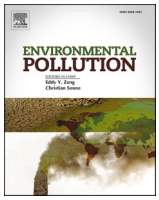
Environmental Pollution 366 (2025) 125405

Contents lists available at ScienceDirect

Environmental Pollution

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



National $PM_{2.5}$ spatiotemporal model integrating intensive monitoring data and land use regression in a likelihood-based universal kriging framework in the United States: 2000–2019☆

Meng Wang [a,b,c,*], Michael Young [c], Julian D. Marshall [d], Logan Piepmeier [c], Jianzhao Bi [c], Joel D. Kaufman [c], Adam A. Szpiro [e]

[a] Department of Epidemiology and Environmental Health, University at Buffalo, Buffalo, NY, USA
[b] RENEW Institute, University at Buffalo, Buffalo, NY, USA
[c] Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA
[d] Department of Civil and Environmental Engineering, University of Washington, Seattle WA, USA
[e] Department of Biostatistics, University of Washington, Seattle, WA, USA

ARTICLE INFO

Keywords:
Fine particulate matters
Spatiotemporal model
Exposure assessment
Fine-scale monitors

ABSTRACT

Nationwide $PM_{2.5}$ exposure models typically rely on regulatory monitoring data as the only ground-level measurements. In this study, we develop a high-resolution spatiotemporal $PM_{2.5}$ model for the contiguous United States from 2000 to 2019 with dense monitoring data at both regulatory and residential sites. Specifically, we combine publicly-available data from 1843 regulatory monitors with our own set of multiple 2-week measurements at 939 residential locations. As we show, these additional data enhance the spatiotemporal prediction capabilities of the model. The model can handle varying data densities and regional variations; it predicts two-week average $PM_{2.5}$ concentrations at fine spatial scale for the contiguous United States. Cross-validation performance indicates a spatial R^2 of 0.93 and a root mean square error (RMSE) of 1.19 (μg/m^3), and a temporal R^2 of 0.85 and RMSE of 2.05 (μg/m^3). Regional spatial R^2 ranged from 0.80 (northwest) to 0.93 (northeast and central). Over time, the average $PM_{2.5}$ across the United Stats decreased from 7.6 μg/m^3 in 2000 to 4.7 μg/m^3 in 2019. Our model effectively captured local $PM_{2.5}$ gradients, highlighting its ability to address fine-scale variations related to local sources and roadways.

1. Introduction

Ambient fine particulate matter ($PM_{2.5}$) is a major risk factor for human morbidity and mortality (Cohen et al., 2017). Epidemiological studies often use advanced exposure models to estimate outdoor $PM_{2.5}$ concentrations because regulatory monitoring stations provide insufficient fine-scale spatial resolution (Kaufman et al., 2016; Lepeule et al., 2014; Di et al., 2017). Spatiotemporal exposure models are valuable in characterizing spatial contrasts of $PM_{2.5}$ both between and within regions as well as temporal variability. Traditional modeling approaches include land use regression (LUR) (Lepeule et al., 2014; Wang et al., 2014; Eeftens et al., 2012), deterministic models (Wang et al., 2015; Laurent et al., 2016) and aerosol optical depth (AOD) from satellite observations (Kloog et al., 2012; Knibbs et al., 2018; Hu et al., 2017), all of which have been widely used for national population studies in the United States. However, all of these methods have limitations. Deterministic modeling and AOD approaches typically have a grid resolution measured in km^{2}s, which typically is not spatially resolved enough to characterize exposures at very local scales. LUR, which employs statistical methods to combine air pollution data with variables from geographic information systems (GIS), is suitable to characterize small-scale spatial variability of $PM_{2.5}$ concentrations. However, some LUR models do not explicitly account for temporal variability (Bechle et al., 2015) and most LUR approaches would not be able to readily incorporate temporally sparse monitoring data, such as data from measurements at residential locations (Jerrett et al., 2005; Di et al., 2019).

More recently, efforts have been made to improve model

☆ This paper has been recommended for acceptance by Prof. Pavlos Kassomenos.

* Corresponding author. University at Buffalo, 270 Farber Hall, Buffalo, NY, USA.
E-mail address: mwang54@buffalo.edu (M. Wang).

https://doi.org/10.1016/j.envpol.2024.125405
Received 30 July 2024; Received in revised form 8 November 2024; Accepted 26 November 2024
Available online 28 November 2024
0269-7491/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

performance by integrating a wider variety of data resources, including monitoring data with LUR and deterministic models and satellite predictions, in one framework (Di et al., 2019). Applications of that modeling approach for have been reported at a national or continental scale such as Europe (de et al., 2016; Chen et al., 2019; de et al., 2018; Vienneau et al., 2013), United States (Hu et al., 2017; Jerrett et al., 2005), Australia (Knibbs et al., 2018) and China (Xu et al., 2019; Xiao et al., 2018; Huang et al., 2021). Epidemiologic studies have suggested stronger associations between $PM_{2.5}$ exposure and mortality using more advanced exposure models (Jerrett et al., 2017).

Despite the fast evolution of modeling algorithms for large-scale exposure assessment, most of the spatiotemporal modeling efforts are based on publicly-available monitoring data as the only ground-level measurements for air pollution. Some researchers have argued that regulatory monitoring data may be less useful for capturing fine-scale concentration variations such as roadside increments of $PM_{2.5}$ caused by traffic emissions in urban areas, resulting in underestimation of $PM_{2.5}$ exposure for some segments of the population samples near traffics (Jerrett et al., 2005). If that is the case, then exposure estimates used in epidemiological studies may not sufficiently capture $PM_{2.5}$ contrast between and within regions, leading to misclassifications of exposure and attenuation of the estimated health effects (Miller et al., 2007). On the other hand, many intra-urban exposure models employ air pollution data from monitors designated to capture human exposure, which provided sufficient details to estimate fine-scale $PM_{2.5}$ variations (Ryan and LeMasters, 2007). In a previous study, we demonstrated that incorporating data from cohort-specific monitors with regulatory monitoring data can provide accurate predictions at the intraurban spatial scale (Keller et al., 2015). However, this approach has not been applied to the development of national $PM_{2.5}$ model in the United States.

