Spatiotemporal calibration of atmospheric nitrogen dioxide concentration estimates from an air quality model for Connecticut

Owais Gilani · Lisa A. McKay · Timothy G. Gregoire · Yongtao Guan · Brian P. Leaderer · Theodore R. Holford

Received: 7 January 2019 / Revised: 17 September 2019 / Published online: 2 November 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
A spatiotemporal calibration and resolution refinement model was fitted to calibrate nitrogen dioxide (NO$_2$) concentration estimates from the Community Multiscale Air Quality (CMAQ) model, using two sources of observed data on NO$_2$ that differed in their spatial and temporal resolutions. To refine the spatial resolution of the CMAQ model estimates, we leveraged information using additional local covariates including total traffic volume within 2 km, population density, elevation, and land use characteristics. Predictions from this model greatly improved the bias in the CMAQ estimates, as observed by the much lower mean squared error (MSE) at the NO$_2$ monitor sites. The final model was used to predict the daily concentration of ambient NO$_2$ over the entire state of Connecticut on a grid with pixels of size 300 $\times$ 300 m. A comparison of the prediction map with a similar map for the CMAQ estimates showed marked improvement in the spatial resolution. The effect of local covariates was evident in the finer spatial resolution map, where the contribution of traffic on major highways to ambient NO$_2$ concentration stands out. An animation was also provided to show the change in the concentration of ambient NO$_2$ over space and time for 1994 and 1995.

Keywords Ambiant air pollution · CMAQ · Integrated exposure modeling · Kalman filter · Resolution refinement · SCARR model

Handling Editor: Pierre Dutilleul.

The authors thank Dr. Lance Waller for useful feedback on the manuscript, as well as the anonymous reviewers whose comments helped improve the manuscript. This research was partially funded by Grant R01ES017416 from the National Institutes of Health.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10651-019-00430-7) contains supplementary material, which is available to authorized users.

Owais Gilani owais.gilani@bucknell.edu

Extended author information available on the last page of the article
1 Introduction

Nitrogen dioxide (NO$_2$) is a highly reactive gas that contributes to the formation of ground-level ozone and fine particle pollution, and is believed to be associated with adverse respiratory health effects (US Environmental Protection Agency 2008). A detailed analysis of the effects of atmospheric pollutants such as NO$_2$ on various health outcomes requires access to data on the concentration of the pollutant on a fine spatial and temporal scale, which is rarely available. However, we often have data on the concentration of atmospheric pollutants for a given region and time period from different sources that differ in their spatial and temporal resolutions, as well as in their measurement accuracy.

Fixed site air quality monitoring stations, such as the US Environmental Protection Agency’s (EPA) monitoring stations (US Environmental Protection Agency 2011), record pollutant concentration data on a dense temporal scale (hourly), but the network of monitoring sites is generally spatially very sparse (e.g. only four sites in Connecticut (CT)), which doesn’t allow for accurate modeling at sites far away from the monitoring sites. On the other hand, data collected at many different spatial locations using passive sampling as part of environmental epidemiological studies, such as the Acid/Aerosol study (Triche et al. 2002), generally provide an aggregate measure of the pollutant concentration over relatively long time periods (1–2 weeks), resulting in spatially dense but temporally sparse data. Such data sources do not allow accurate estimation of pollutant concentrations at a fine temporal scale.

In the absence of a single source of observed data on the concentration of a pollutant that is both spatially and temporally dense, deterministic meteorological air quality models, such as the Community Multiscale Air Quality (CMAQ) model (Byun and Schere 2006; Hogrefe et al. 2009; US Environmental Protection Agency 2014), provide an alternative source of pollutant concentration. Predictions from such models are provided either at the centroids of pixels or as an aggregate measure over the pixel on a regular square grid format, which generally span an extended spatial domain on a dense temporal scale (hourly or daily). However, the grid-cell or pixel sizes are often fairly large (typically 12 km $\times$ 12 km), providing crude spatial resolution. Additionally, these complex models do not use any observed measurements of the pollutant in the modeling process, and can often have significant bias associated with them. To account for these potential biases, various spatiotemporal modeling techniques have been developed that seek to calibrate output from such deterministic models using observed data on the pollutants (Meiring et al. 1998; Brown et al. 2001; Fuentes and Raftery 2005; Li et al. 2008; Berrocal et al. 2010; Alkuwari et al. 2013; Chang et al. 2014).

Most spatiotemporal calibration methods require temporal alignment between the observed data source and output from the deterministic model that needs to be calibrated, in addition to a somewhat dense spatial network of observed sites (Meiring et al. 1998; Brown et al. 2001; Li et al. 2008). Additionally, these models do not address the issue of improving the spatial resolution of the large pixel sizes of deterministic model outputs, known as the “change of support” problem (Cressie 1993). While these methods work well when the pixel sizes are smaller, or if the process being modeled does not exhibit large variability over short distances, they are inadequate in modeling a process such an NO$_2$, which is known to vary considerably over short distances (Jer-
Gilani et al. (2016) developed a two-step modeling strategy, the Spatiotemporal Calibration and Resolution Refinement (SCARR) model, that allows calibrating estimates of a pollutant from a deterministic air quality model available in the form of grid-cell data, while also refining its spatial resolution, using two different sources of measured data that differ in their spatial and temporal resolutions. The modeling strategy was demonstrated by developing a space-time model using partial observations from three sources of data on the concentration of ambient NO₂ over Connecticut in 1994, and its performance was tested using the remaining observations. Additionally, for simplicity in the demonstrative example, the first step of the model was developed as a purely spatial model, without accounting for season as a predictor in the model.

In this paper, we extend the SCARR model and fit it to the same data sources using the complete set of observations to develop a space-time model to estimate the concentration of NO₂ at a fine spatial and temporal resolution over the state of Connecticut for 1994 and 1995. Specifically, estimates of NO₂ from the CMAQ model available in a grid-cell format with relatively large pixel sizes (12 × 12 km) are calibrated in space and time while also refining their spatial resolution using observations from two sources (Acid/Aerosol epidemiologic study data (Triche et al. 2002), and US Environmental Protection Agency monitoring data (US Environmental Protection Agency 2011) measured at different spatial and temporal resolutions. In this analysis, the SCARR model is extended in three ways: (a) the first step of the model is developed as a space-time model instead of a purely spatial model; (b) a parameter is included in the second step of the model that controls the influence of the estimated spatiotemporal calibration bias from the first step; and (c) additional covariates potentially correlated with atmospheric NO₂ are included. The model is then used to predict the daily concentration of ambient NO₂ for 1994 and 1995 over the entire state of Connecticut on a grid with a pixel size of 300 × 300 m.

The remainder of the paper is organized as follows. Section 2 provides a description of the three different sources of available data on concentrations of NO₂, and the additional local covariates included in the model. Section 3 gives details on the two-step SCARR modeling strategy. Results for the fitted model are given in Sect. 4, while Sect. 5 provides predictions of NO₂ for CT for 1994 and 1995 obtained from the fitted SCARR model. Finally, Sect. 6 provides some discussion and directions for future work.

2 Data

2.1 Sources of data on NO₂ concentration

Data on the outdoor concentration of NO₂ for the state of Connecticut (CT) for 1994 and 1995 are available from three different sources—predictions from the Community Multiscale Air Quality (CMAQ) model on a grid-cell format, and observed data from
the Acid/Aerosol study and from EPA monitoring sites, both measured at different spatial and temporal resolutions.

