Atmospheric Environment 102 (2015) 393–405

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

A spatially varying coefficient model for mapping PM10 air quality at the European scale

N.A.S. Hamm [a, *], A.O. Finley [b], M. Schaap [c], A. Stein [a]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7514, Enschede, AE, The Netherlands
[b] Departments of Forestry and Geography, Michigan State University, East Lansing, MI, 48824-1222, USA
[c] Department of Climate, Air and Sustainability, TNO Built Environment and Geosciences, PO Box 80015, 3508, Utrecht, TA, The Netherlands

HIGHLIGHTS

- A novel geostatistical spatially varying coefficients model (SVC).
- Novel geostatistics for combining chemical transport models (CTMs) and observations.
- Evaluate the CTM and identify when and where it performs well and badly.
- Improved estimates of threshold exceedance.

ARTICLE INFO

Article history:
Received 3 March 2014
Received in revised form
19 November 2014
Accepted 21 November 2014
Available online 22 November 2014

Keywords:
PM10
Geostatistics
Model evaluation
Spatially varying coefficient (SVC)
LOTOS-EUROS

ABSTRACT

Particulate matter (PM) air quality in Europe has improved substantially over the past decades, but it still poses a significant threat to human health. Accurate regional scale maps of PM10 concentrations are needed for monitoring progress in mitigation strategies and monitoring compliance with statutory limit values. Chemistry transport models (CTM) use emission databases and simulate the transport and deposition of pollutants. They deliver such maps but are known to be inaccurate. A promising approach is to use geostatistics to model the relationship between the in situ observations and the CTM. This has been shown to be more accurate than using either observations or CTM's alone. This paper presents a spatially varying coefficients (SVC) geostatistical model as an extension of the standard spatially varying intercept (SVI) geostatistical model. SVC allowed the regression coefficient to vary spatially according to a covariance function, the parameters of which were estimated from the data. It was built as a Bayesian hierarchical model and implemented using Markov chain Monte Carlo. The procedure was applied to Airbase PM10 observations and LOTOS-EUROS simulated PM10 for central, southern and eastern Europe. Model-fit diagnostics showed that SVC delivered a better fit to the data than SVI. Mapping the spatially varying coefficients allowed identification of the locations where the CTM performed well or poorly. This could be used for objective CTM evaluation purposes. The posterior predictive simulations were also used to map median PM10 concentrations as well as the probability of exceeding the 50 μg m^{-3} EU daily PM10 concentration threshold. Although posterior median prediction accuracy was similar for SVI and SVC, SVC better modelled the process and yielded narrower credible intervals. As such, SVC was more appropriate for quantifying uncertainty and for mapping threshold exceedances. The resulting maps may be used to guide air quality assessment and mitigation strategies, including those related to health impacts.

1. Introduction

Although particulate matter (PM) air quality in Europe has improved substantially over the past decades it still poses a significant threat to human health (EEA, 2007). Short-term exposure to PM has frequently been associated with increased human morbidity and mortality (Brunekreef and Holgate, 2002). Long-term exposure to PM is considered to have a strong effect on health (Dockery et al., 1993; Pope et al., 1995) and particulate matter has recently been classified as carcinogenic (Loomis et al., 2013). Health impact studies conclude a reduced life expectancy of 3–12 months in large parts of Europe (EEA, 2007). The European air quality standards currently focus on PM10 (PM < 10 μm in

* Corresponding author.
E-mail address: nick@hamm.org (N.A.S. Hamm).

http://dx.doi.org/10.1016/j.atmosenv.2014.11.043
1352-2310/

diameter) and PM2.5. Many European countries have problems adhering to the allowed annual number of exceedances of the daily limit value of 50 μg m^{-3} for PM10.

Accurate regional scale maps of PM10 concentrations are needed for monitoring progress in mitigation strategies and monitoring compliance with statutory limit values. Moreover, accurate background maps are needed for air pollution modelling, management and exposure estimation in urban agglomerations (Beelen et al., 2013). Generation of these maps is challenging because PM concentrations are only measured at a limited number of locations. Moreover, regional scale chemistry transport models (CTM) systematically underestimate observed PM10 concentrations, due to uncertain knowledge on emissions and formation pathways (Stern et al., 2008). Hence, they do not provide accurate regional background maps. Alternatively, empirical models could be used to predict PM10 at locations where it is not measured. These model the relationship between covariate predictor variables and the pollutant of interest and are based typically on regression (Brauer et al., 2011) or geostatistics (Lloyd and Atkinson, 2004). Empirical models may give accurate results, but are restricted to the conditions under which they are developed. A promising new direction is the integration of measured observations with CTM output to provide maps of air quality, which has been shown to be more accurate that using either alone (van de Kassteele and Stein, 2006; Denby et al., 2008; Candiani et al., 2013). In a geostatistical context the CTM is a covariate and the measured observations are the response variable.

A central assumption in geostatistical modelling is stationarity of the mean and covariance. Stationarity of the mean implies the relationship between the measured observations and the CTM does not vary over the geographic domain. Stationarity of the covariance implies residual variance is constant over the domain and that observations at any pairs of points separated by a given geographic distance are equally correlated. The performance of a CTM shows spatial variability, leading to a non-constant relationship with the observations. Furthermore, given the different sources and components of PM and the influence of meteorological conditions, the level of residual variance and the nature of the spatial autocorrelation is likely to vary in space and time. This suggests a regression model for PM10 that uses CTM output as a covariate should accommodate non-stationarity of the mean and perhaps residuals. Non-stationarity in the regression coefficients may be accommodated using a spatially varying coefficients (SVC) model (Gelfand et al., 2003; Finley, 2011). Recent studies have also considered non-stationarity in the covariance (Haskard and Lark, 2010; Anderes and Stein, 2011; Hamm et al., 2012) although, in practice, adopting a spatially varying regression coefficients model often reduces the problem of non-constant residuals variance. Such models are of particular interest because they may yield better fit to observed data and improved inference. A further important point is that modelling the spatial variability in the relationship between the CTM output and the measured observation allows exploration of the spatial variability in the performance of the CTM. This offers improved insight into the CTM.

In this research geostatistics was used to model the relationship between the CTM and the measured observations. Candidate geostatistical regression models were SVC and two simpler submodels: the standard geostatistical spatially varying intercept (SVI) model and simple linear regression (SLR). The CTM used was LOTOS-EUROS (Schaap et al., 2008), which is used widely for regional and continental level modelling. Study objectives were to: (*i*) assess the candidate models' fit to the observed data as well as their prediction accuracy; (*ii*) gain insight into the CTM's performance by quantifying the spatial variability in the relationship between the CTM output and the observations; and (*iii*) develop maps of average PM10, with associated estimates of uncertainty, as well as the probability of exceeding 50 μg m^{-3} on any given day.

2. Methods

2.1. Models

A geostatistical model can be considered an extension of the commonly used linear regression model

$$y(\mathbf{s}) = \beta_0 + \sum_{k=1}^{p} \beta_k x_k(\mathbf{s}) + w_0(\mathbf{s}) + \varepsilon(\mathbf{s}) \quad (1)$$

where y is the response variable, x_k are predictor variables, β_0 is the intercept and the β_k's are the regression slope coefficients. This model differs from simple linear regression (SLR) because both the response and predictor variables are spatially referenced, where $\mathbf{s}$ denotes location. Further, the residual error is partitioned into two components: $w_0(\mathbf{s})$ and ε. Here $w_0(\mathbf{s})$ is a spatially correlated random effect. There are n w_0's, corresponding to n locations and these are assumed to follow a multivariate normal distribution: $\mathbf{w}_0 = (w_0(\mathbf{s}_1), w_0(\mathbf{s}_2),\ldots, w_0(\mathbf{s}_n))^\top \sim \mathrm{MVN}(0, \sigma_0^2\mathbf{R}(\boldsymbol{\phi}_0))$, where 0 is a zero vector, σ_0^2 is variance parameter and $\mathbf{R}(\boldsymbol{\phi}_0)$ is a $n \times n$ spatial correlation matrix. The non-spatial residual component is assumed to be uncorrelated normally distributed: $\varepsilon(\mathbf{s}) \sim N(0, \tau^2)$.

The correlation between $w_0(\mathbf{s}_i)$ and $w_0(\mathbf{s}_j)$ is a function of their geographic separation, hence:

$$\mathbf{R}_{ij}(\boldsymbol{\phi}_0) = \rho(h_{ij}, \boldsymbol{\phi}_0) \quad (2)$$

where $\rho(\cdot)$ is the correlation function, h_{ij} is the Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_j$ and $\boldsymbol{\phi}$ is a vector containing the parameters of the correlation function (e.g., rate of decay, smoothness). The covariance matrix $\sigma_0^2\mathbf{R}(\boldsymbol{\phi}_0)$ must be positive definite, which can be achieved by choosing a valid correlation function (Cressie, 1993). In classical geostatistics, $\beta_0 + \sum_{k=1}^{p}\beta_k x_k(\mathbf{s})$, τ^2, and σ_0^2 are referred to as the trend, nugget, and partial sill, respectively. This model has been applied widely for air pollution mapping (e.g., Lloyd and Atkinson, 2004; van de Kassteele and Stein, 2006; Denby et al., 2008).

The random effects $\mathbf{w}_0$ are spatially correlated and provide local adjustment to the regression intercept. Hence $\beta_0 + w_0(\mathbf{s})$ can be thought of as a *spatially varying intercept* (SVI) and Model (1) is referred to as SVI. Removing the $w_0(\mathbf{s})$ reduces (1) to SLR.

Model (1) assumes the regression coefficients are constant across the domain which, as discussed in Section 1, may not be a realistic. Gelfand et al. (2003) and Finley (2011) addressed this by including slope-coefficient-specific random effects, denoted w_k, to (1) as follows

$$y(\mathbf{s}) = \beta_0 + \sum_{k=1}^{p} \beta_k x_k(\mathbf{s}) + w_0(\mathbf{s}) + \sum_{k=1}^{p} w_k(\mathbf{s}) x_k(\mathbf{s}) + \varepsilon(\mathbf{s}) \quad (3)$$

The $w_k(\mathbf{s})$'s, $k = 1,\ldots,p$, are defined analogous to the w_0's, each with its own set of variance and spatial correlation parameters: σ_k^2 and $\boldsymbol{\phi}_k$. Although the w_k's may vary spatially, they do so according to a correlation structure that is, itself, stationary. In terms of classical geostatistics one can think of different variograms for the intercept and for each slope coefficient. Hence $\beta_k + w_k(\mathbf{s})$ is a *spatially varying coefficient* and (3) is referred to as SVC.

