Scientific Research Publishing

Open Journal of Air Pollution, 2016, 5, 95-108
http://www.scirp.org/journal/ojap
ISSN Online: 2169-2661
ISSN Print: 2169-2653

National Prediction of Ambient Fine Particulates: 2000-2009

David J. Shavlik[1], Sam Soret[1], W. Lawrence Beeson[2], Mark G. Ghamsary[1], Synnove F. Knutsen[2]

[1]Center for Community Resilience, School of Public Health, Loma Linda University, Loma Linda, USA
[2]Center for Nutrition, Healthy Lifestyle and Disease Prevention, School of Public Health, Loma Linda University, Loma Linda, USA
Email: dshavlik@llu.edu, ssoret@llu.edu, lbeeson@llu.edu, mghamsary@gmail.com, sknutsen@llu.edu

How to cite this paper: Shavlik, D.J., Soret, S., Beeson, W.L., Ghamsary, M.G. and Knutsen, S.F. (2016) National Prediction of Ambient Fine Particulates: 2000-2009. *Open Journal of Air Pollution*, **5**, 95-108.
http://dx.doi.org/10.4236/ojap.2016.53008

Received: July 26, 2016
Accepted: August 29, 2016
Published: September 1, 2016

Abstract

A large body of evidence links ambient fine particulates ($PM_{2.5}$) to chronic disease. Efforts continue to be made to improve large scale estimation of this pollutant for within-urban environments and sparsely monitored areas. Still questions remain about modeling choices. The purpose of this study was to evaluate the performance of spatial only models in predicting national monthly exposure estimates of fine particulate matter at different time aggregations during the time period 2000-2009 for the contiguous United States. Additional goals were to evaluate the difference in prediction between federal reference monitors and non-reference monitors, assess regional differences, and compare with traditional methods. Using spatial generalized additive models (GAM), national models for fine particulate matter were developed, incorporating geographical information systems (GIS)-derived covariates and meteorological variables. Results were compared to nearest monitor and inverse distance weighting at different time aggregations and a comparison was made between the Federal Reference Method and all monitors. Cross-validation was used for model evaluation. Using all monitors, the cross-validated R^2 was 0.76, 0.81, and 0.82 for monthly, 1 year, and 5-year aggregations, respectively. A small decrease in performance was observed when selecting Federal Reference monitors only (R^2 = 0.73, 0.78, and 0.80 respectively). For Inverse distance weighting (IDW), there was a significantly larger decrease in R^2 (0.68, 0.71, and 0.73, respectively). The spatial GAM showed the weakest performance for the northwest region. In conclusion, National exposure estimates of fine particulates at different time aggregations can be significantly improved over traditional methods by using spatial GAMs that are relatively easy to produce. Furthermore, these models are comparable in performance to other national prediction models.

Keywords

Long-Term Air Pollution, GAM, Prediction, Fine Particulates

1. Introduction

Evidence has been accumulating from long-term and short-term studies linking ambient air pollution to health outcomes. Fine particulate matter has demonstrated fairly consistent links to cardiovascular and respiratory health [1]-[9]. As a consequence, effort continues to be made to improve large-scale estimation of air pollutants for intra-urban environments and sparsely monitored areas using different methods. Generalized additive models (GAM), which capture spatiotemporal trends, have been used to predict monthly PM_{10}, $PM_{2.5}$, and coarse fraction PM, across the north-eastern US [10]-[12], and nationally [13]. Using Bayesian maximum entropy (BME) one study estimated average $PM_{2.5}$ for the year 2003 [14], and another study used BME in a hybrid approach to estimate national monthly $PM_{2.5}$ [15]. A geographically weighted regression was used to estimate $PM_{2.5}$ in the southeastern US for the year 2003 [16]. In western Europe, land-use regression (LUR) models were developed to predict PM_{10} and NO_2 that included satellite data for the years 2005-2007 [17].

Our goal in this study was to evaluate how spatial GAMs perform and to create national monthly predictions of $PM_{2.5}$ in order to evaluate the effect of aggregating on the monthly, 1-year, and 5-year time metrics for the years 2000 to 2009, to assess different exposure windows. It is also of interest to compare the effect of using a non-reference method (NRM) in contrast to the Federal Reference Method (FRM) alone [18]. Finally, nearest monitor and inverse distance weighted (IDW) methods were used to contrast suggested approaches. The results of these analyses are useful for national epidemiologic investigations of long-term exposure to air pollution that require a moving exposure window with different exposure intervals such as is the case for cardiovascular disease [7].

2. Methods

Spatial monthly generalized additive models (GAM) were developed to predict $PM_{2.5}$ for the contiguous US for the period 2000-2009. These models are a sum of smooth functions of predictor variables that allow for flexible non-linear relationships with the outcome variable [19]. In this research the outcome variable is $PM_{2.5}$ monitoring data from the US Environmental Protection Agency's (EPA) national ambient air quality monitoring network. Meteorological and GIS-derived variables were used as potential predictors. The resulting models makes it possible to predict ambient $PM_{2.5}$ levels at any specified spatial resolution used to represent subject locations, such as street-level residential addresses.

2.1. $PM_{2.5}$ Monitoring Data

Daily and hourly $PM_{2.5}$ data were downloaded from the US EPA's Air Quality System (AQS) online database for the period 2000-2009 [18]. Parameter code 88101 was used for the FRM. The NRM used parameter code 88502 which is comparable to FRM with less than 10% bias [20]. The NRM included the Interagency Monitoring of Protected Visual Environments (IMPROVE) network which monitors national parks and captures $PM_{2.5}$ data in rural settings [21]. There were three inclusion criteria applied to the

air pollution data. First, for hourly $PM_{2.5}$ data, at least 75% of the hourly values during a day must be available to consider them valid for aggregation to daily amounts. Second, only months could be included where observed counts of valid daily values per month, divided by expected counts of daily values per month, were at least 75%. Last, there must be at least three valid observations in a given month for that monthly average to be included. A square root transformation was applied to the $PM_{2.5}$ data and these were back transformed once predictions were made. Performance evaluation was done on back transformed data.