2.1.1 CMAQ model data \( (\hat{Y}_1(s, t)) \)

Data from the Community Multiscale Air Quality (CMAQ) model version 4.7.1 were provided by the Atmospheric Sciences Research Center in Albany, New York (Byun and Schere 2006; Hogrefe et al. 2009; US Environmental Protection Agency 2014). The model uses data from a meteorological forecast model, source emission inventories and chemistry transport modeling to predict hourly NO\(_2\) concentration on a regular grid over CT with each pixel of size 12 km \( \times \) 12 km. These data have an extensive spatial coverage (over the entire state of CT) and are temporally dense, but provide estimates at the centroids of pixels with rather large sizes. Additionally, these estimates have not been calibrated to actual observed measurements of NO\(_2\), and have systematic errors associated with them. The umlaut (¨) above \( Y_1 \) indicates that these data provide estimates at location \( s \) and time \( t \) instead of measured observations.

2.1.2 Acid/Aerosol study data \( (Y_2(s, [t_s])) \)

In the Acid/Aerosol study, 138 families were recruited from mothers delivering babies at seven Connecticut hospitals between 1993 and 1996 (Triche et al. 2002). Of these, 129 families had outdoor NO\(_2\) concentrations measured at their residences by passive sampling using Palmes Tubes (Palmes et al. 1976). At the enrollment home visit, the NO\(_2\) monitoring tube was placed in an inverted funnel-shaped metal weather protector and hung from a tree branch or outdoor clothes line at least 5 ft above the ground and as close to the home as possible. The monitor was left in place for 10–14 days, and the cumulative concentration during that period was recorded. Point locations for the residences were obtained by geocoding each address against ESRI’s® Streets USA database (ESRI 2003); geocoding was unsuccessful for five locations, while two samples were excluded due to equipment contamination. The final analysis utilized 122 (94.6%) samples, all of which were collected at various times between March and December, 1994. These data are spatially dense but temporally sparse as they provide measurements for each residence only once a year, aggregated over a 10–14 day period. The index \([t_s]\) for \( Y_2 \) indicates that data were observed over different lengths of time and at different time points for different Acid/Aerosol locations, \( s \).

2.1.3 EPA data \( (Y_3(s, t)) \)

The US Environmental Protection Agency (EPA) monitors atmospheric NO\(_2\) levels over four locations in CT (Bridgeport, East Hartford, New Haven, and Tolland) and two locations in southern Massachusetts (MA) (Chicopee and Springfield) on an hourly basis (US Environmental Protection Agency 2011), which provide a rather accurate estimate of ambient NO\(_2\) at these locations. While these data are temporally dense, they are spatially very sparse, with only six locations in Connecticut and southern Massachusetts. For each of the six EPA sites, the 24-h mean daily NO\(_2\) concentration was calculated for the 2 years.
The spatial distribution of the Acid/Aerosol and EPA sites with the CMAQ grid overlaid is given in Fig. 1a, while Fig. 1b shows the variation in NO$_2$ concentration over a 30 day period for the EPA monitor in New Haven, CT, March/April 1994. EPA data are missing for some time points, while Acid/Aerosol data provide a 13 day average measurement.

Fig. 1 Spatial and temporal distribution of study data

The spatial distribution of the Acid/Aerosol and EPA sites with the CMAQ grid overlaid is given in Fig. 1a, while Fig. 1b shows the variation in NO$_2$ concentration over a 30 day period for the EPA monitor in New Haven, CT along with the CMAQ estimate and observations from a nearby acid/aerosol site for March/April 1994 (reproduced from Gilani et al. (2016)).

2.2 Variables used for model fitting

2.2.1 CMAQ predictions at Acid/Aerosol and EPA sites—($\tilde{Y}_1(s, t)$)

At each Acid/Aerosol site, the closest CMAQ pixel centroid was identified, and data were extracted for the exact hours that NO$_2$ concentrations were measured at that
Acid/Aerosol site. A 24-h mean CMAQ concentration was then calculated for each of those 10–14 days. Similarly at the EPA sites, the closest CMAQ pixel centroid was identified, and the hourly concentrations were extracted for the 730 days in 1994 and 1995 and converted into 24-h averages.

2.2.2 Traffic measurements

The Department of Transportation for Connecticut, Massachusetts and New York record annual traffic volume in the form of average daily traffic (ADT) for inter-states and numbered highways (Connecticut Department of Transportation 2001; Massachusetts Department of Transportation - Highway Division (2012); New York State Department of Transportation 2012). Following the approach outlined by Holford et al. (2010) and Skene et al. (2010) and using traffic data for 1994 and 1995 provided by the Connecticut, Massachusetts and New York Departments of Transportation, we divided a line file of interstates and numbered highways into approximately 50-m segments, each of which had an associated ADT count. We defined the midpoint of each segment and calculated a measure of traffic volume (TV) on that segment as the product of segment length and ADT. The contribution of a segment to NO$_2$ concentration at any location can be expressed as the product of TV and a dispersion function of distance and direction between the segment midpoint and the site. Holford et al. (2010) found that the dispersion function can be effectively estimated by a step-function with steps of distance 0–0.5 km, 0.5–1 km, 1–2 km, 2–3 km, 3–4 km, 4–5 km and 5–6 km. Beyond 6 kms, the effect of vehicular traffic on NO$_2$ was not statistically significant (Holford et al. 2010; Skene et al. 2010).

To estimate the dispersion function, circular buffers of radii 500 m, 1 km, 2 km, 3 km, 4 km, 5 km and 6 km were created around each Acid/Aerosol and EPA site, and the total traffic volume (TTV) within each concentric buffer ring was calculated by summing the contribution of all point sources within the buffer ring and dividing by 10,000. This gave a measure of 10,000 vehicle-kilometers per day for a given distance range—e.g., a total traffic volume value of 3 at an Acid/Aerosol site for a buffer ring of 0.5–1 km would indicate an average of 30,000 vehicle-kilometers traveled per day within that buffer ring.

2.2.3 Land use covariates

Land use data for Connecticut were obtained from the United States Geological Survey (USGS) ‘National Land Cover Dataset’ (NLCD) for the year 1992 (US Geological Survey 2012). The data were stored as a raster file with 30-m pixels and classified into 17 categories: open water, low intensity residential, high intensity residential, commercial/industrial, bare rock, quarries, transitional barren, deciduous forest, evergreen forest, mixed forest, shrubland, orchards, pastures, row crops, urban/recreational grasses, woody wetlands and emergent herbaceous wetlands.

For each Acid/Aerosol site, we used the “Intersect Points with Raster” tool from the Geostatistical Modeling Environment (Beyer 2012; R Core Team 2013) to count the number of pixels of each land use category within circular buffer rings of size 0–0.5 km, 0.5–1 km and 1–2 km. This total was then multiplied by 900 (area in m$^2$ of a 30
m pixel) and divided by 10,000 to give the area in hectares of each land use category (LUC) within the three buffer rings.

2.2.4 Population density

The 1990 mid-year census tract population and polygon shape file for census tracts was obtained from the US Census Bureau (US Census Bureau 1990, 2012). Population density (per square mile) was assigned to each residence as the mid-year population of the census tract divided by the area of the census tract in square miles.

2.2.5 Elevation

Elevation above sea level, in meters, for each residence was extracted as the raster cell value from the USGS ‘National Map’ for the year 2005 (Gesch 2007; Gesch et al. 2002).

2.2.6 Season

Concentrations of NO\textsubscript{2} follow a seasonal cycle, and are generally higher in the winter months as compared to the summer. To capture the effect of season, a trigonometric and linear function of date was included in the model. Day of the year ratio (DYR) was defined as the midpoint of the days that NO\textsubscript{2} was monitored at an acid/aerosol residence divided by 365, and four covariates were calculated: \( \sin(2 \cdot \pi \cdot \text{DYR}) \), \( \cos(2 \cdot \pi \cdot \text{DYR}) \), \( \sin(4 \cdot \pi \cdot \text{DYR}) \), \( \cos(4 \cdot \pi \cdot \text{DYR}) \).