2.2. Parameter estimation and prediction

In order to use the models it is necessary to estimate the parameter values. These are the regression coefficients,

$\boldsymbol{\beta}=(\beta_0, \beta_1,\ldots, \beta_p)$, and the correlation and variance parameters, $\boldsymbol{\theta} = (\tau^2, \sigma_0^2, \ldots, \sigma_p^2, \boldsymbol{\phi}_0, \ldots, \boldsymbol{\phi}_p)$. Classical geostatistics has proceeded using the method-of-moments approach where a model is fitted to the sample variogram of the regression residuals or by using maximum likelihood estimation. The estimated parameters are then used directly for prediction at unknown locations, a procedure known as spatial interpolation or kriging. These methods are reviewed in standard texts (Cressie, 1993; Diggle and Ribeiro, 2007; Webster and Oliver, 2008).

The research presented in this paper adopted a Bayesian approach for parameter estimation and for prediction at unknown locations. There are two key reasons for this choice. First, the Bayesian approach characterizes uncertainty at all stages in the modelling process and then propagates this through to prediction at unknown locations. In contrast, the classical approaches, mentioned above, use the estimates of $\boldsymbol{\theta}$ directly without considering their associated uncertainty. Ignoring uncertainty in the covariance parameter limits inference about model parameters and prediction (Diggle and Ribeiro, 2007). Second, the Bayesian approach is flexible because it allows the development of complex hierarchical models. SLR and SVI can be specified using the method-of-moments or maximum likelihood approaches; however, the authors are not aware of any such implementation for SVC.

Under Bayes theorem the parameters' posterior distribution is proportional to the product of the likelihood and parameters' prior distribution (Gelman et al., 2013):

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood} \tag{4}$$

This posterior distribution is then used for inference. Where prior knowledge of a parameter is not available a non-informative prior is used. For most real-world problems analytical solutions do not exist and the posterior distribution is simulated using Markov chain Monte Carlo (MCMC) (Lunn et al., 2013; Gelman et al., 2013). Markov chains are simulated for multiple starting values and allowed to converge. The m post-convergence simulated values are used subsequently to predict w_0, w_p and then y at an unknown location, $\mathbf{s}_0$. The simulated values of β, $\boldsymbol{\theta}$, $w_0(\mathbf{s}_0)$, $w_p(\mathbf{s}_0)$ and $y(\mathbf{s}_0)$ are then used for inference.

The reader is referred to the Supplementary material for a summary of the procedures used in this paper.

2.3. Model evaluation

Model evaluation has received extensive attention in both the statistics and air quality literature. Two approaches were used to evaluate the candidate models. These were (1) statistical goodness-of-fit of the model to the observed data and (2) validation of the model predictions against a hold-out dataset.

Three widely used model goodness-of-fit criteria were used. These were the deviance information criterion (DIC) (Spiegelhalter et al., 2002), the predictive model choice criterion (D) (Gelfand and Ghosh, 1998, Equation (5) with $k \to \infty$), and the scoring rule (SR) of Gneiting and Raftery (2007, Equation 27). These criteria evaluate both the closeness of the average of the m predicted values to the observed data as well as the spread of the individual predictions around the observed data. All three criteria also penalize increasing model complexity.

When comparing the candidate models, the preferred model will yield predictions with low variance that are, on average, close to the observed data whilst having low model complexity. Larger values of SR and lower values of DIC and D indicate a better fit to the data.

In the second approach a proportion of the observations were held out when the model parameters were estimated. The posterior predictions at the hold-out locations were then compared to the observations at those locations. General procedures for air quality model evaluation were discussed by Borrego et al. (2008) and Thunis et al. (2012, 2013). The specific case of PM10 was discussed by Pernigotti et al. (2013). In line with those approaches the root mean square (prediction) error (RMSE), bias and the R^2, were calculated.

$$\text{RMSE} = \sum_{i=1}^{n^*} (y(\mathbf{s}_i) - \hat{y}(\mathbf{s}_i))^2 \Big/ n^*$$

$$\text{bias} = \sum_{i=1}^{n^*} (y(\mathbf{s}_i) - \hat{y}(\mathbf{s}_i)) \Big/ n^*$$

where $\hat{y}(\mathbf{s}_i)$ is the median of the posterior predictive distribution and $y(\mathbf{s}_i)$ is the measured observation, both at $\mathbf{s}_i$, and n^* is the number of held-out observations. Thunis et al. (2012, 2013) developed this further to define the model quality objective (MQO):

$$\frac{\text{RMSE}}{2U} < 1 \tag{5}$$

where $U = \sqrt{\sum_{i=1}^{n} (U_r(y(\mathbf{s}_i)) \times y(\mathbf{s}_i))^2}$ is the observation uncertainty, $U_r(y(\mathbf{s}_i))$ is the relative uncertainty for a given value of $y(\mathbf{s}_i)$ and n is the total sample size. For this paper the relative uncertainty was set to a constant $U_r = 0.25$, in line with Thunis et al. (2012). Pernigotti et al. (2013) proposed a more extensive evaluation of the relative uncertainty; however, that lies outside of the scope of this study.

3. Study area and data

3.1. Study area

The study area was mainland European countries with a substantial number of available PM10 observations. The countries included were Portugal, Spain, Italy, France, Switzerland, Belgium, The Netherlands, Germany, Denmark, Austria, Poland, The Czech Republic, Slovakia and Slovenia. Other eastern European countries and the Balkan states were excluded owing to a lack of publicly available data. The United Kingdom, Ireland and Scandinavia (with the exception of Denmark) were excluded because they do not share a land border with the other countries. Additionally, the Scandinavian countries had few observations and are remote from the European mainland. In principle the presented methods could be extended to a wider geographic area; however, for this paper the study area was restricted to the above-listed contiguous countries.

All data were projected to the European Terrestrial Reference System 1989 (ETRS) Lambert Azimuthal Equal-Area (LAEA) projection which gives a coordinate reference system for the whole of Europe. Distance units are specified in kilometres throughout.

3.2. Observed measurements

Air quality observations for the European Economic Area were available via Airbase (*Air* quality data*Base*).[1] Daily PM10 concentrations were extracted for two 2-week periods: 1–14 April 2009 and 1–14 June 2009. June 2009 was characterized by a stable period of low concentration of PM10 (overall maximum: 64 μg m^{-3}; maximum daily median: 17 μg m^{-3} for all countries). PM10

[1] http://acm.eionet.europa.eu/databases/airbase/(accessed 26 September 2014).

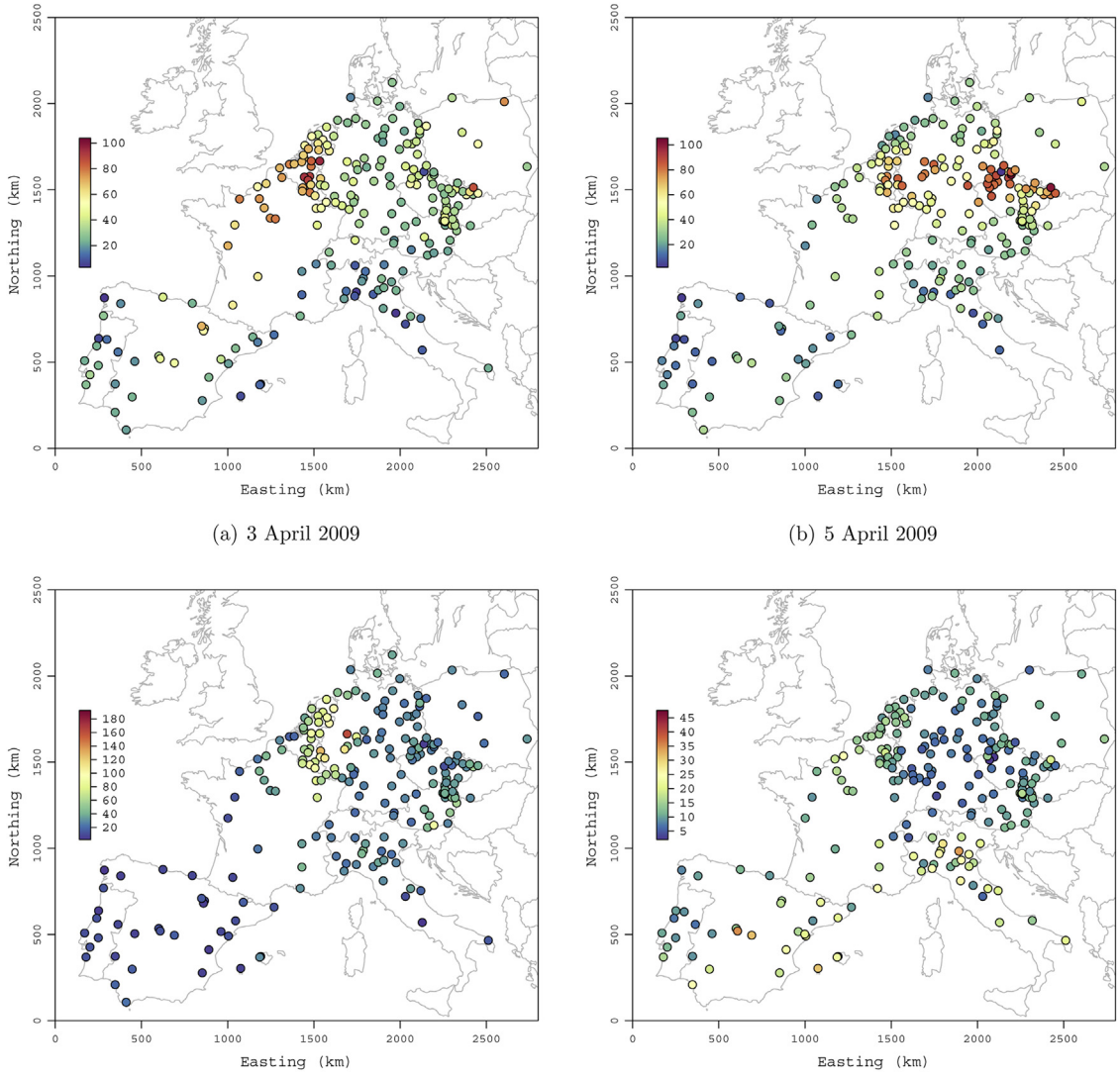
concentrations in April were higher and fluctuated in both space and time (overall maximum: 185 μg m^{-3}; maximum daily median: 42 μg m^{-3}). These two periods were chosen so that the geostatistical models could be evaluated under contrasting conditions. Furthermore, interpretation of the April results is supported by Banzhaf et al. (2013), who described two European air pollution events for 2–7 April 2009 and 11–16 April 2009. Four example days (3, 5, 13 April and 11 June) are shown in Fig. 1. These were chosen to correspond to the April 2009 events whereas 11 June 2009 was a low-pollution day.

