

# How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using Bayesian model averaging



Francesca Pannullo<sup>a,\*</sup>, Duncan Lee<sup>b</sup>, Eugene Waclawski<sup>c</sup>, Alastair H. Leyland<sup>a</sup>

<sup>a</sup> MRC|CSO Social and Public Health Science Unit, University of Glasgow, 200 Renfield Street, Glasgow, G2 3QB, UK

<sup>b</sup> School of Mathematics and Statistics, University of Glasgow, 15 University Gardens, Glasgow, G12 8QW, UK

<sup>c</sup> Public Health, University of Glasgow, 1 Lilybank Gardens, Glasgow, G12 8RZ, UK

## ARTICLE INFO

### Article history:

Received 10 November 2015

Revised 16 February 2016

Accepted 1 April 2016

Available online 14 April 2016

### Keywords:

Bayesian model averaging

Conditional autoregressive models

Nitrogen dioxide

Spatial autocorrelation

## ABSTRACT

The long-term impact of air pollution on human health can be estimated from small-area ecological studies in which the health outcome is regressed against air pollution concentrations and other covariates, such as socio-economic deprivation. Socio-economic deprivation is multi-factorial and difficult to measure, and includes aspects of income, education, and housing as well as others. However, these variables are potentially highly correlated, meaning one can either create an overall deprivation index, or use the individual characteristics, which can result in a variety of pollution-health effects. Other aspects of model choice may affect the pollution-health estimate, such as the estimation of pollution, and spatial autocorrelation model. Therefore, we propose a Bayesian model averaging approach to combine the results from multiple statistical models to produce a more robust representation of the overall pollution-health effect. We investigate the relationship between nitrogen dioxide concentrations and cardio-respiratory mortality in West Central Scotland between 2006 and 2012.

© 2016 The Authors. Published by Elsevier Ltd.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Air pollution has repeatedly been shown to have a detrimental impact on human health, with one of the earliest prominent examples being the London smog episode of 1952 (Health, 1954), which resulted in more than 3000 excess deaths compared with previous years. High pollution episodes such as this have led to the implementation of air pollution legislation, such as the Clean Air Acts in 1956 and 1993 in the UK, and the 2008 Ambient Air Quality Directive in the European Union (EU). These pol-

icy practices have led to a reduction in air pollution concentrations in many parts of the world, however a recent report by the World Health Organisation estimates that outdoor air pollution contributed to 3.7 million premature deaths in people under the age of 60 in 2012 (Organisation, 2014). Air pollution remains a serious public health problem in the UK, as nitrogen dioxide (NO<sub>2</sub>) concentrations currently do not meet the targets set by EU legislation. For example, in West Central Scotland, where the study presented in this paper is based, NO<sub>2</sub> concentrations are predicted to exceed these targets until 2020 (Department for the Environment and Affairs, 2015).

The impacts of both long-term (chronic) and short-term (acute) exposure to pollution have been much researched, with the latter being estimated from daily ecological

\* Corresponding author. Tel.: +44 141 353 7637.

E-mail address: [f.pannullo.1@research.gla.ac.uk](mailto:f.pannullo.1@research.gla.ac.uk) (F. Pannullo).

time-series studies (see Willocks et al., 2012). The long-term health impact of pollution is most often estimated from cohort studies (see Cesaroni et al., 2014), which make use of individual-level pollution and disease data. However, cohort studies are expensive and time consuming to implement due to the required follow-up period. This has led to spatial ecological study designs being used (see Lee et al., 2009; Maheswaran et al., 2005), which make use of routinely available small-area data such as from the Scottish Neighbourhood Statistics (<http://www.sns.gov.uk/>) database, and the Health and Social Care Information Centre (<http://www.hscic.gov.uk/>). Due to their ecological nature these studies cannot be used to determine individual-level causality, but they contribute to and independently corroborate the body of evidence provided by cohort studies.

Spatial ecological studies are based on partitioning the study region into  $n$  contiguous small areas determined by administrative boundaries, such as electoral wards or census tracts. For each small area the response is the number of disease cases in a fixed time period, such as the number of deaths due to respiratory disease in one year. These disease cases are adjusted for varying population demographics across the study region using indirect standardisation, and then regressed against air pollution concentrations and other confounders, such as socio-economic deprivation. Typically, Poisson log-linear models are used to estimate the pollution-health effect, and any residual spatial autocorrelation in the data is accounted for by introducing a set of spatially autocorrelated random effects into the model. This residual spatial autocorrelation could be due to numerous factors, including unmeasured confounding (where an important spatially correlated variable is not included in the model or is unknown), neighbourhood effects (where the behaviour of subjects is influenced by surrounding subjects), and grouping effects (where subjects of similar characteristics group together). The health impact of numerous pollutants have been estimated in these studies, such as particulate matter (PM<sub>2.5</sub>, Lawson et al., 2012 and PM<sub>10</sub>, Rushworth et al., 2014), ozone (O<sub>3</sub>, Wang et al., 2009), sulphur dioxide (SO<sub>2</sub>, Elliott et al., 2007), and nitrogen dioxide (NO<sub>2</sub>, Maheswaran et al., 2012). In this study, we focus on NO<sub>2</sub>, since it is a good marker for traffic-related pollution and because measured NO<sub>2</sub> data, via diffusion tubes which only measure NO<sub>2</sub>, are more spatially prevalent across our study region than for other pollutants.

As with all statistical modelling endeavours, estimating the effects of air pollution on health requires a number of modelling choices to be made, which are likely to affect the results. This variation in effect estimates due to model uncertainty is typically ignored, as results from a single 'final' model are often presented. However, it is likely to be crucial in this context because the estimated effect sizes are small and their significance will depend on the final model chosen (this is highlighted in Table 2), thus it is likely that statistically significant or non-significant results could be presented depending on the choices made by the investigators.

