A Bayesian spatio-temporal model to estimate long term exposure to outdoor air pollution at coarser administrative geographies in England and Wales

Sabyasachi Mukhopadhyay
University of Hohenheim, Germany
Sujit K. Sahu†E-mail: S.K.Sahu@soton.ac.uk
University of Southampton, UK

Summary.
Estimation of long term exposure to air pollution levels over a large spatial domain, such as the mainland UK, entails a challenging modelling task since exposure data are often only observed by a network of sparse monitoring sites with variable amounts of missing data. This article develops and compares several flexible non-stationary hierarchical Bayesian models for the four most harmful air pollutants: NO_2, O_3, PM_{10} and $PM_{2.5}$, in England and Wales during the five year period 2007–2011. The models make use of observed data from the UK's Automatic Urban and Rural Network (AURN) as well as output of an atmospheric air quality dispersion model developed recently especially for the UK. Land use information, incorporated as a predictor in the model, further enhances the accuracy of the model. Using daily data for all four pollutants over the five year period we obtain empirically verified maps which are the most accurate among the competition. Monte Carlo integration methods for spatial aggregation are developed and these allow us to obtain predictions, and their uncertainties, at the level of a given administrative geography. These estimates for local authority areas can readily be used for many purposes such as modelling of aggregated health outcome data and are made publicly available alongside this paper.

1. Introduction

Long term exposure to outdoor air pollution has been associated with many adverse effects on human health, such as respiratory and cardiovascular diseases. The literature establishing this linkage is rapidly growing (Boezen et al., 1999; Dockery and Pope, 1994; Samet et al., 2000; Kassomenos et al., 1995; Bell et al., 2004). Such efforts require accurate measurements of air quality and since the latter is not available everywhere in a large spatial study domain, methods for spatial interpolation are in high demand. Typically, air pollution concentrations are monitored at a handful of sites which are then associated with

†*Address for correspondence:* Sujit K. Sahu, Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom

large surrounding geographical regions. However, the outdoor air we breathe is a toxic mix of airborne particles which travel 100s and 1000s of kilometres – e.g. see recent news headlines that Saharan dust and pollution from continental Europe are triggering high air pollution levels in the UK‡. The absence of physical boundary walls between nations and smaller administrative geographies, and the predominant weather patterns at the time, will always allow movement of air pollution both spatially and temporally. Estimation of the spatial and temporal variation in air pollution levels must then be performed by appropriate spatio-temporal modelling which also directly allows estimation and then reporting of the associated uncertainties.

Recently, there has been a lot of interest in developing statistical space-time models for air pollution levels observed over large spatial domains, e.g. the continental USA, Europe and other parts of the world such as Asia and South America where air pollution is a very prominent health hazard. Methodological developments include: downscaler models; (Berrocal et al., 2010; Alkuwari et al., 2013), data fusion models (Sahu et al., 2010), land use regression (Jerrett et al., 2005; Hoek et al., 2008; Shaddick et al., 2013; Venegas et al., 2014; Morrison et al., 2016), hierarchical space-time autoregressive models (Sahu et al., 2007). Model based methods for real-time forecasting of air pollution have also been developed (Huerta et al., 2004; McMillan et al., 2005; Sahu et al., 2009).

Statistical modelling of UK air pollution data for estimating long term exposure poses a unique set of challenges because of the sparsity of the monitoring sites (about 144 sites covering England and Wales, which is the study region in this paper), irregular monitoring leading to a large percentage of missing data, and the proximity to continental Europe with which the UK exchanges emission and pollution depending on the prevailing weather. Key articles discussing Bayesian modelling for UK data include: Lee and Shaddick (2010); Shaddick and Wakefield (2002) and Pirani et al. (2014). The spatial domain used in these three papers is only the Greater London area with a limited number of monitoring sites. In particular, Lee and Shaddick (2010) model daily data from 30 monitoring sites between 2003 and 2005. They perform a simulation study for four pollutants: carbon monoxide (CO), ozone (O_3), nitrogen di-oxide (NO_2) and PM_{10}. However, they do not validate the air pollution modelling for the real life data from Greater London. Pirani et al. (2014) model daily PM_{10} exposure data from 45 sites for 728 days during 2002-2003. Lastly, Shaddick and Wakefield (2002) consider spatio-temporal modelling of four pollutants measured daily at eight monitoring sites in London over the 4-year period, 1994-1997.

There have been numerous other air pollution modelling and validation efforts which are based on non-Bayesian methods. For example, Gulliver et al. (2011) provide a comparative assessment of methods to predict mean annual PM_{10} concentrations across 52 monitoring sites in London. This article does not, however, model temporal variation of the pollutants. Gulliver and Briggs (2011) discuss GIS-based air pollution dispersion models for city-wide exposure assessment at daily temporal resolution. In a similar vein, the ADMS-Urban model, see e.g. Carruthers et al. (2000) provide city-wide exposure maps. A number of other articles, e.g., Atkinson et al. (2012); Pirani et al. (2014, 2015); Rushworth

‡http://www.bbc.com/news/science-environment-26848489, accessed 10/11/2016, http://www.bbc.co.uk/news/uk-32233922, accessed 10/11/2016

et al. (2014) discuss and use air pollution estimates in order to estimate health effects. However, their main purpose is not exposure modelling and validation, and hence these articles do not report the accuracy of their air pollution estimates.

There has been much less interest in modelling UK-wide air pollution data. Recently, Lee et al. (2016) developed space-time models for air pollution data from England and Wales but only at the monthly temporal resolution to allow linkage with aggregated monthly health outcome data, which was the main purpose of that paper. High resolution spatial maps of UK-wide *annual* air pollution exposure levels for NO_2 and PM_{10} for the year 2001 only are available from the website, `http://www.envhealthatlas.co.uk/` (accessed 27/10/2016) which has been prepared by the Small Area Health Statistics Unit (SAHSU) of the Imperial College, London. They use land-use regression methods for predicting concentrations at a 100m $\times$ 100m spatial resolution. However, the website does not report the accuracy of the estimates nor does it provide estimates of air pollution at the daily temporal resolution.

Lack of statistical modelling for obtaining UK-wide air pollution estimates does not imply a lack of air pollution dispersion modelling using physical and chemical transport models. For example, Savage et al. (2013) develop the Air Quality Unified Model (AQUM) model for the whole of the UK. AQUM is a 3-dimensional weather and chemistry transport model used by the Met Office to deliver the UK national air quality forecast for the Department of Environment Food and Rural Affairs (Defra) and for scientific studies of atmospheric composition and air quality. The model has been run in hindcast mode to re-create hourly varying, UK air pollution concentrations for the period 2007 to 2011. However, the raw outputs of the AQUM are biased (Savage et al., 2013) and there are no associated uncertainties for the air pollution estimates. This drawback excludes the direct use of the AQUM estimates, or their deterministic adjusted values, in rigorous scientific health effects studies where the uncertainties of the air pollution estimates must be taken into account; otherwise the health effect estimates may be inflated (Lee et al., 2016). To overcome the biases in the AQUM, we model monitoring data which are more accurate. The use of the AQUM output frees us from having to use emission data, as in Pirani et al. (2014), and relevant meteorological data as in Sahu et al. (2007). These additional variables, already included as inputs in the AQUM, do not remain significant in the Bayesian model when outputs from a computer simulation model, such as the AQUM, are already present (Sahu and Bakar, 2012).

The primary motivation for this article is to develop space-time models for daily air pollution levels for five years, 2007-2011, in England and Wales. We use Bayesian model selection methods to validate and select the best model for each of the four pollutants NO_2, O_3, PM_{10} and $PM_{2.5}$. Using the selected model we obtain air pollution estimates at a 1 kilometre square grid covering the whole study region. Our modelling at the point level, described by a latitude and a longitude pair, and at the daily temporal scale, allows us to develop air pollution estimates at any coarser level of spatial resolution, e.g. local authority levels and at any aggregated temporal levels, e.g. quarterly and annual. The main advantage of the Bayesian methods, implemented using Markov chain Monte Carlo (MCMC), lies in the ability to estimate the uncertainties associated with the aggregated

pollution levels, as we demonstrate in detail. The resulting aggregated pollution estimates are then available to be used to integrate modelling of health outcome data which are often recorded for administrative geographies, such as the local authority areas, see e.g. Lee et al. (2016).

The remainder of the paper is organised as follows. Section 2 describes the data summaries and explores relationships between the pollutant levels and the AQUM values. In Section 3 we discuss the methods including the models and how to obtain the predictions at point level, and at coarser geographies. Data analysis and model validation results are presented in Section 4. Finally, a few summary remarks are placed in Section 5.

2. Data description

We model daily air pollution data collected from the 144 active AURN stations in England and Wales for the five years 2007 to 2011. AURN is the UK's largest automatic monitoring network and is the main network used for compliance reporting against the Ambient Air Quality Directives, see e.g. `http://uk-air.defra.gov.uk/networks` (accessed 17/11/2016). The stations, see the top panel of Figure 1 for their locations, measure oxides of nitrogen (NO_x), sulphur dioxide (SO_2), ozone (O_3), carbon monoxide (CO) and particles (PM_{10}, $PM_{2.5}$). Data from these stations are publicly available from a wide range of electronic, media and web platforms. The pollutants are measured at a height between 1.5 to 4 meters above the ground level§ depending on the type of pollutant. The data quality is verified and ratified on an ongoing basis as detailed in the cited Defra website.

We obtain the daily concentration data for four pollutants: 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 since these are the most harmful pollutants in the UK which may have health effects, see e.g. Lee et al. (2016). All pollutants are measured in microgram per cubic metre ($\mu g/m^3$) units. In our modelling we have used the daily maximum for NO_2, the daily maximum 8-hour running mean for ozone and the daily mean for both PM_{10} and $PM_{2.5}$ following the 2008 EU directive on air pollution, see Commission (2008).

