	56.00	120.70	16.01	2523
RKS (46)	10.99	49.12	48.58	101.40	17.64	322
All (142)	10.99	57.72	58.08	120.73	16.69	3633
			PM_{10}			
Rural (16)	4.10	16.70	16.82	32.04	5.68	153
Urban (80)	7.16	19.09	20.03	44.84	6.24	1623
RKS (46)	5.79	20.99	22.74	54.01	7.92	988
All (142)	4.10	19.62	20.82	54.01	7.05	2764
			$PM_{2.5}$			
Rural (16)	5.38	10.11	10.72	28.49	3.64	110
Urban (80)	2.66	12.95	13.72	36.45	5.27	1321
RKS (46)	4.11	13.66	14.71	36.22	6.08	620
All (142)	2.66	12.87	13.86	36.45	5.53	2051

Table S.4.2: Summary of the monthly pollution data for the four pollutants from the 16 Rural, 80 Urban and 46 RKS sites over the 5 years. All pollutants are measured in μgm^{-3}.Here Sd stands for standard deviation and N is the number of available monthly averages on which the summaries have been calculated.

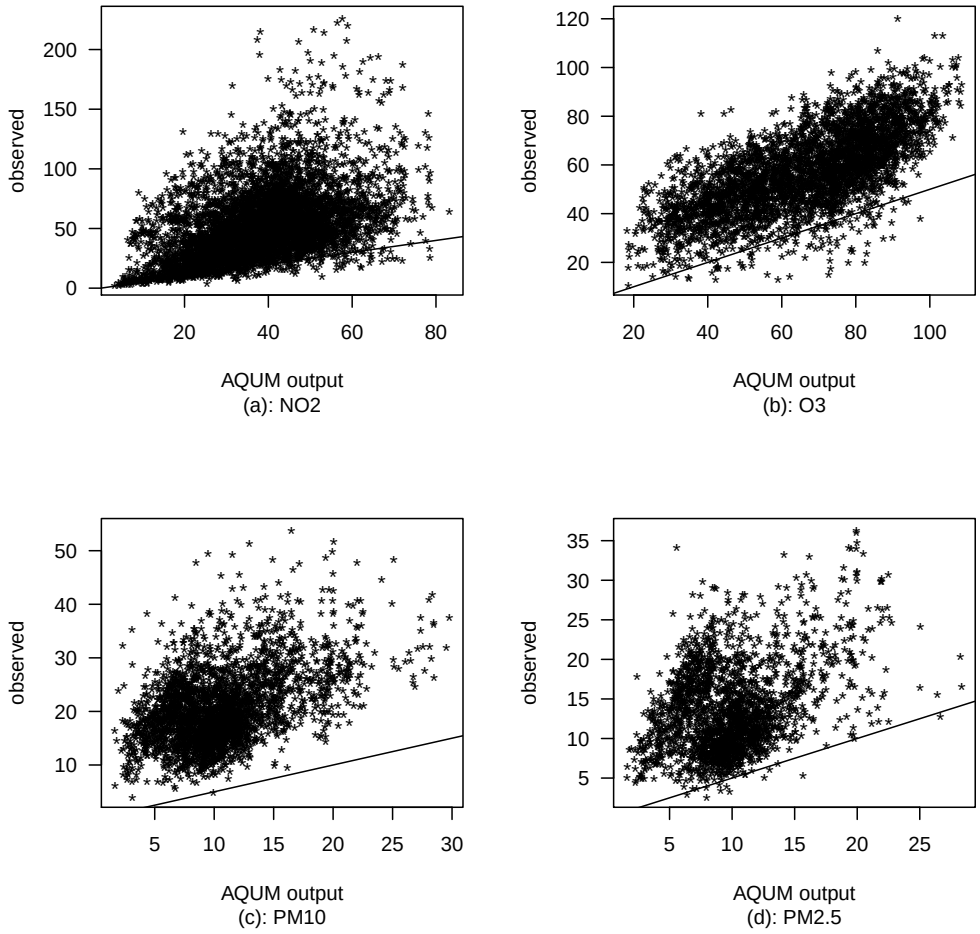


Figure S.4.1: Scatterplot of the observed concentrations against the estimated AQUM model output for each of the four pollutants. The line $y = x$ is superimposed.