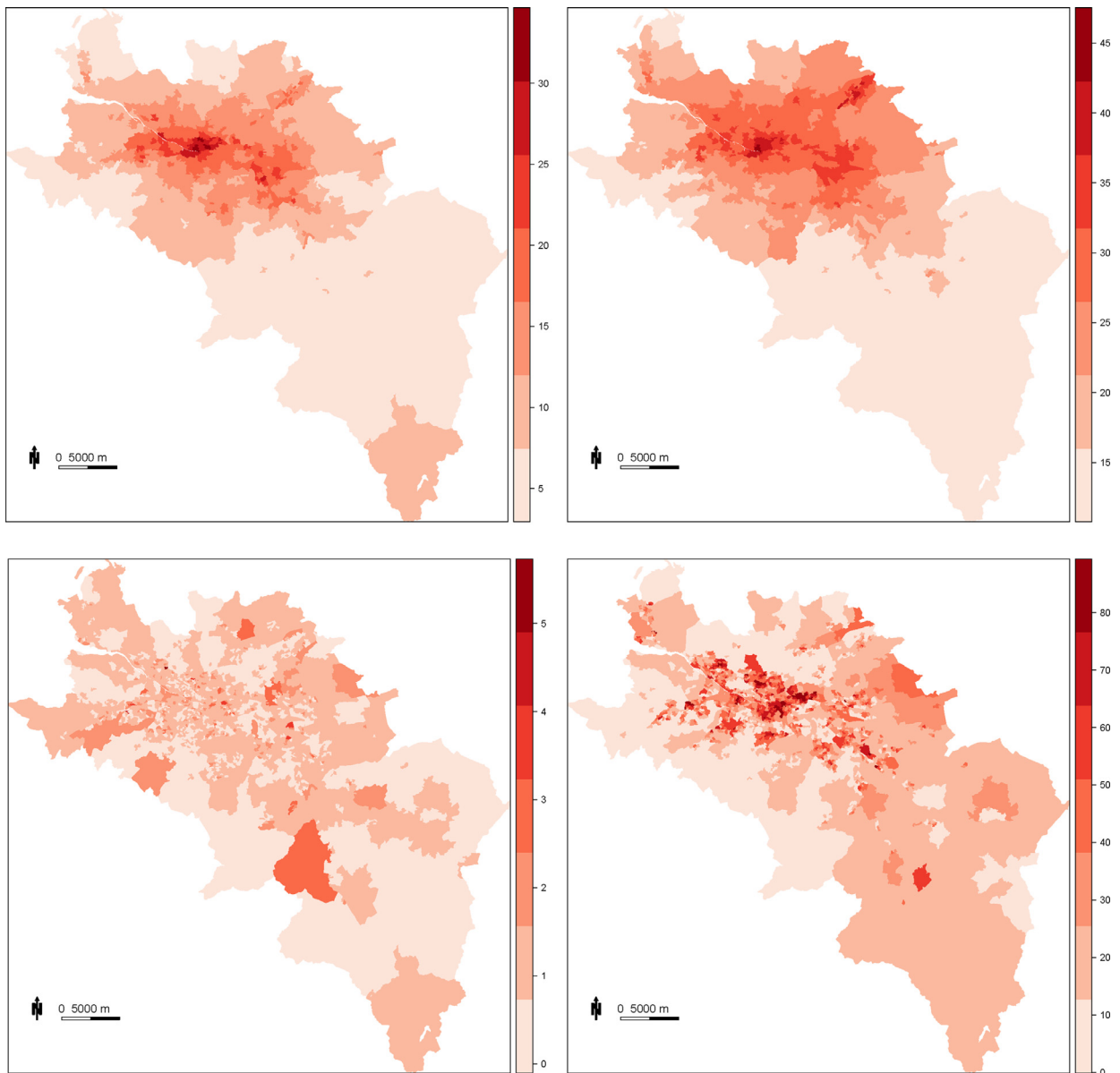
In this paper we investigate the impact of three such modelling choices, namely estimation of NO<sub>2</sub> concentra-

tions, the measure of socio-economic deprivation used, and the method of control for residual spatial autocorrelation. Pollution concentrations have been estimated in small-area studies using atmospheric dispersion models (see Haining et al., 2010; Lee et al., 2009), since they provide complete spatial coverage of the study region. However, these modelled concentrations are known to contain biases, and thus statistical fusion models (Berrocal et al., 2010; Fuentes and Raftery, 2005; McMillan et al., 2009) are increasingly being used, which calibrate the modelled concentrations with observed pollution measurements. Socio-economic deprivation is the major confounder in these studies, and existing studies have either used individual variables such as job seekers allowance or house price (Lee et al., 2014) or composite indexes such as the Townsend index (Maheswaran et al., 2005) to account for it. Finally, residual spatial autocorrelation can be ignored by fitting a simple Poisson log-linear model to the data, while a common adjustment is to add a set of random effects represented by a conditional autoregressive (CAR, Besag et al., 1991) prior to the linear predictor. However, Clayton et al. (1993) and Paciorek (2010) have shown this may lead to collinearity between the fixed and random effects, and a number of extensions have been proposed such as Hughes and Haran (2013) and Lee and Sarran (2015).

Therefore, in this paper we present a new study of NO<sub>2</sub> concentrations and cardio-respiratory mortalities in West Central Scotland, in which we quantify empirically how robust the estimated pollution-health effect sizes are to these factors, and present a Bayesian model averaging (BMA, Hoeting et al., 1999; Raftery, 1995) approach to estimating the overall effect size whilst accounting for model uncertainty. Section 2 describes our motivating study, along with descriptions of the disease, air pollution and deprivation data. Section 3 presents the statistical models described above for taking into account residual spatial autocorrelation and estimating NO<sub>2</sub> concentrations. This section also presents the BMA methodology for combining the estimated air pollution effects from the range of models considered. Our results from the individual models and BMA are presented in Section 4, while Section 5 provides a concluding discussion.

## 2. Motivating study

The methodology developed in this paper is motivated by a new epidemiological study investigating the health impact of long-term exposure to air pollution in West Central Scotland, for the seven year period 2006 to 2012. West Central Scotland is centred around the Greater Glasgow conurbation, which has a population of around 1.1 million people and a land area of 386 km<sup>2</sup>. West Central Scotland is partitioned into  $n = 2089$  non-overlapping data zones, which are the key small area geography in Scotland comprising between 500 and 1000 residents of similar social characteristics. These data zones are described at the Scottish Neighbourhood statistics website (SNS, <http://www.sns.gov.uk/>). The layout of the study region is presented in Fig. 1, where the city of Glasgow is the set of small data zones in the middle north of the figure.



**Fig. 1.** Display of the data. The top left panel shows background NO<sub>2</sub> concentrations provided by DEFRA from an atmospheric dispersion model averaged across 2006–2012, while the top right panel shows estimates from a statistical fusion model. The bottom left panel displays the Standardised mortality ratio (SMR) for cardio-respiratory disease aggregated over 2006–2012, while the bottom right panel displays the SIMD score (without health domain), where a high score indicates deprivation and a low score indicates affluence.

## 2.1. Description of the data

The disease data comprise counts of the numbers of cardio-respiratory mortalities (International Classification of Diseases, 10th Revision: I00–I99, J00–J99) within each of the 2089 data zones during the seven year period 2006 to 2012. These death records were obtained from National Records Scotland. In order to take into account the heterogeneity of the population within each data zone in terms of their size and demographic structure, the expected numbers of cardio-respiratory mortalities were calculated by indirect standardisation, using age- and sex-

specific cardio-respiratory mortality rates for the whole of West Central Scotland. However, due to the low numbers of cardio-respiratory mortalities occurring in a single year (mean of 4.093 for 2006), the cardio-respiratory deaths have been aggregated across the seven year period. The spatial distribution of disease risk is shown in the bottom left panel of Fig. 1, which displays the standardised mortality ratios (SMR, observed numbers/expected numbers) across West Central Scotland for the aggregated years 2006 to 2012. A SMR of 1.2 corresponds to a 20% increase in the risk of disease compared to what is expected. The highest SMRs are found in areas with the highest level of

deprivation, and range between 0 and 4.747, with a mean SMR of 1.066, and a standard deviation of 0.440. Data zones have a SMR of zero when there have been no deaths, which occurs mostly in the centre of Glasgow as it consists mainly of shopping districts.

