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

Sabyasachi Mukhopadhyay

University of Hohenheim, Germany

and Sujit K. Sahu

University of Southampton, UK

[Received December 2015. Final revision April 2017]

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. The paper develops and compares several flexible non-stationary hierarchical Bayesian models for the four most harmful air pollutants, nitrogen dioxide and ozone, and PM$_{10}$ and PM$_{2.5}$ particulate matter, in England and Wales during the 5-year period 2007–2011. The models make use of observed data from the UK’s automatic urban and rural network 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 5-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 enable 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.

Keywords: Bayesian modelling; Exposure to air pollution; Gaussian process; Local-authority-wise UK air pollution levels; Spatial aggregation and prediction

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 these are 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 large surrounding geographical regions. However, the outdoor air that we breathe is a toxic mix of airborne particles which travel hundreds and thousands of kilometres—for example see news headlines that Saharan dust and pollution from continental Europe are triggering high air pollution levels in the UK (http://www.bbc.com/news/science-environment-26848489; http://www.

Address for correspondence: Sujit K. Sahu, Southampton Statistical Sciences Research Institute and Department of Mathematical Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.
E-mail: S.K.Sahu@soton.ac.uk

© 2017 The Authors Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/18/181465
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
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 spatiotemporal modelling which also directly allows estimation and then reporting of the associated uncertainties.

Recently, there has been much 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 auto-regressive models (Sahu et al., 2007). Model-based methods for realtime 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 are 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 references discussing Bayesian modelling for UK data include Lee and Shaddick (2010), Shaddick and Wakefield (2002) and Pirani et al. (2014). The spatial domain that was used in these three references is only the Greater London area with a limited number of monitoring sites. In particular, Lee and Shaddick (2010) modelled daily data from 30 monitoring sites between 2003 and 2005. They performed a simulation study for four pollutants: carbon monoxide, ozone (O₃), nitrogen dioxide (NO₂) and particulate matter PM₁₀. However, they did not validate the air pollution modelling for the real life data from Greater London. Pirani et al. (2014) modelled daily PM₁₀ exposure data from 45 sites for 728 days during 2002–2003. Lastly, Shaddick and Wakefield (2002) considered spatiotemporal 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) provided a comparative assessment of methods to predict mean annual PM₁₀-concentrations across 52 monitoring sites in London. They did not, however, model temporal variation of the pollutants. Gulliver and Briggs (2011) discussed geographical information system based air pollution dispersion models for citywide exposure assessment at daily temporal resolution. In a similar vein, the atmospheric dispersion model proposed by Carruthers et al. (2000) provides citywide exposure maps. Various other references, e.g. Atkinson et al. (2012), Pirani et al. (2014, 2015) and Rushworth et al. (2014), discuss and use air pollution estimates to estimate health effects. However, their main purpose is not exposure modelling and validation, and hence these references 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 enable linkage with aggregated monthly health outcome data, which was their main purpose. High resolution spatial maps of UK-wide annual air pollution exposure levels for NO₂ and PM₁₀ for the year 2001 only are available from http://www.envhealthatlas.co.uk/ which has been prepared by the Small Area Health Statistics Unit of Imperial College London. They use land use regression methods for
predicting concentrations at a 100 m × 100 m spatial resolution. However, the Web site 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) developed the air quality unified model (A QU M) for the whole of the UK. The A QU M is a three-dimensional weather and chemistry transport model used by the Met Office to deliver the UK national air quality forecast for the Department for Environment, Food and Rural Affairs and for scientific studies of atmospheric composition and air quality. The model has been run in hindcast mode to recreate hourly varying, UK air pollution concentrations for the period 2007–2011. However, the raw outputs of the A QU M 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 A QU M 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 A QU M, we model monitoring data which are more accurate. The use of the A QU M 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 A QU M, do not remain significant in the Bayesian model when outputs from a computer simulation model, such as the A QU M, are already present (Sahu and Bakar, 2012).

The primary motivation for this paper 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 NO₂, O₃, PM₁₀ and PM₂.⁵. Using the model selected we obtain air pollution estimates at a 1-km 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, enables 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 by using Markov chain Monte Carlo (MCMC) sampling, lies in the ability to estimate the uncertainties that are 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 local authority areas; see for example Lee et al. (2016).

The remainder of the paper is organized as follows. Section 2 describes the data summaries and explores relationships between the pollutant levels and the A QU M 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.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

http://wileyonlinelibrary.com/journal/rss-datasets

2. Data description

We model daily air pollution data collected from the 144 active automatic urban and rural network (AURN) stations in England and Wales for the five years 2007–2011. The AURN is the UK’s largest automatic monitoring network and is the main network used for compliance
Fig. 1. Map showing the boundaries of 346 local authorities in England and Wales and the locations of 144 AURN monitoring sites: ☆, the 90 modelling sites; ▲, location of the 54 validation sites reporting against the Ambient Air Quality Directives; see for example http://uk-air.defra.gov.uk/networks. The stations (see Fig. 1 for their locations) measure oxides of nitrogen, NO\textsubscript{x}, sulphur dioxide, O\textsubscript{3}, carbon monoxide and particles (PM\textsubscript{10} and PM\textsubscript{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 m and 4 m above ground level (https://uk-air.defra.gov.uk/networks/site-info?site_id=LB) depending on the type of pollutant. The data quality is verified and ratified on an on-going basis as detailed at https://uk-air.defra.gov.uk/networks/site-info?site_id=LB.