3 Model

As outlined by Gilani et al. (2016), the spatiotemporal calibration and resolution refinement (SCARR) model to calibrate the predictions of the concentration of ambient NO\textsubscript{2} from the CMAQ model and to improve its spatial resolution is developed in two steps. Calibration here refers to estimating the systematic additive and multiplicative “bias” present in the CMAQ data so that it can be accounted for. The first step, Calibration and Spatial Refinement, uses the spatially dense but temporally sparse Acid/Aerosol data along with other publicly available data on various local covariates related to the modification and dispersion of NO\textsubscript{2} to calibrate the CMAQ predictions over space and time, while also improving their spatial resolution. This step provides a continuous space representation of the artificially discrete CMAQ data, and estimates the additive and multiplicative calibration bias of the CMAQ data. The second step, Spatiotemporal Calibration using Dynamic Space-Time modeling, improves the temporal resolution of Step I by using the temporally dense but spatially sparse EPA data for estimating the temporal evolution of the additive and multiplicative calibration constants at a finer temporal resolution (daily). We think of the spatial variation in pollutant concentration as consisting of two components—large scale spatial variation over the region \( S \) of interest as well as small scale variation (i.e. there is a global trend and a local trend). The SCARR model assumes that the CMAQ data capture the large scale variation but
not the small scale variation. Based on the inputs that go into the CMAQ model and the scale at which the model is run, this is a reasonable assumption. The SCARR model also assumes that the small scale spatial variation in pollutant concentration is due to local factors such as land use type, traffic density, population density, elevation, and that this small scale spatial variability is similar between the shorter time interval (1 day) of the EPA data and the longer time durations (10–14 days) of the Acid/Aerosol data.

3.1 Step I

We fitted a model to calibrate and refine the granularity of the daily mean CMAQ NO$_2$ concentration using acid/aerosol NO$_2$ data via an Integrated exposure modeling (IEM) approach (Holford et al. 2010). Exposure to a pollutant at a location $s$ is often due to an accumulation of pollutant concentration from multiple nearby sources, including point, line, and area sources. The cumulative exposure at location $s$ from a line source of exposure (such as a highway) can be represented as $\int_{L} h \phi(d) dd$, where $h$ is the pollutant concentration along line segment $L$, and $\phi(d)$ is a dispersion function, which depends on the distance $d$ to the source. Similar expressions for point and area sources are given in Gilani et al. (2016). In our case, we model the dispersion function as a step function with $k$ steps, and estimate it by writing $\phi(d) = g \lambda$, where $g = (g_1, \ldots, g_k)$, and $g_1, \ldots, g_k$ are indicator functions defining the steps. The vector of unknown parameters $\lambda$ can be estimated by including $G = \int h g$ as a regressor in a generalized linear mixed model fitted to observed concentrations of the pollutant. Let $Y_2(s, t)$ be the concentration of NO$_2$ from the Acid/Aerosol data observed at location $s$ averaged over the 10–14 day time interval $[t_s]$; $X$ be a matrix in which each column represents a covariate not related to dispersion, with an additional column of ones for the intercept; $G$ be a matrix in which each column represents a dispersion related covariate (i.e. TTV and land use categories (LUC)) for each step of the step function (as described above); $\hat{Y}_1(s, t)$ be the estimate from the CMAQ data at the pixel centroid nearest to location $s$. Then, a spatiotemporal calibration and refinement model, as described by Gilani et al. (2016), is specified by:

$$
Y_2(s, t) = \int_{[t_s]} (X(s, t) \beta + G(s, t) \lambda + \gamma \hat{Y}_1(s, t)) \, dt = \ ||[t_s]|| + v(s, t) + \epsilon(s, t),
$$

where $v(s, t)$ are independent mean zero Gaussian random errors. Residual spatiotemporal dependence between the sites is modeled using the hierarchical error term, $v(s, t)$, where $v \sim \text{MVN}(0, \sigma^2 \Sigma)$. For the covariates $X$ not related to dispersion, Gilani et al. (2016) included only one variable, population density, in their model. In our analysis, we further include elevation, as well as a trigonometric function of time (as described in Sect. 2.2) to capture the seasonal variation in NO$_2$ concentrations. In this setup, the terms in the mean function not related to the CMAQ data $\hat{Y}_1$, i.e. $X(s, t) \beta + G(s, t) \lambda$, form the spatiotemporally varying additive calibration bias for the CMAQ data, while $\gamma$ is the multiplicative calibration bias.

Given the temporal sparsity of the Acid/Aerosol data, it was difficult to accurately estimate a space-time covariance matrix $\Sigma$. Additionally, aggregation over the 10–14
day time interval accounted for the short-range temporal dependence at a given site \( s \). We therefore assumed that most of the residual dependence between sites was spatial in nature. We considered both spatially dependent and independent error models. For the spatially dependent error models, spherical, exponential and Matérn covariance functions were used to model \( \Sigma \), whereas for the spatially independent error model, the hierarchical error term was removed and a linear regression framework was utilized.

The initial model included all the variables, and the traffic covariates were selected by a backward approach, giving preference to buffer rings closer to the residences. For example, to include the TTV for the buffer 1.0–2.0 km, the buffers 0–0.5 km and 0.5–1.0 km must also be included in the model. This strategy was used to provide estimates of the step dispersion function at all distances included in the model. The same strategy was adopted for land use categories. Nested traffic and land use buffer models were compared using \( F \)-tests, and buffer variables not contributing significantly to the model (at the 5% level) were removed. Leave-one-out cross-validation was also performed on the best few models, and preference was given to models for which the prediction error sum of squares (PRESS) was closer to the error sum of squares.

### 3.2 Step II

We fitted a dynamic space-time model for temporal calibration of the spatially refined CMAQ estimates from Step I. The assumption that the small scale spatial variability is similar between the shorter time interval of the EPA data and the longer time duration of the Acid/Aerosol data allows us to generalize the results of Step I to calibrate estimates similar between the shorter time interval of the EPA data and the longer time duration of CMAQ estimates from Step I. The assumption that the small scale spatial variability is

We therefore assumed that most of the residual dependence between sites was spatial in nature. We considered both spatially dependent and independent error models. For the spatially dependent error models, spherical, exponential and Matérn covariance functions were used to model \( \Sigma \), whereas for the spatially independent error model, the hierarchical error term was removed and a linear regression framework was utilized.

The initial model included all the variables, and the traffic covariates were selected by a backward approach, giving preference to buffer rings closer to the residences. For example, to include the TTV for the buffer 1.0–2.0 km, the buffers 0–0.5 km and 0.5–1.0 km must also be included in the model. This strategy was used to provide estimates of the step dispersion function at all distances included in the model. The same strategy was adopted for land use categories. Nested traffic and land use buffer models were compared using \( F \)-tests, and buffer variables not contributing significantly to the model (at the 5% level) were removed. Leave-one-out cross-validation was also performed on the best few models, and preference was given to models for which the prediction error sum of squares (PRESS) was closer to the error sum of squares.

### 3.2 Step II

We fitted a dynamic space-time model for temporal calibration of the spatially refined CMAQ estimates from Step I. The assumption that the small scale spatial variability is similar between the shorter time interval of the EPA data and the longer time duration of the Acid/Aerosol data allows us to generalize the results of Step I to calibrate estimates of the CMAQ data on a finer temporal resolution in Step II using the temporally dense EPA data. Let \( s_1, \ldots, s_6 \) represent the spatial location of the six EPA monitor sites. The estimated spatiotemporal additive calibration bias \( \tilde{C}(s_i, t) \) at location \( s_i, i = 1, \ldots, 6 \), from Step I is defined as \( \tilde{C}(s_i, t) = X(s_i, t)\hat{\beta} + G(s_i, t)\hat{\lambda} \), where \( \hat{\beta} \) and \( \hat{\lambda} \) are vectors of parameters estimated in the first step. Note that the tilde (\( \tilde{\cdot} \)) on \( C \) signifies that this variable was calculated using parameters estimated in Step I of the model. We further define \( Y_3(t) = (Y_3(s_1, t), \ldots, Y_3(s_6, t))^\prime \); \( \tilde{Y}_1(t) \) a diagonal matrix with \( \text{diag}(\tilde{Y}_1(t)) = (\tilde{Y}_1(s_1, t), \ldots, \tilde{Y}_1(s_6, t)) \); and \( \tilde{C}(t) = (\tilde{C}(s_1, t), \ldots, \tilde{C}(s_6, t))^\prime \).