Airbase daily values are averaged over the within-day hourly values when at least 18 hourly measurements are available, otherwise no data are provided. The total number of daily observations varied between 215 and 238 for the relevant days. Airbase monitors are classified by type of area (rural, urban, suburban) and by type (background, industrial, traffic or unknown). Only rural background monitors were used. This is common for comparing measured observations to coarse resolution CTM simulations (Denby et al., 2008). Monitoring sites above 800 m altitude were also excluded. These tend to be located in areas of variable topography and the accuracy of the CTM for locations that shift from inside to outside the mixing layer is known to be poor. No further quality control was performed on the data.

3.3. LOTOS-EUROS (CTM) data

LOTOS-EUROS (v1.8) is a 3D CTM that simulates air pollution in the lower troposphere. The simulator projection is normal longitude–latitude and the standard grid resolution is 0.50° longitude ×0.25° latitude (approximately 25 km × 25 km). LOTOS-EUROS simulates the evolution of the components of particulate matter separately. Hence, this CTM incorporates the dispersion, formation and removal of sulphate, nitrate, ammonium, sea salt, dust, primary organic and elemental carbon and non-specified primary material, although, it does not incorporate secondary organic aerosol. Hence, a systematic underestimation of observed PM10 levels is present in the LOTOS-EUROS validation, which is a common feature of CTMs (Stern et al., 2008) For a detailed description of LOTOS-EUROS the reader is referred to Hendriks et al. (2013) and the references therein.

The hour-by-hour calculations of European air quality in 2009 were driven by the European Centre for Medium Range Weather



(a) 3 April 2009

(b) 5 April 2009

(c) 13 April 2009

(d) 11 June 2009

Fig. 1. Observed values for 3, 5, 13 April and 11 June 2009. Units: μg m^{-3}.

Forecasting (ECMWF) meteorology. Emissions were taken from the MACC (Monitoring Atmospheric Composition and Climate) emissions database (Pouliot et al., 2012). Emissions from wild fires were neglected. Boundary conditions were taken from the global MACC service (Flemming et al., 2009). The LOTOS-EUROS CTM, in this configuration, has participated in international model comparison studies showing that LOTOS-EUROS is state-of-the-art (Solazzo et al., 2012).

The LOTOS-EUROS hourly model output was averaged to daily mean PM10 concentrations. LOTOS-EUROS grid cells that were spatially coincident with the Airbase observations were extracted and used as the covariate necessary for the geostatistical modelling. The LOTOS-EUROS grids were projected to the ETRS89 LAEA coordinate reference system. The projected grid had a resolution of 25 km × 25 km.

3.4. Model specification and implementation

Following (3) SVC is specified as:

$$y(\mathbf{s}) = \beta_0 + \beta_{LE}x(\mathbf{s}) + w_0(\mathbf{s}) + w_{LE}(\mathbf{s})x(\mathbf{s}) + \varepsilon(\mathbf{s}) \quad (6)$$

where x is the LOTOS-EUROS simulated PM10, y is the square root of the observed PM10 and other terms are as defined previously. The corresponding SVI and SLR sub-models are given as $y(\mathbf{s}) = \beta_0 + \beta_{LE}x(\mathbf{s}) + w_0(\mathbf{s}) + \varepsilon(\mathbf{s})$ and $y(\mathbf{s}) = \beta_0 + \beta_{LE}x(\mathbf{s}) + \varepsilon(\mathbf{s})$, respectively. The observed PM10 was square-root transformed in order to better meet the assumptions of the above linear model. The predictions were back transformed prior to mapping.

In Model (6) $\beta_0 + w_0(\mathbf{s})$ is the spatially varying intercept and allowed the offset between the CTM and the observations to vary spatially. $\beta_{LE} + w_{LE}(\mathbf{s})$ is the spatially varying coefficient that allowed the relation between the CTM and the observations to vary spatially. Where $\beta_0 + w_0(\mathbf{s})$ is close to zero and $\beta_{LE} + w_{LE}(\mathbf{s})$ is close to unity, the CTM is accurate.

The spatial correlation functions for both w_0 and w_{LE} and were assumed to be exponential $\rho(h_{ij},\boldsymbol{\phi}_k)=\exp(-\phi_k h_{ij})$ where ϕ_k is the spatial decay parameter. Exploratory data analysis using other correlation functions yielded comparable results. For the exponential function the *effective spatial range* is the geographic distance at which the spatial correlation is negligible. This is typically taken as $-\log(0.05)/\phi_k \approx 3/\phi_k$ km, which is the geographic distance where the spatial correlation drops to 0.05 (ρ=0.05) (Diggle and Ribeiro, 2007). Note that, for an exponential correlation function, the largest correlations are exhibited at relatively short separations, for example $-\log(0.5)/\phi_k \approx 0.7/\phi_k$ km for ρ=0.5.

The priors for β_0 and β_{LE} were both given normal distributions and set to $N(0,10^5)$. The variance parameters (τ^2, σ_0^2 and σ_{LE}^2) received inverse Gamma (IG) distributed priors with shape hyperparameter equal to 2 and scale obtained from sample variograms of the SLR model residuals. With a shape of 2, the IG prior has mean equal to the scale and an infinite variance. Defining priors with a large variance aims to make them non-informative, whilst the choice of an IG distribution for the variance parameters ensures that they are positive or zero. The spatial decay parameters, ϕ_0 and ϕ_{LE} were both given uniform distributions and set to Unif(0.001,3). This corresponds to an effective ranges between 1 and 3000 km, which is approximately the maximum extent of the domain. Further guidance about the choice of priors is given by Banerjee et al. (2004) and Finley (2011).

For each day SLR and SVI were fit using the spBayes R package (Finley et al., 2014) whereas SVC was written in C++. All models were run on a Linux workstation and used Intel's Math Kernel Library for efficient matrix operations.

For each model, three MCMC chains with dispersed parameter starting values were run for 50,000 iterations. Each SVC chain took approximately 30 min to run for each day. Chain trace plots and Gelman-Rubin diagnostics (Lunn et al., 2013; Gelman et al., 2013) were used to identify convergence at ~ 15,000 iterations. Parameter inference, goodness-of-fit diagnostics (DIC, D and SR), and predictive maps were based on the post-convergence samples. To calculate the model-validation diagnostics (bias, RMSE and R) models were re-run excluding a 25% hold-out sample. The model-validation diagnostics were then calculated using these 25% hold-out samples, whereas U (Equation (5)) was calculated across all observations, n, on a given day.

The median of the posterior distribution and the 95% credible intervals were calculated for all parameters (β_0, β_{LE}, τ^2, σ^2, ϕ). Similarly posterior samples of spatial random effects (w_0's and w_{LE}'s) and predictions (y's) were summarized by the median and 95% credible interval and mapped.

4. Results

4.1. Model evaluation and comparison

The validation diagnostics, evaluated for a 25% hold-out sample, are shown in Tables 1 and 2. As expected, the CTM systematically underestimated observed PM10. The R^2, between the CTM and the observations, was low and the RMSE was large. The RMSE met the MQO (RMSE <$2U$) for all days in April, with the exception of 11, 12 and 13 April. The RMSE for the CTM predictions did not meet the MQO for any day in June. Maps showing the residuals between the measured observations and the CTM output, $y(\mathbf{s}) - x(\mathbf{s})$, for the four exemplar dates are shown in Fig. S1 in the Supplementary material. As expected the CTM generally underestimated observed PM10.

Tables 1 and 2 also show the validation diagnostics for the three regression models. Adopting SLR led to a clear reduction in the bias, an increase in R^2 and a reduction in the RMSE compared to CTM. The MQO objective was met for all days except 14 April and 13 June. For April, adopting the two spatial models (SVI and SVC) did not lead to a consistent change in the bias relative to SLR; however, the R^2 increased and the RMSE decreased. For June adopting the two spatial models also did not lead to a consistent change in the bias relative to SLR; however, the R^2 increased and the RMSE decreased. For RMSE the MQO was achieved for all dates for both SVI and SVC. The reduction in the RMSE was less substantial in June than in April. Overall, these validation diagnostics show that the spatial models are preferred to SLR; however, they did not distinguish between SVI and SVC.

The 25% hold-out sample was selected at random. Choosing a different random hold-out would yield different numeric results but was not expected to lead to different conclusions. To verify this, the above process was repeated for different hold-out samples. This led to small differences in the numeric values but did not change the above conclusions. The results for a second hold-out run are included in the Supplementary material (Tables S1 to S2).

As discussed in Section 2.3, models can be compared using the DIC, D and SR model-fit diagnostics. These are presented in Table 3 (April) and Table 4 (June). In all cases the two spatial models gave a substantially better fit to the data than SLR. For April, all three diagnostics gave substantial support for SVC over SVI. For June all three diagnostics supported SVC over SVI, although for the DIC diagnostic, this evidence was less strong on 2, 7–10 and 13 June, particularly on 8 and 9 June. This suggests that, on these two days, there is weaker evidence to support the notion that the relationship between the simulations and the observations varies spatially. The overall conclusion is that SVC is preferred to SVI.

Table 1
Table showing the validation diagnostics computed for the 25% hold-out sample for April.