2.2. Potential Predictors

Candidate predictors were either GIS derived or meteorological. The following variables were GIS derived: Land use maps from the Multi-Resolution Land Characteristics Consortium allowed calculations of the proportion of imperviousness (2006 data) and tree canopy (2001 data) within 1 km, 2 km, 4 km, and 8 km buffers [22]. The proportion of land use (2006 data) was also calculated for high, medium, and low intensities, and cultivated crops within 1 km, 2 km, and 4 km buffers. US census databases for the year 2000 were used to determine block group population density within 1 km, 2 km, 4 km, and 8 km buffers (Environmental Sciences Research Institute-(ESRI), Redlands, CA). Elevation was extracted from the US Geological Survey (USGS) National Elevation Dataset [23]. US EPA National Emissions Inventory was used for point-source emissions of $PM_{2.5}$ within 1 km, 2 km, 4 km, 8 km, and 12 km buffers (2005 data) [24]. Distance to nearest road by 4 road classes was measured using ESRI's Street Map database (2011 data). Indicator variables were created for the South Coast and San Joaquin Valley air basins in California because of the unique characteristics of the regions and the high number of extreme values [25]. All of the above GIS determined variables were created in Arc Map 10.0 (ESRI), and except for elevation, all of the continuous GIS variables were log transformed.

National meteorological data including average temperature, maximum temperature, dew point temperature, relative humidity, barometric pressure, air stagnation, and wind speed, were downloaded from the National Climate Data Center for the study time period [26].

2.3. Subgroups

Six US regions were defined in this research to evaluate model performance in geographic subgroups (see **Figure 1**). They were the northwest (NW = OR, WA, ID), southwest (SW = CA, NV, UT, AZ), north central (NC = MT, WY, ND, SD,NE, MN, IA), south central (SC = CO, NW, KS, OK, TX, MO, AR, LA), northeast (NE = WI, MI, IL, IN, OH, KY, WV,VA, PA, NY, VT, NH, ME, MA, CT, RI, MD, DE, NJ, and DC), and southeast (SE = TN, MS, AL, NC, SC, GA, and FL).

2.4. Generalized Additive Models

Separate spatial only models were created for each of the 120 months during the 10

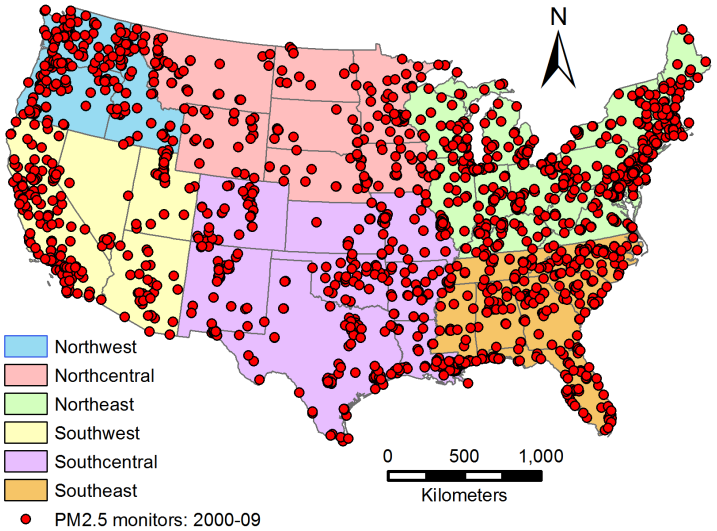

Figure 1. $PM_{2.5}$ monitoring locations and the six regions of the US defined in this study.

year study period (1/2000 to 12/2009). The model was defined as:

$$y_{i,t} = \alpha + c(z_i) + \sum_{j=1}^{J} d_j(x_{i,j}) + \sum_{k=1}^{K} f_k(w_{i,k,t}) + e_{i,t} \quad (1)$$

$$e_{i,t} \sim N(0, \sigma_e^2); t = 1, \cdots, T$$

where $y_{i,t}$ is the square-root transformation of $PM_{2.5}$ for monitoring site i and month t. The function $c()$ is a bivariate thin plate regression spline for the spatial coordinates z_i for each monitoring location. The functions $d()$ and $f()$ are a single cubic regression spline for $j = 1 \ldots J$ time invariant predictors x_j, and for $k = 1 \ldots K$ time varying predictors w_k, respectively. GAM procedures used the gam () function from the *mgcv* package from R [19]. When modeling GAMs, the maximum degree of flexibility of the spline functions can be set by the user by choosing the basis dimension. For the spatial coordinates function the basis dimension was set at 300, and for time invariant and time varying predictors the basis dimension was set at 10. These models were compared to the traditional methods IDW and the nearest monitor approach.

2.5. Model Building

The data were divided into 10 random sets. Cross-validation was used to select covariates and assess model predictive ability within data sets 1 - 9 by taking each out in turn and predicting to the held out set. The computed squared correlation (regression based R^2) between the observed and predicted values helped determined covariate and model choice along with root-mean-squared prediction error (RMSPE). Because selecting variables by cross-validation could bias the performance of the cross-validated statistics, the 10^{th} set was not used for variable selection, but was held out until after the final model was selected to evaluate evidence of over-fitting of this final model [10] [11]. Performance was further evaluated by: 1) the average mean prediction error (MPE), 2) the

regression slope where observed is the dependent variable and the predicted value is the independent variable, and 3) by a mean squared error based R^2 (MSE-R^2) which is a measure of fit to the 1-1 line, defined as MSE-R^2 = 1-RMSPE^2/var (obs) [27].