The observation equation for a dynamic state-space model describing the relationship between \( Y_3(t) \) and \( \tilde{Y}_1(t) \) in Gilani et al. (2016) is extended by adding a parameter \( \beta_c \) (constant over space and time), which controls the influence of \( \tilde{C}(t) \) in the temporal calibration of Step II. The observation equation of the state-space model is thus given by

\[
Y_3(t) = A(t) + \beta_c \tilde{C}(t) + \gamma \tilde{Y}_1(t) + Z(t), \quad t = 1, 2, \ldots, 730, \tag{2}
\]

where \( Z(t) = (Z(s_1, t), \ldots, Z(s_6, t))^\prime \) is a multivariate white noise time series with the \( Z(s_i, t), i = 1, \ldots, 6 \) mutually independent \( N(0, \sigma_Z^2(s)) \) variates. \( A(t) = (A(s_1, t), \ldots, A(s_6, t))^\prime \) is a stochastic process whose evolution over time is described by the state equation

\[
A(t) - \mu_A = \Psi_A(A(t-1) - \mu_A) + F\xi(t), \tag{3}
\]
where \( \xi(t) = (\xi(s_1, t), \ldots, \xi(s_6, t))^T \) is a multivariate stochastic process with \( \xi(s_i, t), i = 1, \ldots, 6 \), mean zero Gaussian variates with variance \( \sigma^2_A(s) \). \( \Psi_A \) is a diagonal matrix with \( \text{diag}(\Psi_A) = (\psi_A(s_1), \ldots, \psi_A(s_6)) \), where \( \psi_A(s_i), i = 1, \ldots, 6 \) are autoregressive parameters lying in the range 0–1.

In the model described above, \( A(t) + \beta_c \tilde{C}(t) \) provides the additive calibration bias on the finer temporal scale, while \( \gamma \) estimated in Step I, provides the multiplicative calibration bias. The matrix \( F \) in Eq. 3 can be used to model the spatial correlation in the additive bias to arrive at an integrated space-time model. However, for this example, with only six locations spread over relatively large distances, accurately estimating the spatial correlation structure of \( A(t) \) is not possible. We therefore assumed spatial independence for \( A(s_i, t) \) between the six sites. Additionally, we assumed that the six sites have identical parameters \( \sigma^2_Z, \sigma^2_A, \psi_A, \text{ and } \mu_A \), allowing us to pool across the six sites to estimate \( A(t) \) common to the entire study region (estimates for these parameters from fitting this model individually to the six sites were not very different, providing some evidence that this was a reasonable assumption). In this setting, \( \beta_c \tilde{C}(s, t) \) captures the spatial and two-week temporal variability in the additive calibration bias, while \( A(t) \) provides the finer-scale temporal evolution of the additive calibration bias that is common for all sites. Under these assumptions, Eqs. 2 and 3 can be rewritten in vector notation as

\[
\begin{align*}
Y_3(t) &= A(t)I_6 + \beta_c \tilde{C}(t) + \gamma \tilde{Y}_1(t) + \sigma_ZZ(t)I_6 \quad (4) \\
A(t) - \mu_A &= \psi_A(A(t-1) - \mu_A) + \sigma_A\xi(t) \quad t = 1, 2, \ldots, 730, \quad (5)
\end{align*}
\]

where \( Z(t) \) and \( \xi(t) \) are standard Gaussian variates, \( I \) a vector of 1’s and \( I \) the identity matrix.

The dynamic space-time model outlined in Eqs. 4 and 5 was fitted to data using the Kalman filter (Kalman 1960) (see Gilani et al. 2016, for details). Both maximum likelihood and Bayesian techniques using Markov Chain Monte Carlo (MCMC) simulations were utilized to estimate the parameters \( \sigma^2_Z, \sigma^2_A, \psi_A, \text{ and } \beta_c \) using the package \( \text{dlm} \) (Petris 2010) in \( \text{R} \) statistical package. Estimates produced by the two methods for the full model were quite similar, and due to the faster computational speed of maximum likelihood estimation (MLE) in this case (less than a second using MLE as compared to about 18 min using MCMC to fit the model and predict at one new location), further model fitting and prediction was conducted using MLE. However, we note that the performance of MCMC algorithms in terms of speed depends on a number of factors, and these should be considered before choosing between the two estimation approaches.

To test the fit of the two-step SCARR model, we calculated the coefficient of correlation, the empirical mean squared error (MSE), the empirical mean absolute error (MAE), and the relative bias (rBIAS) for each of the six sites, comparing the EPA observations with predictions from the SCARR and CMAQ models (calculation details are given in Table 4). We also fit the final SCARR model at the 122 Acid/Aerosol locations (details on fitting the full model at a new location are given in Sect. 5) and calculated the mean square prediction error (MSPE) for the SCARR model predictions and the CMAQ model estimates at these sites. A plot of the standardized residuals
versus fitted values and a $QQ$ plot were used to check for the normality assumption, while an autocorrelation plot of the residuals over various temporal lags was used to test for the AR(1) assumption for $A(t)$. Convergence status in the model fitting function in the package $dlm$ was verified, while multiple initial values for the model parameters were tested to ensure global convergence.

4 Results

4.1 Step I

Table 1 gives a summary of all covariates used to develop the spatial calibration and refinement model. Figure 2 shows the distribution of ADT for interstates and numbered highways for 1994 in Connecticut, while Figs. 3 and 4 show the geographic distribution of population density for 1990 and elevation in 1992, respectively.

We explored the shape of the dispersion function relating NO$_2$ to total traffic volume by fitting a model using just the TTV buffer covariates. The estimated dispersion function is given in Fig. 5a. To check for small-scale spatial anisotropy in the dispersion of NO$_2$ related to traffic, we divided the traffic buffers into four quadrants and calculated the total traffic volume within each quadrant for each buffer ring. Fig. 5b shows the step dispersion function in each direction, estimated by fitting the directional traffic buffer covariates in separate models for each direction. The shape of the four directional dispersion functions appear very similar to each other and to the estimated isotropic dispersion function in Fig. 5a, suggesting small scale spatial isotropy in the dispersion of NO$_2$. We therefore used the omnidirectional total traffic volume covariates in developing the model.

Based on previous results (Skene et al. 2010) and exploratory analysis, we reclassified the land use categories into developed (including low intensity residential, high intensity residential and commercial/industrial), forest (including deciduous, evergreen and mixed forests, pastures, row crops, and urban/recreational grasses), and other (all other categories) (see supplementary material for details). Figure 6 shows the spatial distribution of the reclassified land use categories for Connecticut in 1992.