We obtain the daily concentration data for the four pollutants NO\textsubscript{2}, O\textsubscript{3} and particles less than 10 μm (PM\textsubscript{10}) and 2.5 μm (PM\textsubscript{2.5}) in size since these are the most harmful pollutants in the UK which may have health effects; see for example Lee et al. (2016). All pollutants are measured in micrograms per cubic metre. In our modelling we have used the daily maximum for NO\textsubscript{2}, the daily maximum 8-h running mean for O\textsubscript{3} and the daily mean for both PM\textsubscript{10} and PM\textsubscript{2.5} following the 2008 European Union directive on air pollution; see European Commission (2008).

For each pollutant, the total number of possible daily observations is 262944 (= 144 × 1826). However, in our data set there are 159534 (60.67%), 98628 (37.51%), 79227 (30.13%) and 58177 (22.12%) daily observations respectively for the four pollutants NO\textsubscript{2}, O\textsubscript{3}, PM\textsubscript{10} and PM\textsubscript{2.5}. The remaining observations are missing for 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
Table 1. Percentage of missing daily data out of the total number of observations in a year, which is 52704 (366×144) for 2008 and 52560 (365×144) for the other years

| Pollutant | Results (%) for the following years: |
|-----------|-------------------------------------|
|           | 2007   | 2008   | 2009   | 2010   | 2011   | Overall |
| NO₂       | 39.16  | 39.23  | 38.39  | 38.09  | 35.47  | 38.07   |
| O₃        | 56.56  | 62.87  | 64.05  | 63.05  | 61.81  | 61.67   |
| PM₁₀      | 64.99  | 66.71  | 70.24  | 72.41  | 69.78  | 68.83   |
| PM₂₅      | 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 three site types: ‘rural’, ‘urban’ (which also includes suburban), and ‘road and kerbside’. Summary statistics, as reported in Table 2, show significant variability in all four pollutants across these three site types. As yearwise breakdown of the missing observations. Most notably, more than 90% of the PM₂₅-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 (those which are regulated) until the publication of the 2008 European Union air pollution directive. In spite of the missing values the spatiotemporal models in Section 3 are estimated by using the large number of available observations as noted above.
expected, Table 2 shows that the rural sites are less polluted than the urban and the road-and-kerbside sites for NO$_2$, PM$_{10}$ and PM$_{2.5}$ whereas the converse is true for O$_3$. O$_3$ concentrations are generally lower in urban areas than in the surrounding rural areas because of reaction with NO$_x$ (mostly NO) emissions, which are greatest in urban areas; see for example Grewe et al. (2012) and also page 3 of the report ‘Ozone in the United Kingdom’, written 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 Fig. 2. These plots also show differences in variances for each pollutant by site type, with road-and-kerbside sites having a larger spread for NO$_2$, PM$_{10}$ and PM$_{2.5}$ because of many extreme observations. Finally, boxplots of the pollution concentrations by year (which have been 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 that are proposed in this paper address the sparsity of the observed AURN data by including the hourly hindcasts of the AQUM, which was developed by Savage et al. (2013) especially for the UK. The 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, the AQUM does not use the observed AURN data but it produces output on a 12 km × 12 km square grid covering the whole UK. We use bilinear interpolation methods, which interpolate once horizontally and then vertically (see for example the Wikipedia entry https://en.wikipedia.org/wiki/Bilinear_interpolation), to predict the hourly pollution concentration values at the corners of a 1 km × 1 km square grid that we use for our prediction. These hourly values are subsequently aggregated to daily values for our modelling. The corners of these 1-km grid cells do not coincide with the locations of the 144 AURN monitoring sites and to resolve this misalignment we use bilinear interpolation of the four grid corners of each site to estimate the daily AQUM output for the 144 observation sites. These outputs are moderately correlated with the corresponding observations with correlations of 0.44 for NO₂, 0.68 for O₃, 0.59 for PM₁₀ and 0.61 for PM₂.₅. We also obtained scatter plots showing these correlations but those have been 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 road and kerbside) 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 satellite MODIS; see Schneider et al. (2009). These maps, based on satellite data from 2001–2002, provide land use information (urban–rural) at a 500-m spatial resolution and have an overall accuracy of 93%. Thus the rural–urban classification for each corner of a 1-km predictive grid location is accurately obtained from these maps. However, these data do not provide the road-and-kerbside site type classification for the 1-km predictive grid. To obtain these classifications, we use the detailed open data roads network product from the Ordnance Survey (https://www.ordnancesurvey.co.uk/opendatadownload/products.html). We calculate the distance of each of the 1-km grid locations to the nearest road. The grid locations which are within 400 ft of the nearest road are classified as road and kerbside.