For each pollutant, the total number of possible daily observations is 262,944 ($= 144 \times 1826$). However, in our data set there are 159,534 (60.67%), 98,628 (37.51%), 79,227 (30.13%) and 58,177 (22.12%) daily observations present, respectively, for the four pollutants NO_2, O_3, PM_{10} and $PM_{2.5}$. The remaining observations are missing due to several reasons including: instrument malfunction, discontinuation of some sites and then introduction of new replacement sites during the study period, or the fact that not all sites monitor all pollutants. Table 1 shows a year-wise break-up of the missing observations. Most notably, more than 90% of the $PM_{2.5}$ observations are missing for 2007 and 2008 because this pollutant was not monitored before 2009 since this was not considered to be one of the criteria pollutants (the ones which are regulated) until the publication of the 2008 EU air pollution directive. In spite of the missing values the spatio-temporal models in Section 3 are estimated using the large number of available observations as noted above.

§https://uk-air.defra.gov.uk/networks/site-info?site_id=LB

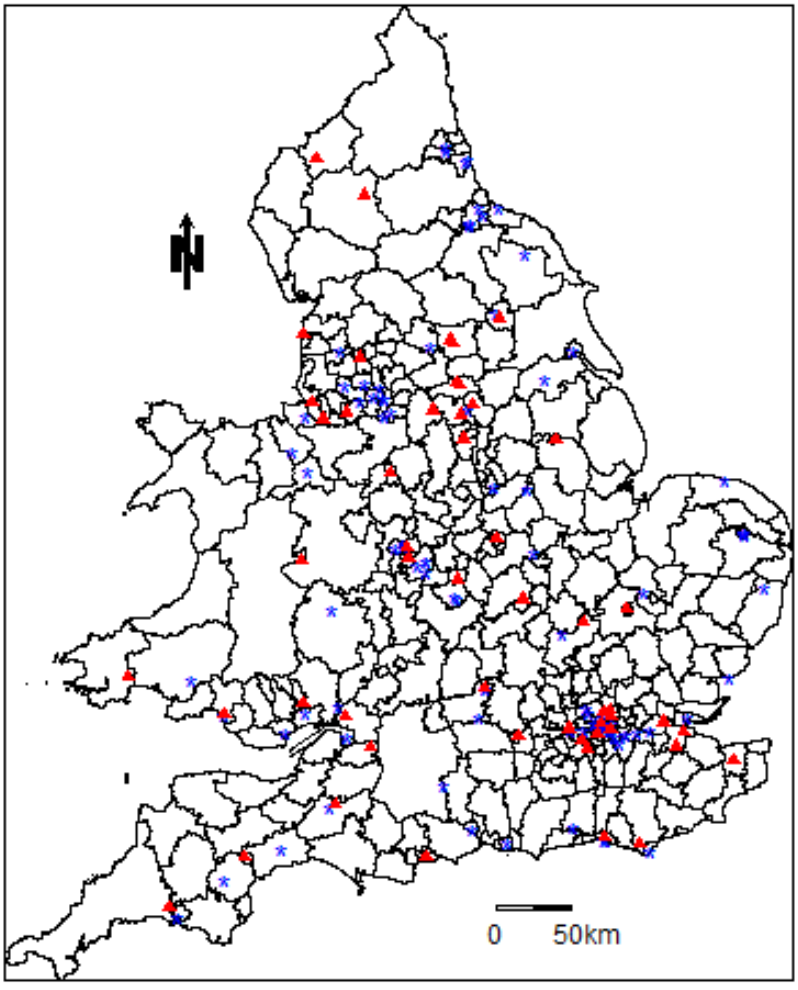

Fig. 1. A map showing the boundaries of 346 local authorities in England and Wales and the locations of 144 AURN monitoring sites. The (blue) stars represent the 90 modelling sites and the (red triangles) locate the 54 validation sites.

Table 1. Percentage of missing daily data out of the total number of observations in a year, which is 52704 (366$\times$ 144) for 2008 and 52560 (365$\times$ 144) for the other years.

Pollutant	2007	2008	2009	2010	2011	Overall
NO_2	39.16	39.23	38.39	38.09	35.47	38.07
O_3	56.56	62.87	64.05	63.05	61.81	61.67
PM_{10}	64.99	66.71	70.24	72.41	69.78	68.83
$PM_{2.5}$	96.66	92.21	67.57	65.18	64.75	77.26

The location of each of the 144 sites is classified into one of 3 site types: "Rural", "Urban" (which also includes suburban), and "Road and Kerbside" (denoted by RKS). Summary statistics, as reported in Table 2, show significant variability in all four pollutants across these three site types. As expected, the table shows that the Rural sites are less polluted than the Urban and the RKS sites for NO_2, PM_{10} and $PM_{2.5}$while the converse is true for O_3. Ozone concentrations are generally lower in urban areas than in the surrounding rural areas due to reaction with NO_X (mostly NO) emissions, which are greatest in urban areas, see e.g. Grewe et al. (2012) and also page 3 of the report, "Ozone in the United Kingdom", prepared by the Air Quality Expert Group and available from `http://www.defra.gov.uk/environment/airquality/aqeg`. Boxplots of the distributions of each pollutant by site type are displayed in Figure 2. These plots also show differences in variances for each pollutant by site type, with RKS sites having a larger spread for NO_2, PM_{10} and $PM_{2.5}$ due to many extreme observations. Finally, boxplots of the pollution concentrations by year (plots omitted for brevity) showed little variability in the median from year to year, although there is considerable variation in the distribution of the extreme values.

The methods proposed in this paper address the sparsity of the observed AURN data by including the hourly hindcasts of the AQUM, developed by Savage et al. (2013) especially for the UK. AQUM is a chemistry transport model based on weather and emission inventory data for which full details are provided in Savage et al. (2013). However, AQUM does not use the observed AURN data but it produces output on a 12 kilometre square grid covering the whole of UK. We use bi-linear interpolation methods, which interpolate once horizontally and then vertically, see e.g. the Wikipedia entry¶, to predict the hourly pollution concentration values at the corners of a one kilometre square grid we use for our prediction purposes. These hourly values are subsequently aggregated to daily values for our modelling purposes. The corners of these 1 kilometre grid cells do not coincide with the locations of the 144 AURN monitoring sites and to resolve this misalignment we use bilinear interpolation, see e.g. of the four grid corners of each site to estimate the daily AQUM output for the 144 observation sites. These output are moderately correlated with the corresponding observations with correlations of 0.44 for NO_2, 0.68 for O_3, 0.59 for PM_{10} and 0.61 for $PM_{2.5}$ respectively. We have also obtained scatter plots showing these correlations but those are omitted for brevity. These moderate values of correlations help us in regression modelling of the observations that we consider in the next section.

Inclusion of the site type classifications (Rural, Urban and RKS) poses a problem for predictions at locations for which site types are unknown. We tackle the Rural/Urban classification problem by using map data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, see Schneider et al. (2009). These maps, based on satellite data from 2001-2002, provide land use information (urban/rural) at a 500m spatial resolution and has an overall accuracy of 93%. Thus the rural/urban classification for each corner of a one kilometre predictive grid locations is accurately obtained from these maps. However, these data do not provide the RKS site type classification for the one kilometre predictive grid. To obtain these classifications, we use the detailed open data roads net-

¶https://en.wikipedia.org/wiki/Bilinear_interpolation

Table 2. Summary statistics of the four pollutants measured at sites in $\mu g/m^3$ units. AD stands for the available number of data points based on which the summaries have been calculated, SD stands for standard deviation. Total size of combined data is 262,944 (144$\times$ 1826).

NO_2						
Site Type	Min	Median	Mean	Max	SD	AD
Rural (16)	0	14.70	19.55	183.9	15.54	20,292
Urban (80)	2	47.40	49.91	401.0	26.76	88,529
RKS (48)	2	69.0	76.02	506.0	43.62	54,005
Combined (144)	0	50.0	54.79	506.0	36.98	162,826
O_3						
Rural (16)	2.25	67.50	68.66	187.5	18.88	20,900
Urban (80)	0	55.75	56.08	171.0	21.95	70,473
RKS (48)	0	49.50	49.44	145.0	22.40	9,406
Combined (144)	0	58.25	58.07	187.50	22.15	100,779
PM_{10}						
Rural (16)	0	15.0	16.99	117.60	9.82	4,364
Urban (80)	0.43	17.08	19.98	194.30	11.48	47,778
RKS (48)	0	19.12	22.54	174.70	12.69	29,817
Combined (144)	0	17.68	20.75	194.30	11.95	81,959
$PM_{2.5}$						
Rural (16)	2.20	8.92	10.72	67.62	6.49	3,199
Urban (80)	0.08	10.71	13.63	113.50	9.53	38,188
RKS (48)	0	11.85	14.66	90.96	9.79	18,402
Combined (144)	0	10.92	13.79	113.50	9.51	59,789

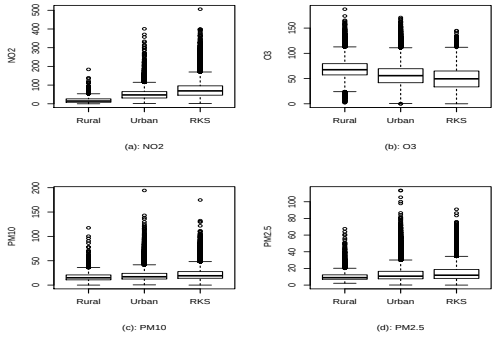

Fig. 2. Boxplots of the daily average concentrations for each pollutant by three site types. The central box in each plot here (and in the other figures below) shows the quartiles, the whiskers show the farthest observation from the median which is still within 1.5 times the inter-quartile range and any observations outside the whiskers are suspected outliers, ploted by the symbol 'o'.

work product from the Ordnance Survey‖. We calculate the distance of each of the one kilometre grid locations to the nearest road. The grid locations which are within 4 metres of the nearest road are classified as RKS.