Concentrations of  $\text{NO}_2$  are available from two sources: measured data from automatic monitors and diffusion tubes, and modelled concentrations from an atmospheric dispersion model. The set of measured data are more spatially dense than corresponding data for other pollutants (mainly thanks to the inclusion of diffusion tubes), but still do not give complete spatial coverage of all 2089 data zones in West Central Scotland. Therefore the simplest approach is to only use the modelled concentrations, which are annual mean background concentrations available at a  $1\text{km} \times 1\text{km}$  resolution for the seven year period from the Department for Environment, Food and Rural Affairs (DEFRA, <http://uk-air.defra.gov.uk/>). These concentrations are temporally aggregated over the seven years by averaging, and are spatially aggregated to the data zone level using

$$\text{NO}_{2_i} = \frac{\sum_{j=1}^{n_i} \exp(-d_{ij}) \tilde{\text{NO}}_{2_j}}{\sum_{j=1}^{n_i} \exp(-d_{ij})}, \quad (1)$$

where  $\text{NO}_{2_i}$  is the averaged concentration for data zone  $i$ . Here  $\tilde{\text{NO}}_{2_1}, \dots, \tilde{\text{NO}}_{2_{n_i}}$  are the modelled  $\text{NO}_2$  concentrations at the  $n_i$  grid squares within data zone  $i$ , and  $d_{ij}$  is the Euclidean distance between the population weighted centroid of data zone  $i$  and the centroid of grid square  $j$ . This is to ensure that the data zone takes a representative value according to the location at which the population density is greatest. Data zones containing no grid square centroids were assigned the  $\text{NO}_2$  concentration nearest the population weighted centroid of the data zone. These data zone averaged modelled concentrations are displayed in the top left panel of Fig. 1, and the city of Glasgow has the highest level of background concentrations, as expected. However, these modelled concentrations are known to contain biases, and in this paper we compare the health effects estimated from using them to those obtained from predicting  $\text{NO}_2$  with a statistical fusion model. A number of statistical fusion models have been proposed for combining measured and modelled pollution data, including Berrocal et al. (2010); Fuentes and Raftery (2005); McMillan et al. (2009). However, the model we use here is that of Pannullo et al. (2015), which uses a similar approach to that of Berrocal et al. (2010) and was developed for the West Central Scotland study region. The general form of the model is given by

$$V(\mathbf{s}_k) \sim N(\mathbf{b}(\mathbf{s}_k)^\top \boldsymbol{\alpha} + \phi(\mathbf{s}_k), v^2 \sigma^2), \quad k = 1, \dots, m, \\ \boldsymbol{\phi} = (\phi(\mathbf{s}_1), \dots, \phi(\mathbf{s}_m)) \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}(\rho)), \quad (2)$$

where  $V(\mathbf{s}_k)$  is the measured  $\text{NO}_2$  concentration at spatial location  $\mathbf{s}_i$  for  $i = 1, \dots, m$  spatial locations. The  $m$  measurements are modelled by a set of covariates  $\mathbf{b}(\mathbf{s}_k)$  with regression parameters  $\boldsymbol{\alpha}$ , and the former include the modelled concentration in the nearest grid square and the local environment in which the site is located (e.g. roadside, urban background, rural). The second term in the mean model is a spatial random effect  $\boldsymbol{\phi} = (\phi(\mathbf{s}_1), \dots, \phi(\mathbf{s}_m))$ , which accounts for residual spatial autocorrelation in the

measured data and are modelled by a Gaussian process with a spatial exponential correlation matrix  $\boldsymbol{\Sigma}(\rho)$  where  $\rho$  denotes the range parameter. Full details of this model are available from Pannullo et al. (2015).

This model calibrates the modelled concentrations according to urban background and rural environments since the West Central Scotland study region has a large proportion of rural areas, while allowing the effect of the modelled concentrations to vary across space. Pannullo et al. (2015) showed that their fusion model produced improved predictions of the measured data in a cross-validation exercise compared to using the modelled concentrations in isolation, and predictions were made from the model on a  $1\text{ km} \times 1\text{ km}$  resolution for each year separately. These predictions were then temporally aggregated by averaging and spatially aggregated up to data zone level using the same form as (1). The resulting concentrations are displayed in the top right panel of Fig. 1, which is structurally similar to the modelled concentrations as expected.

The main confounding factor in ecological health studies is socio-economic deprivation (Mackenbach et al., 1997), in which populations with higher levels of deprivation may be more susceptible to the effects of air pollution (Laurent et al., 2007), for example due to them having worse underlying health on average compared with more affluent communities. However, deprivation is multifactorial and difficult to measure, and is commonly represented by a composite index. Here we make use of the Scottish Index of Multiple Deprivation (SIMD, <http://www.gov.scot/Topics/Statistics/SIMD>), which is a composite index containing seven domains, namely: access to services; crime; education, skills and training; employment; income; health; and housing. The health domain is not used in this study since it contains an indicator of the Comparative Mortality Factor, and therefore includes deaths which are part of the outcome in our study. Correlations between the remaining six domains are displayed in Table 1, where there are high correlations between income, employment, and education; and weak to moderate correlations with the access, housing and crime domains. The overall SIMD index is not appropriate for this study since it contains the health domain. Therefore, we re-weighted the overall index to remove the health domain based on the original index methodology, details of which can be found here <http://www.gov.scot/Publications/2004/10/20089/45173>. The bottom left panel of Fig. 1 displays the re-weighted overall index, where it is clear that the city of Glasgow contains the majority of the most deprived areas, as expected.

Initially, a simple Poisson log-linear model (without any spatial random effects) was fitted to the data with  $\text{NO}_2$  (DEFRA modelled concentrations) and income deprivation as covariates, and the overdispersion parameter was estimated as 4.35, suggesting substantial overdispersion with respect to the Poisson assumption of equal mean and variance. The residuals from this model were then tested for spatial autocorrelation, using a permutation test based on Moran's I statistic (Moran, 1950). The null hypothesis of this test is no spatial autocorrelation, and Moran's I statistic was 0.036 with a p-value of 0.003, suggesting that spatial autocorrelation is present in the residuals.

**Table 1**  
Correlations between the six deprivation measures, where EST denotes the education, skills and training domain.

Variable	Access	Crime	EST	Employment	Income	Housing
Access	1	−0.252	−0.250	−0.287	−0.321	−0.411
Crime	−	1	0.411	0.436	0.430	0.351
EST	−	−	1	0.833	0.860	0.680
Employment	−	−	−	1	0.946	0.436
Income	−	−	−	−	1	0.658
Housing	−	−	−	−	−	1

### 3. Statistical models for estimating air pollution and health effects

The aim of this paper is to estimate the sensitivity of the estimated relationship between NO<sub>2</sub> concentrations and cardio-respiratory mortality in the West Central Scotland region between 2006 and 2012. In doing this we estimate the sensitivity of the effect to changing the estimation of NO<sub>2</sub> concentrations, control for socio-economic deprivation, and allowance for residual spatial autocorrelation in the mortality data after adjusting for the covariate effects. We compare three specific Poisson log-linear models in this sensitivity analysis, which differ in their control for residual spatial autocorrelation. These models are then combined to estimate an overall pollution-health effect using Bayesian model averaging, with inference based on Markov chain Monte Carlo (MCMC) simulation. These models are implemented in the R software environment (R Core Team, 2015), using the CARBayes (Lee, 2013) and ngspatial (Hughes and Cui, 2015) packages, while Bayesian model averaging was implemented using code written by the authors. Sensitivity to the estimation of NO<sub>2</sub> and control for socio-economic deprivation is assessed by fitting different covariate combinations in all of the three models described below.