An initial analysis of the land use categories displayed a high degree of negative correlation between the “developed” and “forest” categories (e.g. $-0.92$ between “developed 0.5–1 km” and “forest 0.5–1 km” rings). Additionally, within each of the three categories, we observed very high positive correlations between the buffer rings (e.g. $0.91$ between 0–0.5 km and 0.5–1 km “forest” rings). This raised the concern of multicollinearity if all the land use category covariates were included in the model together. In fact, inclusion of all of these covariates in the model resulted in unstable and unusual parameter estimates. Due to the high degree of correlation between the different buffer rings within the categories, the buffer rings for each category were combined to derive a single covariate for each category from 0 to 2 km. The new categories “developed 0–2 km” and “forest 0–2 km” had a correlation coefficient of $-0.90$, and including both categories in the model did not provide significant improvement in the model as compared to including just one covariate. Including
Table 1  Summary statistics for Step I model covariates

| Variable                                           | Mean (SD)* | Range          |
|----------------------------------------------------|------------|----------------|
| Acid/Aerosol NO$_2$ (ppb)$^a$                      | 13.6 (5.28) | 4.39–33.1      |
| CMAQ NO$_2$ (ppb)$^b$                             | 10.6 (4.75) | 2.51–24.0      |
| Traffic density (traffic volume/km$^2$)            |            |                |
| Buffer ring (km)$^c$                               |            |                |
| 0.0–0.5                                            | 0.89 (1.92) | 0.00–12.5      |
| 0.5–1.0                                            | 1.00 (1.52) | 0.00–8.00      |
| 1.0–2.0                                            | 1.14 (1.31) | 0.00–5.95      |
| 2.0–3.0                                            | 0.99 (1.02) | 0.02–4.40      |
| 3.0–4.0                                            | 1.11 (1.08) | 0.05–5.38      |
| 4.0–5.0                                            | 0.91 (0.71) | 0.06–3.12      |
| 5.0–6.0                                            | 0.89 (0.62) | 0.05–2.50      |
| Land use density (hectares/km$^2$)                 |            |                |
| Category/Buffer ring (km)$^d$                       |            |                |
| Developed$^e$                                       |            |                |
| 0.0–0.5                                            | 19.89 (15.82) | 0.17–49.27   |
| 0.5–1.0                                            | 53.07 (42.34) | 0.40–146.7    |
| 1.0–2.0                                            | 186.65 (141.35) | 2.64–558.3   |
| Forest$^f$                                         |            |                |
| 0.0–0.5                                            | 8.39 (5.01)  | 0.08–15.91     |
| 0.5–1.0                                            | 26.78 (13.88) | 0.55–48.89    |
| 1.0–2.0                                            | 113.99 (47.69) | 8.54–193.9    |
| Other$^g$                                          |            |                |
| 0.0–0.5                                            | 0.41 (0.47)  | 0.00–3.07      |
| 0.5–1.0                                            | 1.38 (1.39)  | 0.12–10.47     |
| 1.0–2.0                                            | 5.95 (5.67)  | 0.50–42.84     |
| Population density (population/mi$^2$)             | 1682 (2284) | 86.3–11529     |
| Elevation (m)                                      | 106 (68.5)  | 1.60–331.57    |

$^a$ N = 122 for all covariates

$^a$ 24 h mean NO$_2$ concentration sampled outside 122 homes over a 10–14 day period

$^b$ 24 h mean NO$_2$ concentration estimated from the CMAQ model for the same 10–14 day period as the Acid/Aerosol residence

$^c$ Total daily traffic volume (in 10,000 vehicle-km) divided by area (in km$^2$) of buffer ring

$^d$ Total hectares of land use category divided by area (in km$^2$) of buffer ring

$^e$ “Developed” land use category includes low intensity residential, high intensity residential, and commercial/industrial

$^f$ “Forest” land use category includes deciduous, evergreen and mixed forests, pastures, row crops, and urban/recreational grasses

$^g$ “Other” land use category includes open water, bare rock, quarries, transitional barren, shrubland, orchards, woody wetlands and emergent herbaceous wetlands
these two covariates separately in the model, while keeping all other covariates fixed, provided almost exactly the same parameter estimates, with the sign reversed. We, therefore, arbitrarily retained “forest 0–2 km” in the final model and excluded “devel-
Fig. 4 Elevation (meters) in CT (Gesch 2007; Gesch et al. 2002)

oped 0–2 km”. The “other 0–2 km” land use category did not significantly improve the model in the presence of “forest 0–2 km”, and was therefore not included in the final model.

We fit a spatially independent error model using a linear regression framework to identify the best subset of variables to include in the model. Omnidirectional and four-directional variograms of the residuals (not shown) revealed no evidence of spatial correlation in the residuals, which justified the assumption of spatial independence of the errors. However, we fitted spatially dependent error models as well. The covariates from the final regression model were included in a spatially dependent error model with the spherical, exponential, and Matérn covariance functions to model the spatial dependence in the errors using PROC MIXED in SAS 9.3 (SAS Institute, Cary, NC). These models provided similar parameter estimates as the linear regression model, arriving at the maximum likelihood when the estimated covariance parameters provided an effective range of zero.

In the presence of the other covariates, elevation did not appear to improve the fit of the model. CMAQ NO$_2$, population density and the trigonometric and linear function of date were statistically significant, and were included in the final model. None of the TTV buffer rings further than 2 km of the Acid/Aerosol sites were statistically significant, and were removed from the model. A summary of the covariates in the final model and their parameter estimates is given in Table 2, while Fig. 7 shows the observed concentrations of NO$_2$ from the Acid/Aerosol data, along with the estimated trigonometric and linear function of date.
Fig. 5 Estimated (a) isotropic and (b) anisotropic step dispersion functions for total traffic volume (TTV) based on distance $d$ and direction from a residence

4.2 Step II

Table 3 provides summary statistics for NO$_2$ concentrations at the six sites for the EPA and CMAQ data, along with the correlation between these two sources of data at each location. It also gives the estimate of the spatiotemporal additive bias $\tilde{C}(s, t)$ at each site, calculated using parameter estimates from Step I. The EPA site at Springfield had observations for all 730 days, while Tolland had observations only for 369 days between October 1994 and November 1995. The other four sites had EPA data missing for a few days, with the available data ranging between 697 and 727 days. CMAQ data generally underestimates NO$_2$ concentrations, with the difference most pronounced.
Fig. 6  Reclassified land use in CT, 1992 (US Geological Survey 2012)

for New Haven and Springfield. The additive bias $\tilde{C}(s, t)$ from Step I captures this difference, providing highest estimates for these two sites.

The MLE of $\mu_A$ was not significantly different from zero, and was therefore removed from the model. The estimates (and their standard errors) for the remaining parameters in the final model were $\sigma_Z = 22.53 (0.559)$, $\psi_A = 0.593 (0.033)$, $\sigma_A = 30.32 (1.868)$, and $\beta_c = 0.713 (0.013)$. Figure 8 shows the smoothed estimate of $A(t)$, calculated using the Kalman filter, along with its 95% confidence interval. Figure 9a shows the observed EPA and SCARR fitted concentration of NO$_2$ for New Haven, along with the 95% confidence interval, while Fig. 9b gives a comparison between the observed EPA, SCARR model predictions, and CMAQ model estimates of NO$_2$ concentration at this site. While the CMAQ model clearly underestimates NO$_2$ concentrations in the summer, SCARR model estimates more closely follow the observed concentrations. Similar plots for the remaining five EPA sites are given in Figures S1–S4 in the Supplement.