3. Model specification and prediction

In our study region of England and Wales, we have daily data from $n = 144$ sites for $T = 1826$ days. As is common practice in modelling air pollution concentration data we model on the square-root scale to stabilise the variance (Sahu et al., 2007; Berrocal et al., 2010). However, for ease of interpretation, the accuracy of all predictions from the pollution model are assessed on the original scale. Let $z_{(p)}(\mathbf{s}_i, t)$, and $x_{(p)}(\mathbf{s}_i, t)$ respectively denote the measured and modelled AQUM pollution concentration on the square root scale at location $\mathbf{s}_i$ during day t for pollutant p. In what follows we suppress the subscript p for notational clarity.

3.1. Hierarchical model

As part of the Bayesian modelling hierarchy we proceed with the top-level specification

$$Z(\mathbf{s}_i, t) = Y(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \;\; \epsilon(\mathbf{s}_i, t) \sim \mathrm{N}(0, \sigma^2_\epsilon), \tag{1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $Y(\mathbf{s}_i, t)$ is the true process and $\epsilon(\mathbf{s}_i, t)$ is the independent nugget effect absorbing micro-scale variability. (Henceforth, values of the subscripts i and t will always be in the range mentioned here.) At the next stage of the hierarchy we specify:

$$Y(\mathbf{s}_i, t) = \mu(\mathbf{s}_i, t) + \eta(\mathbf{s}_i, t) \tag{2}$$

where $\mu(\mathbf{s}_i, t)$ denotes the mean surface and $\eta(\mathbf{s}_i, t)$ is a space-time process, which we specify later. The mean surface is modelled as:

$$\mu(\mathbf{s}_i, t) = \gamma_0 + \gamma_1 x(\mathbf{s}_i, t) + \sum_{j=2}^{r} \delta_j(\mathbf{s}_i) \left(\gamma_{0j} + \gamma_{1j} x(\mathbf{s}_i, t)\right), \tag{3}$$

where we propose a site type specific regression on the modelled square-root AQUM concentrations $x(\mathbf{s}_i, t)$. Here $r = 3$, corresponding to the three site types (Rural, Urban, RKS), and the rural site type corresponds to $j = 1$ and is the base line level. Thus (γ_0, γ_1) are respectively the slope and intercept terms for the Rural sites, while $(\gamma_{0j}, \gamma_{1j})$ are the incremental adjustments for site type j, $j = 2, 3$. Finally, $\delta_j(\mathbf{s}_i)$ is an indicator function, equalling one if site $\mathbf{s}_i$ is of the jth site type and zero otherwise.

3.2. Specification of the spatio-temporal process

We consider four modelling possibilities for the spatio-temporal process $\eta(\mathbf{s}_i, t)$, which differ in their level of sophistication. The first one is the simplest that assumes $\eta(\mathbf{s}_i, t) =$

‖https://www.ordnancesurvey.co.uk/opendatadownload/products.html, accessed 10/11/2016

0 for all sites $\mathbf{s}_i$ and times t, which renders (1) a simple regression model that is used for comparison purposes with the other two models. The second model for $\eta(\mathbf{s}_i, t)$ is an independent over time Gaussian process (GP) with zero mean and a Matérn covariance function given by:

$$\text{Cov}(\eta(\mathbf{s}, t), \eta(\mathbf{s}', t)) = \frac{\sigma_\eta^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}||\mathbf{s} - \mathbf{s}'||\phi)^\nu K_\nu(2\sqrt{\nu}||\mathbf{s} - \mathbf{s}'||\phi), \tag{4}$$

for $\phi > 0$ and $\nu > 0$, where $\Gamma(\nu)$ is the standard gamma function, K_ν is the modified Bessel function of the second kind with order ν, and $||\mathbf{s} - \mathbf{s}'||$ is the distance between sites $\mathbf{s}$ and $\mathbf{s}'$. The parameter ϕ controls the rate of decay of the correlation as the distance $||\mathbf{s}_i - \mathbf{s}_j||$ increases and the parameter ν controls smoothness of the random field (Banerjee et al., 2004; Cressie, 1993). That is, for each time t,

$$\boldsymbol{\eta}_t = (\eta(\mathbf{s}_1, t), \ldots, \eta(\mathbf{s}_n, t))^\top \sim \text{N}(\mathbf{0}, \sigma_\eta^2 H_\eta(\phi, \nu)), \tag{5}$$

where $H_\eta(\phi, \nu)_{ij} = C(||\mathbf{s}_i - \mathbf{s}_j||; \phi, \nu), j = 1, \ldots, n$, which is assumed to be the Matérn correlation function with decay and smoothness parameters ϕ and ν respectively. The special case of the Matérn correlation function when $\nu = 0.5$ is called the exponential correlation function given by:

$$\text{Corr}(\eta(\mathbf{s}, t), \eta(\mathbf{s}', t)) = \exp(-||\mathbf{s} - \mathbf{s}'||\,\phi).$$

In this case ϕ alone determines the rate of decay of the spatial correlation as the distance $d = ||\mathbf{s} - \mathbf{s}'||$ between two sites increases. A related quantity, called the effective range, best describes the limit of the correlation decay to zero in practical situations. However, $\exp(-d\,\phi) \neq 0$ for any finite value of d and ϕ. Hence the *effective range*, for a given value of ϕ, is defined to be the value of the distance d for which $\exp(-d\phi) = 0.05$, which implies $d = 3/\phi$.

The third model introduces non-stationary covariance structure following Sahu and Mukhopadhyay (2015). This is achieved by first assuming a set of knot-locations, $\mathbf{S}_m^* = (\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*)$, which are specified below, for a value of m which will be chosen by out of sample prediction performance. Given $\mathbf{S}_m^*$, we assume that $\boldsymbol{\eta}_t^* = (\eta(\mathbf{s}_1^*, t), \ldots, \eta(\mathbf{s}_m^*, t))^\top$ is a zero mean GP with the Matérn covariance function (4). The non-stationary modelling proposal is to replace $\eta(\mathbf{s}_i, t)$ in (2) by

$$\tilde{\eta}(\mathbf{s}_i, t) = E\left[\eta(\mathbf{s}_i, t)|\boldsymbol{\eta}_t^*\right]. \tag{6}$$

The $n+m$ dimensional vector $(\boldsymbol{\eta}_t, \boldsymbol{\eta}_t^*)$ is assumed to be a realisation of the same underlying GP as in (4) independently for each t. By writing $\tilde{\boldsymbol{\eta}}_t = (\tilde{\eta}(\mathbf{s}_1, t), \ldots, \tilde{\eta}(\mathbf{s}_n, t))^\top$ and using multivariate Gaussian theory we have

$$\tilde{\boldsymbol{\eta}}_t = C^*(\phi, \nu) H_{\eta^*}^{-1}(\phi, \nu) \boldsymbol{\eta}_t^* \tag{7}$$

where $C^*(\phi, \nu)$ is the $n \times m$ cross-correlation matrix between $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_t^*$, i.e. $(C^*)_{ij} = C(|\mathbf{s}_i - \mathbf{s}_j^*|; \phi, \nu)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ and $H_{\eta^*}(\phi, \nu)$ is an $m \times m$ correlation matrix having elements $H_{\eta^*}(\phi, \nu)_{kj} = C(||\mathbf{s}_k^* - \mathbf{s}_j^*||; \phi, \nu)$, for $k, j = 1, \ldots, m$. Clearly,

$$\text{Var}(\tilde{\boldsymbol{\eta}}_t) = C^*(\phi, \nu) H_{\eta^*}^{-1}(\phi, \nu) C^{*T}(\phi, \nu),$$

which shows non-stationarity of the $\tilde{\boldsymbol{\eta}}_t$ process.

The $\tilde{\boldsymbol{\eta}}_t$ surface, is now based on linear functions of the m-dimensional $\boldsymbol{\eta}_t^*$ instead of the n-dimensional $\boldsymbol{\eta}_t$. This leads to a reduction of computational burden when m is much smaller than n. However, $n = 144$ is not very large in this paper and hence dimension reduction by Gaussian Predictive Process approximation (Banerjee et al., 2008) is not required here. We still use (7) to benefit from having a flexible non-stationary model for the spatial random effects. Sahu and Mukhopadhyay (2015) show that this flexibility in modelling, even when $m > n$, leads to more accurate predictive models, which are also considered as candidate models in this paper.

Finally, we remove the temporal independence assumption in the previous model by introducing an autoregressive model for $\boldsymbol{\eta}_t^*$. Here we assume that

$$\boldsymbol{\eta}_t^* \sim N\left(\varrho\boldsymbol{\eta}_{t-1}^*, \sigma_\eta^2 H_{\eta^*}(\phi,\nu)\right), \quad \text{for } t = 1,\ldots,T, \tag{8}$$

with $\boldsymbol{\eta}_0^* = \mathbf{0}$ as the initial condition and ϱ as the unknown auto-regressive parameter for which we specify a prior distribution. Thus, given m and $\mathbf{S}_m^*$, $\boldsymbol{\eta}_t^*$ is determined from the above auto-regressive process, and then $\tilde{\boldsymbol{\eta}}_t$ is obtained by using (7) for all $t = 1,\ldots,T$.