The first step in model building was to force spatial coordinates of monitoring location into the model (assumed *a-priori* to be the most important predictor of $PM_{2.5}$). The second step added additional covariates that gave the highest cross-validated R^2 and lowest RMSPE in datasets 1 - 9. In the last step covariates were retained only if they improved the R^2 by approximately 1% or more [28]. Cross-validated performance statistics were reported on a monthly basis (calculating statistics for each month and then averaging statistics across 120 months), a 1 year aggregation (averaging statistics across 10 years), a 5 year aggregation (averaging statistics across 2 time periods: 2000-2004, 2005-2009), and no aggregation (e.g. reporting a single R^2 for datasets 1 - 9). Once a final model was selected then performance statistics from datasets 1 - 9 were compared to the holdout dataset 10 to see if there was any evidence of over fitting.

3. Results

3.1. Descriptive Statistics

During the study period all of the monitors (including non-reference) located on the east coast had on average the highest levels of $PM_{2.5}$ (NE = 12.7 and SE = 12.4 μg/m^3) and the northwest had the lowest (7.4 μg/m^3) as seen in Table 1. The highest peak was in the north central region (88.0 μg/m^3). It is however more instructive to look at the

Table 1. Descriptive statistics of mean monthly $PM_{2.5}$ (μg/m^3) by region and season for combined monitors and federal reference method only.

Subgroup	$PM_{2.5}$ μg/m^3							$PM_{2.5}$ μg/m^3 (FRM[a])						
	Mean	SD	Min	5%	95%	Max	N	Mean	SD	Min	5%	95%	Max	N
Total	10.9	5.0	0.0	3.6	19.7	88.0	1892	11.8	4.8	0.0	5.1	20.3	88.0	1580
NW[b]	7.4	4.6	0.3	2.1	16.5	50.5	201	8.8	5.0	0.8	3.7	19.3	50.5	134
SW[c]	9.8	6.9	0.0	2.2	23.8	69.5	224	11.3	6.9	0.0	4.3	25.5	69.5	165
NC[d]	8.0	4.2	0.3	2.2	14.9	88.0	198	8.8	4.0	1.2	4.0	15.4	88.0	161
SC[e]	10.0	3.9	0.5	3.8	16.5	82.5	353	10.6	3.7	0.7	5.0	17.0	82.5	280
NE[f]	12.7	4.4	1.6	6.3	20.7	68.4	621	13.0	4.3	1.6	6.8	20.8	68.4	577
SE[g]	12.4	4.1	1.8	6.8	20.0	48.0	295	12.6	4.1	2.5	7.1	20.2	48.0	263
Mar-May	9.8	4.1	0.0	3.6	16.7	48.0	1832	10.5	3.8	0.0	4.7	17.0	48.0	1535
Jun-Aug	12.3	5.5	0.0	4.4	21.8	88.0	1829	13.2	5.3	0.0	5.4	22.1	88.0	1524
Sep-Nov	10.5	4.8	0.0	3.6	18.7	82.5	1848	11.3	4.7	0.0	5.1	19.2	82.5	1543
Dec-Feb	11.1	5.3	0.0	2.4	19.8	65.3	1866	12.1	4.8	0.0	5.7	20.4	65.3	1562

[a]Federal reference method only; [b]OR, WA, & ID; [c]CA, NV, UT, & AZ; [d]MT, WY, ND, SD, NE, MN, & IA; [e]CO, NM, KS, OK, TX, MO, AR, & LA; [f]WI, MI, IL, IN, OH, KY, WV, VA, PA, NY, VT, NH, ME, MA, CT, RI, MD, DE, NJ, & DC; [g]TN, MS, AL, NC, SC, GA, & FL.

95th percentile where the north central region had the lowest value (15.0 μg/m^3) and the southwest had the highest value (23.2 μg/m^3). Non-reference monitors had a much lower overall mean (6.8 μg/m^3) when compared to the reference method (11.8 μg/m^3) (Table 2). The non-reference monitoring locations tend to be in rural areas and the largest number of these (n = 133), were found in the northwest which may help explain the lowest average levels observed for this area (Table 2).

Temporal trends for combined monitors and FRM show a steady decline in average $PM_{2.5}$ and the 95th percentile over the study period (see Figure 2). Non-reference monitors show an increase through 2007 followed by a decline. Seasonally, it appears that $PM_{2.5}$ has highs in the summer months and in the winter months, with July and August having the highest peaks followed by January. April has the lowest valley with a dramatic drop for the 95th percentile (see Figure 3). The number of active monitors was not very different for the different seasons (Table 1).

3.2. Prediction

The final model contained only variables that improved the cross-validation R^2 among data sets 1 - 9 by approximately 1% or more. The retained variables were monitoring spatial coordinates and two GIS predictors; population block group density at a 4-km buffer and elevation.

When reporting overall performance results, we chose to report performance statistics four ways to reflect different time aggregations. For the spatial GAMs, that included all monitors, performance was strong with an R^2 = 0.80, 0.76, 0.81 and 0.82 for categories *None*, *Month*, 1-*Year*, and 5-*Year* time aggregations, respectively (Table 3). The first

Table 2. Descriptive statistics of mean monthly NRM[a] $PM_{2.5}$ (μg/m^3) by region and season.