In this paper, we develop a national scale spatiotemporal model for $PM_{2.5}$ that leverages an advanced statistical framework, an extensive database of geographic covariates, nationwide $PM_{2.5}$ regulatory monitoring stations, and intra-urban research monitoring sites collected at selected fixed sites and individual homes to accurately predict concentrations at a very finely resolved (<10 m resolution) spatial scale.

2. Methods

2.1. Study regions

We developed a $PM_{2.5}$ exposure model for the 48 contiguous United States (i.e., excluding Alaska and Hawaii). To account for heterogeneity in $PM_{2.5}$ distributions attributable to variations in climate conditions, pollution sources, and terrain, we subdivided the country into nine modeling regions, coded as northwest, west, southwest, west north central, south, east north central, central, northeast, and southeast, based on National Oceanic and Atmospheric Administration climate regions (Fig. S1). Therefore, each climate region may have homogenous climate over space. This approach allows for different model parameters optimized for specific regions of the country, including kriging parameters and regression coefficients. A regionalized modeling approach for national air pollution assessment has been used previously for nitrogen dioxide and was shown to improve predictive performance (Young et al., 2016). To make predictions near the boundaries of climate regions, within 100 km buffer zones we averaged regional predictions using inverse distance weighting (IDW).

2.2. Monitoring data

Monitoring data include (1) publicly-available continuous long-term measurements from the Air Quality System (AQS) including sites from the Interagency Monitoring of Protected Visual Environments (IMPROVE) of the U.S. Environmental Protection Agency (EPA), (2) spatially dense but temporally sparse residential monitoring data from cohort-specific campaigns, including the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) (Cohen et al., 2009) and the Subpopulations and Intermediate Outcome Measures in COPD Air Study (SPIROMICS Air) (Hansel et al., 2017), and (3) fixed-site monitoring from the same studies (MESA Air; SPIROMICS Air), designed to capture important intra-urban features such as the effect of being near roadways (Fig. S2). Herein, we refer to these three datasets as AQS or regulatory monitors, residential monitors, and fixed-site monitors, respectively. "Non-regulatory" refers to the combined dataset of residential and fixed site monitors.

We included 1843 p.m.$_{2.5}$ AQS monitoring stations with data between 2000 and 2019. Daily averages of $PM_{2.5}$ data at the AQS monitoring sites were aggregated at a 2-week time scale to account for sampling every 3 or 6 days at some sites and to match the temporal scale of the residential monitors. We excluded monitoring sites with less than 2 months of continuous data.

Non-regulatory monitors were at 911 participants' addresses with between one and three 2-week averages per site, and at 28 fixed sites operated continuously for between one and four years in ten metropolitan areas (Seattle, WA; San Francisco, CA; Los Angeles and Riverside, CA; St. Paul, MN; New York, NY; Winston-Salem, NC; Chicago, IL; Ann Arbor, MI; Salt Lake City, UT; Baltimore, MD). These data captured $PM_{2.5}$ residential exposure in the US communities, providing additional information to strengthen the model. The data have shown good agreement and little bias with those from the located AQS sites (Keller et al., 2015).

Monitoring sites were assigned to one of the nine climate regions, with sites in the overlapping buffer zone belonging to multiple regions.

2.3. Geographic covariates

More than 220 geographic covariates were included for the regionalized spatiotemporal model development (Table S1). Variables included geographic features – such as road networks (e.g., distances to nearby major roads and, within buffers, length of roads and truck routes, and counts of intersections), industrial and port emissions, population density, land use (e.g., commercial space), and land cover (e.g., green spaces) – averaged over multiple radial buffer distances. Similarly, we incorporated annual averages of specific emission sources for NOx, SO2, CO, $PM_{2.5}$ and PM_{10} from the U.S. EPA Emission Inventory Groups. We also included annual average satellite-measured ground-level NO_2 for 2005 to 2007. The ground-level measurement data were estimated by columnar satellite data processed by GEOS-Chem reactive transport model and was used to capture traffic emissions with a resolution of 13 $\times$ 24 km^2 [29]. Details of the geographic variables and their selection criteria are described in a previous publication (Keller et al., 2015; Wang et al., 2016).

2.4. Spatiotemporal modeling strategy Overview

We developed a hierarchical modeling strategy to fully accommodate the unique features of our monitoring data. First, we developed separate models for each region to generate region-specific modeling parameters, optimized using restricted maximum likelihood (REML). Second, we evaluated the overall performance of the regionalized models in the contiguous United States by combining the predictions from cross-validation across the nine climate regions. This data integration takes the differences of predictions within the boundary buffer area into account using the inverse weights based on distance to the boundary to average the predictions at the same site that belongs to different neighborhood climate regions. Finally, we estimated long-term $PM_{2.5}$ concentrations in each region and stitched the predictions to the national scale with the overlapping area averaged by inverse distance weighing as mentioned above. Technical details of implementation, including the model structures and principles, evaluation approaches and applications of the models for $PM_{2.5}$ have previously been reported for citywide studies, and are described briefly below (Keller et al., 2015;

Wang et al., 2016).

2.5. Model development

Each regionalized model is comprised of a spatio-temporal trend model and spatio-temporal residuals, written as

$$C(s,t) = \mu(s,t) + \nu(s,t), \quad (1)$$

where $C(s,t)$ denotes the two-week average concentration of $PM_{2.5}$ at location s and time t; $\mu(s,t)$ represents the spatio-temporal mean surface; and $\nu(s,t)$ represents the zero-mean spatio-temporal residual variation which has a spatial correlation structure, but is here assumed independent across the two-week periods. The mean $\mu(s,t)$ is further decomposed into a long-term average $\beta_0(s)$ at location s, a linear combination of temporal trend basis functions $f_i(t)$ with spatially-varying coefficient fields $\beta_i(s)$ as follows:

$$\mu(s,t) = \beta_0(s) + \sum_{i=1}^{m} \beta_i(s) f_i(t). \quad (2)$$

We derive the time trends $f_i(t)$ from AQS and non-regulatory long-term fixed sites. Specifically, the trends are computed from a singular value decomposition (SVD) of the space-time data matrix for the AQS and fixed sites using the SpatioTemporal R package (Lindstrom et al., 2014), dealing with missing or incomplete monitoring data using an iterative expectation-maximization (EM) procedure described in detail by Sampson et al. (2011).