3.3. Specifying the knot locations

Now we return to specifying the knot locations $\mathbf{S}_m^*$ for a given m where m is to be chosen by cross-validation. Sahu and Mukhopadhyay (2015) show that a random selection of $\mathbf{S}_m^*$ is preferable to a space filling design that distributes the m locations evenly within the study region. In our implementation, we discretise by M points, which are the corners of a set of 1-kilometre grid squares covering the study region. To add flexibility we assume a probability surface $p(\mathbf{s}_j^*), j = 1,\ldots,M$ such that $\sum_{j=1}^M p(\mathbf{s}_j^*) = 1$ and $p(\mathbf{s}_j^*) \geq 0$ for $j = 1,\ldots,M$. The $p(\mathbf{s}_j^*)$ can be a normalised population density surface that will guarantee knots being placed at high density areas. A random sample of m locations is proposed to be used as the knot-locations $\mathbf{S}_m^*$. In effect, this implies a discrete prior distribution for $\mathbf{S}_m^*$ and MCMC model fitting is easily accomplished by using a Metropolis-Hastings step for $\mathbf{S}_m^*$ where the proposal samples are drawn from the prior itself. Throughout, we choose $m = 25$ which was chosen by an out of sample root mean square prediction error criterion (see Section 3.6 below) among the possible values of 16, 25, 36, 49 and 100. A complete square value, e.g. 25, are put forward as candidates so that an equal number of points are chosen in the two co-ordinate directions.

3.4. Specification of the prior distributions

Our proposed Bayesian model is completed by assigning vague but proper prior distributions for the remaining model parameters. These include zero-mean Gaussian priors for the regression parameters γ's in (3) with large variances of 10^4, inverse gamma prior distribution for the variance parameters $(\sigma_\epsilon^2, \sigma_\eta^2)$ with hyper-parameters $(2, 1)$ parameterised to have mean 1 and infinite variance. Informative (and thus proper) prior distributions must be specified for the two parameters ν and ϕ describing the Matérn correlation function, since

these are weakly identified in the likelihood, see e.g. Zhang (2004), who shows that consistent estimation is not possible for these parameters. In addition, sparsely observed spatial data are not very informative for the smoothness parameter ν, see e.g. Stein (1999). Hence we shall adopt the exponential correlation function corresponding to $\nu = 0.5$ following many authors, e.g. Berrocal et al. (2010); Sahu et al. (2007).

Now it remains to specify a proper prior distribution for ϕ. Here, besides fixing ϕ at several plausible effective ranges we entertain two proper prior distributions: (1) a uniform prior distribution in the interval $(0.001, 0.01)$ corresponding to having an effective range between 300 and 3000 kilometres and (2) a gamma prior distribution with parameters 2 and 1, parameterised to have mean 2 but with infinite variance. The implied effective range between 300 and 3000 kilometres corresponding to the uniform prior distribution provides opportunities for the models to have adequate spatial correlation but avoiding singularity corresponding to an infinite range. However, the gamma prior distribution does not bound ϕ within any finite range unlike the uniform prior distribution. In Section 4, suitable prediction validation criteria will be used to choose between these contrasting specifications.

3.5. Prediction details

The hierarchical space-time model allows us to predict pollutant concentrations at any location $\mathbf{s}'$ and at any time point $t, 1 \leq t \leq T$. We first consider prediction of the large percentages of the missing data as noted in Section 2. Suppose that $Z(\mathbf{s}_i, t)$ is missing for particular values of $\mathbf{s}_i$ and t. By virtue of the Bayesian model, implemented using MCMC, we have a value of the $\mu(\mathbf{s}_i, t)^{(\ell)}$ at each iteration ℓ for $\ell = 1, \ldots, L$ of the implemented MCMC algorithm. (Henceforth ℓ will denote the MCMC iteration index and L will denote the retained number of MCMC iterates we use for inference.) Using the sampled values of the spatio-temporal process $\eta(\mathbf{s}_i, t)^{(\ell)}$ we obtain $Y(\mathbf{s}_i, t)^{(\ell)}$ following Equation (2). Subsequently, we sample $Z(\mathbf{s}_i, t)^{(\ell)}$ using (1) which also needs $\sigma_\epsilon^{2(\ell)}$. At the end of the MCMC run, we utilse the samples $Z(\mathbf{s}_i, t)^{(\ell)}, \ell = 1, \ldots, L$ to esimate the missing $Z(\mathbf{s}_i, t)$ value and its variability.

Prediction of $Z(\mathbf{s}', t)$ at a new location $\mathbf{s}'$, where $\mathbf{s}'$ is not one of $\mathbf{s}_1, \ldots, \mathbf{s}_n$, using the Bayesian model, proceeds by evaluating the posterior predictive distribution:

$$\pi(z(\mathbf{s}', t)|\mathbf{z}) = \int \pi(z(\mathbf{s}', t)|\mathbf{S}_m^*, \boldsymbol{\eta}^*, \boldsymbol{\theta}, \mathbf{z}) \pi(\mathbf{S}_m^*, \boldsymbol{\eta}^*, \boldsymbol{\theta}|\mathbf{z}) d\mathbf{S}_m^* d\boldsymbol{\eta}^* d\boldsymbol{\theta}, \tag{9}$$

where $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \varrho, \sigma_\epsilon^2, \sigma_\eta^2, \phi)^T$ denotes the model parameters and $\boldsymbol{\eta}^* = (\boldsymbol{\eta}_1^*, \ldots, \boldsymbol{\eta}_T^*)$. Here $\pi(z(\mathbf{s}', t)|\mathbf{S}_m^*, \boldsymbol{\eta}^*, \boldsymbol{\theta})$ requires $\tilde{\eta}(\mathbf{s}', t)$, see (1), which is calculated as $\mathbf{c}^{*T}(\mathbf{s}'; \phi, \nu) H_{\eta^*}^{-1}(\phi, \nu) \boldsymbol{\eta}_t^*$, analogous to (7), where $\mathbf{c}^*(\mathbf{s}'; \phi, \nu)$ is $m \times 1$ with the ith element given by $\mathbf{c}_i^*(\mathbf{s}'; \phi, \nu) = C(|\mathbf{s}_i^* - \mathbf{s}'|; \phi, \nu)$. Composition sampling is used to simulate from (9) as follows. At the ℓth MCMC iteration samples of $(\mathbf{S}_m^{*(\ell)}, \boldsymbol{\eta}^{*(\ell)}, \boldsymbol{\theta}^{(\ell)})$ are drawn from the joint posterior distribution. Now $z^{(\ell)}(\mathbf{s}', t)$ is to be sampled from the Gaussian distribution given by (1). However, this requires $Y^{(\ell)}(\mathbf{s}', t)$ which in turn requires $\mu^{(\ell)}(\mathbf{s}', t)$ and $\eta^{(\ell)}(\mathbf{s}', t)$, see (2). The mean term $\mu^{(\ell)}(\mathbf{s}', t)$ is evaluated using the $\boldsymbol{\gamma}^{(\ell)}$ in (3). This evaluation also requires

the site type indicators $\delta_j(\mathbf{s}'), j = 2, \ldots, r$ which we take from the MODIS satellite data as discussed in Section 2. Also note that (3) requires $x(\mathbf{s}', t)$ which we obtain using bilinear interpolation as mentioned in Section 2.

For $\eta^{(\ell)}(\mathbf{s}', t)$ we note that according to (7) we have $\tilde{\eta}(\mathbf{s}', t) = \mathbf{c}^{*T}(\mathbf{s}'; \phi, \nu) H_{\eta^*}^{-1}(\phi, \nu) \boldsymbol{\eta}_t^*$. Hence $\tilde{\eta}^{(\ell)}(\mathbf{s}', t)$ is obtained as:

$$\tilde{\eta}^{(\ell)}(\mathbf{s}', t) = \mathbf{c}^{*T}(\phi^{(\ell)}, \nu^{(\ell)}) H_{\eta^*}^{-1}(\phi^{(\ell)}, \nu^{(\ell)}) \boldsymbol{\eta}_t^{*(\ell)}.$$

Now we obtain $Y^{(\ell)}(\mathbf{s}', t) = \mu^{(\ell)}(\mathbf{s}', t) + \tilde{\eta}^{(\ell)}(\mathbf{s}', t)$ using (2). Finally, $z^{(\ell)}(\mathbf{s}', t)$ is obtained as a draw from the Gaussian distribution with mean $Y^{(\ell)}(\mathbf{s}', t)$ and variance $\sigma_\epsilon^{2(\ell)}$.

3.6. Evaluating the predictive performance

The predictive performance of each pollution model is assessed by a cross-validation exercise, where data at sites $(\mathbf{s}_1, \ldots, \mathbf{s}_{n_0})$ are used to fit the model while data at sites $(\mathbf{s}_{n_0+1}, \ldots, \mathbf{s}_n)$ are held out to assess predictive performance. The root mean square prediction error (RMSPE) and mean absolute prediction error (MAPE) are used to quantify prediction accuracy, which are given by

$$\text{RMSPE} = \sqrt{\frac{1}{N_v} \sum_{j=n_0+1}^{n} \sum_{t=1}^{T} (z(\mathbf{s}_j, t) - \hat{z}(\mathbf{s}_j, t))^2},$$

$$\text{MAPE} = \frac{1}{N_v} \sum_{j=n_0+1}^{n} \sum_{t=1}^{T} |z(\mathbf{s}_j, t) - \hat{z}(\mathbf{s}_j, t)|,$$

where $\hat{z}(\mathbf{s}_j, t)$ is the posterior median from the predictive distribution. Here N_v is the total number of available (i.e. non-missing) observations from these validation sites over the $T = 1826$ days. These criteria are used to select the best model. Once the best model has been chosen, we use all the available data from $n = 144$ sites for inferential purposes.