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 stabilize the variance (Sahu et al., 2007; Berrocal et al., 2010). However, for ease of interpretation, the accuracies of all predictions from the pollution model are assessed on the original scale. Let \( z(p)(s_i, t) \) and \( x(p)(s_i, t) \) respectively denote the measured and modelled AQUM pollution concentration on the square-root scale at location \( s_i \) 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

\[
Z(s_i, t) = Y(s_i, t) + \epsilon(s_i, t), \quad \epsilon(s_i, t) \sim N(0, \sigma^2_{\epsilon}),
\]

for \( i = 1, \ldots, n \) and \( t = 1, \ldots, T \), where \( Y(s_i, t) \) is the true process and \( \epsilon(s_i, t) \) is the independent nugget effect absorbing microscale variability. (Henceforth, values of the subscripts \( i \) and \( t \) will always be in the range that is mentioned here.) At the next stage of the hierarchy we specify

\[
Y(s_i, t) = \mu(s_i, t) + \eta(s_i, t)
\]
where \( \mu(s_i, t) \) denotes the mean surface and \( \eta(s_i, t) \) is a space–time process, which we specify later. The mean surface is modelled as

\[
\mu(s_i, t) = \gamma_0 + \gamma_1 x(s_i, t) + \sum_{j=2}^{r} \delta_j(s_i) \{ \gamma_{0j} + \gamma_{1j} x(s_i, t) \},
\]

where we propose a site-type-specific regression on the modelled square-root AQUM concentrations \( x(s_i, t) \). Here \( r = 3 \), corresponding to the three site types (rural, urban, road and kerbside), and the rural site type corresponds to \( j = 1 \) and is the baseline level. Thus \((\gamma_0, \gamma_1)\) are respectively the slope and intercept terms for the rural sites, whereas \((\gamma_{0j}, \gamma_{1j})\) are the incremental adjustments for site type \( j, j = 2, 3 \). Finally, \( \delta_j(s_i) \) is an indicator function, equalling 1 if site \( s_i \) is of the \( j \)th site type and 0 otherwise.

### 3.2. Specification of the spatiotemporal process

We consider four modelling possibilities for the spatiotemporal process \( \eta(s_i, t) \), which differ in their level of sophistication. The first is the simplest that assumes \( \eta(s_i, t) = 0 \) for all sites \( s_i \) and times \( t \), which renders model (1) a simple regression model that is used for comparison with the other two models. The second model for \( \eta(s_i, t) \) is an independent-over-time Gaussian process (GP) with zero mean and a Matérn covariance function given by

\[
\text{cov}\{\eta(s, t), \eta(s', t)\} = \frac{\sigma_n^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}\|s - s'\|) \nu K_\nu(2\sqrt{\nu}\|s - s'\|),
\]

for \( \phi > 0 \) and \( \nu > 0 \), where \( \Gamma(\nu) \) is the standard gamma function, \( K_\nu \) is the modified Bessel function of the second kind with order \( \nu \) and \( \|s - s'\| \) is the distance between sites \( s \) and \( s' \). The parameter \( \phi \) controls the rate of decay of the correlation as the distance \( \|s_i - s_j\| \) increases and the parameter \( \nu \) controls smoothness of the random field (Banerjee et al., 2004; Cressie, 1993), i.e., for each time \( t \),

\[
\eta_t = (\eta(s_1, t), \ldots, \eta(s_n, t))^T \sim N\{0, \sigma_n^2 H_\eta(\phi, \nu)\},
\]

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

\[
\text{corr}\{\eta(s, t), \eta(s', t)\} = \exp(-\|s - s'\|/\phi).
\]

In this case \( \phi \) alone determines the rate of decay of the spatial correlation as the distance \( d = ||s - s'|| \) between two sites increases. A related quantity, which is 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 \( \phi \). Hence the effective range, for a given value of \( \phi \), is defined to be the value of the distance \( d \) for which \( \exp(-d\phi) = 0.05 \), which implies that \( d = 3/\phi \).

The third model introduces non-stationary covariance structure following Sahu and Mukhopadhyay (2015). This is achieved by first assuming a set of knot locations, \( S_m = (s_1^*, \ldots, s_m^*) \), which are specified below, for a value of \( m \) which will be chosen by out-of-sample prediction performance. Given \( S_m^* \), we assume that \( \eta_i^* = (\eta(s_1^*, t), \ldots, \eta(s_m^*, t))^T \) is a zero-mean GP with the Matérn covariance function (4). The non-stationary modelling proposal is to replace \( \eta(s_i, t) \) in equation (2) by

\[
\tilde{\eta}(s_i, t) = E[\eta(s_i, t)|\eta_i^*].
\]
The \((n+m)\)-dimensional vector \((\eta_t, \eta_t^*)\) is assumed to be a realization of the same underlying GP as in equation (4) independently for each \(t\). By writing \(\tilde{\eta}_t = (\tilde{\eta}(s_1, t), \ldots, \tilde{\eta}(s_n, t))^T\) and using multivariate Gaussian theory we have

\[
\tilde{\eta}_t = C^*(\phi, \nu)H_{\eta t}^{-1}(\phi, \nu)\eta_t^*
\]

where \(C^*(\phi, \nu)\) is the \(n \times m\) cross-correlation matrix between \(\eta_t^*\) and \(\eta_t^*\), i.e. \((C^*)_{ij} = C(||s_i - s_j^*||; \phi, \nu)\) for \(i = 1, \ldots, n\) and \(j = 1, \ldots, m\) and \(H_{\eta t}(\phi, \nu)\) is an \(m \times m\) correlation matrix having elements \(H_{\eta t}(\phi, \nu)_{kj} = C(||s_k^* - s_j^*||; \phi, \nu)\), for \(k, j = 1, \ldots, m\). Clearly,

\[
\text{var}(\eta_t) = C^*(\phi, \nu)H_{\eta t}^{-1}(\phi, \nu)C^T(\phi, \nu),
\]

which shows non-stationarity of the \(\eta_t\)-process.