Subgroup	$PM_{2.5}$ μg/m^3 NRM[a]						
	Mean	SD	Min	5%	95%	Max	N
Total	6.8	4.1	0.3	1.5	14.3	42.2	431
NW[b]	6.2	3.7	0.3	1.6	13.5	42.2	133
SW[c]	4.9	3.4	0.3	1.1	11.2	36.1	63
NC[d]	4.7	3.4	0.3	0.9	11.1	35.4	42
SC[e]	7.9	3.9	0.5	1.9	14.4	29.1	95
NE[f]	8.5	4.0	1.8	3.3	15.9	26.5	58
SE[g]	10.3	3.8	1.8	5.1	17.5	29.1	40
Mar-May	6.4	3.5	0.6	1.9	12.8	24.7	400
Jun-Aug	7.9	4.3	1.3	3.0	16.1	40.4	413
Sep-Nov	6.7	3.8	0.4	1.9	13.8	34.0	412
Dec-Feb	6.2	4.4	0.3	0.8	14.2	42.2	415

[a]Non-federal reference method; [b]OR, WA, & ID; [c]CA, NV, UT, & AZ; [d]MT, WY, ND, SD, NE, MN, & IA; [e]CO, NM, KS, OK, TX, MO, AR, & LA; [f]WI, MI, IL, IN, OH, KY, WV, VA, PA, NY, VT, NH, ME, MA, CT, RI, MD, DE, NJ, & DC; [g]TN, MS, AL, NC, SC, GA, & FL.

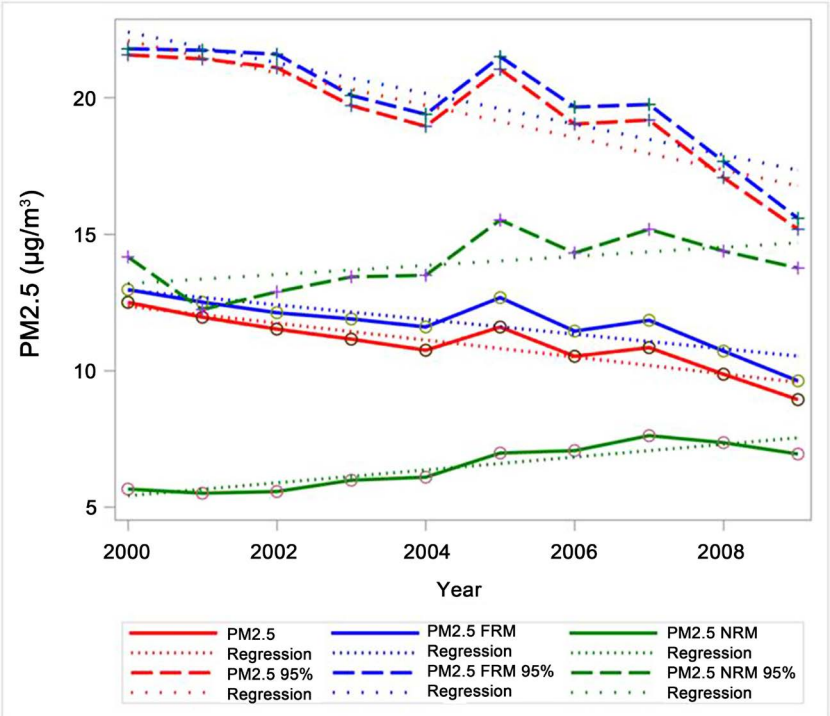

Figure 2. Average and upper 95th percentile Yearly trends for $PM_{2.5}$ including all monitors, federal reference method (FRM), non-reference method (NRM), and regression lines, for monitoring locations during the period 2000 to 2009.

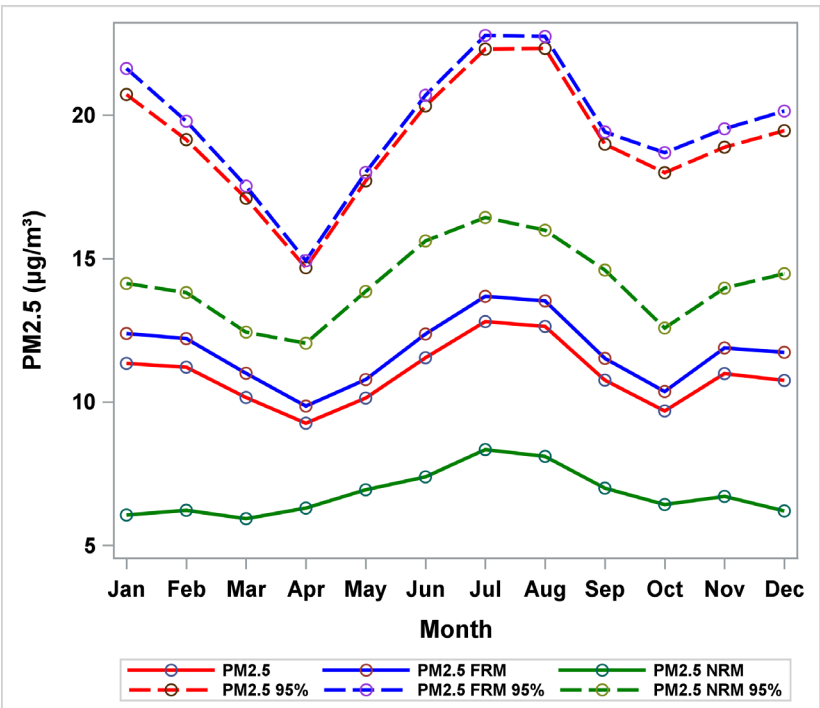

Figure 3. Average and upper 95th percentile Monthly seasonal trends for $PM_{2.5}$ including all monitors, federal reference method (FRM), and non-reference method (NRM), for monitoring locations during the period 2000 to 2009.

category None (for no aggregation) is a single R^2 for all the data of observed and predicted values during the 10 year period (see Table 3). Although a convenient number to report, this number does not accurately represent the way the data are used in health effects models, hence the presentation of other time aggregations. The MSE-R^2s were very close to the regression based R^2s showing little bias, *i.e.* deviation from the 1-1 line. The regression slopes similarly were close to 1.0 also indicating little bias from the 1-1 line. The GAMs were significantly better than the IDW methods by 7% - 10% in absolute differences and even more so than the nearest monitors. The nearest monitors showed a distinctive drop in MSE-R^2s when compared to regression based R^2s and a deviation of the slopes from 1.0. RMSPE decreased when the time aggregation was increased and had the smallest values for the GAM method. FRM monitors alone had a 2% - 5% absolute drop in R^2s values compared to all monitors.