#### 3.1. Data and likelihood model

The vector of the observed numbers of cardio-respiratory mortalities is denoted by  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , while the expected numbers of mortalities are computed using indirect standardisation based on age- and sex-specific cardio-respiratory mortality rates in West Central Scotland. These expected counts are denoted by  $\mathbf{E} = (E_1, \dots, E_n)$ , where for data zone  $i$ ,  $E_i = \sum_r N_{ir} \gamma_r$ , where  $N_{ir}$  is the number of people in age-sex group  $r$  in data zone  $i$ , and  $\gamma_r$  denotes the region-wide age-sex mortality rate. The vector of NO<sub>2</sub> concentrations is denoted by  $\mathbf{x} = (x_1, \dots, x_n)$  for all  $n$  data zones, while each measure of socio-economic deprivation is denoted by  $\mathbf{u} = (u_1, \dots, u_n)$ . Thus, for the  $i$ th data zone the vector of covariates is given by  $\mathbf{z}_i^\top = (1, x_i, u_i)$  while the corresponding parameters are given by  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ , so that  $\beta_1$  is the intercept term while  $\beta_2$  is the key parameter in this model, namely the effect of NO<sub>2</sub> on cardio-respiratory mortality risk. A general Bayesian Poisson log-linear model for these data is given by

$$Y_i | E_i, R_i \sim \text{Poisson}(E_i R_i) \quad \text{for } i = 1, \dots, n,$$

$$\ln(R_i) = \mathbf{z}_i^\top \boldsymbol{\beta} + \phi_i,$$

$$\boldsymbol{\beta} \sim N(\mathbf{m}, \mathbf{V}), \tag{3}$$

where  $R_i$  is the risk of disease in data zone  $i$ . The regression parameters  $\boldsymbol{\beta}$  are assigned a weakly informative multivariate Gaussian prior with hyperparameters  $(\mathbf{m}, \mathbf{V})$ , typically with mean zero and a large diagonal variance matrix. The final term in the linear predictor  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)$  controls for the residual spatial autocorrelation in the data after accounting for the covariate effects, and we consider three modelling specifications here.

#### 3.2. Model 1 – no spatial autocorrelation

The simplest approach is to ignore the presence of any residual spatial autocorrelation and assume  $\phi_i = 0$  for all data zones  $i$ , which is equivalent to fitting a Poisson generalised linear model to the data. This model naively assumes the cardio-respiratory counts are independent conditional on the covariates, which as illustrated in Section 2 is not true for our case study. Additionally, the model does not allow for overdispersion relative to the Poisson likelihood, and thus makes the restrictive assumption that  $\mathbb{E}[Y_i] = \text{Var}[Y_i]$ , which was shown in Section 2 to be unrealistic. This model is thus included here for comparison purposes with the other models described below.

#### 3.3. Model 2 – globally smooth spatial autocorrelation

The standard approach to accounting for residual spatial autocorrelation and overdispersion in this context is to model  $\boldsymbol{\phi}$  by a set of globally spatially smooth (autocorrelated) random effects  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)$ . A number of models can be specified for these random effects, including conditional autoregressive (CAR), simultaneous autoregressive (SAR) or geostatistical models. However, CAR priors are the most common in this field, and examples of their use include Maheswaran et al. (2005) and Lee et al. (2009). A number of globally smooth CAR priors have been proposed, and a review by Lee (2011) concluded that the model proposed by Leroux et al. (1999) was the most appealing. This model can be specified by a set of  $n$  univariate full conditional distributions  $f(\phi_i | \boldsymbol{\phi}_{-i})$ , where  $\boldsymbol{\phi}_{-i} = (\phi_1, \dots, \phi_{i-1}, \phi_{i+1}, \dots, \phi_n)$ . Spatial autocorrelation is imposed using a binary  $n \times n$  neighbourhood matrix  $\mathbf{W}$ , whose  $ij$ th element  $w_{ij} = 1$  if areas  $(i, j)$  share a common border, and  $w_{ij} = 0$  otherwise. This specification asserts that neighbouring areas have random effects that are partially autocorrelated, otherwise the random effects are

conditionally independent. The model has the form

$$\phi_i | \phi_{-i} \sim N\left(\frac{\rho \sum_{j=1}^n w_{ij} \phi_j}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}\right), \quad (4)$$

where  $\rho$  controls the level of spatial autocorrelation, with  $\rho = 0$  corresponding to spatial independence with mean zero and constant variance, and  $\rho = 1$  corresponding to strong spatial autocorrelation (and simplifying to the intrinsic CAR model). Weakly informative hyperpriors are assigned for  $\tau^2$  and  $\rho$ ; typically an inverse-gamma( $a, b$ ) distribution for  $\tau^2$ , and a uniform distribution on the unit interval for  $\rho$ .

### 3.4. Model 3 – orthogonal smoothing

One problem with traditional CAR models such as (4), is that the spatially smooth random effects have been shown to be correlated with the spatially smooth fixed effects, such as air pollution (Paciorek, 2010). This spatial confounding between the fixed and random effects leads to variance inflation and the model parameters becoming uninterpretable. Much research has been conducted on controlling for this spatial confounding, in which the random effects are instead modelled with a series of basis functions that are orthogonal to the covariates, thus mitigating this confounding (Hughes and Haran, 2013; Reich et al., 2006). In this paper, we utilise the orthogonal smoothing model proposed by Hughes and Haran (2013) due to the low dimensionality of the random effects, which leads to fast computation. This model replaces the random effects  $\phi_i$  in Eq. (3) with a linear combination of basis functions that are orthogonal to the fixed effects. Let the matrix of  $p$  covariates be denoted by  $\mathbf{Z} = (\mathbf{z}_1^\top, \dots, \mathbf{z}_n^\top)^\top$ , then the orthogonal projection matrix (hat matrix) onto the column space of the design matrix  $\mathbf{Z}$  is defined by

$$\mathbf{P} = \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top, \quad (5)$$

and let the residual projection matrix onto the space orthogonal to  $\mathbf{Z}$  be defined by

$$\mathbf{P}' = \mathbf{I}_n - \mathbf{P}. \quad (6)$$

The residual projection matrix is then used to create a set of eigenvectors, from the matrix product  $\mathbf{P}'\mathbf{W}\mathbf{P}'$ , which combines covariate orthogonality given by  $\mathbf{P}'$  with spatial adjacency given by  $\mathbf{W}$ . The eigenvectors of  $\mathbf{P}'\mathbf{W}\mathbf{P}'$  contain all possible mutually distinct spatial patterns of clustering orthogonal to  $\mathbf{Z}$ . Furthermore, spatial dependence is related to both positive and negative eigenvalues, where positive eigenvalues correspond to positive spatial autocorrelation. The size of the eigenvalue associated with a given eigenvector determines the relative importance of its spatial pattern, so Hughes and Haran (2013) suggest only selecting the first  $q < n$  eigenvectors corresponding to the largest positive eigenvalues. This matrix is denoted by  $\mathbf{M}$ , where  $\mathbf{m}_i^\top = (m_{i1}, \dots, m_{iq})$ . The chosen number of eigenvectors  $q$  acts as a tuning parameter, which determines the extent of dimensionality reduction in the model, and the authors suggest using  $q = 50$  as a default choice. The orthogonal smoothing model replaces the random effects in the linear predictor in Eq. (3) by

$$\ln(R_i) = \mathbf{z}_i^\top \boldsymbol{\beta} + \mathbf{m}_i^\top \boldsymbol{\delta},$$

$$\boldsymbol{\delta} \sim N(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W})_s^{-1}), \quad (7)$$

where the random effects  $\boldsymbol{\delta}$  are assigned a Gaussian prior with mean  $\mathbf{0}$ , and precision matrix given by  $\mathbf{Q}(\mathbf{W})_s = \mathbf{M}^\top \mathbf{Q}(\mathbf{W}) \mathbf{M}$ , where  $\mathbf{Q}(\mathbf{W}) = \text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}$  corresponds to the precision matrix for the intrinsic CAR prior (Besag et al., 1991).