3.7. Aggregating predictions to administrative geographies

The proposed geo-statistical models are able to predict at any unmonitored location $\mathbf{s}'$ using the methodology described above. In this section we consider spatial aggregation to any coarser level administrative geography, e.g., local authorities or electoral wards, which may be required to align with aggregated health outcome data. For the kth administrative region, denoted by A_k, for $k = 1, \ldots, K$, where K is the total number of such regions, we define the average pollution concentration at time t by

$$Z_{kt} = \frac{1}{|A_k|} \int_{A_k} Z(\mathbf{s}, t) d\mathbf{s}, \tag{10}$$

where $|A_k|$ is the area of the region A_k. We approximate (10) using numerical integration as

$$\bar{Z}_{kt} = \frac{1}{n_k} \sum_{j=1}^{n_k} Z(\mathbf{s}_{kj}, t), \tag{11}$$

where $(\mathbf{s}_{k1}, \ldots, \mathbf{s}_{kn_k})$ forms a fine grid of prediction locations within the kth region. As mentioned in the Introduction, we perform the predictions at a 1 kilometre square grid covering the study region. Hence, the n_k locations, $(\mathbf{s}_{k1}, \ldots, \mathbf{s}_{kn_k})$, are simply taken as the corners of the 1 kilometre square grid which fall within A_k.

Under the Bayesian inference setting, $\bar{Z}_{kt}$ will have a posterior predictive distribution given the observed data $\mathbf{z}$ since each $Z(\mathbf{s}_{kj}, t)$ also has such a distribution. This posterior predictive distribution can be summarised using MCMC samples drawn from it using composition sampling. Corresponding to each draw from the joint posterior distribution we draw a sample $z^{(\ell)}(\mathbf{s}_{kj}, t)$ from the posterior predictive distribution (9). At each MCMC iteration ℓ, we form the average

$$\bar{Z}_{kt}^{(\ell)} = \frac{1}{n_k} \sum_{j=1}^{n_k} z^{(\ell)}(\mathbf{s}_{kj}, t).$$

These MCMC iterates, $\bar{Z}_{kt}^{(\ell)}, \ell = 1, \ldots, L$ are summarised to estimate Z_{kt}. Uncertainties in these estimates are also easily estimated using the MCMC iterates. Note that temporal aggregation can easily be performed using the MCMC iterates. For example, if it is required to estimate $\bar{Z}_k = \frac{1}{T} \sum_{t=1}^{T} \bar{Z}_{kt}$, then we simply obtain

$$\bar{Z}_{k}^{(\ell)} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{n_k} \sum_{j=1}^{n_k} z^{(\ell)}(\mathbf{s}_{kj}, t) \tag{12}$$

for each $\ell = 1, \ldots, L$ and summaries e.g, mean and median of these MCMC iterates are used. Moreover, uncertainties in these estimates expressed through, for example, credible intervals and standard deviations are also obtained using the MCMC iterates. Prediction details for $z^{(\ell)}(\mathbf{s}_{kj}, t)$ for any arbitrary site $\mathbf{s}_{kj}$ have been noted in the previous sub-section. Lastly, we also note that temporal aggregation for a sub-period (e.g. annual) of the whole time domain can be performed similarly by simply taking the average similar to the one in (12) but only using the simulations $z^{(\ell)}(\mathbf{s}_{kj}, t)$ for all the individual time points t which fall in that sub-period. For example, we can find annual averages from modelling five years' data.

4. Results

4.1. Implementation details and models setup

All model fitting and model choice results are based on $L = 5000$ MCMC iterations, after discarding first 5000 iterations at which point convergence was assessed. We first consider the issue of model choice and validation for each pollutant separately. These tasks are performed by fitting the models with data from 90 (62.5%) randomly selected sites and then validating the available observations from the remaining 54 sites. Thus model fitting is done using a maximum of 164,340 $(= 90 \times 1826)$ observations and validation is done using a maximum of 98,604 $(= 54 \times 1826)$ observations. For individual pollutants the actual numbers of fitting and validation observations will be less than these maximums

because of missing data. The 90 fitting sites contained 50 Urban, 11 Rural and 29 RKS sites while the 54 validations sites contained 29 Urban, 7 Rural and 18 RKS sites. Thus all three site types are adequately represented in both the fitting and validation set of sites. For model comparison purposes we consider the RMSPE, MAPE, bias, actual coverage of the 95% prediction intervals, since spatial prediction is the main objective here.

For each pollutant, in addition to the AQUM, we entertain nine modelling possibilities as follows. The first one is simple Kriging, performed independently for each time point (day here), using the well known `fields` package (Furrer et al. (2013)) in the `R` programming language. The second model is a simple linear regression model, implemented using MCMC in the Bayesian framework, that does not take care of the spatio-temporal dependence in the data, viz. models (1), (2) and (3) with $\eta(\mathbf{s}_i, t) = 0$ for all i and t. To introduce spatial dependence only we consider the independent in time GP model described as the second model in Section 3. These three models are chosen solely for benchmarking purposes since these off-the-shelf methods are often used in practice and also these are simple stationary GP Versions of our proposed models.

The remaining seven Bayesian models are all based on the most complex non-stationary spatio-temporal models described in Section 3. The first five of the seven models are obtained for five different fixed values of ϕ, the decay parameter. The five values, denoted by $\phi_i, i = 1, \ldots, 5$ correspond to effective ranges of 3500 (ϕ_1), 3000 (ϕ_2), 600 (ϕ_3), 300 (ϕ_4) and 100 (ϕ_5) kilometres respectively, and these choices are guided by the need to include large to moderate amounts of spatial correlation into the model. In the final two modelling attempts we specify the uniform and gamma prior distributions for the decay parameter ϕ as mentioned in Section 3.1.

4.2. Model validation and comparison results

Tables 3 and 4 report the validation results for AQUM and all nine modelling scenarios noted above. The tables show that the raw AQUM outputs are far worse than all nine modelling strategies including Kriging. This is expected as AQUM does not use the actual observations and those output are heavily biased as noted previously. All six Bayesian models, based on the non-stationary spatio-temporal models, perform much better than the three bench marking strategies: Kriging, Linear and GP. The differences between the performances of the six Bayesian models are not very large which is due to the fact that the dominant linear model part is same for all six models. Slight differences in performance are observed by varying the spatial correlation structure. For NO_2 the model with ϕ fixed at ϕ_2 is chosen to be the best; the model with the Uniform prior distribution is best for O_3 and the model with the Gamma prior distribution is best for PM_{10} and $PM_{2.5}$. Each of the four best models reports an R^2 value, which is the value of the sample correlation coefficient between the observations and their predictions, between 80 to 90% which shows very good agreement between the model based predictions and the held out observations. Model adequacy can be checked by the achieved coverages of the 95% prediction intervals which are seen to be more than 90% for NO_2, O_3 and PM_{10}. However, the achieved coverages are between 80 to 84.6% for $PM_{2.5}$. This is most likely because of the smaller sample sizes available both for fitting and validation for this pollutant. However, RMSPE

of the best model is 4.30, which is less than half of the standard deviation of 9.52 for the full data (see Table 2) and hence the model does perform quite well in reducing prediction variability.

Next we investigate if the best chosen model for each of the four pollutants performs similarly in validating the three site types: Rural, Urban and RKS. Recall that there were 29 Urban, 7 Rural and 18 RKS sites among the 54 validation sites. Table 5 shows the RMSPE's aggregated by the site types. The table shows that the RMSPE is slightly smaller for the Urban sites for NO_2, PM_{10} and $PM_{2.5}$. This is expected since the Urban sites outnumber the other site types both among the fitting and validation sites, which leads to more accurate estimation and more stable value of the RMSPE. For O_3, the Rural sites have a smaller RMSPE value, which shows that the higher pollution values are estimated with on average more accuracy.

So that we can compare the accuracy of the proposed models with that of the previous modelling efforts of Pirani et al. (2014) we consider fitting our models to data from monitoring sites in Greater London only. We only perform this experiment for PM_{10} because comparable model validation results for daily data are available from Pirani et al. (2014) for this pollutant only. Table 6 reports the results for their best model and the best performing model in this paper. Our model has much lower RMSPE with a better R^2 value than the Pirani et al. (2014) modelling effort, although this is not an exact comparison since the data used in Pirani et al. (2014) and in this paper are different. However, some justification for the comparison comes from the fact that the error rates are for daily data observed in the same spatial domain of Greater London.

4.3. Parameter estimates

Parameter estimates for the best Bayesian models are presented in Table 7. These parameter estimates have been obtained by fitting the models with all the available data from 144 sites, see the last column of Table 2. The models rightly find significantly elevated levels of NO_2 concentration in the urban and roadside sites compared to the rural sites since all the incremental slopes and intercepts are positively significant in Table 7. However, this behaviour reverses for O_3 which is known to be lower in urban and roadside areas due to its negative correlation with NO_2. The estimates for the same parameters for models fitted to the two species of particulate matter show a mixed behaviour regarding the three site types. The estimates show moderately significant temporal correlation, but spatial correlation seems to be more dominant since the estimates of ϕ, $\hat{\phi}$ show large effective ranges $(3/\hat{\phi})$.

The spatial variance, σ_η^2 is estimated to be higher than the nugget effect σ_ϵ^2 except for the case of modelling NO_2. This is plausible since the variability of NO_2 is much higher than the other three pollutants (see Table2) and the model absorbs the extra variability using the nugget effect parameter, σ_ϵ^2. Overall, all of the linear parameters show significantly different linear relationships between the AQUM model output and the observations in the square-root transformed scale. Hence, this implies considerable modelling success with the AQUM outputs that results in very low out of sample root mean square prediction errors.