The \(\tilde{\eta}_t\)-surface is now based on linear functions of the \(m\)-dimensional \(\eta_t^*\) instead of the \(n\)-dimensional \(\eta_t\). This leads to a reduction in the computational burden when \(m\) is much smaller than \(n\). However, \(n = 144\) is not very large in this paper and hence dimension reduction by the Gaussian predictive process approximation (Banerjee et al., 2008) is not required here. We still use equation (7) to benefit from having a flexible non-stationary model for the spatial random \(\tilde{\eta}_t\).

Finally, we remove the temporal independence assumption in the previous model by introducing an auto-regressive model for \(\eta_t^*\). Here we assume that

\[
\eta_t^* \sim N(\rho \eta_{t-1}^*, \sigma_{\eta}^2 H_{\eta t}^{-1}(\phi, \nu)), \quad \text{for } t = 1, \ldots, T,
\]

with \(\eta_0^* = 0\) as the initial condition and \(\rho\) as the unknown auto-regressive parameter for which we specify a prior distribution. Thus, given \(m\) and \(S_m^*\), \(\eta_t^*\) is determined from the above auto-regressive process, and then \(\tilde{\eta}_t\) is obtained by using equation (7) for all \(t = 1, \ldots, T\).

### 3.3. Specifying the knot locations

Now we return to specifying the knot locations \(S_m^*\) for a given \(m\) where \(m\) is to be chosen by cross-validation. Sahu and Mukhopadhyay (2015) showed that a random selection of \(S_m^*\) is preferable to a space filling design that distributes the \(m\) locations evenly within the study region. In our implementation, we discretize by \(M\) points, which are the corners of a set of 1-km grid squares covering the study region. To add flexibility we assume a probability surface \(p(s_j^*)\), \(j = 1, \ldots, M\), such that \(\Sigma_{j=1}^{M} p(s_j^*) = 1\) and \(p(s_j^*) \geq 0\) for \(j = 1, \ldots, M\). The \(p(s_j^*)\) can be a normalized population density surface that will guarantee that knots are placed at high density areas. A random sample of \(m\) locations is proposed to be used as the knot locations \(S_m^*\). In effect, this implies a discrete prior distribution for \(S_m^*\) and MCMC model fitting is easily accomplished by using a Metropolis–Hastings step for \(S_m^*\) 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) among the possible values of 16, 25, 36, 49 and 100. A complete square value, e.g. 25, is put forward as a candidate so that an equal number of points is 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 $\gamma$ in equation (3) with large variances of $10^4$, and an inverse gamma prior distribution for the variance parameters $(\sigma_\gamma^2, \sigma_\phi^2)$ with hyperparameters $(2, 1)$ parameterized to have mean 1 and infinite variance. Informative (and thus proper) prior distributions must be specified for the two parameters $\nu$ and $\phi$ describing the Matérn correlation function, since these are weakly identified in the likelihood; see for example Zhang (2004), who showed that consistent estimation is not possible for these parameters. In addition, sparsely observed spatial data are not very informative for the smoothness parameter $\nu$; see for example Stein (1999). Hence we shall adopt the exponential correlation function corresponding to $\nu = 0.5$ following many researchers, e.g. Berrocal et al. (2010) and Sahu et al. (2007).

Now it remains to specify a proper prior distribution for $\phi$. Here, besides fixing $\phi$ at several plausible effective ranges we entertain two proper prior distributions:

(a) a uniform prior distribution in the interval $(0.001, 0.01)$ corresponding to having an effective range between 300 and 3000 km and

(b) a gamma prior distribution with parameters 2 and 1, parameterized to have mean 2 but with infinite variance.

The implied effective range between 300 and 3000 km 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 $\phi$ 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 $s'$ 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 $Z(s_i, t)$ is missing for particular values of $s_i$ and $t$. By virtue of the Bayesian model, implemented by using MCMC sampling, we have a value of the $\mu(s_i, t)^{(l)}$ at each iteration $l$ for $l = 1, \ldots, L$ of the implemented MCMC algorithm. (Henceforth $l$ will denote the MCMC iteration index and $L$ will denote the retained number of MCMC iterates that we use for inference.) Using the sampled values of the spatiotemporal process $\eta(s_i, t)^{(l)}$ we obtain $Y(s_i, t)^{(l)}$ following equation (2). Subsequently, we sample $Z(s_i, t)^{(l)}$ by using model (1) which also needs $\sigma^2_{\gamma}$. At the end of the MCMC run, we utilize the samples $Z(s_i, t)^{(l)}$, $l = 1, \ldots, L$, to estimate the missing $Z(s_i, t)$ value and its variability.

Prediction of $Z(s', t)$ at a new location $s'$, where $s'$ is not one of $s_1, \ldots, s_n$, using the Bayesian model, proceeds by evaluating the posterior predictive distribution

$$
\pi\{z(s', t)|z\} = \int \pi\{z(s', t)|S_m^*, \eta^*, \theta, z\} \pi(S_m^*, \eta^*, \theta|z) dS_m^* d\eta^* d\theta,
$$