The spatially-varying long-term average $\beta_0(s)$ and time trend coefficients $\beta_i(s)$ are modeled in a universal kriging framework fitted based on all monitoring data, with a LUR mean and either an exponential or independent covariance structure. Each LUR mean model is comprised of component scores computed as linear combinations of geographic covariates and satellite-based ground-level NO_2 estimated by Partial Least Squares (PLS) (Abdi and Williams, 2013). To avoid overfitting risk while provide adequate feasibility for model development by regions, we selected 4 temporal trends for each region and the LUR model for each $\beta_i(s)$ included 3 region-specific PLS scores.

2.6. Model estimation and prediction

Parameter estimation was done via REML using the SpatioTemporal package in R, version 3.1.1. The first step in REML is a nonlinear optimization to estimate the covariance models in the various kriging covariances embedded in the spatiotemporal model described above for all regions (i.e., range, nugget, and sill for each of the $\beta_i(s)$ and for the spatiotemporal residual field $\nu(s,t)$). For each regional model, we repeated the estimation procedure ten times on random subsets of the data, with each subset comprised of 100 AQS sites and all of the non-regulatory data. For each random subset of the data, we tried multiple initial conditions, and we selected the kriging parameter estimates that maximized the REML criterion across all initial conditions and random subsets of the data, subject to the restriction that the optimization identified a confirmed local maximum. We used this approach rather than optimization with all of the data at one time to balance computational burden with the need to assess multiple initial conditions. The second stage of REML combines estimating the regression coefficients in the mean models for the $\beta_i(s)$ with prediction at new sites, and for this step we used all of the monitoring data together along with the covariance parameters selected from the first stage.

When making prediction at locations near the boundary of multiple modeling regions, we used inverse distance weighting as follows:

$$C_t = \sum_{i=n} \frac{\widetilde{W}_{i,t}}{\sum_{i=n} \widetilde{W}_{i,t}} \times C_{i,t} \quad (3)$$

Here, C_i denotes the $PM_{2.5}$ concentration at time t and climate region i. $\widetilde{W}_{i,t}$ refers to the weight of $PM_{2.5}$ exposure under one of the conditions: (1) $\widetilde{W}_{i,t} = 0$, where the site is out of the climate region with a distance to the inter-region boundary greater than 100 km; (2) $\widetilde{W}_{i,t} = 1$, where the site is within the inter-region boundary of the climate region; (3) $\widetilde{W}_{i,t} = \frac{100-x}{100}$, where the site is out of the climate region but within the 100 km buffer zone whose distance is x km away from the inter-region boundary.

2.7. Model evaluation and predictions

We used 10-fold cross-validation to evaluate the spatiotemporal model prediction performance at the AQS and fixed sites with long-term data averaged across the study period, holding fixed the pre-computed time trends and PLS scores as well as the estimated covariance parameters. In each cross-validation run, we left out 10% of the AQS and fixed sites (all data from each site held out at the same time), trained the model on the remaining data (90% of AQS and fixed sites; 100% of residential sites), and made predictions at the held-out locations. We computed root mean square error (RMSE) and cross-validation R^2 (for bi-weekly $PM_{2.5}$ predictions versus observations across the AQS and fixed sites) to quantify accuracy (RMSE) and precision (R^2) of each model in predicting the spatiotemporal variability of $PM_{2.5}$. We evaluated performance in representing spatial variability using the long-term and annual averages for the AQS and fixed sites, and evaluated the model performance in representing temporal variability using the median value of the agreements between the predictions and observations across individual sites at the two-week scale over the entire study period. To assess the impact of wildfire event on our model, we restricted the evaluation among 100 p.m.$_{2.5}$ monitors that were operating within 5 km of a fire that lasted two weeks or more. To investigate model performance away from air monitors, we estimated within-monitor RMSE as a function of distance to a nearest air monitor.

Using a version of the model trained on all of the monitoring data, we generated prediction maps to illustrate the features of our model. These maps include annual and seasonal averages at the national scale (10 km grid) and detailed maps in Memphis, TN and Minneapolis-St. Paul, MN comparing predictions at the 10 m, 100m, and 1 km scales.

Table 1
Descriptive statistics for the measurement of $PM_{2.5}$ concentrations ($\mu g/m^3$) at AQS and home locations based on two-week averages by study regions (N: number of sites).

Regions	AQS + Fixed			Home		
	N	Median (IQR)	Range	N	Median (IQR)	Range
Northwest	242	5.4 (5.2)	(0.1, 102.5)	25	5.3 (0.9)	(3.4, 7.4)
West	218	7.4 (7.1)	(0.2, 109.3)	185	13.3 (9.5)	(0.2, 55.2)
Southwest	220	5.5 (4.4)	(0.2, 86.5)	36	5.8 (5.5)	(1.7, 29.2)
Northwest central	256	6.4 (5.1)	(0.2, 102.5)	36	5.8 (5.5)	(1.7, 29.2)
Central	660	11.3 (5.9)	(0.9, 154.5)	272	10.9 (6.9)	(1.2, 58.8)
Northeast central	380	10.1 (6.3)	(0.8, 54.8)	297	9.4 (6.2)	(1.2, 58.8)
Southeast	456	10.6 (5.8)	(0.9, 76.9)	256	12.3 (6.1)	(2.6, 51.7)
South	349	10.0 (5.1)	(0.6, 154.5)	–	–	–
Northeast	380	10.1 (6.3)	(0.8, 54.8)	257	12.7 (6.7)	(2.5, 63.6)

3. Results

Table 1 presents the descriptive statistics for the $PM_{2.5}$ 2-week averaged observations at the AQS and non-regulatory monitoring locations by study regions. The AQS and fixed monitoring sites which captured long-term temporal trends of $PM_{2.5}$ exposure varied substantially within and across the regions. As expected, the median concentration of $PM_{2.5}$ is higher in the central and east regions (e.g., northeast, southeast, northeast central and south regions) (median: 10.0–11.3 μg/m^3) than in the west regions (e.g., northwest, west, southwest and northwest central regions) (median: 5.4–7.4 μg/m^3). The $PM_{2.5}$ monitors at the residential locations which characterized fine-scale variations within a city showed the highest median levels and largest spatial variations among the regions.