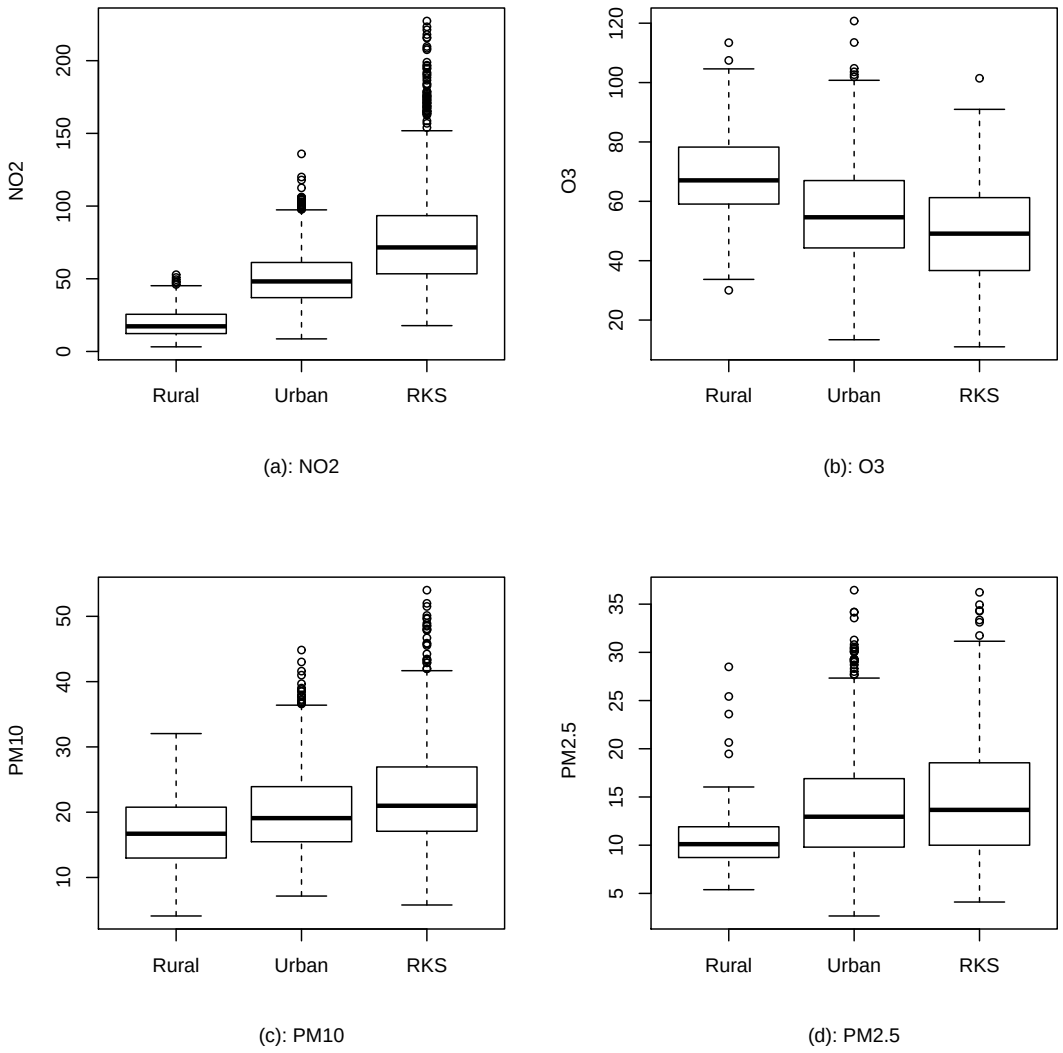
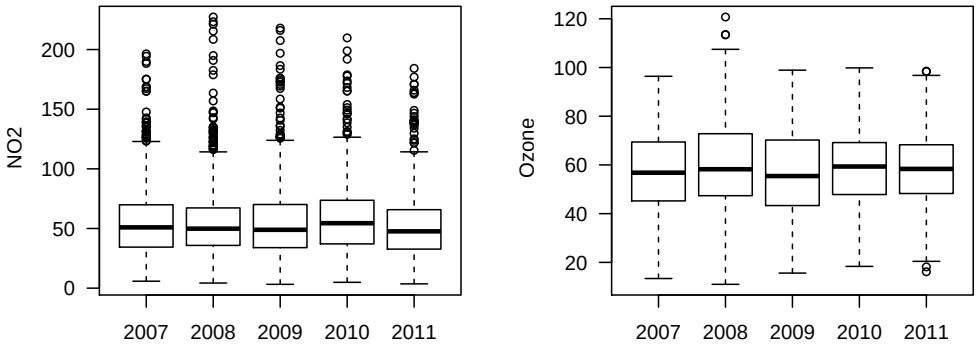
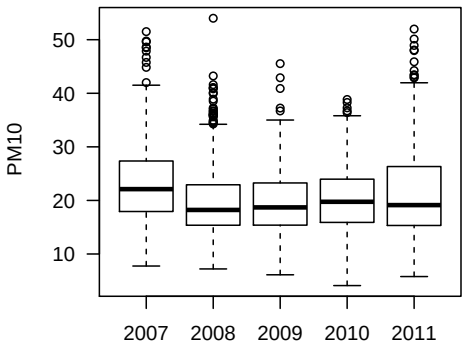


Figure S.4.2: Boxplots of the monthly average concentrations for each pollutant by three site types.

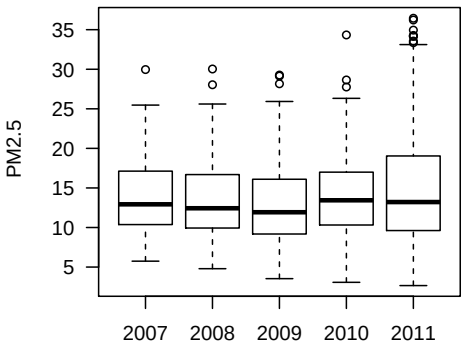


(a): NO2

(b): O3



(c): PM10



(d): PM2.5

Figure S.4.3: Boxplots of the monthly average concentrations for each pollutant by year.