where $\theta = (\gamma, \phi, \sigma_\gamma^2, \sigma_\phi^2, \phi)^T$ denotes the model parameters and $\eta^* = (\eta_1^*, \ldots, \eta_n^*)$. Here $\pi\{z(s', t)|S_m^*, \eta^*, \theta\}$ requires $\bar{\eta}(s', t)$ (see model (1)), which is calculated as $e^{*T}(s'; \phi, \nu)H_{\eta^{-1}}(\phi, \nu)\eta^*$, analogous to equation (7), where $c_s^*(s'; \phi, \nu)$ is $m \times 1$ with the $i$th element given by $c_s^*(s'; \phi, \nu) = C(|s_i^* - s'|; \phi, \nu)$. Composition sampling is used to simulate from distribution (9) as follows. At the $l$th MCMC iteration samples of $(S_m^{(l)}, \eta^{(l)}, \theta^{(l)})$ are drawn from the joint posterior distribution. Now $z^{(l)}(s', t)$ is to be sampled from the Gaussian distribution given by model (1). However, this requires $Y^{(l)}(s', t)$ which in turn requires $\mu^{(l)}(s', t)$ and $\eta^{(l)}(s', t)$; see equation (2). The mean term $\mu^{(l)}(s', t)$ is evaluated by using the $\gamma^{(l)}$ in equation (3). This evaluation also requires the site type indicators $b_j(s')$, $j = 2, \ldots, r$, which we take from the MODIS satellite
Long-term Exposure to Outdoor Air Pollution

data as discussed in Section 2. Also note that equation (3) requires \(x(s', t)\) which we obtain by using bilinear interpolation as mentioned in Section 2.

For \(n^{(l)}(s', t)\) we note that according to equation (7) we have

\[\tilde{\eta}(s', t) = e^{*T}(s', \phi, \nu)H_{n^{(l)}}^{-1}(\phi, \nu)\eta_t^{*}.\]

Hence \(n^{(l)}(s', t)\) is obtained as

\[n^{(l)}(s', t) = e^{*T}(\phi^{(l)}, \nu^{(l)})H_{n^{(l)}}^{-1}(\phi^{(l)}, \nu^{(l)})\eta_t^{*}.\]

Now we obtain \(Y^{(l)}(s', t) = \eta^{(l)}(s', t) + \tilde{\eta}(s', t)\) by using equation (2). Finally, \(z^{(l)}(s', t)\) is obtained as a draw from the Gaussian distribution with mean \(Y^{(l)}(s', t)\) and variance \(\sigma^2 e^{(l)}\).

3.6. Evaluating the predictive performance

The predictive performance of each pollution model is assessed by a cross-validation exercise, where data at sites \(s_1, \ldots, s_n\) are used to fit the model whereas data at sites \(s_{n_0 + 1}, \ldots, s_n\) 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

\[\text{RMSPE} = \sqrt{\frac{1}{N_v} \sum_{j=n_0+1}^n \sum_{t=1}^T \{z(s_j, t) - \tilde{z}(s_j, t)\}^2},\]

\[\text{MAPE} = \frac{1}{N_v} \sum_{j=n_0+1}^n \sum_{t=1}^T |z(s_j, t) - \tilde{z}(s_j, t)|,\]

where \(\tilde{z}(s_j, t)\) 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 inference.

3.7. Aggregating predictions to administrative geographies

The proposed geostatistical models can predict at any unmonitored location \(s'\) by using the methodology that was 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 \(k\)th 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

\[Z_{kt} = \frac{1}{|A_k|} \int_{A_k} Z(s, t) ds,\]  

(10)

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

\[\tilde{Z}_{kt} = \frac{1}{n_k} \sum_{j=1}^{n_k} Z(s_{kj}, t),\]  

(11)

where \((s_{k1}, \ldots, s_{kn_k})\) forms a fine grid of prediction locations within the \(k\)th region. As mentioned in Section 1, we perform the predictions at a 1-km square grid covering the study region. Hence, the \(n_k\) locations \((s_{k1}, \ldots, s_{kn_k})\) are simply taken as the corners of the 1-km 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 $z$ since each $Z(s_{kj},t)$ also has such a distribution. This posterior predictive distribution can be summarized by using MCMC samples drawn from it using composition sampling. Corresponding to each draw from the joint posterior distribution we draw a sample $z^{(l)}(s_{kj},t)$ from the posterior predictive distribution (9). At each MCMC iteration $l$, we form the average

$$\bar{Z}^{(l)}_{k}=\frac{1}{n_k}\sum_{j=1}^{n_k} z^{(l)}(s_{kj},t).$$

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

$$\bar{Z}^{(l)}_{k} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{n_k} \sum_{j=1}^{n_k} z^{(l)}(s_{kj},t)$$

(12)

for each $l=1,\ldots,L$ and summaries, e.g. the 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 by using the MCMC iterates. Prediction details for $z^{(l)}(s_{kj},t)$ for any arbitrary site $s_{kj}$ have been noted in the previous subsection. Lastly, we also note that temporal aggregation for a subperiod (e.g. annual) of the whole time domain can be performed similarly by simply taking the average similar to that in equation (12) but only using the simulations $z^{(l)}(s_{kj},t)$ for all the individual time points $t$ which fall in that subperiod. For example, we can find annual averages from modelling 5 years’ data.

4. Results

4.1. Implementation details and models set-up

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 164340 ($= 90 \times 1826$) observations and validation is done using a maximum of 98604 ($= 54 \times 1826$) 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 road-and-kerbside sites whereas the 54 validations sites contained 29 urban, seven rural and 18 road-and-kerbside sites. Thus all three site types are adequately represented in both the fitting and the validation set of sites. For model comparison we consider the RMSPE, MAPE, bias and 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 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 by using MCMC sampling in the Bayesian framework, that does not take care of the spatiotemporal dependence in the data, i.e. models (1)–(3) with $\eta(s_i,t) = 0$ 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 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 spatiotemporal models described in Section 3. The first five of the seven models are obtained for five different fixed values of $\phi$, the decay parameter. The five values, which are denoted by $\phi_i, i = 1, \ldots, 5$, correspond to effective ranges of 3500 ($\phi_1$), 3000 ($\phi_2$), 600 ($\phi_3$), 300 ($\phi_4$) and 100 ($\phi_5$) km, and these choices are guided by the need to include large-to-moderate amounts of spatial correlation in the model. In the final two modelling attempts we specify the uniform and gamma prior distributions for the decay parameter $\phi$ as mentioned in Section 3.1.