Table 2 shows the measures of cross-validated prediction accuracy at AQS and fixed sites with long-term monitoring data. The national model shows excellent performance, with overall spatiotemporal R^2 of 0.87 and an RMSE of 2.66 μg/m^3 (analogous statistics for annual averages range from 0.81 to 0.88 and 1.95–3.45 μg/m^3) and the overall spatial-only model performance of an R^2 of 0.93 and RMSE of 1.19 μg/m^3 (analogous statistics for annual averages range from 0.88 to 0.94 and 1.04–1.66 μg/m^3). Our model showed moderate performance at predicting within-site temporal variability with a median R^2 over the entire modeling period of 0.85 and RMSE of 2.05 μg/m^3; analogous statistics for annual averages ranged from 0.80 to 0.87 and 1.24–2.34 μg/m^3, respectively. There was some variability in prediction accuracy by modeling region (Table S2 and Fig. S3); the southeast region showed the best model performance (spatiotemporal $R^2 = 0.87$, RMSE = 2.16 μg/m^3), the northwest region showed the worst model performance (spatiotemporal $R^2 = 0.82$, RMSE = 3.60 μg/m^3). The model performed equally well in terms of RMSE regardless the distance from a nearest air monitor (see Fig. S4). Moreover, the predictions in areas with active fires were underestimated compared to those without active fires (see Fig. S5).

The first PLS component, which explains the majority of the long-term mean $PM_{2.5}$ variation, was negatively associated with elevation and green space (e.g., grass and shrub), and was positively associated with the features indicating primary emissions (e.g., traffic and anthropogenic factors, including population and impervious surfaces; see Fig. S6 characterizing the relationships between the first PLS component and the GIS covariates).

As illustration of the spatiotemporal patterns of $PM_{2.5}$, spatial maps of annual average $PM_{2.5}$ concentrations in the United States are shown in Fig. 1 in the years of 2000, 2008 and 2016 based on the best regional model predictions (see all year maps in Fig. S7). $PM_{2.5}$ concentration are highest in the central and southeastern U.S., and are relatively low but most variable in the southwest (Table S3). Regarding the temporal trends, the national long-term average dropped from 7.6 μg/m^3 in 2000 to 4.7 μg/m^3 in 2019, with most substantial reduction occurring in the eastern parts of the United States, demonstrating the effectiveness of the Clean Air Act. By seasons, average $PM_{2.5}$ levels are highest in summer (July–September 2008; population average: 7.8 μg/m^3) and lowest in Autumn (October–December 2008; 5.8 μg/m^3).

To investigate $PM_{2.5}$ predictions at fine-scale, we visualize high-resolution $PM_{2.5}$ maps for 2015 in two typical US cities (Memphis, TN, Saint Paul, MN) and compare the predictions in a 1000 m area with spatial resolution of 10, 100, or 1000 m (Fig. 2). We observed larger spatial variations of $PM_{2.5}$ concentrations predicted with 10-m resolution than those with 100- and 1000-m resolution, suggesting that the model can predict concentration differences at a 10-m scale. The 10-m resolution predictions reflect, e.g., the dispersion patterns of traffic emissions along the roadways; they provide spatially more precise predictions of the residential exposure variations than do the predictions using 100 or 1000 m resolutions. This was further evidenced in the transect plots in which substantial variations of $PM_{2.5}$ exposure were observed across the cities.

Table 2
Model performances from regular cross-validation (R^2 and RMSE in μg/m^3) of all the study areas stitched to the national scale in the United States including spatiotemporal, spatial and temporal R^2 respectively, at the AQS and fixed monitoring sites.

Year	Spatiotemporal		Spatial		Temporal	
	R (Kaufman et al., 2016)	RMSE	R (Kaufman et al., 2016)	RMSE	R (Kaufman et al., 2016)	RMSE
2000	0.86	3.45	0.92	1.66	0.82	2.34
2001	0.84	3.41	0.92	1.65	0.80	2.46
2002	0.88	2.92	0.93	1.47	0.87	1.93
2003	0.86	2.86	0.93	1.40	0.84	2.08
2004	0.86	2.77	0.93	1.37	0.82	2.06
2005	0.88	2.91	0.94	1.45	0.86	2.15
2006	0.85	2.82	0.92	1.45	0.82	2.03
2007	0.86	2.97	0.92	1.46	0.85	2.06
2008	0.84	2.76	0.90	1.40	0.82	1.83
2009	0.84	2.42	0.90	1.23	0.85	1.54
2010	0.86	2.32	0.93	1.14	0.82	1.62
2011	0.86	2.35	0.92	1.15	0.85	1.62
2012	0.83	2.20	0.90	1.11	0.84	1.43
2013	0.84	2.39	0.89	1.18	0.82	1.37
2014	0.84	2.28	0.90	1.18	0.83	1.43
2015	0.82	2.32	0.89	1.11	0.82	1.50
2016	0.81	2.05	0.89	1.06	0.80	1.28
2017	0.84	2.61	0.88	1.13	0.83	1.31
2018	0.84	3.00	0.88	1.24	0.85	1.39
2019	0.83	1.95	0.90	1.04	0.82	1.24
Total	0.87	2.66	0.93	1.19	0.85	2.05

4. Discussion

In this study, we developed a spatiotemporal modeling framework with hierarchical approaches to combine regionalized predictions into a national scale. The model incorporated $PM_{2.5}$ observational datasets from the AQS monitoring sites and existing fine-scale non-regulatory monitors (non-regulatory: residential and fixed-site), as well as multiple predictor variables. The modeling approach performed consistently well over a 20-year period at the AQS sites. This approach allows the estimation of both short-term and long-term exposures to $PM_{2.5}$ for epidemiological studies.