For Table 4, results are shown for the 10th hold out set. When compared to Table 3, there is little evidence that there was over-fitting due to model selection in datasets 1 - 9.

When graphically comparing observed versus predicted values for all monitors in datasets 1 - 9, the model is a good fit except for a small number of extreme observations that under predicted (see Figure 4). Eight observations had observed values over 55 μg/m^3 which should be considered exceptional events (e.g. fires) since each one of these moni-

Table 3. Cross-validated statistics on datasets 1 - 9 predicting $PM_{2.5}$ (μg/m^3) for NM[a], IDW[b], and spatial GAM[c], by four aggregations of time from 2000-2009.

Model	Time	$PM_{2.5}$ μg/m^3				
		Reg-R^2[d]	MSE-R^2[e]	MPE[f]	RMSPE[g]	Slope
NM	None	0.64	0.60	−0.299	3.19	0.80
	Month	0.60	0.53	−0.300	3.03	0.76
	1 Year	0.64	0.60	−0.309	2.28	0.82
	5 Year	0.67	0.63	−0.282	2.14	0.83
IDW	None	0.73	0.72	−0.212	2.64	0.99
	Month	0.68	0.68	−0.213	2.51	0.99
	1 Year	0.71	0.70	−0.232	1.97	1.02
	5 Year	0.73	0.72	−0.195	1.85	1.01
GAM	None	0.80	0.80	0.066	2.23	0.99
	Month	0.76	0.76	0.066	2.16	0.99
	1 Year	0.81	0.81	0.080	1.55	0.98
	5 Year	0.82	0.82	0.116	1.49	0.97
GAM (FRM[h])	None	0.78	0.78	0.041	2.24	0.97
	Month	0.72	0.72	0.041	2.15	0.95
	1 Year	0.76	0.76	0.039	1.49	0.96
	5 Year	0.79	0.79	0.041	1.40	0.96

[a]Nearest monitor; [b]inverse distance weighting; [c]generalized additive model; [d]regression based R^2; [e]mean squared error based R^2; [f]mean prediction error; [g]root mean squared prediction error; [h]federal reference method only.

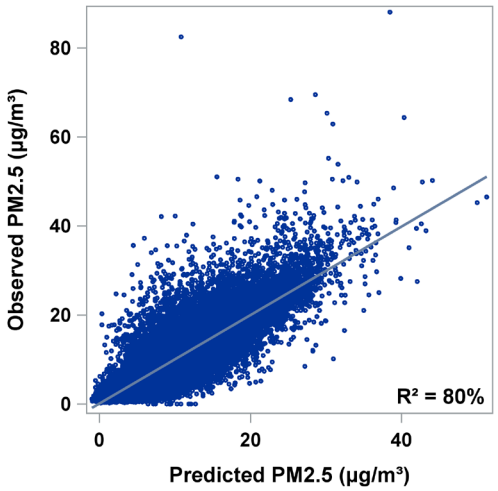

Figure 4. Observed $PM_{2.5}$ vs. predicted $PM_{2.5}$ for data sets 1 - 9.

Table 4. Evaluating model over-fit on dataset 10 for spatial GAM[c], by three aggregations of time from 2000-2009.

Model	Time	$PM_{2.5}$ µg/m³				
		Reg-R^2[d]	MSE-R^2[e]	MPE[f]	RMSPE[g]	Slope
GAM	Month	0.75	0.75	0.046	2.29	1.02
	1 Year	0.81	0.80	0.086	1.66	1.03
	5 Year	0.84	0.83	0.113	1.51	1.03
GAM (FRM[h])	Month	0.71	0.70	0.009	2.28	0.99
	1 Year	0.74	0.74	0.036	1.60	1.02
	5 Year	0.79	0.79	0.003	1.49	1.02

[c]generalized additive model; [d]regression based R^2; [e]mean squared error based R^2; [f]mean prediction error; [g]root mean squared prediction error; [h]federal reference method only.

tors did not have similar values (over 55 µg/m³) at other times during the study period.

In Figure 5, a map of the US is shown of surface predictions for $PM_{2.5}$ for the year 2005. The impact of population density is readily seen with the increases in $PM_{2.5}$ in the population centers of the US. The effect of elevation is also easily seen in the mountainous regions of the nation.

3.3. Subgroup Analysis

The northwest had much poorer model performance judged by a monthly R^2 = 0.37 and a 1 year R^2 = 0.46 (Table 5). The regression slopes were also lower than 1.0. Furthermore, without use of the non-reference monitors, the performance was noticeably worse with an R^2 = 0.23 and 0.29, MSE-R^2 = 0.06 and 0.02, and a regression slope of 0.59 and 0.69 for monthly and 1-year aggregations, respectively. It is interesting to note that the north central region had a similar number of monitors compared with the northwest over a larger geographic area and yet it performed well like the other regions.

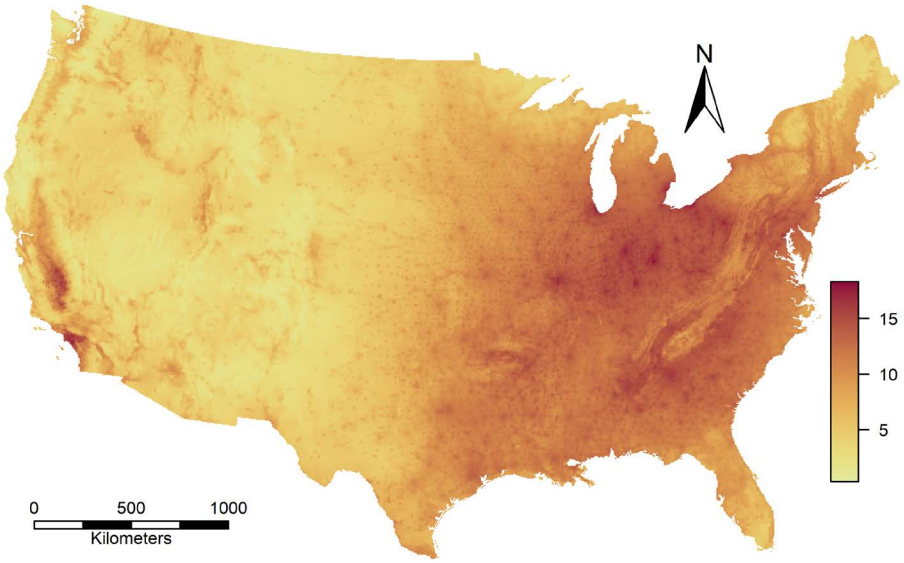

Figure 5. Surface predictions of $PM_{2.5}$ (μg/m^3) across the contiguous US for the year 2005.