### 4.2. Model validation and comparison results

Tables 3 and 4 report the validation results for the AQUM and all nine modelling scenarios noted above. They show that the raw AQUM outputs are far worse than all nine modelling strategies including kriging. This is expected as the AQUM does not use the actual observations and those outputted are heavily biased as noted previously. All six Bayesian models, based on the non-stationary spatiotemporal models, perform much better than the three benchmarking strategies: kriging, linear and GP. The differences between the performances of the six Bayesian models are not very large, which is because the dominant linear model part is the same for all six models. Slight differences in performance are observed by varying the spatial correlation structure. For NO$_2$ the model with $\phi$ fixed at $\phi_2$ is chosen to be the best; the model with the

| Model          | RMSPE | MAPE | Bias  | Coverage (%) | $R^2$ |
|----------------|-------|------|-------|--------------|-------|
| NO$_2$: fitting SS = 100138; validation SS = 62688 |        |      |       |              |       |
| AQUM           | 39.12 | 25.50| -17.06| 94.43        | 0.60  |
| Kriging        | 32.87 | 22.88| 2.56  | 98.08        | 0.58  |
| Linear         | 30.46 | 19.63| -5.09 | 97.43        | 0.89  |
| GP             | 31.46 | 22.08| 1.73  | 97.04        | 0.58  |
| $\phi_1$       | 18.30 | 13.60| 1.22  | 96.64        | 0.88  |
| $\phi_2$       | 17.65 | 12.99| 0.41  | 97.42        | 0.89  |
| $\phi_3$       | 17.90 | 12.86| -0.70 | 97.43        | 0.89  |
| $\phi_4$       | 20.85 | 14.56| 0.05  | 96.59        | 0.84  |
| Uniform        | 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 = 64373; validation SS = 36406 |        |      |       |              |       |
| AQUM           | 20.80 | 16.04| -2.95 | 93.86        | 0.69  |
| Kriging        | 13.30 | 9.86 | -2.95 | 99.52        | 0.66  |
| Linear         | 16.0  | 12.42| 8.47  | 91.61        | 0.89  |
| GP             | 16.53 | 12.81| 3.76  | 91.18        | 0.88  |
| $\phi_1$       | 10.22 | 7.62 | 0.09  | 91.72        | 0.89  |
| $\phi_2$       | 10.33 | 7.68 | 0.11  | 91.72        | 0.89  |
| $\phi_3$       | 10.27 | 7.65 | 0.22  | 91.72        | 0.89  |
| $\phi_4$       | 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  |

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

| Model  | RMSPE | MAPE | Bias  | Coverage (%) | $R^2$ |
|--------|-------|------|-------|--------------|-------|
| PM$_{10}$; fitting SS = 52625; validation SS = 29334 |
| AQUUM  | 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  |
| $\phi_1$ | 5.49  | 3.58 | −0.62 | 89.23        | 0.80  |
| $\phi_2$ | 5.51  | 3.63 | −0.46 | 89.07        | 0.80  |
| $\phi_3$ | 5.59  | 3.61 | −0.80 | 90.75        | 0.80  |
| $\phi_4$ | 6.21  | 3.97 | −0.57 | 87.32        | 0.79  |
| Uniform| 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 = 38481; validation SS = 21308 |
| AQUUM  | 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  |
| $\phi_1$ | 4.45  | 3.25 | −1.23 | 82.74        | 0.85  |
| $\phi_2$ | 4.32  | 2.66 | −1.17 | 83.84        | 0.85  |
| $\phi_3$ | 4.38  | 2.72 | −1.03 | 84.61        | 0.85  |
| $\phi_4$ | 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  |

†SS, sample size, which is the number of daily observations. $R^2$ denotes the sample correlation coefficient between the predictions and actual observations.

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% and 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% and 84.6% for PM$_{2.5}$. This is most likely to be because of the smaller sample sizes that are available both for fitting and for validation for this pollutant. However, the 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 whether the best chosen model for each of the four pollutants performs similarly in validating the three site types: rural, urban, and road and kerbside. Recall that there were 31 urban, seven rural and 18 road-and-kerbside sites among the 54 validation sites. Table 5 shows the RMSPEs aggregated by the site types. It 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 among both the fitting and the validation sites, which leads to more accurate estimation and more stable values 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
Table 5. Aggregated RMSPEs of individual pollutants according to the three site types†

| Pollutant | Rural RMSPE | Urban RMSPE | Road-and-kerbside RMSPE |
|-----------|-------------|-------------|-------------------------|
| NO₂       | 18.69       | 15.63       | 19.53                   |
| O₃        | 9.59        | 10.35       | 11.40                   |
| PM₁₀      | 5.52        | 5.15        | 6.13                    |
| PM₂.₅     | 4.42        | 4.16        | 4.55                    |

†The results correspond to the best model for each pollutant as reported in Table 3 and Table 4.