Predicting national $PM_{2.5}$ concentrations across the contiguous United States has been reported in several studies with various modeling approaches and varying model performances (Hu et al., 2017; Di et al., 2019; Jerrett et al., 2017). Notably, a recent study by Di. et al. (2019) (Di et al., 2019) reported a new generation $PM_{2.5}$ model based on their existing model (Di et al., 2017) which included sophisticated ensemble-based machine learning approach with satellite-derived AOD, land-use variables, chemical transport model predictions, and several meteorological variables as major predictor variables. The model included daily $PM_{2.5}$ observations from the available AQS monitoring sites and showed an overall cross-validation spatial R^2 of 0.89. In comparison, our model performed better in terms of the cross-validation spatial R^2 of 0.93 at the AQS sites and showed a similar declining trend of prediction ability towards the most recent years. This trend likely reflects the reduction of variability in $PM_{2.5}$ concentrations over time affected by the success of national long-term air pollution regulations as well as the fact that several geographical variables (e.g., land use types, emissions, road networks) may better reflect historical exposure than those from more recent years (see Fig. S8 for comparison of prediction maps). Furthermore, both of the studies showed consistently higher cross-validation R^2 in the north eastern regions, perhaps because of better coverage there of air monitors (though the definition of the sub-regions may differ). Our model outperformed many other national and regional $PM_{2.5}$ models in the United States when taking the spatiotemporal coverage, the resolution and the performances into account (Hu et al., 2017; Beckerman et al., 2013; van et al., 2019; Reid

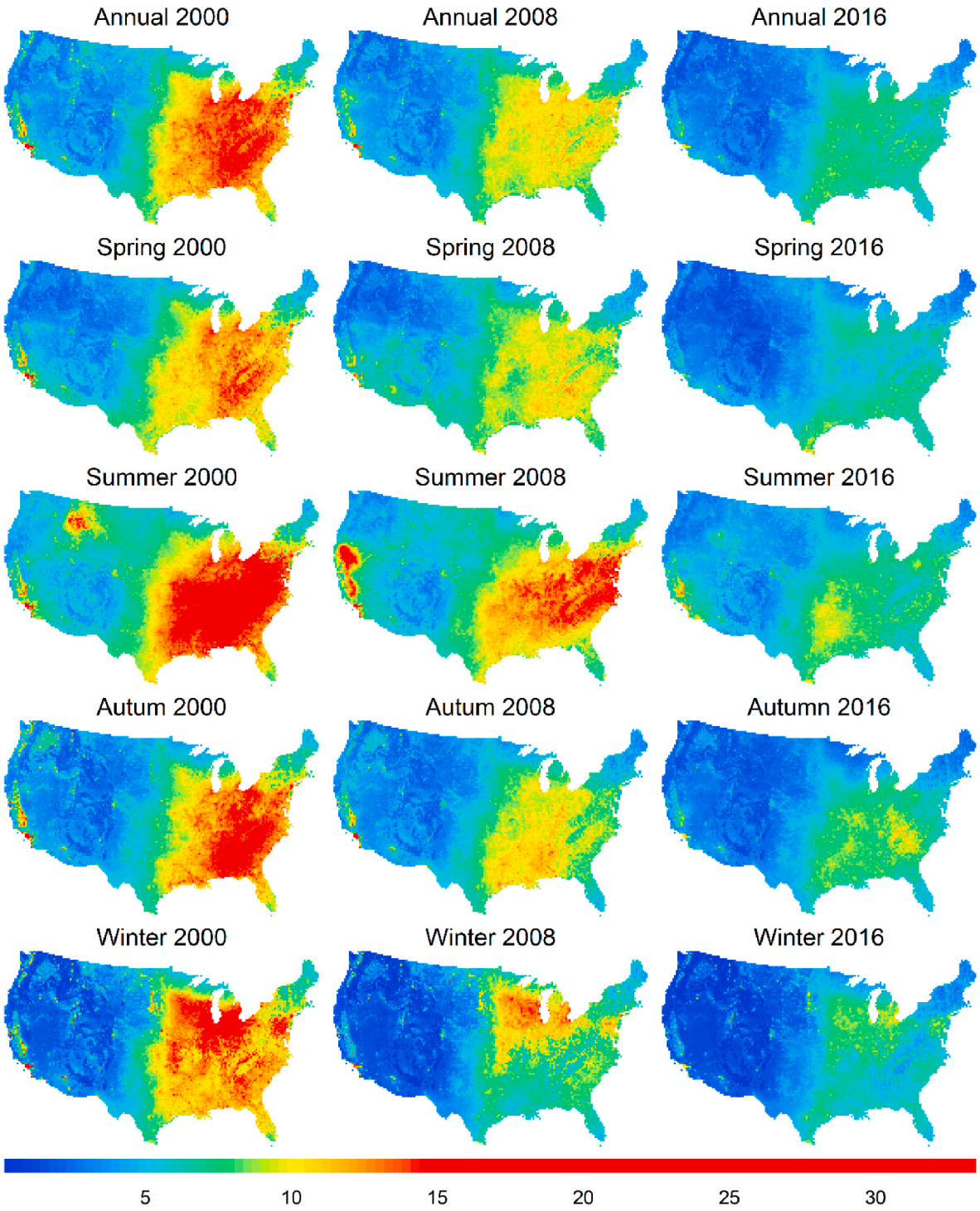


Fig. 1. Annual and seasonal average spatial mapping of $PM_{2.5}$ in 3 selected years (i.e., 2000, 2008, and 2016) based on the best regional model predictions stitched to the national scale in the United States. The best model included all the monitoring sites from both the AQS and non-regulatory monitors. The color scales for all the maps are the same. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

et al., 2021; Just et al., 1994; Brokamp, 2022; Swanson et al., 2022).

Our hierarchical spatiotemporal approach to modeling $PM_{2.5}$ at a national scale has several strengths. First, our statistical model decomposes the spatiotemporal air pollution field into spatially varying long-term averages and temporal trends (Lindstrom et al., 2014; Sampson et al., 2011; Szpiro et al., 2009). This enables us to leverage highly unbalanced monitoring data to learn about both spatial and temporal variability at all locations, which is especially important for incorporating short-term (e.g., two-week average) research monitoring data at home sites that better represents locations where study subjects live than long-term regulatory monitoring data (Cohen et al., 2009; Kirwa et al., 2021; Keller et al., 2017). As with most LUR models, our predictions are at points, meaning that predictions can be made at arbitrarily precise geolocated coordinates, and those predictions reflect the available information from fine-scale geographic features (Fig. 2). Although some studies applied downscaling approach to refine exposure at finer scale, an advantage of our approach is that our model is intrinsically at the point location scale, which means it uses all available