Day	R^2				Bias				RMSE				2*U*
	CTM	SLR	SVI	SVC	CTM	SLR	SVI	SVC	CTM	SLR	SVI	SVC	
1	0.22	0.33	0.63	0.59	7.60	−1.43	1.33	2.01	14.64	11.08	8.28	9.16	15.92
2	0.35	0.45	0.78	0.77	9.39	−0.94	0.46	0.95	15.08	10.19	6.48	6.64	17.26
3	0.47	0.67	0.74	0.74	12.12	2.09	2.25	2.71	19.06	12.02	10.45	10.70	21.16
4	0.56	0.67	0.82	0.82	17.62	3.94	4.10	4.45	25.25	14.30	10.77	10.96	25.46
5	0.59	0.66	0.80	0.82	13.96	−0.32	0.55	2.03	20.95	12.62	9.73	9.41	23.41
6	0.38	0.46	0.77	0.75	11.58	−0.08	2.18	2.67	19.04	13.96	9.48	9.87	21.25
7	0.32	0.47	0.74	0.73	9.25	1.69	1.02	0.73	16.09	13.17	8.79	9.00	17.34
8	0.41	0.56	0.72	0.71	5.90	−1.66	−0.64	−0.24	11.75	9.99	7.62	7.72	13.87
9	0.47	0.38	0.54	0.59	4.24	1.18	1.28	1.71	9.79	8.10	7.01	6.94	12.28
10	0.34	0.16	0.57	0.63	5.36	2.65	1.80	1.49	10.57	9.08	6.55	5.94	12.74
11	0.11	0.57	0.71	0.71	8.75	1.14	0.65	1.15	20.76	10.16	7.81	7.62	16.12
12	0.47	0.48	0.81	0.85	14.47	1.13	0.01	−0.01	23.06	17.11	10.58	9.46	20.20
13	0.55	0.52	0.74	0.77	13.67	1.04	−1.02	−1.53	22.60	16.34	12.22	11.79	21.69
14	0.43	0.27	0.41	0.43	10.50	−0.38	−1.35	−1.46	17.56	20.57	17.65	17.37	19.60

Table 2
Table showing the validation diagnostics computed for the 25% hold-out sample for June.

Day	R^2				Bias				RMSE				2*U*
	CTM	SLR	SVI	SVC	CTM	SLR	SVI	SVC	CTM	SLR	SVI	SVC	
1	0.35	0.39	0.64	0.60	8.24	0.16	0.39	0.38	10.74	6.37	4.85	5.14	8.87
2	0.36	0.42	0.67	0.64	8.27	−1.84	−1.00	−0.72	10.49	6.28	4.59	4.78	9.62
3	0.16	0.26	0.44	0.43	9.16	−0.17	−0.34	−0.22	11.47	6.60	5.82	6.08	9.68
4	0.18	0.33	0.53	0.52	7.24	−1.21	−1.23	−0.68	10.20	7.34	6.03	6.26	9.20
5	0.14	0.20	0.58	0.57	6.57	0.18	0.11	0.39	10.13	7.34	5.36	5.36	9.16
6	0.33	0.46	0.58	0.49	8.10	1.44	1.30	1.51	10.15	5.00	4.44	5.06	8.79
7	0.15	0.01	0.21	0.22	6.09	−0.11	−0.37	−0.31	8.51	6.13	5.24	5.31	6.95
8	0.11	0.11	0.32	0.32	5.19	1.39	1.36	1.48	7.52	4.25	3.78	3.97	6.56
9	0.10	0.12	0.35	0.37	6.74	−0.80	−0.77	−0.57	8.56	5.22	4.53	4.41	7.00
10	0.12	0.19	0.41	0.33	7.18	−1.56	−0.95	−0.60	9.65	6.67	5.66	5.82	7.81
11	0.22	0.16	0.50	0.52	5.66	−1.26	−0.67	−0.16	7.67	6.00	4.67	4.43	6.76
12	0.13	0.11	0.65	0.68	5.63	−1.05	−0.69	−0.57	8.72	7.20	4.66	4.32	7.65
13	0.18	0.16	0.59	0.60	6.99	−1.20	−0.97	−0.36	10.93	9.24	6.64	6.31	8.95
14	0.31	0.37	0.64	0.62	6.89	−1.47	−1.67	−1.49	9.50	6.81	5.63	5.36	9.18

4.2. Estimation of the model parameters

The posterior estimates of the regression coefficients are shown in Fig. 2 (April) and Fig. 3 (June) for SLR, SVI and SVC. For June the credible intervals overlapped and there was no significant difference in the estimated values between the three models. Exceptions were observed in April, where on several days β_{LE} was larger for SLR. Recall that SVI and SVC both modelled spatial autocorrelation in the residuals whereas SLR did not. This may lead to differences in the posterior regression coefficient estimates, depending on the spatial structure on any given day.

If the credible interval for β_{LE} would include zero for SLR or SVI this would imply that there is no significant relationship between the observations and the CTM. The same interpretation cannot be applied for SVC, where it is necessary to consider maps of the spatially varying coefficient $\beta_{LE} + w_{LE}(\mathbf{s})$ and whether the location-specific credible intervals include zero.

Fig. 4 shows the posterior estimates of $\beta_{LE} + w_{LE}(\mathbf{s})$ for 3 and 5 April. On 3 April the observations and the CTM showed a significant relationship over much of the domain, although there is considerable variability in the value of $\beta_{LE} + w_{LE}(\mathbf{s})$. This should be close to 1 if the CTM is accurate. There was a clear band of high

Table 3
Table showing the goodness-of-fit diagnostics for April.

Day	DIC			D			SR		
	SLR	SVI	SVC	SLR	SVI	SVC	SLR	SVI	SVC
1	681	494	474	569	164	132	−274	47	103
2	670	426	400	551	111	89	−267	138	190
3	726	468	452	682	138	114	−314	96	141
4	759	590	543	770	265	188	−345	−58	39
5	708	600	503	631	288	148	−299	−77	93
6	697	532	509	646	212	162	−299	−16	53
7	685	494	466	583	174	129	−280	32	108
8	607	475	447	422	166	122	−208	33	114
9	589	534	527	377	243	189	−179	−58	16
10	580	488	444	364	184	123	−177	12	123
11	646	464	425	486	147	105	−240	71	152
12	740	563	548	699	233	193	−322	−31	22
13	711	585	561	654	258	207	−307	−52	3
14	641	556	533	493	249	187	−242	−43	22

Table 4
Table showing the goodness-of-fit diagnostics for June.

Day	DIC			D			SR		
	SLR	SVI	SVC	SLR	SVI	SVC	SLR	SVI	SVC
1	529	332	313	291	83	65	−128	187	249
2	514	381	374	262	104	88	−103	145	192
3	524	375	315	270	97	62	−105	175	281
4	544	395	360	292	110	79	−125	143	229
5	589	450	439	351	147	115	−168	77	145
6	497	431	414	240	143	109	−83	70	144
7	555	442	433	314	139	119	−143	81	129
8	482	415	413	230	139	113	−75	66	126
9	474	418	416	222	143	117	−70	56	115
10	539	448	441	293	151	124	−125	58	111
11	471	357	346	216	100	79	−62	150	215
12	548	377	365	305	104	83	−137	148	204
13	606	475	469	375	158	135	−183	61	104
14	513	428	409	258	138	104	−98	79	155

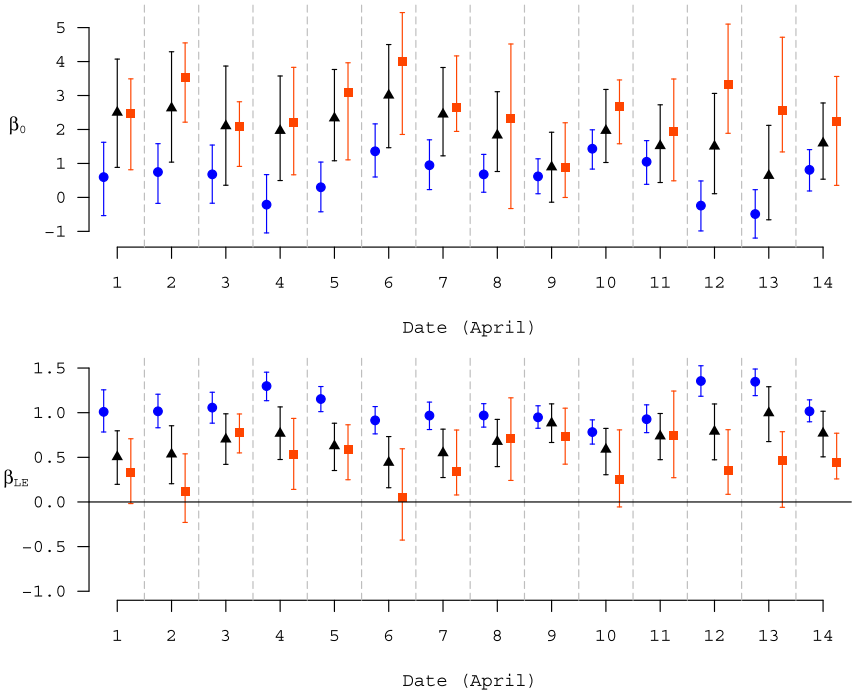


Fig. 2. Estimates of the intercept, β_0 (top) and slope, β_{LE} (bottom), coefficients for April 2009. Blue circles indicate SLR, black triangles indicate SVI and orange squares indicate SVC. The symbol indicates the posterior median value and the bars the 95% credible interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

values across Spain, France and Belgium, as well as Poland. By 5 April this band of high values had shifted to lie across central Germany, The Czech Republic and Poland. There was no significant relationship between the CTM and the observations over most of Spain, Italy, Portugal and western France where the observed PM10 concentrations were low compared to central and eastern Europe. Results from other days can be interpreted in a similar way. An example for 11 June, a low pollution day, is given in the Supplementary material (Fig. S2).

Fig. 5 shows the posterior estimates of $\beta_{LE} + w_{LE}(\mathbf{s})$ for 13 April for both SVI and SVC. On this day there was a localized high-pollution event centred over north-west Germany and the Netherlands. A significant relationship between the CTM and the observations occurred mainly in a band covering this area.