Table 6. Model validation measures for PM₁₀ within London†

| Model          | RMSPE | MAPE | Bias | R²    | Coverage (%) |
|----------------|-------|------|------|-------|--------------|
| 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  | —            |

†PM₁₀: fitting SS = 11828; validation SS = 1393.

sites in Greater London only. We perform this experiment only for PM₁₀ 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 a much lower RMSPE with a better R²-value than the Pirani et al. (2014) modelling effort, although this is not an exact comparison since the data that were 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 NO₂ concentration in the urban and roadside sites compared with the rural sites since all the incremental slopes and intercepts are positively significant in Table 7. However, this behaviour reverses for O₃, which is known to be lower in urban and roadside areas because of its negative correlation with NO₂. 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, ̂φ, show large effective ranges (3/̂φ).

The spatial variance σ²η is estimated to be higher than the nugget effect σ² except for the case of modelling NO₂. This is plausible since the variability of NO₂ is much higher than that of the other three pollutants (see Table 2) and the model absorbs the extra variability by using the
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$_{2.5}$.

| Parameter | Estimates for NO$_2$ | Estimates for O$_3$ | Estimates for PM$_{10}$ | Estimates for PM$_{2.5}$ |
|-----------|----------------------|---------------------|-------------------------|-------------------------|
| $\gamma_0$ | 3.66 (3.51, 3.80)    | 6.28 (6.16, 6.39)   | 4.03 (3.92, 4.13)       | 3.21 (3.14, 3.28)       |
| $\gamma_1$ | 0.23 (0.22, 0.24)    | 0.22 (0.21, 0.23)   | 0.11 (0.09, 0.12)       | 0.077 (0.070, 0.084)    |
| $\gamma_02$ | 1.27 (1.20, 1.35)    | 0.95 (0.99, 0.88)   | 0.09 (0.11, 0.07)       | 0.11 (0.12, 0.09)       |
| $\gamma_12$ | 0.07 (0.06, 0.08)    | 0.05 (0.04, 0.06)   | 0.03 (0.02, 0.04)       | 0.043 (0.038, 0.047)    |
| $\gamma_03$ | 2.52 (2.44, 2.60)    | 0.53 (0.59, 0.48)   | 0.04 (0.06, 0.02)       | 0.10 (0.12, 0.09)       |
| $\gamma_13$ | 0.03 (0.02, 0.04)    | 0.007 (0.003, 0.013) | 0.03 (0.02, 0.04)       | 0.041 (0.036, 0.045)    |
| $\rho$     | 0.11 (0.09, 0.12)    | 0.27 (0.26, 0.28)   | 0.20 (0.18, 0.21)       | 0.16 (0.15, 0.18)       |
| $\sigma_{\epsilon}^2$ | 1.97 (1.95, 1.99)  | 0.29 (0.28, 0.31)   | 0.12 (0.11, 0.14)       | 0.0749 (0.0743, 0.0754) |
| $\sigma_{\eta}^2$ | 1.12 (1.06, 1.19)  | 0.57 (0.55, 0.59)   | 0.48 (0.46, 0.51)       | 0.51 (0.48, 0.51)       |
| $\phi$     | 0.001 fixed          | 0.005 (0.0049, 0.0051) | 0.0008 (0.0007, 0.0009) | 0.00030 (0.00022, 0.00035) |

†Here the rural sites have been set as the base category. Hence $\gamma_{0j}$ and $\gamma_{1j}$ are the incremental intercept and slope for the urban sites when $j = 2$ and for the road-and-kerbside sites when $j = 3$. 

nugget effect parameter $\sigma^2$. Overall, all the linear parameters show significantly different linear relationships between the AQUM 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 RMSPEs.

4.4. Spatiotemporal 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 Fig. 1.
Within each local authority, \( k \), where \( k = 1, \ldots, 346 \), we evaluate the Bayesian predictions \( z^{(l)}(s_{kj}, t) \) for each grid point \( s_{kj} \) of a 1-km 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 summarized to produce the predictive maps and their standard deviations are shown in Figs 3–6.

The annual pollution maps plotted, along with their uncertainties, show considerable spatial variation. As expected, the NO\(_2\) levels are higher in London and other urban areas than in 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 Figs 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 the 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 Fig. 2 reveal much more variability for the suspected extreme
observations than the observations lying in the central box. The square-root transformation
adopted reduces such a 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 accompanying on-line 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 5-year period 2007–2011. 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 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 realizations 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 can 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 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 coefficient
model for the AQUM outputs. However, our preliminary results by using those models did
not show any gain in predictive ability and, as a result, we have not considered those in this
paper.

The models proposed 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 researchers model annual average data,
and error rates from such modelling are available. For example, Shaddick et al. (2013) modelled
the 2001 annual average NO$_2$ data from 934 sites in the then 15 European Union countries and
their best model has a root-mean-squared error of 8. This is not directly comparable with our
model RMSPE for the daily data as provided in Table 3 since daily data and annual average
data have different variabilities. However, their root-mean-squared error of 8 is relative to the
standard deviation of 12 for the annual average as reported in their Table 1. A 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 with the standard deviation of the full data, than what was
reported in Shaddick et al. (2013). For PM$_{10}$, Gulliver et al. (2011) modelled annual mean data
from 52 monitoring sites in London and their best land use regression model, PMLUR, reported
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 model of Gulliver et al. (2011) 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 models proposed. Thus using variants
of the models we obtain the most accurate empirically verified maps of daily air pollution levels for England and Wales over the 5-year period.