Table 5. Cross-validated performance measures of GAM[a] for predicting two sets of $PM_{2.5}$ monitors, by three aggregations of time and by regions, on datasets 1 - 9 during 2000-2009.

Time	Region	$PM_{2.5}$ μg/m^3[b]					$PM_{2.5}$ μg/m^3 (FRM[c])				
		Reg-R^2[d]	MSE-R^2[e]	MPE[f]	RMSPE[g]	Slope	Reg-R^2	MSE-R^2	MPE	RMSPE	Slope
Month	NW[h]	0.37	0.33	0.177	2.70	0.83	0.23	0.06	0.154	2.98	0.56
	SW[i]	0.69	0.68	0.166	3.25	1.04	0.61	0.60	0.064	3.50	0.95
	NC[j]	0.68	0.67	−0.026	1.95	0.93	0.61	0.58	−0.026	1.88	0.86
	SC[k]	0.69	0.67	0.027	1.88	0.90	0.63	0.61	0.013	1.88	0.88
	NE[l]	0.70	0.69	0.063	1.77	0.97	0.67	0.67	0.054	1.75	0.95
	SE[m]	0.66	0.66	0.031	1.56	0.93	0.67	0.66	0.029	1.49	0.92
1 Year	NW	0.46	0.45	0.193	2.11	0.89	0.29	0.02	0.029	2.21	0.62
	SW	0.76	0.75	0.178	2.45	1.06	0.65	0.65	0.081	2.67	0.98
	NC	0.75	0.75	0.012	1.52	0.97	0.70	0.70	−0.017	1.30	0.91
	SC	0.77	0.76	0.053	1.40	0.92	0.71	0.70	0.022	1.38	0.92
	NE	0.79	0.78	0.070	1.16	0.97	0.76	0.76	0.053	1.13	0.97
	SE	0.81	0.81	0.029	1.00	0.97	0.85	0.85	0.021	0.87	0.98
5 Year	NW	0.44	0.41	0.180	1.84	0.83	0.25	0.07	−0.072	2.21	0.58
	SW	0.74	0.74	0.225	2.43	1.04	0.64	0.63	0.088	2.58	0.95
	NC	0.66	0.65	0.165	1.84	0.95	0.73	0.72	0.019	1.19	0.91
	SC	0.82	0.81	0.085	1.23	0.91	0.79	0.79	0.017	1.17	0.93
	NE	0.82	0.82	0.071	1.07	0.98	0.80	0.80	0.061	1.04	0.98
	SE	0.83	0.82	0.088	0.96	0.93	0.88	0.88	0.037	0.77	0.98

[a]generalized additive model; [b]all monitors; [c]federal reference method only; [d]regression based R^2; [e]mean squared error based R^2; [f]mean prediction error; [g]root mean squared prediction error; [h]OR, WA, & ID; [i]CA, NV, UT, & AZ; [j]MT, WY, ND, SD, NE, MN, & IA; [k]CO, NM, KS, OK, TX, MO, AR, & LA; [l]WI, MI, IL, IN, OH, KY, WV, VA, PA, NY, VT, NH, ME, MA, CT, RI, MD, DE, NJ, & DC; [m]TN, MS, AL, NC, SC, GA, & FL.

However there was a drop in the 5-year aggregation R^2 (0.66), probably related to the effect of aggregation on a reduced sample size and range. For a 1-year aggregation the southeastern part of the US had the best model performance (R^2 = 0.81; RMSPE = 1.00). Consistently, December through February had the lowest R^2 and the largest RMSPE, and June through August had the highest R^2 and the lowest RMSPE (Table 6).

4. Discussion

We created spatial GAMs that are simple and effective in modeling $PM_{2.5}$ across large areas like the US with R^2 = 0.80, 0.76, 0.81 and 0.82 for categories *None*, *Month*, 1 *Year*, and 5 *Year* time aggregations, respectively. There have been a few attempts at modeling $PM_{2.5}$ over the contiguous US. Beckerman *et al.* used a hybrid approach to model national monthly $PM_{2.5}$ estimates using Federal Reference Method only with an R^2 = 0.79. They created a spatiotemporal model by combining LUR models as a first stage and Bayesian maximum entropy interpolation of the LUR residuals as a second stage [15]. Our analysis including only the federal reference method monitors with no time aggregation had a similar value (R^2 = 0.78, No aggregation). Another analysis, using the Bayesian maximum entropy framework with a moving window, estimated $PM_{2.5}$ across the US for the year 2003 with a Pearson's correlation of 0.90 (R^2 = 0.81) [14]. Using combined monitors our analysis had an R^2 = 0.81 for a comparable 1-year aggregation. Considering only FRM monitors, the R^2 dropped to 0.78. It was not clear from the description which set of monitors was used in the previously mentioned moving window approach. A study using spatiotemporal GAMs for the northeastern US had an R^2 =

Table 6. Cross-validated performance measures of GAM[a] for predicting two sets of $PM_{2.5}$ monitors, by three aggregations of time and by seasons, on datasets 1 - 9 during 2000-2009.