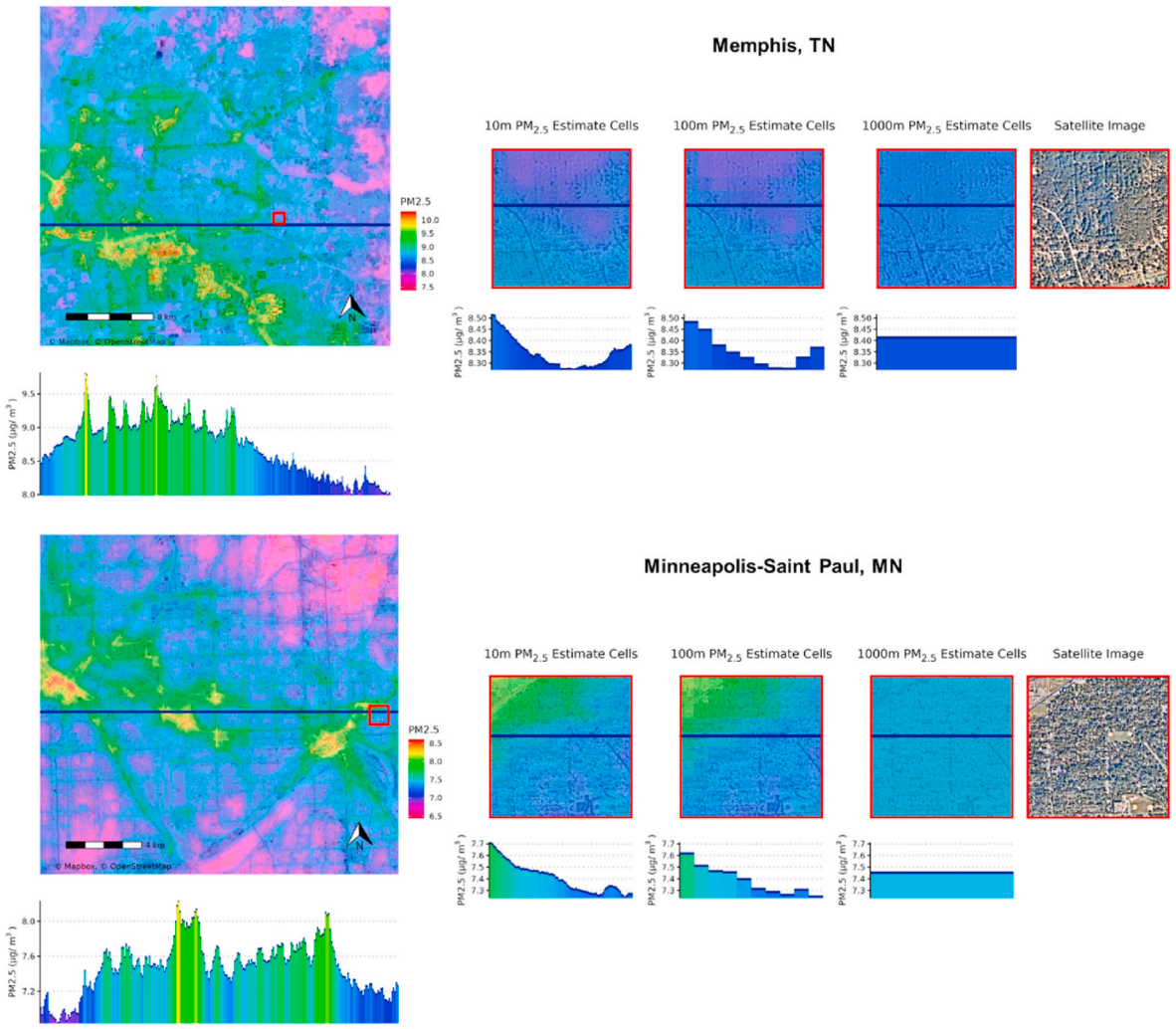


Fig. 2. Annual average predictions of $PM_{2.5}$ modeled with all sites (AQS + non-regulatory sites) in 2015 in Memphis, TN and Minneapolis-Saint Paul, MN and comparison of model predictions in a small area (red box) with spatial resolutions of 10m, 100m, and 1000m. The transect plots indicate the spatial variations of $PM_{2.5}$ concentrations across the blue lines in the cities and the small grid cells. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

information to optimize performance at that scale (Di et al., 2019; Huang et al., 2021). Our model incorporates reduced-dimension temporal basis functions and PLS loadings as inputs into a parametric modeling framework, which enables our model to flexibly learn from complex relationships among hundreds of geographic covariates temporal patterns at thousands of monitoring locations while also allowing for diagnostic assessment of variable importance and model structure that ensure face validity of the predictions. Finally, by decomposing the United States into regions, our model exploits national-scale variability in pollution sources, chemical composition, topography, and meteorology that lead to different regional predictive models that can be combined to optimize regional and national prediction accuracy.

Our study has some noted differences compared to the other studies. Unlike some other recent papers, we do not directly use satellite measures of aerosol optical depth (AOD) and the relation between AOD and $PM_{2.5}$[11] as a primary predictor variable (Di et al., 2019). Some of the challenges in using AOD include informative missingness due to meteorology (e.g., missing data in cloudy days) as well as relatively large spatial scale (a km of more). Instead, our model focuses on learning the spatial structure available for AOD from hundreds of geographic covariates (see Table S1), a large database of publicly-available regulatory and research monitoring data, geostatistical smoothing of long-term averages and spatially varying temporal trends, and satellite-based estimates of other features including greenspace and NO_2 concentrations. In addition, our modeling is conducted at a two-week timescale, which is necessary to include research monitoring including filter-based sites on a one-in-three schedule and home sites that measure two-week averages. While there is some loss of temporal information compared to daily monitoring or daily estimates, exposures of interest for many epidemiological studies range from months to years, and there is significant benefit in being able to incorporate the monitoring data at home locations, which cannot readily be incorporated in a daily model. Finally, our regionalized modeling approach require combining estimates from multiple models near regional boundaries. We do this using a version of inverse-distance weighting that ensures smooth transitions between regions. Lastly, our model underestimates $PM_{2.5}$ concentrations in areas that are frequently impacted by wildfires due to lack of adequate information for wildfires. Future plan will be to incorporate fire emissions and trajectory data in the model.

In conclusion, using rich monitoring data from AQS sites and dedicated, non-regulatory spatial monitoring campaigns, we were able to develop long-term spatiotemporal models for $PM_{2.5}$ that perform well in terms of both prediction accuracy and precision in the contiguous United States from 2000 to 2019.