### 3.5. Bayesian model averaging

Bayesian model averaging provides a coherent framework for combining the estimates of the same quantity of interest from a number of different Bayesian models into a single overall estimate, which accounts for model uncertainty. Such model uncertainty is often ignored in existing studies, and as we show in the next section can have a large impact on the results. Recall that  $\beta_2$  is the key parameter of interest in this model, namely the effect of NO<sub>2</sub> concentrations on cardio-respiratory mortality risk. Consider the case of having  $K$  candidate models, where in our study we have  $K = 42$  (see next section for details). Denote these models by  $(M_1, \dots, M_K)$  and their respective sets of model parameters by  $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ . Let  $\beta_2$  denote the true unknown parameter of interest and  $\hat{\beta}_{2_k}$  denote the estimate (posterior median) from the  $k$ th model. Then the posterior distribution of interest is

$$f(\beta_2 | \mathbf{Y}) = \sum_{k=1}^K f(\beta_2 | M_k, \mathbf{Y}) f(M_k | \mathbf{Y}). \quad (8)$$

Here  $f(\beta_2 | M_k, \mathbf{Y})$  is the posterior distribution of  $\beta_2$  from model  $K$ , and  $f(M_k | \mathbf{Y})$  is the posterior probability of model  $M_k$ . This equation essentially averages the posterior distributions for  $\beta_2$  under each model weighted by their posterior model probabilities. The posterior probability for model  $M_k$  is given by

$$f(M_k | \mathbf{Y}) = \frac{f(\mathbf{Y} | M_k) f(M_k)}{\sum_{l=1}^K f(\mathbf{Y} | M_l) f(M_l)}, \quad (9)$$

where  $f(M_k)$  is the prior probability for model  $M_k$ . We specify our prior ignorance via a discrete uniform prior for  $f(M_k)$ , that is  $f(M_k) = 1/K$ . This specification simplifies the posterior probability for model  $M_k$  in Eq. (9) to

$$f(M_k | \mathbf{Y}) = \frac{f(\mathbf{Y} | M_k)}{\sum_{l=1}^K f(\mathbf{Y} | M_l)}. \quad (10)$$

The marginal (averaged over the parameters) probability of the data given model  $M_k$  is computed by

$$f(\mathbf{Y} | M_k) = \int_{\boldsymbol{\theta}_k} f(\mathbf{Y} | \boldsymbol{\theta}_k, M_k) f(\boldsymbol{\theta}_k | M_k) d\boldsymbol{\theta}_k, \quad (11)$$

which can be approximated by  $J$  McMC samples as

$$f(\mathbf{Y} | M_k) \approx \frac{1}{J} \sum_{j=1}^J f(\mathbf{Y} | \boldsymbol{\theta}_k^{(j)}, M_k) f(\boldsymbol{\theta}_k^{(j)} | M_k), \quad (12)$$

where  $\boldsymbol{\theta}_k^{(j)}$  is the  $j$ th McMC sample for model  $M_k$ . Once these quantities have been computed the posterior mean and variance of  $\beta_2$  are given by:

$$\mathbb{E}(\beta_2 | \mathbf{Y}) = \sum_{k=1}^K \hat{\beta}_{2_k} f(M_k | \mathbf{Y}), \quad (13)$$

$$\text{Var}(\beta_2|\mathbf{Y}) = \sum_{k=1}^K [\text{Var}(\beta_2|M_k, \mathbf{Y}) + \hat{\beta}_{2k}^2] f(M_k|\mathbf{Y}) - \mathbb{E}(\beta_2|\mathbf{Y})^2, \quad (14)$$

where  $\text{Var}(\beta_2|M_k, \mathbf{Y})$  is the posterior variance of  $\beta_2$  from model  $M_k$ . Based on a normal approximation to the posterior, an approximate 95% credible interval can be obtained for  $\beta_2$  that accounts for model uncertainty.

#### 4. Results from the West Central Scotland study

We now present the results of our study investigating the long-term effects of NO<sub>2</sub> concentrations on cardio-respiratory mortality in West Central Scotland between 2006 and 2012. Section 4.1 describes the set of results obtained from fitting the range of models described in this paper, which illustrates the sensitivity of the results due to model choice. Then Section 4.2 presents our overall estimate of the effect of NO<sub>2</sub> on cardio-respiratory mortality using the Bayesian model averaging approach outlined in the previous section. Inference for all models described in this section is based on running 5 parallel Markov chains for 120,000 iterations, which included a burn-in period of 20,000 iterations. The remaining samples were thinned by 10 to reduce their autocorrelation, thus producing a final set of 50,000 posterior samples across the five chains.

##### 4.1. Results – sensitivity to model choice

We empirically investigate the sensitivity of the estimated pollution-health effect to three modelling choices. The first is the estimation of spatially averaged NO<sub>2</sub> concentrations for each data zone, and we compare averaging the raw output from the atmospheric dispersion model used by DEFRA (<http://uk-air.defra.gov.uk/>, denoted *DEFRA*) to averaging predictions from the fusion model proposed by Pannullo et al. (2015) (denoted *Fusion*). The second modelling choice is how to control for the confounding effects of socio-economic deprivation, and we compare using the composite Scottish Index of Multiple Deprivation (SIMD, minus the health domain), with individual indicators from its sub-domains, namely access to services, crime, education, employment, housing and income. Finally, we compare three approaches to controlling for residual spatial autocorrelation, ignoring it (denoted *GLM*), modelling it using random effects represented by the globally smooth model proposed by Leroux et al. (1999) (denoted *Leroux*), and modelling it using a set of orthogonal random effects proposed by Hughes and Haran (2013) (denoted *OS*).

All combinations of these factors gives a set of 42 possible models, and the results are summarised in Table 2 and 3, which respectively display the posterior median relative risks and 95% credible intervals, and the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) together with the effective number of parameters ( $p_D$ ) for each model and the root mean square error (RMSE) of the data  $\mathbf{Y}$ . All pollution-health effects are presented on the relative risk scale for a 5  $\mu\text{gm}^{-3}$  increase in NO<sub>2</sub> concentrations (for both *DEFRA* and *Fusion*), as this is a realistic change in

**Table 2**

Posterior median relative risks (RR) and 95% credible intervals for a 5  $\mu\text{gm}^{-3}$  increase in NO<sub>2</sub> concentrations on cardio-respiratory mortality. The results displayed relate to models varying in their estimation of NO<sub>2</sub>, control for deprivation and allowance for residual spatial autocorrelation. The results in bold are substantial effects at the 5% level.