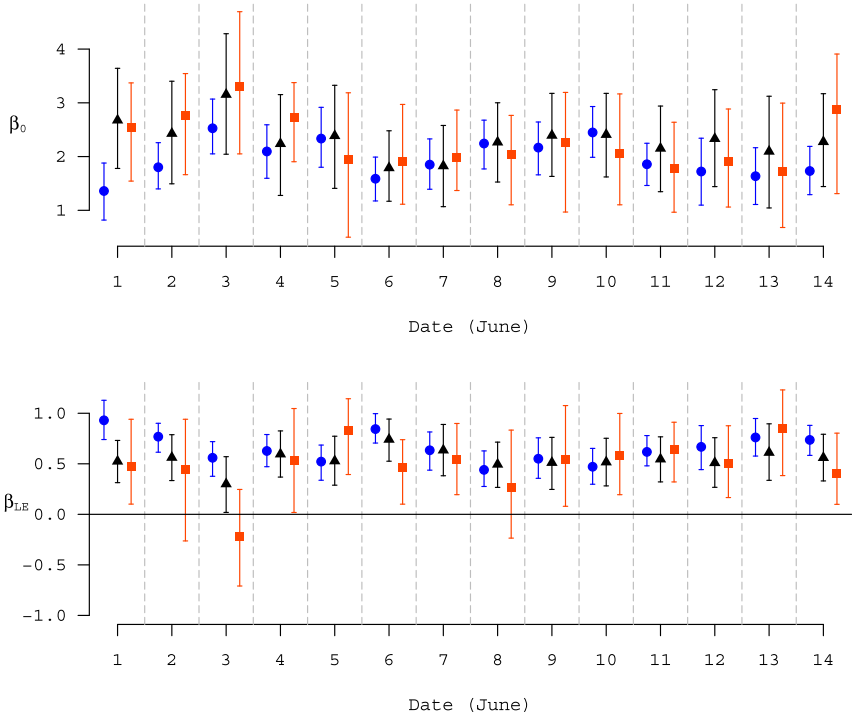


Fig. 3. Estimates of the intercept, β_0 (top) and slope, β_{LE}, coefficients for June 2009. Other details as for Fig. 2.

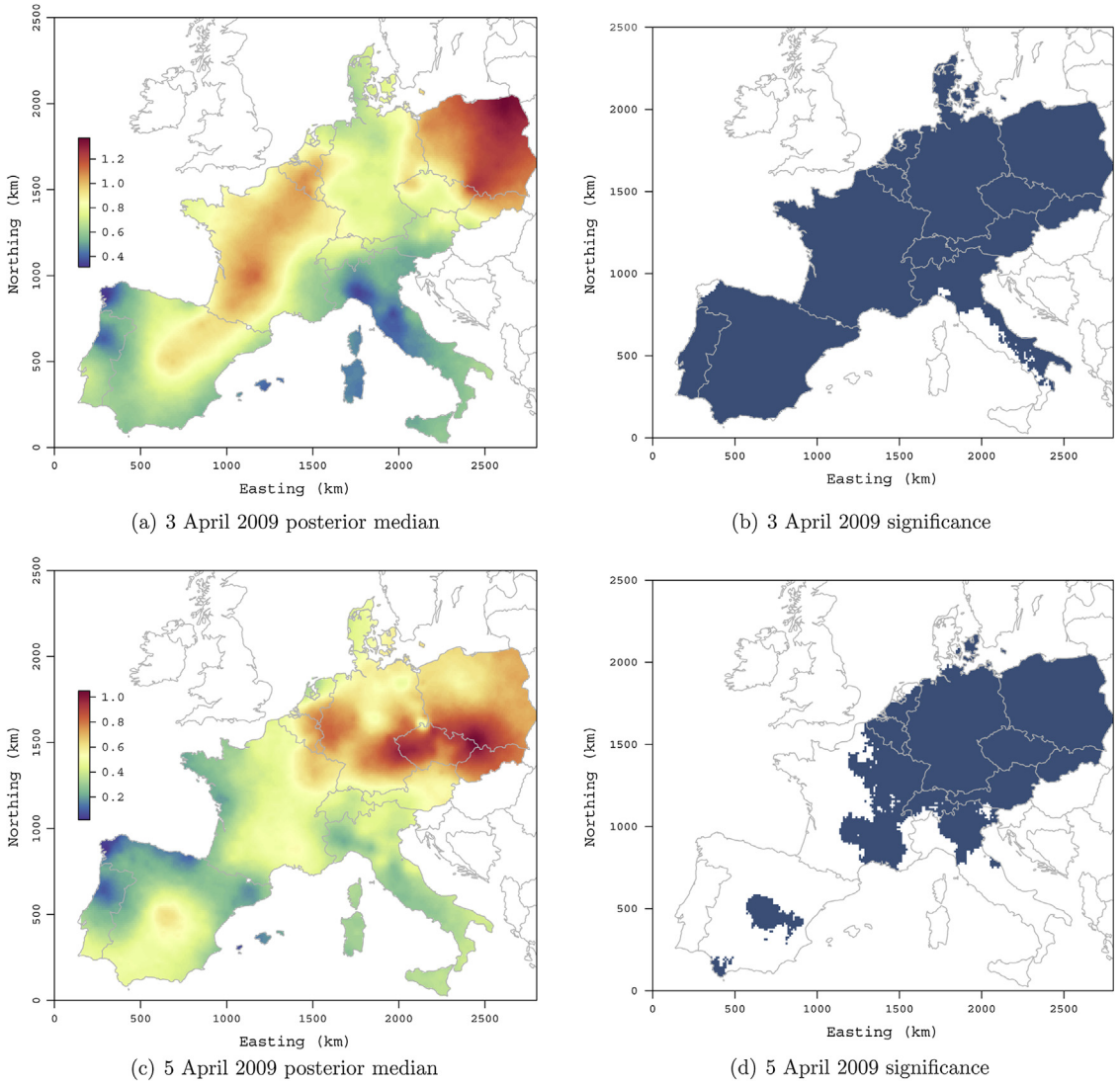
Estimates of $\beta_0 + w_0(\mathbf{s})$ for SVI (Fig. 5(c)) and SVC (Fig. 5(d)) are also shown. The pattern for $\beta_0 + w_0(\mathbf{s})$ (SVI) was similar to $\beta_{LE} + w_{LE}(\mathbf{s})$. This is because residual variation was modelled by w_0 for SVI but was partitioned between w_0 and w_{LE} for SVC. A similar phenomenon was observed for other days.

The posterior estimates of the uncorrelated residual variance, τ^2, are shown in Fig. 6. These were clearly largest for SLR. SVI and SVC incorporated a spatially correlated random effect, w_0 with variance σ_0^2. SVC incorporated an additional spatially correlated random effect, w_{LE} with variance σ_{LE}^2. The w_0 and w_{LE} accounted for much of the residual variability, hence τ^2 was reduced. This led to more accurate prediction and, in part, this also explains why the goodness-of-fit diagnostics favoured SVI and SVC over SLR. The posterior median of τ^2 was lower for SVC than for SVI, but the credible intervals overlapped. Further discussion on σ_0^2 and σ_{LE}^2 is not central to this paper. The estimates of σ_0^2 and σ_{LE}^2 are presented in the Supplementary material (Figs. S3 and S4) but are not discussed further here.

The median value of ϕ_0 for the SVI model were consistent in both months, with the effective range lying between 500 and 1500 km. SVC tended to have a longer median effective range than SVI (between 1000 and 2500 km). SVC gave a more flexible autocorrelation structure because both the intercept and slope were allowed to vary and this may account for these differences; however, the 95% credible regions overlapped so strong conclusions cannot be drawn. The wide 95% credible intervals, up to 3000 km, are consistent with other studies. For SVC, the median effective range for ϕ_{LE} varied between 2600 and 2800 km across all days in both months, with credible intervals up to 1200 km wide. The posterior estimates of the effective range are illustrated in Figs. S5 and S6 in the Supplementary material. Finally, the geostatistical definition of the effective range should be clarified. This is the spatial separation for which the spatial correlation between two points becomes small (ρ=0.05). The exponential decay means that the strongest correlation is shown at much shorter distances. Taking ρ=0.5 yields a correlation length of ~110, 350 and 580 km, corresponding to an effective range (ρ=0.05) of 500, 1500 and 2500 km respectively. These correlation lengths correspond approximately to the size of the structures shown in Figs. 4, 5, 7 and 8.

4.3. Mapping

Fig. 7 shows the posterior predicted median, 95% credible interval and probability of exceeding 50 μg m^{-3} for both SVI and SVC.