The Bayesian prediction methodology developed has been applied to a 1-km square grid covering England and Wales. This enables us to aggregate the predictions to any administrative geographies that are coarser than the 1-km square grid, e.g. to electoral wards or local authority areas. These predictions are on a daily timescale 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 on line alongside this paper from the Web site of the ‘Medical and environmental data mash-up infrastructure’ project (https://www.data-mashup.org.uk/research-projects/statistical-downscaling-of-gridded-air-quality-data/). The accompanying on-line 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), 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 with air pollution regulation can also be judged at any unmonitored 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 European Union regulations, see European Commission (2008), state that the 24-h average PM$_{10}$, the quantity modelled 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 that is developed here can estimate the number of days for 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 paper is concerned only 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 5 years. There are many references discussing such issues; see for example Shaddick et al. (2008) and Zidek et al. (2005).

Finally, the methodology that is presented in this paper can be improved in several ways. For example, to account for correlation in the pollutants one can use multivariate spatiotemporal 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 paper. We also thank Natalia Tejedor for providing us the relevant MODIS satellite data.
References

Alkuwari, F. A., Guillans, S. and Wang, Y. (2013) Statistical downscaling of an air quality model using Fitted Empirical Orthogonal Functions. Atmos. Environ., 81, 1–10.

Atkinson, R. W., Yu, D., Armstrong, B. G., Pettenden, S., Wilkinson, P., Doherty, R. M., Heal, M. R. and Anderson, H. R. (2012) Contribution response function for ozone and daily mortality: results from five urban and five rural U.K. populations. Environ. Health Perspect., 120, 1411–1417.

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

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008) Gaussian predictive process models for large spatial data sets. J. R. Statist. Soc. B, 70, 825–848.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M. and Dominici, F. (2004) Ozone and short-term mortality in 95 US urban communities, 1987–2000. J. Am. Med. Ass., 292, 2372–2378.

Berrocal, V. J., Gelfand, A. E. and Holland, D. M. (2010) A spatio-temporal downscaler for outputs from numerical models. J. Agric. Biol. Environ. Statist., 15, 176–197.

Boezen, H., van der Zee, S., Postma, D., Vonk, J., Gerritsen, J. and Hoek, G. (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., Edmunds, H. A., Lester, A. E., McHugh, C. and Singles, R. (2000) Use and validation of ADMS urban in contrasting urban and industrial locations. Int. J. Environ. Pollut., 14, 364–374.

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

Dockery, D. and Pope, C. A. (1994) Acute respiratory effects of particulate air pollution. A. Rev. Publ. Hlth, 15, 107–132.

European Commission (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.

Fuller, G. W., Temper, A. H., Baker, T. D., Yttri, K. E. and Butterfield, D. (2014) Contribution of wood burning to PM10 in London. Atmos. Environ., 87, 87–94.

Furrer, R., Nychka, D. and Sain, S. (2013) fields: tools for spatial data. R Package Version 8.4.1. University Corporation for Atmospheric Research, Boulder.

Grewe, V., Dahlmann, K., Matthes, S. and Steinbrecht, W. (2012) Attributing ozone to NOx emissions: implications for climate mitigation measures. Atmos. Environ., 59, 102–107.

Gulliver, J. and Briggs, D. (2011) Stems-air: a simple GIS-based air pollution dispersion model for city-wide exposure assessment. Sci. Total Environ., 409, 2419–2429.

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

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

Huerta, G., Sanso, B. and Stroud, J. R. (2004) A spatiotemporal model for Mexico City ozone levels. Appl. Statist., 53, 231–248.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahuwaroglu, T., Morrison, J. and Giovis, C. (2005) A review and evaluation of intraurban air pollution exposure models. J. Expo. Anal. Environ. Epidem., 15, 185–204.

Kassomenos, P., Kotroni, V. and Kallos, G. (1995) Analysis of climatological and air quality observations from greater Athens area. Atmos. Environ., 29, 3671–3688.

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

Lee, D. and Shaddock, G. (2010) Spatial modeling of air pollution in studies of its short-term health effects. Biometrics, 66, 1238–1246.

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

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

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

Pirani, M., Gulliver, J., Fuller, G. W. and Blangiardo, M. (2014) Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. J. Expos. Sci. Environ. Epidem., 24, 319–327.

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

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

Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2007) High-resolution space-time ozone modeling for assessing trends. J. Am. Statist. Ass., 102, 1221–1234.
Sahu, S. K., Gelfand, A. E. and Holland, D. M. (2010) Fusing point and areal level space–time data with application to wet deposition. *Appl. Statist.*, **59**, 77–103.

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

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

Samet, J., Dominici, F., Curriero, F., Zeger, S. and Coursac, I. (2000) Fine particulate air pollution and mortality in 20 US cities. *New Engl. J. Med.*, **343**, 1742–1749.

Savage, N. H., Agnew, P., Davis, L. S., Ordóñez, C., Thorpe, R., Johnson, C. E., O’Connor, F. M. and Dalvi, M. (2013) Air quality modelling using the Met Office Unified Model (AQM OS24-26): model description and initial evaluation. *Geoscient. Model Devpmnt*, **6**, 353–372.

Schneider, A., Friedl, M. A. and Potere, D. (2009) A new map of global urban extent from modis satellite data. *Environ. Res. Lett.*, **4**, 1–11.

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

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

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

Zhang, H. (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Statist. Ass.*, **99**, 250–261.

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

---

**Supporting information**

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary materials for the paper: “A Bayesian spatio-temporal model to estimate long term exposure to outdoor air pollution at coarser administrative geographies in England and Wales”’.