Deprivation	Model	RR (95% CI)	
		Fusion	DEFRA
Access	GLM	<b>1.036 (1.016, 1.056)</b>	<b>1.050 (1.026, 1.075)</b>
	Leroux	<b>1.033 (1.006, 1.059)</b>	<b>1.045 (1.015, 1.075)</b>
	OS	<b>1.029 (1.020, 1.039)</b>	<b>1.041 (1.030, 1.051)</b>
Crime	GLM	<b>1.038 (1.018, 1.058)</b>	<b>1.053 (1.033, 1.074)</b>
	Leroux	<b>1.039 (1.015, 1.063)</b>	<b>1.053 (1.027, 1.079)</b>
	OS	<b>1.034 (1.025, 1.043)</b>	<b>1.046 (1.037, 1.057)</b>
Education	GLM	1.006 (0.988, 1.024)	1.019 (0.999, 1.039)
	Leroux	1.007 (0.991, 1.024)	<b>1.019 (1.002, 1.041)</b>
	OS	1.006 (0.998, 1.015)	<b>1.021 (1.011, 1.030)</b>
Employment	GLM	1.010 (0.990, 1.030)	<b>1.020 (1.000, 1.040)</b>
	Leroux	1.015 (0.998, 1.033)	<b>1.025 (1.007, 1.044)</b>
	OS	<b>1.014 (1.006, 1.023)</b>	<b>1.025 (1.016, 1.036)</b>
Housing	GLM	0.992 (0.973, 1.012)	0.989 (0.968, 1.011)
	Leroux	0.990 (0.971, 1.009)	0.980 (0.959, 1.002)
	OS	0.992 (0.983, 1.002)	<b>0.987 (0.977, 0.997)</b>
Income	GLM	1.003 (0.985, 1.021)	1.010 (0.990, 1.030)
	Leroux	1.008 (0.992, 1.108)	1.012 (0.995, 1.030)
	OS	1.007 (0.998, 1.015)	<b>1.013 (1.004, 1.023)</b>
SIMD	GLM	1.007 (0.989, 1.025)	1.017 (0.997, 1.037)
	Leroux	1.013 (0.997, 1.030)	<b>1.021 (1.003, 1.040)</b>
	OS	<b>1.011 (1.003, 1.020)</b>	<b>1.021 (1.011, 1.030)</b>

long-term exposure. Overall, there is evidence that increasing NO<sub>2</sub> concentrations is associated with small but positive increases in the risk of cardio-respiratory mortality, as 36 out of the 42 models estimate the relative risk to be greater than 1. However, the range of the effects estimated across the 42 models is large, being between a 2% decreased risk (0.980) to a 5.3% increased risk (1.053) associated with a 5  $\mu\text{gm}^{-3}$  increase in NO<sub>2</sub>. This suggests that the results are highly sensitive to model choice, and that if we had presented results from a single model then we could have shown either a positive or a negative effect of NO<sub>2</sub> on mortality risk. Focusing on the 95% credible intervals shows that 23 of the 42 intervals are wholly above the null risk of 1, which is just over 57% of the models considered.

The three modelling choices considered here all appear to have the potential to substantially effect the estimated relative risks, as the estimates from varying one factor at a time can lead to large changes in risk. For example, changing the NO<sub>2</sub> metric from that produced by the fusion model to that produced by DEFRA resulted in the risks changing by between -1% and 5.3%, and in all but 3 cases these changes were positive. This indicates that overall using the DEFRA concentrations resulted in increased risks compared with using the predictions from the fusion model. Changing the control for socio-economic deprivation also had a large impact on the results, with changes in relative risk of between 3.8% and 7.3% across the 7 measures considered depending on the combination of *DEFRA/Fusion* and *GLM/Leroux/OS*. In general, using the hous-

**Table 3**

Model fit for each of the 42 models, measured by the Deviance Information Criterion (DIC), the effective number of parameters ( $p_D$ ), and the root mean square error (RMSE).

Deprivation	Model	DIC ( $p_D$ )		RMSE	
		Fusion	DEFRA	Fusion	DEFRA
Access	GLM	20219 (2)	20182 (2)	13.560	13.519
	Leroux	13797 (1508)	13799 (1507)	2.518	2.525
	OS	19130 (76)	19115 (74)	12.614	12.604
Crime	GLM	20017 (2)	19967 (2)	13.471	13.429
	Leroux	13793 (1498)	13791 (1497)	2.511	2.510
	OS	19222 (67)	19201 (66)	12.707	12.697
Education	GLM	18240 (2)	18224 (2)	12.742	12.724
	Leroux	13601 (1369)	13600 (1367)	2.687	2.688
	OS	17964 (62)	17942 (62)	12.336	12.319
Employment	GLM	18373 (2)	18359 (2)	12.811	12.812
	Leroux	13600 (1378)	13597 (1377)	2.655	2.658
	OS	18010 (66)	17996 (65)	12.323	12.318
Housing	GLM	19107 (2)	19106 (2)	12.989	12.993
	Leroux	13737 (1451)	13736 (1450)	2.522	2.522
	OS	18336 (69)	18352 (69)	12.302	12.323
Income	GLM	18139 (2)	18135 (2)	12.638	12.623
	Leroux	13589 (1362)	13589 (1362)	2.693	2.692
	OS	17743 (66)	17729 (60)	12.128	12.129
SIMD	GLM	18277 (2)	18267 (2)	12.701	12.694
	Leroux	13609 (1374)	13606 (1373)	2.672	2.670
	OS	17900 (62)	17898 (64)	12.242	12.240

ing indicator resulted in the lowest effect sizes, while using crime resulted in the highest estimates. Finally, varying the control for spatial autocorrelation had a slight effect on the results, with differences in relative risk between the three models considered ranging between 0.1% and 0.9%. The only pattern of note is that the effect sizes are attenuated for the OS models compared to the Leroux models in 10 out of the 14 models, with 2 models comprising the same effect size.

Finally, Table 3 summarises the fit of each model via the DIC and RMSE, which shows that in all cases the Leroux model fits the data best compared with the other alternatives. This is surprising considering the globally smooth model has the potential for correlation between the fixed and random effects and thus one would expect the OS model to outperform the Leroux model. The OS model has many fewer effective number of parameters compared to the Leroux model, which makes it more parsimonious. However, it is this reduction in dimensionality that has resulted in a poorer fit to the data (in terms of DIC). In most cases, the DIC is lower for the DEFRA concentrations compared to the Fusion concentrations, while the income domain provides the best fit to the data of all the socio-economic indicators considered here. Furthermore, the RMSE allows us to assess the closeness of the models fitted values to the observed health outcomes (with the lowest values indicating better performance), and to give a sense of scale the 25th and 75th percentiles of the observed cardio-respiratory deaths were 14 and 33 respectively. The Leroux models have the lowest RMSE values compared to the GLMs and OS models, which is due to their increased number of effective parameters. For the Leroux models the relationship between RMSE and depriva-

tion is opposite to that observed for the GLMs and OS models, with income, the best deprivation covariate in terms of DIC, having the highest RMSE compared to the other deprivation measures (2.693 compared to 2.518 for access). The reason is that after adjusting for income deprivation there is less residual spatial variation in the model compared to using the other deprivation covariates. Thus, the random effects have less spatial variation, and thus less impact on the fitted values. For example, the variance  $\tau^2$  is 0.230 for the access covariate compared to 0.094 for income. This is also observed in the effective number of parameters  $p_D$ , which is smallest for the model with income. In contrast, the DIC is an overall measure of model quality that penalises more complex models containing more parameters, hence higher DIC values for access (and others) compared to income.