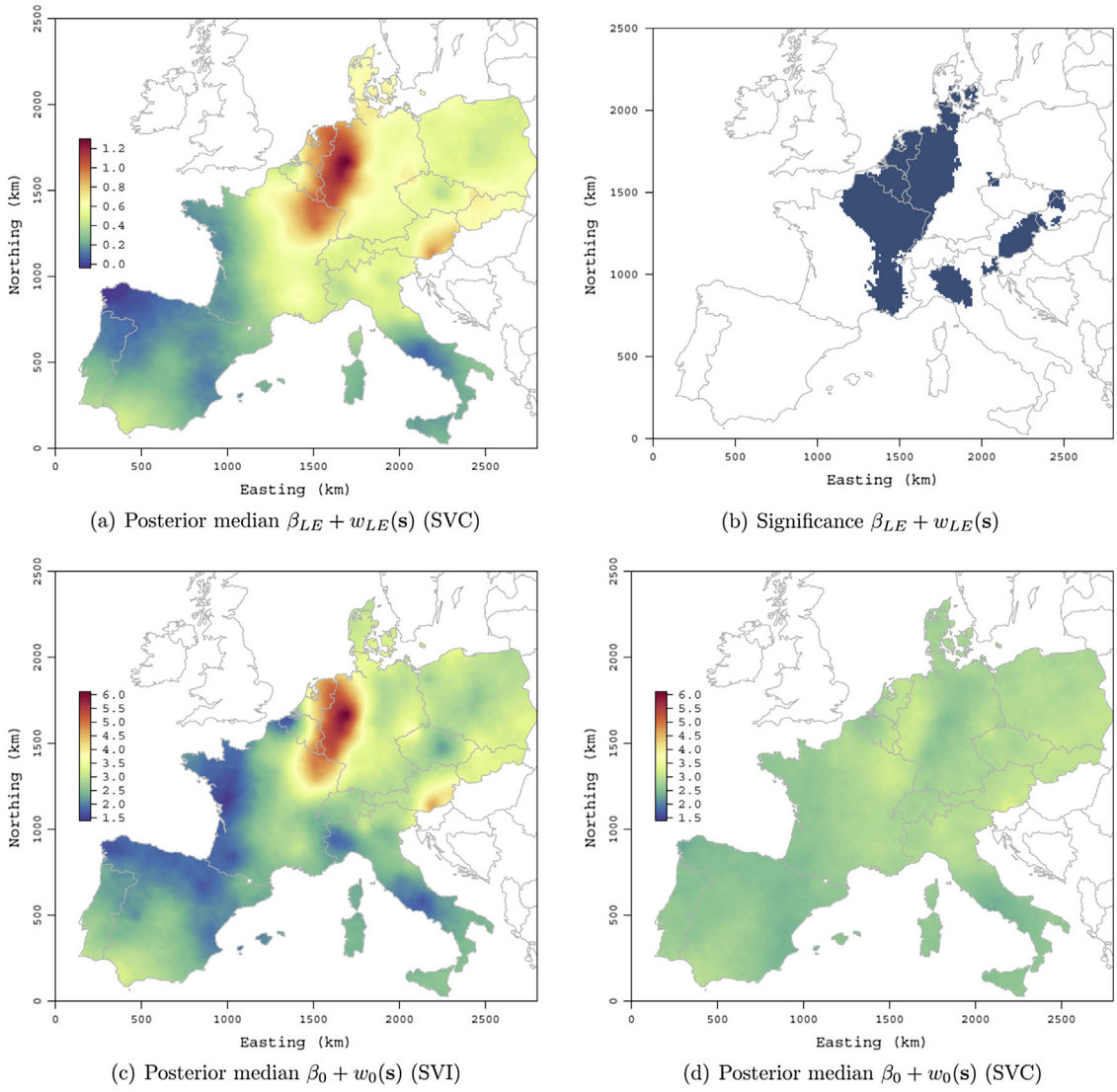
(a) 3 April 2009 posterior median

(b) 3 April 2009 significance

(c) 5 April 2009 posterior median

(d) 5 April 2009 significance

Fig. 4. Maps showing the posterior median of $\beta_{LE} + w_{LE}(\mathbf{s})$ for 3 & 5 April 2009. Blue (right panel) indicates that the 95% credible interval for $\beta_{LE} + w_{LE}(\mathbf{s})$ does not include zero. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Posterior median $\beta_{LE} + w_{LE}(\mathbf{s})$ (SVC)

(b) Significance $\beta_{LE} + w_{LE}(\mathbf{s})$

(c) Posterior median $\beta_0 + w_0(\mathbf{s})$ (SVI)

(d) Posterior median $\beta_0 + w_0(\mathbf{s})$ (SVC)

Fig. 5. Maps showing the posterior median of $\beta_{LE}+w_{LE}(\mathbf{s})$ and $\beta_0 + w_0(\mathbf{s})$ for 13 April 2009. Also shown (b) is whether the credible interval for $\beta_{LE} + w_{LE}(\mathbf{s})$ includes zero.

The mapped posterior median was similar for the two maps. The 95% credible interval was narrower for SVC than for SVI. The area with a high probability of exceeding the 50 μg m^{-3} was more tightly focused for SVC. In general, it was observed that the posterior medians were similar for SVI and SVC although the 95% interval was wider for SVI. Probability-of-exceedance depends on both the posterior median and the posterior variance. Because SVC fits the data better it is expected to give more realistic estimates of exceedence probabilities. This figure illustrates several important benefits of the hierarchical Bayesian methodology presented in this paper: (1) it provides a framework for building SVC, (2) it allows uncertainty in the predictions to be fully quantified and (3) following from the two previous points, exceedence probabilities can be calculated properly.

Fig. 8 shows the posterior median and probability-of-exceedance maps for 3 and 5 April for the SVC model. This clearly shows the high pollution event that developed over northern France and Belgium in early April and was then transported east through 4–6 April. An example for a low-pollution day, is given in the Supplementary material (Fig. S7).

5. Discussion

The results presented in Tables 1 and 2 show that the raw CTM output was biased and relatively inaccurate. Combining the observations and CTM output is therefore recommended since this substantially reduced the bias in the predictions, increased the accuracy and increased the R^2 between the observations and predictions. Three regression models were used to achieve this: the non-spatial SLR and the spatial SVI and SVC models. All showed an improvement relative to the raw CTM output. The goodness-of-fit and validation diagnostics both showed that the two spatial models were preferred to SLR. The goodness-of-fit diagnostics showed that SVC fitted the data better than SVI; however, for predicting the median PM10 concentration the validation diagnostics showed that SVI and SVC were similar. On the other hand, the single random effect in the SVI soaked up enough of the residual spatial variability to deliver accurate median predictions. There is a further point to note. The validation diagnostics were evaluated at the median prediction and did not take into account the variance of the posterior predictions. Indeed discussion of

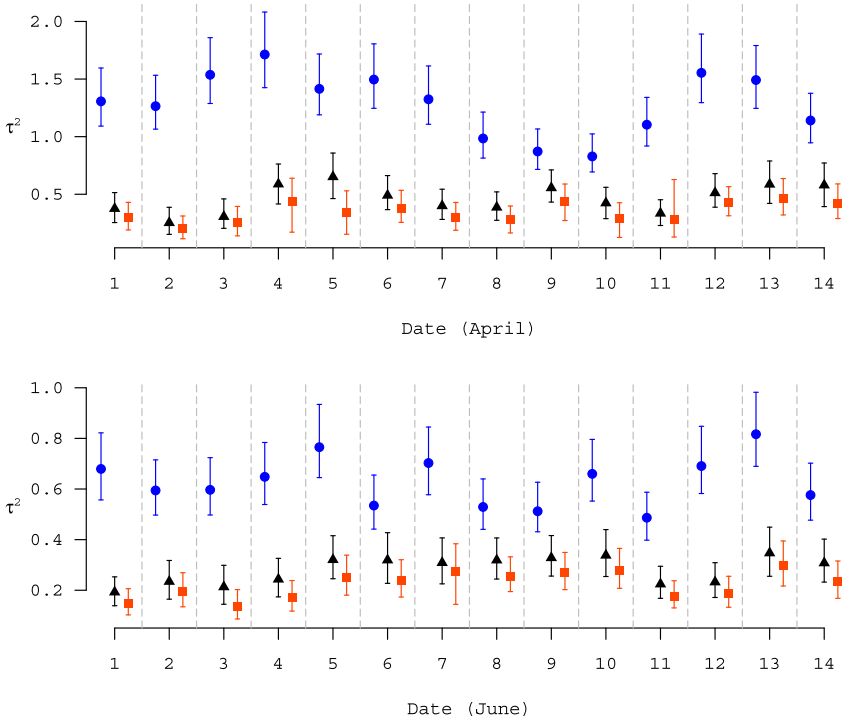


Fig. 6. Posterior estimates of τ^2 for April and June. Other details as for Fig. 2.

model prediction uncertainty was given almost no attention in the air quality model evaluation literature cited in Section 2.3. This was addressed by the goodness-of-fit diagnostics, which considered also the deviance. Two models that deliver similar median predictions but with different variances would not be distinguished by the validation diagnostics. If mapping the median value is the objective then both models will deliver similarly accurate results and the more parsimonious SVI would suffice.

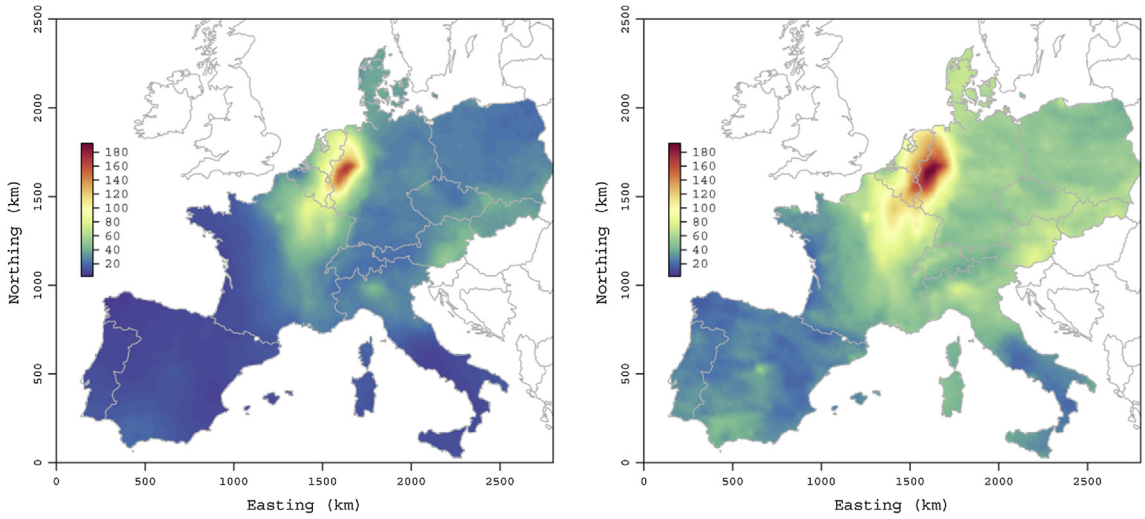
The flexible structure of SVC allowed the residual variability to be partitioned between the random intercept and random slope. This offered two additional benefits that SVI did not offer. First, it allowed exploration of the relationship between the observations and the CTM. Second SVC better represented the processes and yielded predictions with lower variance. These points are explored further below.

Maps of the spatially varying coefficient were used to identify geographic areas where the CTM performed relatively well or relatively badly. Of particular interest is $\beta_{LE} + w_{LE}(\mathbf{s})$ since this shows whether and where the CTM reproduces the variability in the observations. The intercept, $\beta_0 + w_0(\mathbf{s})$ adjusts for bias. A systematic SVC analysis over a longer time series would provide independent information on model performance, which can vary per region or per season. Hence, these maps could be used to target development of the CTM and associated emission databases and could be recreated for future versions of the CTM and used as part of its evaluation.

In this study, low or insignificant values of $\beta_{LE} + w_{LE}(\mathbf{s})$ were generally associated with low PM10 concentrations. There are several reasons why, at low PM10 concentrations, simulation of the actual pollution gradients is more difficult. First, during conditions with low PM10 concentrations, the uncertainty associated with the measurements (at least 15% (Hitzenberger et al., 2004)) is large relative to the spatial gradients. Second, the impact of station representativeness is lower during high pollution episodes. This is because secondary inorganic aerosols dominate the PM10 mass and the impact of primary pollutants, due to local emissions, are relatively small (Weijers et al., 2011). Evidence for this comes from the smaller relative difference between regional, urban background and traffic stations during high-pollution episodes. Third, CTMs generally are more accurate for secondary inorganic components, which have a relatively low contribution at low PM10 concentrations. During clean air conditions spatial gradients are normally low and determined by natural PM components such as dust, sea salt and organics which are associated with the largest uncertainties in their source strength and formation processes. The detailed chemical composition data for these components are few, limiting the possibility to validate the model. Finally, meteorological conditions during times of low PM10 concentration are characterized by wet and windy conditions. Simulation of precipitation amounts and timing of rainfall events remain challenging. The SVC offers a complementary approach to evaluate whether or not the model developments lead to more accurate characterization of background conditions, based on a large number of available PM10 measurements.