Table 3. Assessment of predictive performance for a range of models for NO_2 and O_3. SS stands for sample size which is the number of daily observations. R^2 denotes the sample correlation coefficient between the predictions and actual observations.

Model	**RMSPE**	**MAPE**	**Bias**	**Coverage (%)**	R^2
NO_2: Fitting SS = 100,138, validation SS=62,688					
AQUM	39.12	25.50	–17.06	–	0.42
Kriging	32.87	22.88	2.56	69.59	0.53
Linear	30.46	19.63	–5.09	94.43	0.60
GP	31.46	22.08	1.73	98.08	0.58
ϕ_1	18.30	13.60	1.22	96.64	0.88
ϕ_2	17.65	12.99	0.41	97.42	0.89
ϕ_3	17.90	12.86	–0.70	97.43	0.89
ϕ_4	20.85	14.56	0.05	96.59	0.84
Unif	17.82	13.07	0.32	97.35	0.89
Gamma	19.23	13.64	–1.81	95.36	0.87
O_3: Fitting SS = 64,373, validation SS=36,406					
AQUM	20.80	16.04	–2.95	–	0.68
Kriging	13.30	9.86	–2.95	78.25	0.80
Linear	16.0	12.42	8.47	93.86	0.69
GP	16.53	12.81	3.76	99.52	0.66
ϕ_1	10.22	7.62	0.09	91.61	0.89
ϕ_2	10.33	7.68	0.11	91.18	0.88
ϕ_3	10.27	7.65	0.22	91.93	0.88
ϕ_4	11.31	8.41	0.29	92.38	0.86
Uniform	10.17	7.59	0.07	91.72	0.89
Gamma	10.42	7.75	0.20	91.03	0.88

Table 4. Assessment of predictive performance for a range of models for PM_{10} and $PM_{2.5}$. SS stands for sample size which is the number of daily observations. R^2 denotes the sample correlation coefficient between the predictions and actual observations.

Model	RMSPE	MAPE	Bias	Coverage (%)	R^2
PM_{10}: Fitting SS = 52,625, validation SS=29,334					
AQUM	14.70	11.36	–10.74	–	0.59
Kriging	7.34	4.75	–0.75	64.96	0.77
Linear	9.98	6.74	–1.74	93.70	0.61
GP	10.24	7.75	1.89	99.79	0.57
ϕ_1	5.49	3.58	–0.62	89.23	0.80
ϕ_2	5.51	3.63	–0.46	89.07	0.80
ϕ_3	5.59	3.61	–0.80	90.75	0.80
ϕ_4	6.21	3.97	–0.57	87.32	0.79
Unif	5.65	3.17	–0.52	89.02	0.80
Gamma	5.48	3.56	–0.65	90.03	0.81
$PM_{2.5}$: Fitting SS = 38,481, validation SS=21,308					
AQUM	9.29	6.35	–4.53	–	0.58
Kriging	4.63	2.96	–0.72	67.84	0.81
Linear	8.03	5.30	–1.87	92.73	0.60
GP	8.38	6.59	2.08	99.92	0.85
ϕ_1	4.45	3.25	–1.23	82.74	0.85
ϕ_2	4.32	2.66	–1.17	83.84	0.85
ϕ_3	4.38	2.72	–1.03	84.61	0.85
ϕ_4	4.99	3.05	–1.54	83.30	0.81
Uniform	4.56	2.85	–1.13	84.21	0.85
Gamma	4.30	2.66	–0.97	82.38	0.85

Table 5. Aggregated RMSPEs of individual pollutants according to the three site types. The results corresponds to the best model for each pollutant as reported in Table 3 and Table 4.

Pollutant	Rural	Urban	RKS
NO_2	18.69	15.63	19.53
O_3	9.59	10.35	11.40
PM_{10}	5.52	5.15	6.13
$PM_{2.5}$	4.42	4.16	4.55

Table 6. Model validation measures for PM_{10} within London.

Model	RMSPE	MAPE	Bias	R^2	Coverage (%)
PM_{10}: Fitting SS = 11,828, validation SS=1,393					
AQUM	14.48	12.72	3.39	0.43	–
Gamma	3.81	2.85	0.87	0.85	89.37
Pirani et al. (2014)	4.75	–	–	0.63	–

Table 7. Estimates (posterior median) and the 95% credible interval for the parameters of the best model for NO_2, O_3, PM_{10} and $PM_{2.5}$. Here the rural sites have been set as the base category. Hence γ_{0j} and γ_{1j} are incremental intercept and slope for the urban sites when $j = 2$ and for the RKS sites when $j = 3$.

Parameter	Estimates for NO_2	Estimates for O_3
γ_0	3.66 (3.51, 3.80)	6.28 (6.16, 6.39)
γ_1	0.23 (0.22, 0.24)	0.22 (0.21, 0.23)
γ_{02}	1.27 (1.20, 1.35)	–0.93 (–0.99, –0.88)
γ_{12}	0.07 (0.06, 0.08)	0.05 (0.04, 0.06)
γ_{03}	2.52 (2.44, 2.60)	–0.53 (–0.59, –0.48)
γ_{13}	0.03 (0.02, 0.04)	0.007 (0.003, 0.013)
ρ	0.11 (0.09, 0.12)	0.27 (0.26, 0.28)
σ^2_ϵ	1.97 (1.95, 1.99)	0.29 (0.28, 0.31)
σ^2_η	1.12 (1.06, 1.19)	0.57 (0.55, 0.59)
ϕ	0.001 (fixed)	0.005 (0.0049, 0.0051)
	Estimates for PM_{10}	Estimates for $PM_{2.5}$
γ_0	4.03 (3.92, 4.13)	3.21 (3.14, 3.28)
γ_1	0.11 (0.09, 0.12)	0.077 (0.070, 0.084)
γ_{02}	–0.09 (–0.11, –0.07)	–0.11 (–0.12, –0.09)
γ_{12}	0.03 (0.02, 0.04)	0.043 (0.038, 0.047)
γ_{03}	–0.04 (–0.06, –0.02)	–0.10 (–0.12, –0.09)
γ_{13}	0.03 (0.02, 0.04)	0.041 (0.036, 0.045)
ρ	0.20 (0.18, 0.21)	0.16 (0.15, 0.18)
σ^2_ϵ	0.12 (0.11, 0.14)	0.0749 (0.0743, 0.0754)
σ^2_η	0.48 (0.46, 0.51)	0.51 (0.48, 0.51)
ϕ	0.0008 (0.0007, 0.0009)	0.00030 (0.00022, 0.00035)

4.4. Spatio-temporal aggregation results

Aggregation of the point level predictions to any given administrative geographies is performed by using block averaging at each MCMC iteration as detailed in Section 3.7. We also perform annual aggregation as detailed in Section 3.7. Here we illustrate annual aggregation for the 346 local authorities in England and Wales for 2011, boundaries of which have been shown in Figure 1.

Within each local authority, k, where $k = 1, \ldots, 346$, we evaluate the Bayesian predictions $z^{(\ell)}(\mathbf{s}_{kj}, t)$ for each grid point $\mathbf{s}_{kj}$ of a one kilometre square grid. At each MCMC iteration ℓ, these predictions are spatially and temporally aggregated to produce average local authority specific pollution estimates. These are then summarised to produce the predictive maps and their standard deviations, which are shown in Figures 3 to 6.

The plotted annual pollution maps, along with their uncertainties, show considerable spatial variation. As expected, the NO_2 levels are higher in London and other urban areas than the rural areas. However, the reverse happens for O_3. The particulate matter levels are consistently higher in urban areas. However, these values are lower in a few local authorities in central London according to Figures 5 and 6. This is very plausible as particulate matter pollution is linked to wood burning and such activities are not very common in central London, see Fuller et al. (2014). As expected, predictive uncertainties, (see the maps of standard deviation) are generally higher for the local authorities where there are not many monitoring sites. Also higher pollutant levels are generally associated with higher prediction standard deviations. This is expected since the boxplots in Figure 2 reveal much more variability for the suspected extreme observations than the observations lying in the central box. The adopted square-root transformation reduces such mean-variance relationship, but does not completely eliminate it in the empirical modelling (Sahu et al., 2007). Further illustrations of the predictive maps are provided in the accompanied supplemental material.