#### 4.2. Results – BMA

The previous section shows clear sensitivity of the results to the model fitted, and one solution would be to choose a single 'best' model, for example by minimising the DIC. However this clearly ignores model uncertainty, which can be accounted for using BMA as described in Section 3.5 by combining the estimated effect sizes from the 42 models considered here. When this was done the overall estimated relative risk was 1.011 together with associated 95% uncertainty interval of (0.993, 1.029). This small but positive effect indicates that for a  $5 \mu\text{g}\text{m}^{-3}$  increase in  $\text{NO}_2$  concentrations cardio-respiratory deaths increase by an estimated 1.1%, although the lower end of the 95% credible interval is below the null risk of 1. In fact, the posterior probability that the relative risk is greater



than 1 is 0.884. This result is essentially a mixture between the effect estimates from the *Leroux* model including income and DEFRA NO<sub>2</sub> concentrations, and the effect estimate from the *Leroux* model including income and Fusion model NO<sub>2</sub> concentrations. The former had the most influence on the overall effect size since its posterior model probability  $f(M_k|Y)$  was 67.82%, whilst it was 32.17% for the latter. So in this example the large differences in fit across the 42 models has resulted in only two models contributing to the overall effect estimate.

## 5. Discussion

In this paper we investigated the sensitivity of the pollution-health relationship in West Central Scotland to the impact of three modelling choices: the estimation of NO<sub>2</sub> concentrations, control for socio-economic deprivation, and control for residual spatial autocorrelation after accounting for the covariate effects. Our main finding is that the choice of these three factors can have a major impact on the resulting pollution-health effects, which means that if only a single model was presented researchers could show a wide range of effect sizes depending on which model is chosen. The estimated pollution-health effect in this study varies considerably across the 42 models (effect sizes range from 0.980–1.053), highlighting the estimated pollution-health effect sizes are not robust to the three aforementioned factors.

We then utilised BMA to combine the results from all 42 models into an overall pollution-health effect size, whilst taking into account model uncertainty. Our final estimated effect size shows a 5  $\mu\text{gm}^{-3}$  increase in NO<sub>2</sub> concentrations is associated with 1.1% higher cardio-respiratory deaths in West Central Scotland between 2006 and 2012. However, this effect is (borderline) not substantial at the 5% level, as the resulting 95% credible interval contains the null risk of 1. This could be due to the fact that the majority of the NO<sub>2</sub> concentrations are relatively low, and thus greater variation in the exposure would be needed to observe substantial health impacts.

A second finding from our study is the attenuation of the pollution-health effects when the NO<sub>2</sub> concentrations were estimated using the geostatistical fusion model, compared to when the NO<sub>2</sub> concentrations were estimated by the DEFRA diffusion model. The estimated health effects changed between –1% and 5.3%, indicating that increased risks are observed when the DEFRA concentrations are utilised. This is an interesting result considering the majority of spatial ecological studies in Scotland (and indeed in the UK) make use of modelled concentrations because of their wide availability and fine scale spatial coverage. Furthermore, the correlation between residual disease (after adjustment from income deprivation) and pollution from both the fusion and DEFRA models is 0.041 and 0.029 respectively. This highlights that the DEFRA pollution concentrations are more correlated with residual disease, thus explaining why it has a stronger effect size (see Table 2) compared to the pollution concentrations from the fusion model. However, in terms of pollution predictive performance, Pannullo et al. (2015) show that the DEFRA data are not as good at predicting measured pollution concen-

trations at the point level, since the root mean square prediction error (RMSPE) is 0.337 compared to 0.255 for the fusion model. Furthermore, a recent study conducted in mainland Scotland by Huang et al. (2015) concluded that the estimated health effects of NO<sub>2</sub> were largely consistent when estimated from a fusion model compared to modelled concentrations from the DEFRA.

A third finding from our study is that the global spatial autocorrelation model comprising the DEFRA concentrations and income deprivation dominated the overall pollution-health effect size when combining models using BMA. The posterior model probability was 67.82%, while for the model with the fusion model concentrations it was 32.17%. It is interesting to note that only 2 out of the 42 models had a considerable influence, suggesting that the global spatial autocorrelation model with the income deprivation and DEFRA concentrations are the most important factors for investigating the impact of air pollution on health in West Central Scotland. In addition, the global spatial autocorrelation model, which has been under much scrutiny by Reich et al. (2006) and others, outperformed the orthogonal smoothing model proposed by Hughes and Haran (2013) in terms of model fit via the DIC.

There are a few drawbacks to our study. Firstly, we had to aggregate our cardio-respiratory deaths over a 7 year period to ensure there was enough variation in the disease data. This meant only a purely spatial study could be performed and it would be important to investigate how the pollution-health risks have changed over time. In addition, there may not be enough cardio-respiratory deaths at the data zone level, thus upgrading to a larger spatial resolution, such as intermediate geographies (average population of 4300 inhabitants), may improve the power of the study to detect an association. Furthermore, hospital admission data could be used instead as the health outcome since there will be more events and thus no need to aggregate over multiple years. Secondly, the DEFRA modelled concentrations come with no measure of uncertainty and this could impact the analysis. The predicted concentrations from the geostatistical fusion model do have measures of prediction uncertainty, however, in this study we treated the predicted NO<sub>2</sub> concentrations as the known and true values, which again could impact our results. Therefore, an avenue for future work will be to incorporate the uncertainty surrounding the predicted NO<sub>2</sub> concentrations in a combined Bayesian framework, which estimates the exposures and health risks simultaneously. Finally, our key message is that due to the small effect sizes, a range of pollution-health effects can be estimated depending on the modelling choices made, and thus accounting for this is vital to provide robust evidence of the damaging effect of air pollution on health.

## Acknowledgements

The authors thank the joint editor and the reviewer for their helpful comments on our paper. FP is funded by a Medical Research Council studentship (1321719). FP and AL are part of the Inequalities programme at the Social and Public Health Sciences Unit, University of Glasgow, funded by the Medical Research Council and the Chief Scientist

Office of the Scottish Government Health and Social Care Directorate (MC\_UU\_12017/13).