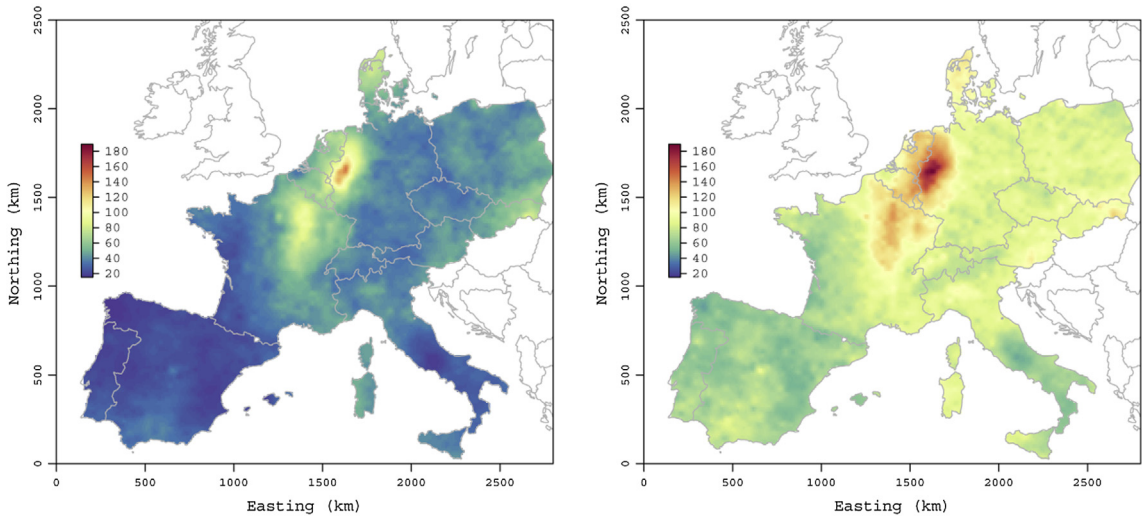
The model evaluation showed that SVC fitted the data better than SVI, but that the posterior median predictions were similar for both models. This was reflected in the final mapping which showed that the posterior median (kriged) surfaces were similar for both models but that the SVI generally had wider posterior prediction credible intervals. The fit, and hence the predictions were tighter for the SVC. This had a further consequence for the threshold exceedance maps, which were shown to be more accurate for SVC. The potential health impacts of PM10 were outlined in Section 1. The 50 μg m^{-3} is an EU limit set for health reasons. This is an important result when such maps are used for air quality assessment and management. The analysis could be extended to other limit values and other pollutants.

Classical geostatistics has also been used to calculate exceedance probabilities (e.g., Denby et al., 2008). Such approaches may underestimate the exceedance probability because they do not account for variability in the estimates of the variance parameters,



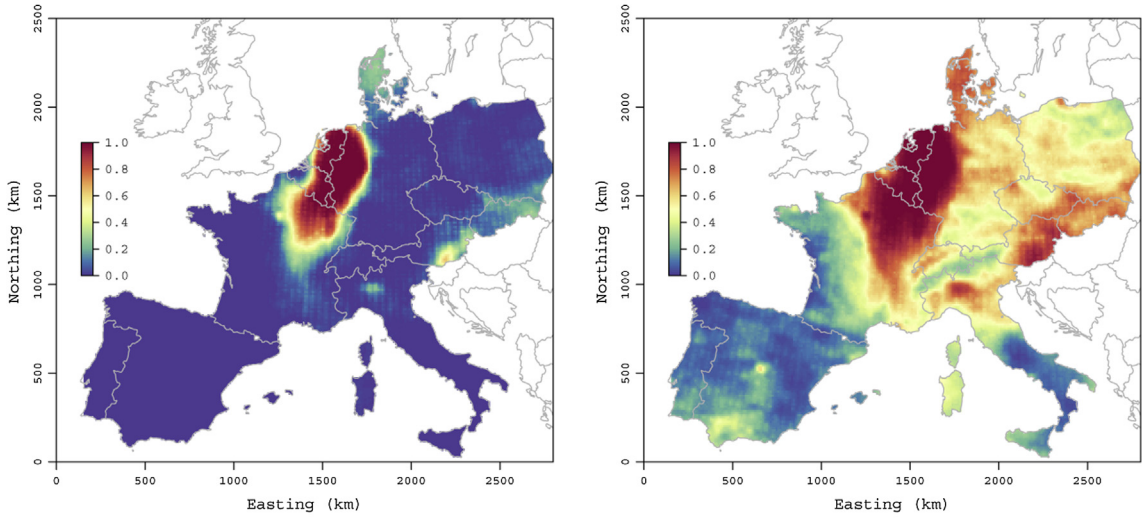
(a) Posterior median SVC

(b) Posterior median SVI



(c) Width 95% CI SVC

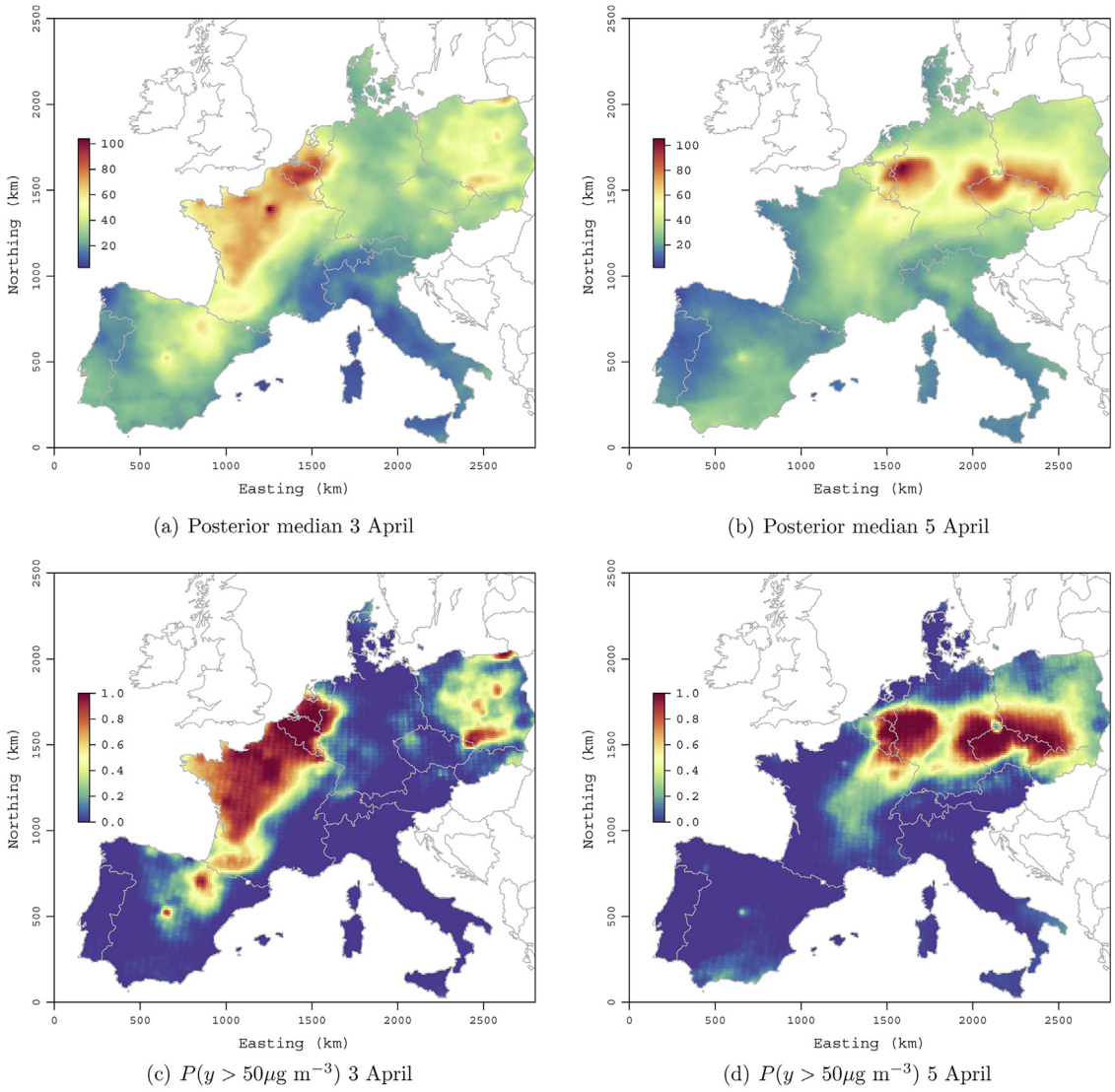
(d) Width 95% CI SVI



(e) $P(y > 50\mu g\ m^{-3})$ SVC

(f) $P(y > 50\mu g\ m^{-3})$ SVI

Fig. 7. Posterior median predicted (kriged) values, width of the 95% credible interval and probability of exceeding 50 μg m^{-3} for SVI and SVC for 13 April 2009. Units (a–d): μg m^{-3}.



(a) Posterior median 3 April

(b) Posterior median 5 April

(c) $P(y > 50\mu g\ m^{-3})$ 3 April

(d) $P(y > 50\mu g\ m^{-3})$ 5 April

Fig. 8. Posterior median predicted (kriged) values and probability of exceeding 50 μg m^{-3}. Units (a–b): μg m^{-3}.

$\boldsymbol{\theta} = (\tau^2, \sigma_0^2, \boldsymbol{\phi})$. In contrast, the Bayesian approach presented here does exactly that. Further, the authors are not aware of a comparable model where the parameters can be estimated using method-of-moments or maximum likelihood estimation. The implementation presented here took advantage of the flexibility of Bayesian hierarchical modelling.