Time	Season	$PM_{2.5}$ μg/m^3[b]					$PM_{2.5}$ μg/m^3 (FRM[c])				
		Reg-R^2[d]	MSE-R^2[e]	MPE[f]	RMSPE[g]	Slope	Reg-R^2	MSE-R^2	MPE	RMSPE	Slope
Month	Mar-May	0.79	0.78	0.044	1.73	0.99	0.74	0.73	0.026	1.75	0.96
	Jun-Aug	0.86	0.86	0.029	1.85	0.99	0.85	0.85	0.016	1.82	0.99
	Sep-Nov	0.74	0.74	0.059	2.18	0.99	0.71	0.71	0.036	2.18	0.96
	Dec-Feb	0.66	0.66	0.133	2.89	0.98	0.60	0.59	0.086	2.87	0.91
1 Year	Mar-May	0.82	0.82	0.044	1.46	0.99	0.77	0.77	0.023	1.46	0.96
	Jun-Aug	0.89	0.89	0.029	1.52	1.00	0.89	0.89	0.015	1.45	0.99
	Sep-Nov	0.77	0.77	0.069	1.84	0.99	0.74	0.73	0.037	1.82	0.97
	Dec-Feb	0.68	0.68	0.149	2.60	0.98	0.61	0.60	0.089	2.53	0.92
5 Year	Mar-May	0.85	0.85	0.074	1.30	0.98	0.83	0.83	0.030	1.26	0.97
	Jun-Aug	0.91	0.91	0.071	1.35	0.99	0.93	0.93	0.029	1.17	0.99
	Sep-Nov	0.80	0.80	0.091	1.67	0.98	0.79	0.79	0.017	1.57	0.97
	Dec-Feb	0.68	0.68	0.174	2.43	0.97	0.60	0.59	0.074	2.37	0.91

[a]generalized additive model; [b]all monitors; [c]federal reference method only; [d]regression based R^2; [e]mean squared error based R^2; [f]mean prediction error; [g]root mean squared prediction error.

0.77 for monthly estimations during a three-year period [12] and when extended to the contiguous US and for a longer time period (through 2007) had the same R^2 value while including non-reference methods as well [13].

The current research compares well to previous research by others. Although covering the contiguous US like the previously mentioned papers, there are some unique characteristics of this analysis. First, we showed the effects of time aggregation and the amount of improvement that is gained by longer time aggregations. This is useful when considering different time windows for long-term exposure for health effects like cardiovascular disease [7]. Second, we attempted to quantify the difference between using all available monitors, versus using the Federal Reference Method only. Because of the scarcity of monitors in particular areas like the northwest, using good quality, non-reference method monitors may be an appropriate approach for these areas. It is not clear what the impact is on the exposure estimates and ultimately to the health effects models by introducing methods that potentially have more measurement error. Third, an analysis by region is helpful to understand where there are regional weaknesses in model performance. Knowing the number of monitors in a region doesn't necessarily help with this prediction. The north central region had a similar number of monitors as the northwest and even though the network covered a larger geographic area, predictions were much better. This result suggests that special attention should be given to the northwest to find ways to improve the relatively poor results. Last, while comparing well with other methodologies, spatial GAMs are relatively simple models that are easier to set up than spatiotemporal models that have two stages. These results, although not perfectly comparable to other research (somewhat overlapping time periods), would seem to suggest that almost all of the predictive performance comes from spatial considerations alone.

Three predictors were retained for the final model: spatial coordinates, population density with 4-km buffer, and elevation. It should be noted that the results were not substantially different from a 1-km buffer or imperviousness with a 4-km buffer. It is possible that using a 1-km buffer is useful when evaluating inter-urban spatial heterogeneity.

Further analysis should look for ways to improve the modeling of fine particulates in the northwest. Possibly different predictors are important for this region. It would seem that using non-reference monitors for this region is particularly important. Another possible way to improve these models is using daily predictions. Last, using spatial GAMs may be a good input for a second stage using spatiotemporal residuals for time trend effects.

5. Conclusion

There are a number of appropriate methods that can be used to predict national $PM_{2.5}$. The method advocated in this paper (*i.e.* spatial GAM) is useful in that while it had strong model performance, similar to other recently described methods, its implementation is relatively simple. We feel that spatial GAMs have an important contribution to the discussion of and research for the modeling of $PM_{2.5}$ across large-scale areas for

chronic exposure.

Acknowledgements

DJS thanks Ed Santos and Ben Becerra for their assistance with creating maps.

References

[1] Lepeule, J., Laden, F., Dockery, D. and Schwartz, J. (2012) Chronic Exposure to Fine Particles and Mortality: An Extended Follow-Up of the Harvard Six Cities Study from 1974 to 2009. *Environmental Health Perspectives*, **120**, 965-970. http://dx.doi.org/10.1289/ehp.1104660

[2] Jerrett, M., *et al.* (2005) Spatial Analysis of Air Pollution and Mortality in Los Angeles. *Epidemiology,* **16**, 727-736. http://dx.doi.org/10.1097/01.ede.0000181630.15826.7d

[3] Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K. and Thurston, G.D. (2002) Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution. *JAMA—Journal of the American Medical Association*, **287**, 1132-1141. http://dx.doi.org/10.1001/jama.287.9.1132

[4] Miller, K.A., Siscovick, D.S., Sheppard, L., Shepherd, K., Sullivan, J.H., Anderson, G.L. and Kaufman, J.D. (2007) Long-Term Exposure to Air Pollution and Incidence of Cardiovascular Events in Women. *New England Journal of Medicine*, **356**, 447-458. http://dx.doi.org/10.1056/NEJMoa054409

[5] Beelen, R., *et al.* (2008) Long-Term Effects of Traffic-Related Air Pollution on Mortality in a Dutch Cohort (NLCS-AIR Study). *Environmental Health Perspectives*, **116**, 196-202. http://dx.doi.org/10.1289/ehp.10767