5. Discussion

This paper has developed statistical modelling and prediction methodologies for long term exposure to outdoor air pollution levels and illustrated these for England and Wales with daily data from the five year period 2007-11. We have considered the four most important pollutants: NO_2, O_3, PM_{10} and $PM_{2.5}$. Novelty in the statistical model has been introduced through space-time random effects by means of a temporally correlated Gaussian Process (GP) that is allowed to be anchored at a random set of locations within the study region. The random effect at any arbitrary location is obtained as a Kriged value of the realisations of the GP at the anchoring points known as knots. By also varying the number of knots we obtain a highly flexible non-stationary random effects surface that is able to match the non-stationarity in the daily air pollution surface. Further local information has been injected into the model through a site classification variable that describes the location of the monitoring sites. We have found that the site classifier taking three possible values best contrasts the pollution data according to land use. A site type specific regression model with site type varying intercept and slope has been put forward at the heart of the Bayesian

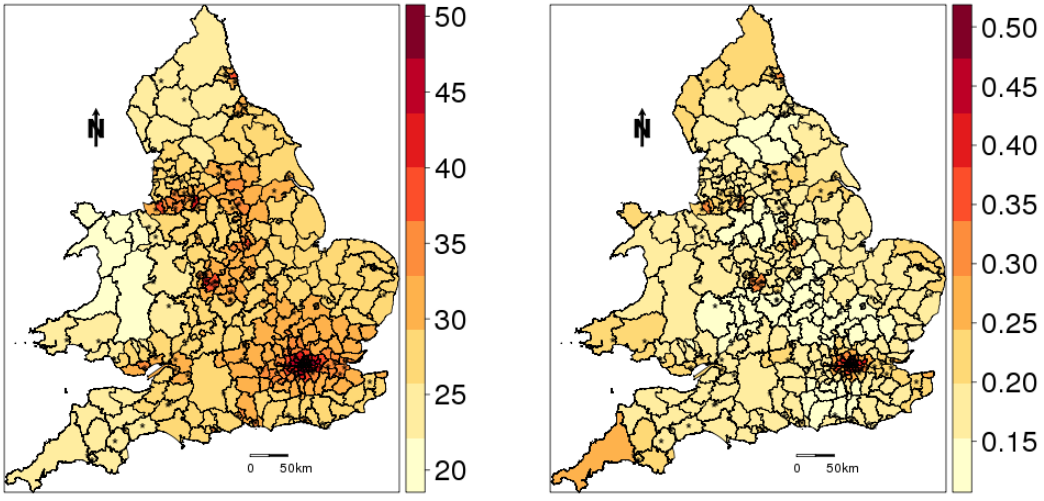

Fig. 3. Local authority-wise annual prediction plot for NO_2 using the 1 kilometre predictive grid (left panel) and their standard deviations (right panel) for 2011. The locations of the 144 AURN sites are superimposed as points on both the panels.

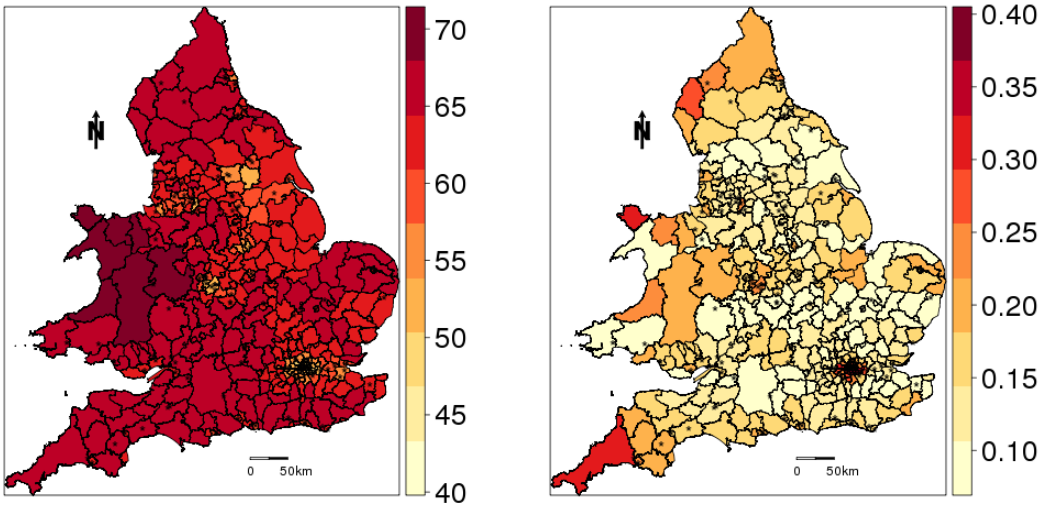

Fig. 4. Local authority-wise annual prediction plot for O_3 using the 1 kilometre predictive grid (left panel) and their standard deviations (right panel) for 2011. The locations of the 144 AURN sites are superimposed as points on both the panels.

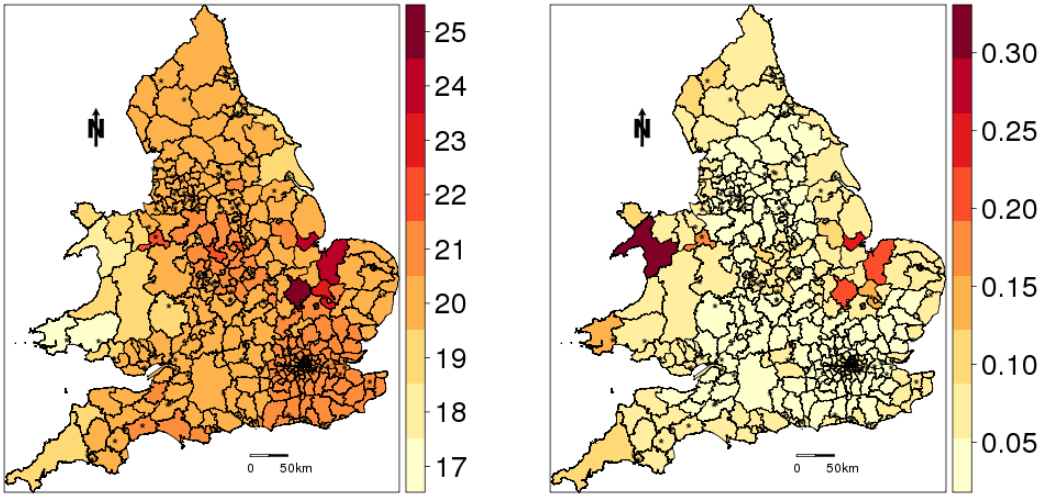

Fig. 5. Local authority-wise annual prediction plot for PM_{10} using the 1 kilometre predictive grid (top panel) and their standard deviations (bottom panel) for 2011. The locations of the 144 AURN sites are superimposed as points on both the panels.

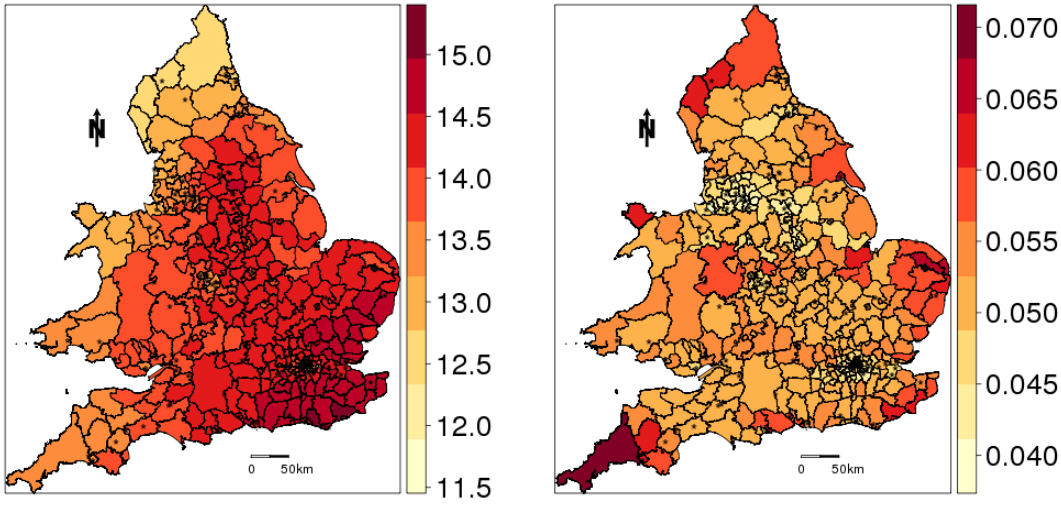

Fig. 6. Local authority-wise annual prediction plot for $PM_{2.5}$ using the 1 kilometre predictive grid (top panel) and their standard deviations (bottom panel) for 2011. The locations of the 144 AURN sites are superimposed as points on both the panels.

space-time model. Empirical results show significance of the site specific regression model for all four pollutants. It is also possible to incorporate a spatially varying co-efficient model for the AQUM outputs. However, our preliminary results using those models did not show any gain in predictive capability and as a result, we have not considered those in this paper.

The proposed models have been empirically verified with hold-out data and have achieved the least, as far as we are aware, root mean square error rates for *daily* PM_{10} data over the 5 year period. Comparable error rates for *daily data* for the other pollutants are not available for modelling data from England and Wales. However, many authors model annual average data and error rates from such modelling are available. For example, Shaddick et al. (2013) model the 2001 annual average NO_2 data from 934 sites in EU-15 countries and their best model has an RMSE of 8. This is not directly comparable to our model RMSPE for the daily data as provided in Table 3 since daily data and annual average data have different variabilities. However, their RMSE of 8 is relative to the standard deviation of 12 for the annual average as reported in their Table 1. Similar comparison for our best model for NO_2 shows the RMSPE value of 17.65 which is relative to the overall data standard deviation of 37.19 as reported in Table 2 here. Thus our model has much lower relative root mean square error, compared to the standard deviation of the full data, than what has been reported in Shaddick et al. (2013). For PM_{10}, Gulliver et al. (2011) model annual mean data from 52 monitoring sites in Londom and their best land use regression model, PMLUR, report an R^2 value of 0.58. This is much lower than the Gamma model R^2 of 0.85 in Table 6, although the R^2 values are not fully comparable since our models are at the daily temporal resolution and the Gulliver et. al. model is at the annual temporal resolution. We have also performed, though not reported here, other cross-validation studies with hold out data from one site at a time and those also provide favourable verdicts for the proposed models. Thus using variants of the proposed models we obtain the most accurate empirically verified maps of daily air pollution levels for England and Wales over the five year period.