Future research should be directed towards modelling a larger set of dates. Interest also lies in mapping other air quality components as well as joint mapping of multiple pollutants (e.g., PM10 and PM 2.5). Challenges include working with smaller data sets as well as the development of the joint model.

Finally, there are two data quality issues that have not been addressed, but which provide worthwhile and challenging subjects for future research. First, automatic measurement devices for PM10 often measure PM10 under different conditions (higher temperature and lower relative humidity) than ambient, which leads to variable losses of semi-volatile PM-components. In general, these effects are dependent on device, PM10 composition, ambient conditions and, thus, space and time. Correction factors are used to avoid systematic differences with the gravimetric reference method, although these are not always perfect. The complexity of sampling PM10 and the lack of a true references will cause additional variability that may violate the stationarity assumption. This has received limited attention in the literature (van de Kassteele et al., 2006) but should be addressed in future. Second, this paper did not account for the spatial support (resolution) of the simulations and the observations. The support of the simulations is nominally 25 km, but the support of the observations is unclear. Rural observations were deliberately chosen because they represent the ambient background conditions over a large area, but the actual size of this area is not clearly defined. This issue should receive future attention.

6. Conclusions

This paper presented a geostatistical spatially varying coefficients (SVC) model for modelling the relationship between in situ observations and the LOTOS-EUROS simulator output. This showed a better fit to the data than the standard spatially varying intercept (SVI) model. Both SVI and SVC gave a better fit than simple linear regression (SLR). SVC allowed exploration of the locations where the simulator performed well or poorly that was not possible with SVI. This gave new insight into the performance of the CTM and could be used for future CTM evaluation. A similar rationale could be applied to other models. SVC gave a

better fit to the data, whereas the accuracy for predicting the posterior median PM10 concentration at unknown locations was similar to SVI. SVC was preferred for mapping threshold exceedance, because its fit yielded narrower credible intervals. Exceedance probabilities relate both to the median and the variance. This was applied to estimate the probability of exceeding the 50 $\mu g\ m^{-3}$ EU daily PM10 concentration threshold. Hence such maps could be used to support air quality assessment and management, including those related to health impacts. It is concluded that SVC is a promising geostatistical method for coupling CTM output with in situ observations to map air quality at the European scale.

Acknowledgements

The authors acknowledge Arjo Segers (TNO) for his support with the LOTOS-EUROS CTM. Finley's work was supported by USA National Science Foundation grants DMS-1106609, EF-1137309, EF-1241874, and EF-1253225, as well as NASA Carbon Monitoring System grants.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atmosenv.2014.11.043.

References

Anderes, E.B., Stein, M.L., 2011. Local likelihood estimation for nonstationary random fields. J. Multivar. Anal. 102 (3), 506–520.
Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability 101. Chapman and Hall/CRC, London.
Banzhaf, S., Schaap, M., Kruit, R.J.W., van der Gon, H., Stern, R., Builtjes, P.J.H., 2013. Impact of emission changes on secondary inorganic aerosol episodes across Germany. Atmos. Chem. Phys. 13 (23), 11675–11693.
Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.Y., Kunzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Toumi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nador, G., Varro, M.J., Dedele, A., Grazuleviciene, R., Molter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Kraemer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Stromgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. Atmos. Environ. 72, 10–23.
Borrego, C., Monteiro, A., Ferreira, J., Miranda, A.I., Costa, A.M., Carvalho, A.C., Lopes, M., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. Environ. Int. 34 (5), 613–620.
Brauer, M., Amann, M., Burnett, R.T., Cohen, A., Dentener, F., Ezzati, M., Henderson, S.B., Krzyzanowski, M., Martin, R.V., Van Dingenen, R., van Donkelaar, A., Thurston, G.D., 2011. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. Environ. Sci. Technol. 46 (2), 652–660.
Brunekreef, B., Holgate, S.T., 2002. Air pollution and health. Lancet 360 (9341), 1233–1242.
Candiani, G., Carnevale, C., Finzi, G., Pisoni, E., Volta, M., 2013. A comparison of reanalysis techniques: applying optimal interpolation and ensemble Kalman filtering to improve air quality monitoring at mesoscale. Sci. Total Environ. 458–460 (0), 7–14.
Cressie, N.A.C., 1993. Statistics for Spatial Data, Revised Edition. John Wiley and Sons, New York.
Denby, B., Schaap, M., Segers, A., Builtjes, P., Horalek, J., 2008. Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. Atmos. Environ. 42 (30), 7122–7134.
Diggle, P.J., Ribeiro Jr., P.J., 2007. Model-based Geostatistics. Springer Verlag, New York.
Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay Jr., M.E., B., G.F., Speizer, F.E., 1993. An association between air pollution and mortality in six U.S. cities. N. Engl. J. Med. 329 (24), 1753–1759.
EEA, 2007. Air Pollution in Europe 1990–2004. European Environment Agency. Tech. rep., EEA Report 2/2007.
Finley, A.O., 2011. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. Methods Ecol. Evol. 2 (2), 143–154.
Finley, A.O., Banerjee, S., Gelfand, A.E., 2014. spBayes for large univariate and multivariate point-referenced spatio-temporal data models. J. Stat. Softw. preprint at http://arxiv.org/abs/1310.8192.
Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M.G., Stein, O., 2009. Coupling global chemistry transport models to ECMWF's integrated forecast system. Geosci. Model Dev. 2 (2), 253–265.
Gelfand, A.E., Ghosh, S.K., 1998. Model choice: a minimum posterior predictive loss approach. Biometrika 85 (1), 1–11.
Gelfand, A.E., Kim, Y.-J., Sirmans, C.F., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. J. Am. Stat. Assoc. 98 (462), 387–396.
Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, third ed. Chapman and Hall, London.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.
Hamm, N.A.S., Atkinson, P.M., Milton, E.J., 2012. A per-pixel, non-stationary mixed model for empirical line atmospheric correction in remote sensing. Remote Sens. Environ. 124, 666–678.
Haskard, K.A., Lark, R.M., 2010. Spectral tempering to model non-stationary variation of soil properties: sensitivity to the initial stationary model. Geoderma 159 (3–4), 350–357.
Hendriks, C., Kranenburg, R., Kuenen, J., van Gijlswijk, R., Kruit, R.W., Segers, A., van der Gon, H.D., Schaap, M., 2013. The origin of ambient particulate matter concentrations in the Netherlands. Atmos. Environ. 69, 289–303.
Hitzenberger, R., Berner, A., Galambos, Z., Maenhaut, W., Cafmeyer, J., Schwarz, J., Muller, K., Spindler, G., Wieprecht, W., Acker, K., Hillamo, R., Makela, T., 2004. Intercomparison of methods to measure the mass concentration of the atmospheric aerosol during INTERCOMP2000-influence of instrumentation and size cuts. Atmos. Environ. 38 (38), 6467–6476.
Lloyd, C.D., Atkinson, P.M., 2004. Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data. Int. J. Appl. Earth Obs. Geoinf. 5 (4), 293–305.
Loomis, D., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Baan, R., Mattock, H., Straif, S., 2013. The carcinogenicity of outdoor air pollution. Lancet Oncol. 14 (13), 1262–1263.
Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2013. The BUGS Book: a Practical Introduction to Bayesian Analysis. CRC Press, London.
Pernigotti, D., Gerboles, M., Belis, C.A., Thunis, P., 2013. Model quality objectives based on measurement uncertainty. Part II: NO_2 and PM10. Atmos. Environ. 79, 869–878.
Pope, C.A., Dockery, D.W., Schwartz, J., 1995. Review of epidemiological evidence of health effects of particulate air pollution. Inhal. Toxicol. 7, 1–18.
Pouliot, G., Pierce, T., van der Gon, H.D., Schaap, M., Moran, M., Nopmongcol, U., 2012. Comparing emission inventories and model-ready emission datasets between Europe and North America for the AQMEII project. Atmos. Environ. 53, 4–14.
Schaap, M., Timmermans, R.M.A., Roemer, M., Boersen, G.A.C., Builtjes, P., Sauter, F., Velders, G., Beck, J., 2008. The LOTOS-EUROS model: description, validation and latest developments. Int. J. Environ. Pollut. 32 (2), 270–290.
Solazzo, E., Bianconi, R., Pirovano, G., Matthias, V., Vautard, R., Moran, M.D., Wyat Appel, K., Bessagnet, B., Brandt, J., Christensen, J.H., Chemel, C., Coll, I., Ferreira, J., Forkel, R., Francis, X.V., Grell, G., Grossi, P., Hansen, A.B., Miranda, A.I., Nopmongcol, U., Prank, M., Sartelet, K.N., Schaap, M., Silver, J.D., Sokhi, R.S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S.T., Galmarini, S., 2012. Operational model evaluation for particulate matter in Europe and North America in the context of AQMEII. Atmos. Environ. 53, 75–92.
Spiegelhalter, D.J., Best, N.G., Carlin, B.R., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B-Stat. Methodol. 64, 583–616.
Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., Kerschbaumer, A., 2008. A model inter-comparison study focussing on episodes with elevated PM10 concentrations. Atmos. Environ. 42 (19), 4567–4588.
Thunis, P., Pederzoli, A., Pernigotti, D., 2012. Performance criteria to evaluate air quality modeling applications. Atmos. Environ. 59, 476–482.
Thunis, P., Pernigotti, D., Gerboles, M., 2013. Model quality objectives based on measurement uncertainty. Part I: Ozone. Atmos. Environ. 79, 861–868.
van de Kassteele, J., Koelemeijer, R.B.A., Dekkers, A.L.M., Schaap, M., Homan, C.D., Stein, A., 2006. Statistical mapping of PM10 concentrations over western Europe using secondary information from dispersion modeling and MODIS satellite observations. Stoch. Environ. Res. Risk Assess. 21 (2), 183–194.
van de Kassteele, J., Stein, A., 2006. A model for external drift kriging with uncertain covariates applied to air quality measurements and dispersion model output. Environmetrics 17 (4), 309–322.
Webster, R., Oliver, M., 2008. Geostatistics for Environmental Scientists, second ed. John Wiley and Sons, Chichester.
Weijers, E.P., Schaap, M., Nguyen, L., Matthijsen, J., van der Gon, H., ten Brink, H.M., Hoogerbrugge, R., 2011. Anthropogenic and natural constituents in particulate matter in the Netherlands. Atmos. Chem. Phys. 11 (5), 2281–2294.