Table 4 gives a comparison of the coefficient of correlation, mean squared error (MSE), mean absolute error (MAE), and the relative bias (rBIAS) between the EPA data and estimates from the SCARR and CMAQ models. For all sites except Tolland, SCARR model provides a better fit than CMAQ, with a marked improvement observed at New Haven and Springfield. Figure 10 shows a comparison for the observed NO$_2$ concentration with the SCARR and CMAQ model predictions at 20 randomly selected Acid/Aerosol sites. The black horizontal line is the mean NO$_2$ recorded over the 10–14 day period at each site, while the blue and red lines show the daily SCARR and CMAQ model predictions, respectively, for those sites for the same time duration. It also shows the approximate month in which the data were collected. SCARR model predictions are generally closer to the observed NO$_2$ concentrations and provide a much smaller
Table 2: Variables in the final model for Step I, along with their parameter estimates and 95% confidence intervals

| Parameter          | Estimate | (95% CI)       |
|--------------------|----------|----------------|
| \( \beta \):       |          |                |
| Intercept          | 11.891   | (8.894, 14.888)|
| Population density (10,000) | 5.508    | (2.656, 8.359)|
| Season             |          |                |
| \( \sin(2 \cdot \pi \cdot \text{DYR}) \) | 1.245    | (0.424, 2.065)|
| \( \cos(2 \cdot \pi \cdot \text{DYR}) \) | 1.727    | (−0.207, 3.662)|
| \( \sin(4 \cdot \pi \cdot \text{DYR}) \) | 1.792    | (0.883, 2.702)|
| \( \cos(4 \cdot \pi \cdot \text{DYR}) \) | 2.706    | (1.685, 3.728)|
| \( \lambda \):     |          |                |
| Total traffic vol. (10,000 v-km) |          |                |
| 0.0–0.5 km         | 0.851    | (0.496, 1.207)|
| 0.5–1.0 km         | −0.138   | (−0.310, 0.035)|
| 1.0–2.0 km         | 0.084    | (0.031, 0.136)|
| Land use (1000 hectares) |          |                |
| Forest             |          |                |
| 0.0–2.0 km         | −5.430   | (−7.606, −3.254)|
| \( \gamma \):     |          |                |
| CMAQ NO\(_2\)      | 0.487    | (0.333, 0.642)|
| \( R^2 \)         | 0.777    |                |
| \( R^2 \) (Adjusted) | 0.757    |                |
| RMSE               | 2.60     |                |
| RMSPE*             | 2.77     |                |

Final model: Acid/Aerosol NO\(_2\) = Population Density + Season + TTV (0–0.5 km, 0.5–1.0 km, 1.0–2.0 km) + LUC (Forest: 0–2 km) + CMAQ NO\(_2\)

*Leave-one-out cross-validation root mean squared prediction error

Fig. 7: Observed concentrations of NO\(_2\) along with the estimated trigonometric and linear function of date to represent season
Table 3  Summary statistics for NO₂ concentration (ppb) for EPA and CMAQ data at six EPA sites, along with the spatiotemporal additive bias \( \tilde{C} \) at these sites, calculated using parameter estimates from Step I

| Site       | EPA \((Y_3)\) Mean (SD) | Range  | CMAQ \((\tilde{Y}_1)\) Mean (SD) | Range  | \(r^a\) | \(\tilde{C}\) |
|------------|-------------------------|--------|----------------------------------|--------|--------|----------|
| Bridgeport | 25.3 (10.5)             | 5.3–74.2 | 20.2 (9.7)                      | 3.9–55.7 | 0.693  | 16.7     |
| Chicopee   | 15.9 (9.8)              | 1.0–71.7 | 9.1 (7.6)                       | 1.0–50.1 | 0.749  | 10.8     |
| E. Hartford| 18.5 (9.8)              | 1.0–63.4 | 16.1 (8.6)                      | 2.3–53.3 | 0.763  | 13.6     |
| New Haven  | 27.5 (10.2)             | 5.9–74.9 | 19.7 (9.0)                      | 3.5–55.0 | 0.658  | 24.6     |
| Springfield| 26.0 (11.1)             | 4.1–84.0 | 14.6 (9.0)                      | 2.0–9.90 | 0.681  | 20.2     |
| Tolland    | 9.5 (6.5)               | 1.0–41.9 | 9.1 (7.0)                       | 1.1–46.6 | 0.819  | 6.13     |

\(a\) Coefficient of correlation between observed EPA data and CMAQ data

Fig. 8  Smoothed estimate of \(A(t)\) (solid blue) with 95% CI (dotted red)

Fig. 9  Comparison of the observed EPA, SCARR model fit, and CMAQ model estimate
Table 4: Comparing the coefficient of correlation ($r$) and mean squared error (MSE) between EPA and (i) SCARR predictions and (ii) CMAQ predictions

| Site         | $r$  | MSE SCARR | MSE CMAQ | MAE SCARR | MAE CMAQ | rBias SCARR | rBias CMAQ |
|--------------|------|-----------|----------|-----------|-----------|-------------|------------|
| Bridgeport   | 0.700| 0.693     | 60.62    | 88.68     | 5.96      | 7.41        | 0.082      | 0.201      |
| Chicopee    | 0.789| 0.749     | 45.00    | 88.59     | 4.56      | 7.49        | 0.151      | 0.428      |
| E. Hartford | 0.797| 0.763     | 37.05    | 46.78     | 4.56      | 5.25        | -0.026     | 0.130      |
| New Haven   | 0.687| 0.658     | 55.83    | 123.8     | 5.80      | 9.07        | -0.037     | 0.281      |
| Springfield | 0.766| 0.681     | 64.36    | 197.4     | 6.05      | 11.98       | 0.118      | 0.439      |
| Tolland     | 0.726| 0.819     | 21.38    | 18.91     | 3.33      | 3.07        | -0.039     | 0.032      |

$\text{MSE}_s = \frac{1}{m_s} \sum_{t=1}^{m_s} (\hat{Y}(s, t) - Y_3(s, t))^2$; $\text{MAE}_s = \frac{1}{m_s} \sum_{t=1}^{m_s} |\hat{Y}(s, t) - Y_3(s, t)|$

$r_{\text{Bias}} = \frac{1}{m_s} \sum_{t=1}^{m_s} \frac{Y_3(s, t)}{\sum_{t=1}^{m_s} Y_3(s, t)} (\hat{Y}(s, t) - Y_3(s, t))$

Fig. 10: Observed Acid/Aerosol (black), SCARR predicted (blue), and CMAQ estimated (red) NO$_2$ concentration at 20 randomly selected Acid/Aerosol sites

MSPE (36.30) as compared to CMAQ estimates (75.83). A similar plot for all 122 Acid/Aerosol sites is given in Figure S5 in the Supplement.

A plot of the standardized residuals against the fitted values from the model as well as a QQ plot of the standardized residuals (not shown) did not reveal any violations of the assumption of normality of the errors. An autocorrelation plot of the residuals over various temporal lags also suggested that the AR(1) assumption for $A(t)$ was justified.
The final model was used to predict the daily concentration of NO₂ for the state of Connecticut in 1994 and 1995 over a fine grid with pixels of size 300 × 300 m. A raster with 300 m square pixels was created that covered the entire state of Connecticut, and the latitude and longitude of the centroid of each pixel was extracted. For each of the 143,050 centroids, the corresponding CMAQ pixels were identified, and the covariates \( \tilde{Y}_1(s, t) \) and \( \tilde{C}(s, t) \) were calculated at each location, as detailed in Sects. 2.2 and 3.2. To predict a calibrated concentration at a new location \( s' \), the vectors \( \tilde{C}(s', t) \) and \( \tilde{Y}_1(s', t) \) from Eq. 4 are augmented to include \( \tilde{C}(s', t) \) and \( \tilde{Y}_1(s', t) \), while the EPA observation for this location \( Y_3(s', t) \) is treated as missing. Then, Eq. 4 is rewritten as

\[
Y_3(t) = A(t)I_7 + \beta_c \tilde{C}(t) + \gamma \tilde{Y}_1(t) + \sigma_z Z(t)I_7 \quad t = 1, 2, \ldots, 730, \tag{6}
\]

and,

\[
E(Y_3(s', t)|Y_3(1), \ldots, Y_3(t)) = E\left(A(s', t)|Y_3(1), \ldots, Y_3(t)\right) + E(\beta_c) \tilde{C}(s', t) + \gamma \tilde{Y}_1(s', t),
\]

where \( E(\beta_c) \) is given by its MLE and \( E\left(A(s', t)|Y_3(1), \ldots, Y_3(t)\right) \) is evaluated by rerunning the Kalman filter on the augmented data vectors using the MLE of the model parameters estimated earlier. The predicted values for each pixel for each day were reassigned to the original raster grid to create 729 raster images. Prediction maps of the daily concentration of NO₂ for Connecticut are displayed for a summer, winter and fall day (June 1, 1994; December 27, 1994; and October 19th, 1995, respectively) in Fig. 11 alongside the corresponding maps created using the predictions from the CMAQ model. An animation that shows the change over time in the spatial distribution of ambient NO₂ in CT for 1994 and 1995 is available online at “https://ogilani.shinyapps.io/CTNO2”.