CRediT authorship contribution statement

Meng Wang: Writing – original draft, Visualization, Methodology, Formal analysis. **Michael Young:** Writing – review & editing, Formal analysis, Data curation. **Julian D. Marshall:** Writing – review & editing, Data curation. **Logan Piepmeier:** Writing – review & editing, Visualization, Formal analysis. **Jianzhao Bi:** Writing – review & editing. **Joel D. Kaufman:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Adam A. Szpiro:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This publication was developed under a STAR research assistance agreements RD831697 (MESA Air), RD-83830001 (MESA Air Next Stage), RD83479601 (UW Center for Clean Air Research), and R83374101 (MESA Coarse), awarded by the U.S. Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication. Research reported in this publication was also supported by ECHO PATHWAYS (NIH grants: 1UG3OD023271-01 and 4UH3OD023271-03), and by grants R56ES026528, R01ES023500, R01ES02588, and P30ES007033 from NIEHS and P01AG055367 from NIA. Funding was also provided by Kresge Foundation Grant No. 243365.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2024.125405.

Data availability

Data will be made available on request.

References

Abdi, H., Williams, L.J., 2013. Partial least squares methods: partial least squares correlation and partial least square regression. Methods Mol. Biol. 930, 549–579.

Bechle, M.J., Millet, D.B., Marshall, J.D., 2015. National spatiotemporal exposure surface for NO2: monthly scaling of a satellite-derived land-use regression, 2000-2010. Environ. Sci. Technol. 49 (20), 12297–12305.

Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.J., van Donkelaar, A., Ross, Z., Su, J., Burnett, R.T., 2013. A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States. Environmental science & technology 47 (13), 7233–7241.

Brokamp, C., 2022. A high resolution spatiotemporal fine particulate matter exposure assessment model for the contiguous United States. Environmental Advances 7, 100155.

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A.H., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Vermeulen, R., Brunekreef, B., Hoek, G., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environ. Int. 130, 104934.

Cohen, M.A., Adar, S.D., Allen, R.W., Avol, E., Curl, C.L., Gould, T., Hardie, D., Ho, A., Kinney, P., Larson, T.V., Sampson, P., Sheppard, L., Stukovsky, K.D., Swan, S.S., Liu, L.J., Kaufman, J.D., 2009. Approach to estimating participant pollutant exposures in the multi-ethnic study of atherosclerosis and air pollution (MESA air). Environmental science & technology 43 (13), 4687–4693.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C.A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C.J.L., Forouzanfar, M.H., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 389 (10082), 1907–1918.

de Hoogh, K., Gulliver, J., Donkelaar, A.V., Martin, R.V., Marshall, J.D., Bechle, M.J., Cesaroni, G., Pradas, M.C., Dedele, A., Eeftens, M., Forsberg, B., Galassi, C., Heinrich, J., Hoffmann, B., Jacquemin, B., Katsouyanni, K., Korek, M., Kunzli, N., Lindley, S.J., Lepeule, J., Meleux, F., de Nazelle, A., Nieuwenhuijsen, M., Nystad, W., Raaschou-Nielsen, O., Peters, A., Peuch, V.H., Rouil, L., Udvardy, O., Slama, R., Stempfelet, M., Stephanou, E.G., Tsai, M.Y., Yli-Tuomi, T., Weinmayr, G., Brunekreef, B., Vienneau, D., Hoek, G., 2016. Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10.

de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Klompmaker, J., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., Hoek, G., 2018. Spatial PM2.5, NO2, O3 and BC models for western Europe - evaluation of spatiotemporal stability. Environ. Int. 120, 81–92.

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Dominici, F., Schwartz, J.D., 2017. Air pollution and mortality in the medicare population. N. Engl. J. Med. 376 (26), 2513–2522.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz, J., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909.

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dedele, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Grazuleviciene, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Molter, A., Nador, G., Nieuwenhuijsen, M., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E., Sugiri, D., Tsai, M.Y., Yli-Tuomi, T., Varro, M.J., Vienneau, D., Klot, S., Wolf, K., Brunekreef, B., Hoek, G., 2012. Development of Land Use Regression models for PM (2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. Environmental science & technology 46 (20), 11195–11205.

Hansel, N.N., Paulin, L.M., Gassett, A.J., Peng, R.D., Alexis, N., Fan, V.S., Bleecker, E., Bowler, R., Comellas, A.P., Dransfield, M., Han, M.K., Kim, V., Krishnan, J.A., Pirozzi, C., Cooper, C.B., Martinez, F., Woodruff, P.G., Breysse, P.J., Barr, R.G., Kaufman, J.D., 2017. Design of the Subpopulations and intermediate Outcome measures in COPD (SPIROMICS) AIR study. BMJ open respiratory research 4 (1), e000186.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM2.5 concentrations in the conterminous United States using the random forest approach. Environmental science & technology 51 (12), 6936–6944.

Huang, C., Hu, J., Xue, T., Xu, H., Wang, M., 2021. High-resolution spatiotemporal modeling for ambient PM(2.5) exposure assessment in China from 2013 to 2019. Environmental science & technology 55 (3), 2152–2162.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. J. Expo. Anal. Environ. Epidemiol. 15 (2), 185–204.

Jerrett, M., Turner, M.C., Beckerman, B.S., Pope, C.A., van Donkelaar, A., Martin, R.V., Serre, M., Crouse, D., Gapstur, S.M., Krewski, D., Diver, W.R., Coogan, P.F., Thurston, G.D., Burnett, R.T., 2017. Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. Environmental health perspectives 125 (4), 552–559.

Just, A.C., Arfer, K.B., Rush, J., Dorman, M., Shtein, A., Lyapustin, A., Kloog, I., 1994. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM(2.5)) using satellite data over large regions. Atmos. Environ. 2020, 239.

Kaufman, J.D., Adar, S.D., Barr, R.G., Budoff, M., Burke, G.L., Curl, C.L., Daviglus, M.L., Diez Roux, A.V., Gassett, A.J., Jacobs Jr., D.R., Kronmal, R., Larson, T.V., Navas-Acien, A., Olives, C., Sampson, P.D., Sheppard, L., Siscovick, D.S., Stein, J.H., Szpiro, A.A., Watson, K.E., 2016. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. Lancet 388 (10045), 696–704.

Keller, J.P., Olives, C., Kim, S.Y., Sheppard, L., Sampson, P.D., Szpiro, A.A., Oron, A.P., Lindstrom, J., Vedal, S., Kaufman, J.D., 2015. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. Environmental health perspectives 123 (4), 301–309.