The developed Bayesian prediction methodology has been applied to a 1-kilometre square grid covering England and Wales. This allows us to aggregate the predictions to any administrative geographies coarser than the 1-kilometre square grid, e.g. to electoral wards or local authority areas. These predictions are on daily time scale and hence can be aggregated to monthly, seasonal and annual scales. For each aggregation task, one needs to average the MCMC iterates of the predictions. These aggregated MCMC iterates enable us to provide uncertainty estimates in the aggregated predictions. We have illustrated the local authority wise prediction maps for the year 2011. All the predictions and data sets are published online alongside this paper from the website of the Medical & Environmental Data Mash-up Infrastructure project.** The accompanying supplement contains the descriptions of the files and further illustrative maps.

The published predictive values and maps can be used in many different types of scientific studies. For example, health effects estimation studies, such as Lee et al. (2016),

**https://www.data-mashup.org.uk/research-projects/statistical-downscaling-of-gridded-air-quality-data/

can benefit from accurate air pollution estimates at any spatial and temporal resolution for any spatial domain (such as Greater London) within England and Wales. Compliance to air pollution regulation can also be judged at any un-monitored site by performing spatial prediction at that site of interest. Compliance summaries, such as the number of days exceeding a threshold level of any pollutant in a year can be estimated, along with the associated uncertainties. For example, the EU regulations, see Commission (2008), state that the 24-hour average PM_{10}, the modelled quantity here, should not exceed the upper assessment threshold 35 $\mu g/m^3$ for more than 35 days in a year. For O_3, the regulation states that the modelled O_3 metric should not exceed 120 $\mu g/m^3$ for more than 25 days per calendar year averaged over 3 years. The Bayesian modelling framework developed here can estimate the number of days any particular site, monitored or unmonitored, for which such a threshold is exceeded and can also estimate maps of probability of non-compliance, as has been demonstrated in Sahu et al. (2007). Such assessments for England and Wales for all four pollutants and for all the five years require further investigation and will be considered elsewhere.

This article only concerns with exposure to outdoor pollution. Assessing the true personal exposure, including indoor air pollution, is a much more demanding task since people move between different environments, travel and relocate over a long study period of five years. There are many articles discussing such issues, see e.g. Shaddick et al. (2008) and Zidek et al. (2005).

Finally, the methodology presented in this paper can be improved in several ways. For example, to account for correlation in the pollutants one can use multivariate spatio-temporal models. However, multivariate modelling will only be required if it is desired to study the correlations between the pollutants. Further methodological development is also required to produce air pollution estimates, and their associated uncertainties, for user defined geographies and temporal windows so that air pollution estimates are available on demand for the spatial domain at the desired temporal resolution as required by the user.

Acknowledgements

The authors gratefully acknowledge the UK Met Office, Duncan Lee, Alastair Rushworth and Mark Bass for many helpful comments on an earlier version of this manuscript. We would also like to thank Natalia Tejedor for providing us the relevant MODIS satellite data.

References

Alkuwari, F. A., S. Guillas, and Y. Wang (2013, DEC). Statistical downscaling of an air quality model using Fitted Empirical Orthogonal Functions. *Atmospheric Environment 81*, 1–10.

Atkinson, R. W., D. Yu, B. G. Armstrong, S. Pattenden, P. Wilkinson, R. M. Doherty, M. R. Heal, and H. R. Anderson (2012). Concentrationresponse function for ozone and daily mortality: Results from five urban and five rural u.k. populations. *Environmental Health Perspectives 120*, 10.1289/ehp.1104108.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall/CRC.

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

Bell, M. L., A. McDermott, S. L. Zeger, J. M. Samet, and F. Dominici (2004). Ozone and short-term mortality in 95 us urban communities, 1987–2000,. *Journal of American Medical Association 292*, 2372–2378.

Berrocal, V. J., A. E. Gelfand, and D. M. Holland (2010). A spatio-temporal downcsaler for outputs from numerical models. *Journal of Agricultural, Biological and Environmental Statistics 15*, 176–197.

Boezen, H., S. van der Zee, D. Postma, J. Vonk, J. Gerritsen, and G. Hoek (1999). Effects of ambient air pollution on upper and lower respiratory symptoms and peak expiratory flow in children. *Lancet 353*, 874–878.

Carruthers, D. J., H. A. Edmunds, A. E. Lester, C. McHugh, and R. Singles (2000). Use and validation of adms urban in contrasting urban and industrial locations. *International Journal of Environment and Pollution 14*, 364–374.

Commission, E. (2008). Directive 2008/50/ec of the european parliament and of the council on ambient air quality and cleaner air for europe. Technical report, Commission of the European Community, Brussels.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.

Dockery, D. and C. A. Pope (1994). Acute respiratory effects of particulate air pollution. *Annu. Rev. Public Health 15*, 107–32.

Fuller, G. W., A. H. Tremper, T. D. Baker, K. E. Yttri, and D. . Butterfield (2014). Contribution of wood burning to pm_10 in london. *Atmospheric Environment 87*, 87–94.

Furrer, R., D. Nychka, and S. Sain (2013). *fields: Tools for Spatial Data*. University Corporation for Atmospheric Research. R package version 8.4.1.

Grewe, V., K. Dahlmann, S. Matthes, and W. Steinbrecht (2012). Attributing ozone to no_x emissions: Implications for climate mitigation measures. *Atmospheric Environment 59*, 102–107.

Gulliver, J. and D. Briggs (2011). Stems-air: A simple gis-based air pollution dispersion model for city-wide exposure assessment. *Science of the Total Environment 409*, 2419–2429.

Gulliver, J., K. de Hoogh, D. Fecht, D. Vienneau, and D. Briggs (2011). Comparative assessment of gis-based methods and metrics for estimating long-term exposures to air pollution. *Atmospheric Environment 45*, 7072–7080.

Hoek, G., R. Beelen, K. de Hoogh, D. Vienneau, J. Gulliver, P. Fischer, and D. J. Briggs (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment 42*, 7561–7578.

Huerta, G., B. Sanso, and J. R. Stroud (2004). A spatio-temporal model for maxico city ozone levels. *Journal of the Royal Statistical Society, Series C 53*, 231–248.

Jerrett, M., A. Arain, P. Kanaroglou, B. Beckerman, D. Potoglou, T. Sahsuvaroglu, M. J., and C. Giovis (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology 15*, 185–204.

Kassomenos, P., V. Kotroni, and G. Kallos (1995). Analysis of climatological and air quality observations from greater athens area. *Atmospheric Environment 29*, 3671–3688.

Lee, D., S. Mukhopadhyay, A. Rushworth, and S. K. Sahu (2016). A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. *Biostatistics 18*(2), 370–385.

Lee, D. and G. Shaddick (2010). Spatial Modeling of Air Pollution in Studies of Its Short-Term Health Effects. *Biometrics 66*(4), 1238–1246.

McMillan, N., S. Bortnick, M. Irwin, and L. Berliner (2005, MAR). A hierarchical Bayesian model to estimate and forecast ozone through space and time. *Atmospheric Environment 39*(8), 1373–1382.

Morrison, K. T., G. Shaddick, S. B. Henderson, and D. L. Buckeridge (2016). A latent process model for forecasting multiple time series in environmental public health surveillance. *Statistics in Medicine 35*, 3085–3100.

Pirani, M., N. Best, M. Blangiardo, S. Liverani, R. W. Atkinson, and G. W. Fuller (2015). Analysing the health effects of simultaneous exposure to physical and chemical properties of airborne particles. *Environment International 79*, 56–64.

Pirani, M., J. Gulliver, G. W. Fuller, and M. Blangiardo (2014, MAY-JUN). Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. *Journal of Exposure Science and Environmental Epidemiology 24*(3), 319–327.

Rushworth, A., D. Lee, and R. Mitchell (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and Spatio-temporal Epidemiology 10*, 29–38.

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

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

Sahu, S. K., A. E. Gelfand, and D. M. Holland (2010). Fusing point and areal level space-time data with application to wet deposition. *Journal of the Royal Statistical Society C 59*, 77–103.

Sahu, S. K. and S. Mukhopadhyay (2015). On generating a flexible class of anisotropic spatial models using gaussian predictive processes. Technical report, University of Southampton.

Sahu, S. K., S. Yip, and D. M. Holland (2009). Improved space-time forecasting of next day ozone concentrations in the eastern u.s. *Atmospheric Environment 43*, 494–501.

Samet, J., F. Dominici, F. Curriero, S. Zeger, and I. Coursac (2000). Fine particulate air pollution and mortality in 20 us cities. *New England Journal of Medicine 343*, 1742–1749.

Savage, N. H., P. Agnew, L. S. Davis, C. Ordóñez, R. Thorpe, C. E. Johnson, F. M. O'Connor, and M. Dalvi (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.

Schneider, A., M. A. Friedl, and D. Potere (2009). A new map of global urban extent from modis satellite data. *Environmental Research Letters 4*(4), 044003.

Shaddick, G., D. Lee, J. Zidek, and R. Salway (2008). Estimating exposure response functions using ambient pollution concentrations. *Annals of Applied Statistics 2*(4), 1249–1270.

Shaddick, G. and J. Wakefield (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the royal Statistical Society, Series C - Applied Statistics 51*(3), 351–372.

Shaddick, G., H. Yan, and D. Vienneau (2013). A bayesian hierarchical model for assessing the impact of human activity on nitrogen dioxide concentrations in europe. *Environmental Ecological Statistics 20*, 553–570.

Stein, M. L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

Venegas, L. E., N. A. Mazzeo, , and M. C. Dezzutti (2014). A simple model for calculating air pollution within street canyons. *Atmospheric Environment 87*, 77–86.

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

Zidek, J., G. Shaddick, R. White, J. Meloche, and C. Chatfield (2005). Using a probabilistic model (pcnem) to estimate personal exposure to air pollution. *Environmetrics 16*, 481–493.