[6] Puett, R.C., Schwartz, J., Hart, J.E., Yanosky, J.D., Speizer, F.E., Suh, H., Paciorek, C.J., Neas, L.M. and Laden, F. (2008) Chronic Particulate Exposure, Mortality, and Coronary Heart Disease in the Nurses' Health Study. *American Journal of Epidemiology*, **168**, 1161-1168. http://dx.doi.org/10.1093/aje/kwn232

[7] Puett, R.C., Hart, J.E., Yanosky, J.D., Paciorek, C., Schwartz, J., Suh, H., Speizer, F.E. and Laden, F. (2009) Chronic Fine and Coarse Particulate Exposure, Mortality, and Coronary Heart Disease in the Nurses' Health Study. *Environmental Health Perspectives*, **117**, 1697-1701. http://dx.doi.org/10.1289/ehp.0900572

[8] Ostro, B., Lipsett, M., Reynolds, P., Goldberg, D., Hertz, A., Garcia, C., Henderson, K.D. and Bernstein, L. (2010) Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study. *Environmental Health Perspectives*, **118**, 363-369. http://dx.doi.org/10.1289/ehp.0901181

[9] Chen, L.H., Knutsen, S.F., Shavlik, D., Beeson, W.L., Petersen, F., Ghamsary, M. and Abbey, D. (2005) The Association between Fatal Coronary Heart Disease and Ambient Particulate Air Pollution: Are Females at Greater Risk? *Environmental Health Perspectives*, **113**, 1723-1729. http://dx.doi.org/10.1289/ehp.8190

[10] Yanosky, J.D., Paciorek, C.J., Schwartz, J., Laden, F., Puett, R. and Suh, H.H. (2008) Spatio-Temporal Modeling of Chronic PM10 Exposure for the Nurses' Health Study. *Atmospheric Environment*, **42**, 4047-4062. http://dx.doi.org/10.1016/j.atmosenv.2008.01.044

[11] Paciorek, C.J., Yanosky, J.D., Puett, R.C., Laden, F. and Suh, H.H. (2009) Practical Large-Scale Spatio-Temporal Modeling of Particulate Matter Concentrations. *Annals of Applied Statistics*, **3**, 370-397. http://dx.doi.org/10.1214/08-AOAS204

[12] Yanosky, J.D., Paciorek, C.J. and Suh, H.H. (2009) Predicting Chronic Fine and Coarse Particulate Exposures Using Spatiotemporal Models for the Northeastern and Midwestern Unit-

ed States. *Environmental Health Perspectives*, **117**, 522-529. http://dx.doi.org/10.1289/ehp.11692

[13] Yanosky, J.D., Paciorek, C.J., Laden, F., Hart, J.E., Puett, R.C., Liao, D. and Suh, H.H. (2014) Spatio-Temporal Modeling of Particulate Air Pollution in the Conterminous United States Using Geographic and Meteorological Predictors. *Environmental Health*, **13**, 63. http://dx.doi.org/10.1186/1476-069X-13-63

[14] Akita, Y., Chen, J.C. and Serre, M.L. (2012) The Moving-Window Bayesian Maximum Entropy Framework: Estimation of PM(2.5) Yearly Average Concentration across the Contiguous United States. *Journal of Exposure Science and Environmental Epidemiology,* **22**, 496-501. http://dx.doi.org/10.1038/jes.2012.57

[15] Beckerman, B.S., Jerrett, M., Serre, M., Martin, R.V., Lee, S.J., van Donkelaar, A., Ross, Z., Su, J. and 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**, 7233-7241.

[16] Hu, X.F., Waller, L.A., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Sarnat, J.A. and Liu, Y. (2013) Estimating Ground-Level PM2.5 Concentrations in the Southeastern US Using Geographically Weighted Regression. *Environmental Research*, **121**, 1-10. http://dx.doi.org/10.1016/j.envres.2012.11.003

[17] Vienneau, D., de Hoogh, K., Bechle, M.J., Beelen, R., van Donkelaar, A., Martin, R.V., Millet, D.B., Hoek, G. and Marshall, J.D. (2013) Western European Land Use Regression Incorporating Satellite- and Ground-Based Measurements of NO_2 and PM10. *Environmental Science & Technology*, **47**, 13555-13564. http://dx.doi.org/10.1021/es403089q

[18] US Environmental Protection Agency. Technology Transfer Network (TTN) Air Quality System (AQS) Data Mart. http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm

[19] Wood, S.N. (2006) Generalized Additive Models: An Introduction with R. Texts in Statistical Science. Chapman & Hall/CRC, Boca Raton.

[20] US EPA (2006) Technical Note on Reporting PM2.5 Continuous Monitoring and Speciation Data to the Air Quality System (AQS). 6.

[21] IMPROVE. Interagency Monitoring of Protected Visual Environments. http://vista.cira.colostate.edu/improve/

[22] US Geological Survey. Multi-Resolution Land Characteristics Consortium (MRLC). http://www.mrlc.gov/index.php

[23] US Geological Survey. National Elevation Datset. http://ned.usgs.gov/

[24] US Environmental Protection Agency. Technology Transfer Network, Clearinghouse for Inventories & Emissions Factors. http://www.epa.gov/ttn/chief/eiinformation.html

[25] Cal/EPA. ARB's Geographical Information System (GIS) Library. http://www.arb.ca.gov/ei/gislib/gislib.htm

[26] National Oceanic and Atmospheric Administration. National Climatic Data Center. http://www.ncdc.noaa.gov/cdo-web/

[27] Keller, J.P., *et al.* (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**, 301-309.

[28] Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G. and Brunekreef, B. (2012) Systematic Evaluation of Land Use Regression Models for NO(2). *Environmental Science & Technology*, **46**, 4481-4489. http://dx.doi.org/10.1021/es204183v