6 Discussion

In the first step of the model, we spatially calibrated and refined the CMAQ NO₂ estimates using observed concentrations available from the Acid/Aerosol data, borrowing spatial information from various local covariates while controlling for time in the model. However, while the resultant model improved the spatial resolution of the CMAQ estimates, the calibration was done over very few time points, which were aggregated over long durations: between 10 and 14 days. Therefore, it didn’t provide effective temporal calibration of the daily variability in the CMAQ data. Given the availability of another source of observed data on the concentration of NO₂ (EPA data) that is temporally dense, in the second step we developed a dynamic spatiotemporal model to use the EPA data for temporal calibration of the spatially refined CMAQ estimates from the first step.

The final model from Step I includes population density, total traffic volume buffers up to 2 km, “forest 0–2 km” buffer for land use type, and a trigonometric function of
date. The model explains about 76% of the variability in the observed data, with a leave-one-out cross validation PRESS statistic (936) close to the residual sum of squares (752), suggesting it does a decent job at predicting the concentration of NO$_2$ at a new location. The coefficient for total traffic volume buffer 0.5–1 km was less than zero, suggesting that traffic volume within that buffer reduces the concentration of NO$_2$, which is unexpected. However, the estimate is not statistically significant, and was included in the final model to allow the next outer buffer ring, 1–2 km, to be included, which was statistically significant. Similar “U shaped” estimates for the dispersion function of NO$_2$ have been observed in previous studies (Holford et al. 2010), and may be explained by the complex process that produces NO$_2$ from nitrogen oxide (NO) and ozone (O$_3$) in the atmosphere. NO$_2$ is not produced directly by combustion in automobiles, and elevated levels of NO$_2$ are often observed at farther distances, perhaps reflecting the time needed to convert NO to NO$_2$. The estimated trigonometric function (Fig. 7) shows two peaks within a 12 month cycle—a higher peak in the winter and a lower one in the summer. While the fit of the curve seems appropriate for these data, comparing the time trend of NO$_2$ for the EPA data suggests that the summer peak estimated by the trigonometric function might be artificially high in the Acid/Aerosol data. However, an advantage of the two stage modeling strategy is that this possible over estimation in the summer months in Step I is countered during the temporal calibration in Step II, as seen by a dip in the mean trend of the estimate of $A(t)$ (Fig. 8) during the corresponding summer period.

The MLE for the parameter $\mu_A$ in Step II of the model was not statistically significantly different from zero. Given the presence of $\tilde{C}(s, t)$ in the model, this was to be expected as $\tilde{C}(s, t)$ captures the mean spatial additive calibration bias for the CMAQ.
data. The parameter $\beta_c$ controls the influence of $\tilde{C}(s, t)$ in Step II. The MLE for $\beta_c$ was 0.713, which suggests a slight mitigation of the spatial additive calibration bias $\tilde{C}(s, t)$, estimated in Step I of the model, when evaluating the temporal evolution of the additive calibration bias $A(t)$ in the second step.

As discussed in Gilani et al. (2016), the CMAQ model does a reasonable job of predicting the ambient NO$_2$ concentrations in the winter months when the concentrations are generally high. But it does not provide very accurate predictions during the low concentration periods in the summer, when it generally underpredicts the true concentration. As seen in Fig. 9b, the SCARR model improves the prediction during the summer months and, overall, appears to provide a better prediction as compared to the CMAQ model. The coefficients of correlation comparing the EPA observations with predictions from the SCARR and CMAQ model, and the empirical mean squared errors (MSE), mean absolute errors (MAE), and the relative bias (rBIAS) for these two models (Table 4), show that the SCARR model provides better predictions at five of the six sites, with a remarkable improvement in the prediction for New Haven and Springfield. The CMAQ model, on the other hand, appears to provide better predictions at Tolland. However, as mentioned in Sect. 4.2, the site at Tolland was missing EPA data for the summer of 1994, and the correlation and MSE for this site reflect the performance of the two models only for the winter period, during which time the CMAQ model generally performs well. For almost all of the Acid/Aerosol sites, predictions from the SCARR model more accurately reflect the truth than the CMAQ model (Figs. 10 and S5). The empirical MSPE for the SCARR model predictions (36.30) at the Acid/Aerosol sites was much smaller than the MSPE for the CMAQ model prediction (75.83).

There are a few limitations to the modeling strategy presented here. The model assumes that the NO$_2$ observations recorded at the Acid/Aerosol and EPA sites accurately reflect the true ambient concentrations. However, qualitatively that might not necessarily be the case as different monitoring equipments record NO$_2$ concentrations at varying levels of accuracy. Additionally, the EPA monitors are typically placed in high concentration locations near major roadways. However, the two step modeling strategy may actually help in balancing this effect by also including observations from the Acid/Aerosol data, whose locations were sampled independent of pollutant concentrations and were therefore more evenly distributed between high and low concentration areas. Another limitation is due to the fact that Step II of the model includes variables that were estimated in Step I ($\tilde{C}$), and the resulting estimates of the model errors do not capture the additional uncertainty of including these estimated variables, leading to somewhat conservative prediction errors. Methods developed for accounting for measurement errors in models can be applied to this model to provide more accurate error estimates (Carroll et al. 2006). Given the Markov property of Kalman filters, a Bayesian approach can also be utilized to explicitly account for the uncertainty in the estimated variables included in the model. However, prediction of NO$_2$ on a fine spatiotemporal resolution, with 143,050 spatial locations and 730 time points, using a Bayesian approach (including a Gibbs sampler for the variance components) is computationally rather intensive. Active research in developing Bayesian methods using parallel computing does provide potential solutions to this problem.
The NO2 prediction maps for CT for 1994 and 1995 (Fig. 11) show a clear improvement in the spatial resolution as compared to the estimates provided by the CMAQ model. The effect of local covariates is evident in the finer spatial resolution map, where the contribution of traffic on major highways to ambient NO2 concentration stands out. These maps provide more accurate estimates for points within the CMAQ pixel. For example, the concentration of NO2 appears rather high at all points in the pan-handle of CT (south-west corner of the map) in the CMAQ model estimates, whereas the SCARR model predictions show that the concentration is high primarily along the major highway (Interstate 95), while locations away from the highway have a significantly lower concentration. These maps with more accurate estimates at the centroid and finer spatial resolution can be very useful in assigning mean daily exposure to NO2 for participants in epidemiologic studies, which is usually not possible to do using available observed sources of data on the concentration of NO2. Predictions from this model significantly contribute to exposure assessment at a fine spatial and temporal resolution for CT in 1994 and 1995.