## References

- Berrocal VJ, Gelfand AE, Holland DM. A spatio-temporal downscaler for output from numerical models. *J Agric Biol Environ Stat* 2010;15:176–97. doi:[10.1007/s13253-009-0004-z](https://doi.org/10.1007/s13253-009-0004-z).
- Besag J, York J, Mollie A. Bayesian image-restoration, with 2 applications in spatial statistics. *Ann Inst Stat Math* 1991;43:1–20. doi:[10.1007/BF00116466](https://doi.org/10.1007/BF00116466).
- Cesaroni G, Forastiere F, Stafoggia M, Andersen ZJ, Badaloni C, Beelen R, et al. Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE project. *BMJ (Clin Res Ed)* 2014;348. doi:[10.1136/bmj.f7412](https://doi.org/10.1136/bmj.f7412).
- Clayton DG, Bernardinelli L, Montomoli C. Spatial correlation in ecological analysis. *Int J Epidemiol* 1993;22:1193–202.
- Department for the Environment, Food and Rural Affairs, Air quality plans for the achievement of EU air quality limit values for nitrogen dioxide (NO<sub>2</sub>) in the UK. 2015. URL: <https://uk-air.defra.gov.uk/library/no2ten/>. [accessed 01.11.15].
- Elliott P, Shaddick G, Wakefield JC, de Hoogh C, Briggs DJ. Long-term associations of outdoor air pollution with mortality in Great Britain. *Thorax* 2007;62:1088–94. doi:[10.1136/thx.2006.076851](https://doi.org/10.1136/thx.2006.076851).
- Fuentes M, Raftery AE. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* 2005;61:36–45. doi:[10.1111/j.0006-341X.2005.030821.x](https://doi.org/10.1111/j.0006-341X.2005.030821.x).
- Haining R, Li G, Maheswaran R, Blangiardo M, Law J, Best N, Richardson S. Inference from ecological models: estimating the relative risk of stroke from air pollution exposure using small area data. *Spat Spat Temporal Epidemiol* 2010;1:123–31. doi:[10.1016/j.sste.2010.03.006](https://doi.org/10.1016/j.sste.2010.03.006).
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Stat Sci* 1999;14:382–417. doi:[10.2307/2676803](https://doi.org/10.2307/2676803).
- Huang G, Lee D, Scott M. An integrated Bayesian model for estimating the long-term health effects of air pollution by fusing modelled and measured pollution data: A case study of nitrogen dioxide concentrations in Scotland. *Spat Spat Temporal Epidemiol* 2015 <http://dx.doi.org/10.1016/j.sste.2015.09.002>.
- Hughes J, Cui X. *ngspatial: Fitting the centered autologistic and sparse spatial generalized linear mixed models for areal data*. 2015.R package version 1.0–5. Minneapolis, MN
- Hughes J, Haran M. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J R Stat Soc Ser B* 2013;75:139–59. <http://EconPapers.repec.org/RePEc:bla:jorssb:v:75:y:2013:i:1:p:139-159>
- Laurent O, Bard D, Filleul L, Segala C. Effect of socioeconomic status on the relationship between atmospheric pollution and mortality. *J Epidemiol Commun Health* 2007;61:665–75. doi:[10.1136/jech.2006.053611](https://doi.org/10.1136/jech.2006.053611).
- Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu J. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat* 2012;17:417–41. doi:[10.1007/s13253-012-0100-3](https://doi.org/10.1007/s13253-012-0100-3).
- Lee D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat Spat Temporal Epidemiol* 2011;2:79–89. doi:[10.1016/j.sste.2011.03.001](https://doi.org/10.1016/j.sste.2011.03.001).
- Lee D. CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. *J Stat Softw* 2013;55:1–24. <http://www.jstatsoft.org/v55/i13/> [accessed 26.10.15].
- Lee D, Ferguson C, Mitchell R. Air pollution and health in Scotland: a multicity study. *Biostatistics* 2009;10:409–23. doi:[10.1093/biostatistics/kxp010](https://doi.org/10.1093/biostatistics/kxp010).
- Lee D, Rushworth A, Sahu SK. A Bayesian localized conditional autoregressive model for estimating the health effects of air pollution. *Biometrics* 2014;70:419–29. doi:[10.1111/biom.12156](https://doi.org/10.1111/biom.12156).
- Lee D, Sarran C. Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics* 2015. doi:[10.1002/env.2348](https://doi.org/10.1002/env.2348).
- Leroux B, Lei X, Breslow N. Statistical models in epidemiology, the environment and clinical trials. Estimation of disease rates in small areas: a new mixed model for spatial dependence. New York: Springer-Verlag; 1999. p. 135–78.
- Mackenbach JP, Kunst AE, Cavelaars AE, Groenhof F, Geurts JJ. Socio-economic inequalities in morbidity and mortality in Western Europe. The EU working group on socioeconomic inequalities in health. *Lancet* 1997;349:1655–9.
- Maheswaran R, Haining RP, Brindley P, Law J, Pearson T, Fryers PR, et al. Outdoor air pollution and stroke in Sheffield, United Kingdom: a small-area level geographical study. *Stroke* 2005;36:239–43. doi:[10.1161/01.STR.0000151363.71221.12](https://doi.org/10.1161/01.STR.0000151363.71221.12).
- Maheswaran R, Pearson T, Smeeton NC, Beevers SD, Campbell MJ, Wolfe CD. Outdoor air pollution and incidence of ischemic and hemorrhagic stroke: a small-area level ecological study. *Stroke* 2012;43:22–7. doi:[10.1161/STROKEAHA.110.610238](https://doi.org/10.1161/STROKEAHA.110.610238).
- McMillan NJ, Holland DM, Morara M, Feng J. Combining numerical model output and particulate data using Bayesian space-time modeling. *Environmetrics* 2009. doi:[10.1002/env.984](https://doi.org/10.1002/env.984).
- Ministry of Public Health, Mortality and morbidity during the London Smog of December 1952, 95, Her Majesty's Stationery Office; 1954.
- Moran P. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17–23.
- Paciorek CJ. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Stat Sci Rev J Inst Math Stat* 2010;25:107–25. doi:[10.1214/10-STS326](https://doi.org/10.1214/10-STS326).
- Pannullo F, Lee D, Waclawski E, Leyland AH. Improving spatial nitrogen dioxide prediction using diffusion tubes: a case study in West Central Scotland. *Atmos Environ* 2015;118:227–35. doi:[10.1016/j.atmosenv.2015.08.009](https://doi.org/10.1016/j.atmosenv.2015.08.009).
- Raftery AE. Bayesian model selection in social research. *Sociol Methodol* 1995;25:111–63. doi:[10.2307/271063](https://doi.org/10.2307/271063).
- Reich BJ, Hodges JS, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* 2006;62:1197–206. doi:[10.1111/j.1541-0420.2006.00617.x](https://doi.org/10.1111/j.1541-0420.2006.00617.x).
- Rushworth A, Lee D, Mitchell R. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spat Spat Temporal Epidemiol* 2014;10:29–38. doi:[10.1016/j.sste.2014.05.001](https://doi.org/10.1016/j.sste.2014.05.001).
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria; 2015. URL: <http://www.R-project.org/>. [accessed 10.10.15].
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol* 2002;64:583–616. doi:[10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353).
- Wang XY, Hu W, Tong S. Long-term exposure to gaseous air pollutants and cardio-respiratory mortality in Brisbane, Australia. *Geospatial Health* 2009;3:257–63.
- Willlocks LJ, Bhaskar A, Ramsay CN, Lee D, Brewster DH, Fischbacher CM, et al. Cardiovascular disease and air pollution in Scotland: no association or insufficient data and study design? *BMC Public Health* 2012;12:227. doi:[10.1186/1471-2458-12-227](https://doi.org/10.1186/1471-2458-12-227).
- World Health Organisation Ambient (outdoor) air quality and health: Fact sheet 313. 2014. URL: <http://www.who.int/mediacentre/factsheets/fs313/en/> [accessed 16.10.15].