Keller, J.P., Chang, H.H., Strickland, M.J., Szpiro, A.A., 2017. Measurement error correction for predicted spatiotemporal air pollution exposures. Epidemiology 28 (3), 338–345.

Kirwa, K., Szpiro, A.A., Sheppard, L., Sampson, P.D., Wang, M., Keller, J.P., Young, M.T., Kim, S.Y., Larson, T.V., Kaufman, J.D., 2021. Fine-scale air pollution models for epidemiologic research: insights from approaches developed in the multi-ethnic study of atherosclerosis and air pollution (MESA air). Curr Environ Health Rep 8 (2), 113–126.

Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2012. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the Mid-Atlantic states. Environmental science & technology 46 (21), 11913–11921.

Knibbs, L.D., van Donkelaar, A., Martin, R.V., Bechle, M.J., Brauer, M., Cohen, D.D., Cowie, C.T., Dirgawati, M., Guo, Y., Hanigan, I.C., Johnston, F.H., Marks, G.B., Marshall, J.D., Pereira, G., Jalaludin, B., Heyworth, J.S., Morgan, G.G., Barnett, A.G.,

2018. Satellite-based land-use regression for continental-scale long-term ambient PM2.5 exposure assessment in Australia. Environmental science & technology 52 (21), 12445–12455.
Laurent, O., Hu, J., Li, L., Kleeman, M.J., Bartell, S.M., Cockburn, M., Escobedo, L., Wu, J., 2016. Low birth weight and air pollution in California: which sources and components drive the risk? Environ. Int. 92–93, 471–477.
Lepeule, J., Litonjua, A.A., Coull, B., Koutrakis, P., Sparrow, D., Vokonas, P.S., Schwartz, J., 2014. Long-term effects of traffic particles on lung function decline in the elderly. Am. J. Respir. Crit. Care Med. 190 (5), 542–548.
Lindstrom, J., Szpiro, A.A., Sampson, P.D., Oron, A.P., Richards, M., Larson, T.V., Sheppard, L., 2014. A flexible spatio-temporal model for air pollution with spatial and spatio-temporal covariates. Environ. Ecol. Stat. 21 (3), 411–433.
Miller, K.A., Siscovick, D.S., Sheppard, L., Shepherd, K., Sullivan, J.H., Anderson, G.L., Kaufman, J.D., 2007. Long-term exposure to air pollution and incidence of cardiovascular events in women. N. Engl. J. Med. 356 (5), 447–458.
Reid, C.E., Considine, E.M., Maestas, M.M., Li, G., 2021. Daily PM(2.5) concentration estimates by county, ZIP code, and census tract in 11 western states 2008-2018. Sci. Data 8 (1), 112.
Ryan, P.H., LeMasters, G.K., 2007. A review of land-use regression models for characterizing intraurban air pollution exposure. Inhal. Toxicol. 19 (Suppl. 1), 127–133.
Sampson, P.D., Szpiro, A.A., Sheppard, L., Lindstrom, J., Kaufman, J.D., 2011. Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. Atmos. Environ. 45 (36), 6593–6606.
Swanson, A., Holden, Z.A., Graham, J., Warren, D.A., Noonan, C., Landguth, E., 2022. Daily 1 km terrain resolving maps of surface fine particulate matter for the western United States 2003-2021. Sci. Data 9 (1).
Szpiro, A.A., Sampson, P.D., Sheppard, L., Lumley, T., Adar, S.D., Kaufman, J., 2009. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. Environmetrics 21 (6), 606–631.
van Donkelaar, A., Martin, R.V., Li, C., Burnett, R.T., 2019. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. Environmental science & technology 53 (5), 2595–2611.
Vienneau, D., de Hoogh, K., Bechle, M.J., Beelen, R., van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G., Marshall, J.D., 2013. Western European land use regression incorporating satellite- and ground-based measurements of NO2 and PM10. Environ. Sci. Technol. 47 (23), 13555–13564.
Wang, M., Beelen, R., Bellander, T., Birk, M., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., Declercq, C., Dimakopoulou, K., Eeftens, M., Eriksen, K.T., Forastiere, F., Galassi, C., Grivas, G., Heinrich, J., Hoffmann, B., Ineichen, A., Korek, M., Lanki, T., Lindley, S., Modig, L., Molter, A., Nafstad, P., Nieuwenhuijsen, M.J., Nystad, W., Olsson, D., Raaschou-Nielsen, O., Ragettli, M., Ranzi, A., Stempfelet, M., Sugiri, D., Tsai, M.Y., Udvardy, O., Varro, M.J., Vienneau, D., Weinmayr, G., Wolf, K., Yli-Tuomi, T., Hoek, G., Brunekreef, B., 2014. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. Environmental health perspectives 122 (8), 843–849.
Wang, M., Gehring, U., Hoek, G., Keuken, M., Jonkers, S., Beelen, R., Eeftens, M., Postma, D.S., Brunekreef, B., 2015. Air pollution and lung function in Dutch children: a comparison of exposure estimates and associations based on land use regression and dispersion exposure modeling approaches. Environmental health perspectives 123 (8), 847–851.
Wang, M., Sampson, P.D., Hu, J., Kleeman, M., Keller, J.P., Olives, C., Szpiro, A.A., Vedal, S., Kaufman, J.D., 2016. Combining land-use regression and chemical transport modeling in a spatiotemporal geostatistical model for ozone and PM2.5. Environmental science & technology 50 (10), 5111–5118.
Xiao, Q., Chang, H.H., Geng, G., Liu, Y., 2018. An ensemble machine-learning model to predict historical PM2.5 concentrations in China from satellite data. Environmental science & technology 52 (22), 13260–13269.
Xu, H., Bechle, M.J., Wang, M., Szpiro, A.A., Vedal, S., Bai, Y., Marshall, J.D., 2019. National PM2.5 and NO2 exposure models for China based on land use regression, satellite measurements, and universal kriging. The Science of the total environment 655, 423–433.
Young, M.T., Bechle, M.J., Sampson, P.D., Szpiro, A.A., Marshall, J.D., Sheppard, L., Kaufman, J.D., 2016. Satellite-based NO2 and model validation in a national prediction model based on universal kriging and land-use regression. Environmental science & technology 50 (7), 3686–3694.