References

Alkuwari FA, Guilleras S, Wang Y (2013) Statistical downscaling of an air quality model using fitted empirical orthogonal functions. Atmos Environ 81:1–10

Berrocal VJ, Gelfand AE, Holland DM (2010) A spatio-temporal downscaler for output from numerical models. J Agric Biol Environ Stat 15(2):176–197

Beyer HL (2012) Geospatial Modeling Environment (Version 0.7.2.1). (software). http://www.spatalecolgy.com/gme. Accessed 10 Apr 2012

Brown P, Diggle P, Lord M, Young P (2001) Space-time calibration of radar rainfall data. J R Stat Soc Ser C 50(2):221–241

Byun D, Schere K (2006) Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl Mech Rev 59:51–77

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. CRC Press, Boca Raton

Chang HH, Hu X, Liu Y (2014) Calibrating MODIS aerosol optical depth for predicting daily PM2.5 concentrations via statistical downscaling. J Expo Sci Environ Epidemiol 24(4):398–404

Connecticut Department of Transportation (2001) 2000 Traffic Volumes. State Maintained Highway Network, Bureau of Policy and Planning

Cressie N (1993) Statistics for spatial data: Wiley series in probability and statistics. Wiley, New York

ESRI (2003) StreetMapUSA. Redlands, CA, Environmental Systems Research Institute

Fuentes M, Raftery AE (2005) Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61(1):36–45

Gesch D (2007) The national elevation dataset. In: Maune D (ed) Digital elevation model technologies and applications: the DEM users manual. Asprs Publications, Bethesda

Gesch D, Omoen M, Greenlee S, Nelson C, Steuck M, Tyler D (2002) The national elevation dataset. Photogramm Eng Remote Sens 68(1):5–11

Gilani O, McKay LA, Gregoire TG, Guan Y, Leaderer BP, Holford TR (2016) Spatiotemporal calibration and resolution refinement of output from deterministic models. Stat Med. https://doi.org/10.1002/sim.6867

Hogrefe C, Lynn B, Goldberg R, Rosenzweig C, Zalewsky E, Hao W, Drasaiwam P, Civerolo K, Ku JY, Sistla G, Kinney PL (2009) A combined model–observation approach to estimate historic gridded fields of PM2.5 mass and species concentrations. Atmos Environ 43(16):2561–2570

Holford T, Ebisu K, McKay L, Gent J, Triche E, Bracken M, Leaderer B (2010) Integrated exposure modeling: a model using GIS and GLM. Stat Med 29(1):116–129
Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsvuvaroglu T, Morrison J, Giovis C (2004) A review and evaluation of intraurban air pollution exposure models. J Expo Sci Environ Epidemiol 15(2):185–204

Kalman R (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35–45

Li B, Eriksson M, Srinivasan R, Sherman M (2008) A geostatistical method for Texas NexRad data calibration. Environmetrics 19(1):1–19

Massachusetts Department of Transportation - Highway Division (2012) Traffic Volume Counts. http://www.mhd.state.ma.us/default.asp?pgid=content/traffic01&sid=about. Accessed 29 Mar 2013

Meiring W, Sampson P, Guttrop P (1998) Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. Environ Ecol Stat 5(3):197–222

New York State Department of Transportation (2012) Department of Transportation - Historic Traffic Data 1977-2011. https://www.dot.ny.gov/divisions/engineering/technical-services/hds-respository/. Accessed 29 Mar 2013

Palmes E, Gunnison A, DiMattio J, Tomczyk C (1976) Personal sampler for nitrogen dioxide. Am Ind Hyg Assoc J 37(10):570–577

Petris G (2010) An R package for dynamic linear models. J Stat Softw 36(12):1–16

R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Skene KJ, Gent JF, McKay LA, Belanger K, Leaderer BP, Holford TR (2010) Modeling effects of traffic and landscape characteristics on ambient nitrogen dioxide levels in Connecticut. Atmos Environ 44(39):5156–5164. https://doi.org/10.1016/j.atmosenv.2010.08.058. http://www.sciencedirect.com/science/article/B6VH3-511G1YK-9/2/4258d4a4b1934748a7a99324663259f8

Triche E, Belanger K, Beckett W, Bracken M, Holford T, Gent J, Jankun T, McSharry J, Leaderer B (2002) Infant respiratory symptoms associated with indoor heating sources. Am J Respir Crit Care Med 166(8):1105–1111

US Census Bureau (1990) US Census Data 1990. ftp.census.gov/census_1990/. Accessed 01 Apr 2013

US Census Bureau (2012) Cartographic Boundary Files - U.S. Census Bureau. http://www.census.gov/geo/www/cob/tr1990.html. Accessed 01 Apr 2013

US Environmental Protection Agency (2008) Integrated Science Assessment for oxides of nitrogen—Health criteria. National Center for Environmental Assessment, Office of Research and Development, Research Triangle Park Durham, NC

US Environmental Protection Agency (2011) Technology Transfer Network (TTN) Air Quality System (AQS)—Nitrogen Dioxide. http://www.epa.gov/ttn/airs/airsaqsdetaildata/downloadaqsdatalist.htm. Accessed 14 Apr 2011

US Environmental Protection Agency (2014) Community Multi-scale Air Quality Model (CMAQ). http://www.epa.gov/amad/Research/RIA/cmaq.html. Accessed 03 Feb 2014

US Geological Survey (2012) USGS National Land Cover Dataset 1992. http://landcover.usgs.gov/natlndcover.php. Accessed 29 Mar 2013

WHO (2003) World Health Organization, Bonn, Germany. Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. http://www.euro.who.int/document/e79097.pdf. Accessed 21 Apr 2011

Owais Gilani is an assistant professor of Mathematics (Statistics) at Bucknell University. He received his Ph.D. in Biostatistics from Yale University in 2014. His research work focuses on developing spatial and spatiotemporal statistical methods for applications in air pollution modeling and environmental epidemiology, as well as quantifying human mobility patterns to study their role in various applications including air pollution exposure and spread of infectious diseases.

Lisa A. McKay is a programmer/analyst in the Biostatistics division of the Yale School of Public Health. She has supported faculty research in developing statistical models which describe cancer incidence and mortality trends among other things, has worked on several projects involving the spatial modeling of environmental effects on childhood asthma, and is currently working on a project which aims to provide evidence-based modeling of the public health impacts of Food and Drug Administration (FDA) tobacco regulatory actions.
Timothy G. Gregoire is the J. P. Weyerhaeuser Professor of Forest Management at the School of Forestry and Environmental Studies, Yale University. His main research interests for the past 40 years have been the application of sampling methods to natural and ecological resources, and the modeling of longitudinally and spatially correlated data. He is a Fellow of the American Statistical Association as well as the American Association for the Advancement of Science. He is an elected member of the ISI, and has served as President of the North American Region of the International Biometric Society.

Yongtao Guan is Leslie O. Barnes Professor of Management Science at the University of Miami Herbert Business School. Yongtao earned his Ph.D. in Statistics from Texas A&M University in 2003. He specializes in modeling spatial, recurrent event and longitudinal data that commonly arise in business, environmental, medical, and social media applications. He was a recipient of the CAREER Award from the National Science Foundation in 2009 and two R01 awards from the National Institute of Health in 2011 and 2013. He is a fellow of the American Statistical Association.

Brian P. Leaderer is the Susan Dwight Bliss Professor of Epidemiology at the Yale School of Public Health and Professor at the Yale School Forestry and Environmental Studies. His research interests, resulting in over 300 publications, are interdisciplinary in nature with a focus on assessing exposures (measured and modeled in both environmental chamber and field studies) to air contaminants (indoor and outdoor) and assessing the health impact resulting from those exposures in epidemiological studies.

Theodore R. Holford is Susan Dwight Bliss Professor of Biostatistics at Yale School of Public Health, New Haven, CT, USA. He got his Ph.D. in Biometry at Yale University in 1973. He has published extensively in the development and the application of statistical methods for environmental epidemiology.

Affiliations

Owais Gilani, Lisa A. McKay, Timothy G. Gregoire, Yongtao Guan, Brian P. Leaderer, Theodore R. Holford

1 Department of Mathematics, Bucknell University, 1 Dent Drive, Lewisburg, PA 17837, USA
2 Center for Perinatal, Pediatric and Environmental Epidemiology, Yale School of Public Health, Yale University, One Church Street, 6th Floor, New Haven, CT 06510, USA
3 Yale School of Forestry & Environmental Studies, Yale University, 195 Prospect Street, New Haven, CT 06511, USA
4 Department of Management Science, School of Business Administration, University of Miami, Coral Gables, FL 33124, USA