UW Biostatistics Working Paper Series

8-6-2015

Historical Prediction Modeling Approach for Estimating Long-Term Concentrations of PM in Cohort Studies Before the 1999 Implementation of Widespread Monitoring

Sun-Young Kim
University of Washington - Seattle Campus, puha0@uw.edu

Casey Olives
University of Washington - Seattle Campus, colives@uw.edu

Lianne Sheppard
University of Washington - Seattle Campus, sheppard@u.washington.edu

Paul D. Sampson
University of Washington - Seattle Campus, pds@stat.washington.edu

Timothy V. Larson
University of Washington - Seattle Campus, tlarson@uw.edu

See next page for additional authors

Authors
Sun-Young Kim, Casey Olives, Lianne Sheppard, Paul D. Sampson, Timothy V. Larson, and Joel Kaufman

Historical prediction modeling approach for estimating long-term concentrations of $PM_{2.5}$ in cohort studies before the 1999 implementation of widespread monitoring

Sun-Young Kim,[1,2], Casey Olives,[1], Lianne Sheppard,[1,3], Paul D. Sampson,[4], Timothy V. Larson,[1,5], Joel D. Kaufman,[1,6,7]

[1]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, USA

[2]Institute of Health and Environment, Seoul National University, Seoul, Korea

[3]Department of Biostatistics, University of Washington, Seattle, USA

[4]Department of Statistics, University of Washington, Seattle, USA

[5]Department of Civil and Environmental Engineering, University of Washington, Seattle, USA

[6]Department of Epidemiology, University of Washington, Seattle, USA

[7]Department of Medicine, University of Washington, Seattle, USA

Corresponding author: Joel D. Kaufman

Department of Environmental and Occupational Health Sciences

University of Washington

4225 Roosevelt Way NE

Seattle, WA 98105

Tel: 206-616-3501

Fax: 206-897-1991

Running title: Predicting fine particulate concentrations since 1980

Acknowledgement: This work was primarily supported by the Multi-Ethnic Study of Atherosclerosis and Air Pollution by the U.S. Environmental Protection Agency (EPA) (RD 831697). Although this publication was developed under a Science to Achieve Results (STAR) research assistance agreement, No. RD831697, awarded by the EPA, it has not been formally reviewed by the EPA. The views expressed in this document are solely those of the University of Washington and the EPA does not endorse any products or commercial services mentioned in this publication. Additional support was provided by the U.S. EPA (CR-834077101-0 and RD-83479601-0) and the National Research Foundation of Korea (Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education: 900-20130078).

CFI statement: None of the authors has any actual or potential competing financial interests.

ABSTRACT

Introduction: Recent cohort studies use exposure prediction models to estimate the association between long-term residential concentrations of $PM_{2.5}$ and health. Because these prediction models rely on $PM_{2.5}$ monitoring data, predictions for times before extensive spatial monitoring present a challenge to understanding long-term exposure effects. The Environmental Protection Agency (EPA) Federal Reference Method (FRM) network for $PM_{2.5}$ was established in 1999. We evaluated a novel statistical approach to produce high quality exposure predictions from 1980-2010 for epidemiological applications.

Methods: We developed spatio-temporal prediction models using geographic predictors and annual average $PM_{2.5}$ data from 1999 through 2010 from the FRM and the Interagency Monitoring of Protected Visual Environments (IMPROVE) networks. The model consists of a spatially-varying long-term mean, a spatially-varying temporal trend, and spatially-varying and temporally-independent spatio-temporal residuals structured using a universal kriging framework. Temporal trends in annual averages of $PM_{2.5}$ before 1999 were estimated by using a) extrapolation based on $PM_{2.5}$ data for 1999-2010 in FRM/IMPROVE, b) $PM_{2.5}$ sulfate data for 1987-2010 in the Clean Air Status and Trends Network, and c) visibility data for 1980-2010 across the Weather-Bureau-Army-Navy network. We validated the resulting models using $PM_{2.5}$ data collected before 1999 from IMPROVE, California Air Resources Board dichotomous sampler monitoring (CARB dichot), the Southern California Children's Health Study (CHS), and the Inhalable Particulate Network (IPN).

Results: The $PM_{2.5}$ prediction model performed well across three trend estimation approaches when validated using IMPROVE and CHS data (R^2= 0.84–0.91). Model performance using CARB dichot and IPN data was worse than those in IMPROVE most likely due to inconsistent sampling methods and smaller numbers of monitoring sites.

Discussion: Our prediction modeling approach will allow health effects estimation associated with long-term exposures to $PM_{2.5}$ over extended time periods of up to 30 years.

INTRODUCTION

Many cohort studies of the long-term effects of fine particulate matter ($PM_{2.5}$) air pollution on health have used exposure prediction models to estimate individual-level long-term concentrations at cohort residences (e.g., Eeften et al. 2012; Paciorek et al. 2009; Puett et al. 2009; Beelen et al. 2014; Sampson et al. 2013; Young et al. 2014). These exposure prediction models rely on $PM_{2.5}$ monitoring data collected from a spatially-distributed monitoring networks. $PM_{2.5}$ predictions are generally infeasible for times before comprehensive spatial monitoring began in the late 1990s or 2000s depending on the county. However, many cohorts were enrolled before these extensive monitoring networks began operating. Many studies thus use $PM_{2.5}$ estimates based on monitoring data from later time periods than cohort follow-up for their health analyses (e.g., Beelen et al. 2008; Cesaroni et al. 2013; Weichenthal et al. 2014). This temporal misalignment of $PM_{2.5}$ predictions with health data could affect study results.

Other studies have developed historical prediction models to temporally align exposure estimates with health outcomes. They used back-extrapolation, historically available large particle data, or physical or chemical models complemented by visibility, emission, meteorology, and satellite data (Beelen et al. 2014; Brauer et al. 2012; Lall et al. 2004; Molnar et al. 2015; Paciorek et al. 2009; Yanosky et al. 2009; Hogrefe et al. 2009; Ozkanak et al. 1985). However, most these studies estimated historical $PM_{2.5}$ concentrations in limited areas and/or for relatively short time periods. Furthermore, the model evaluation for the period prior to extensive monitoring was restricted to small datasets or poorly reported.

In the U.S., many populations of great value for assessment of $PM_{2.5}$ health effects collected data well before 1999, when reliable long-term regulatory monitoring data for $PM_{2.5}$

began to be available. We aimed to develop a national prediction model to estimate annual average concentrations of $PM_{2.5}$ in the continental U.S. for the entire time period between 1980 through 2010. We evaluated our historical predictions from 1980 through 1998 using available external validation datasets and investigated residential historical predictions using a multi-city cohort.

METHODS

$PM_{2.5}$ data

We obtained daily $PM_{2.5}$ concentrations collected in the two national $PM_{2.5}$ monitoring networks: Environmental Protection Agency (EPA) Federal Reference Method (FRM) and Interagency Monitoring of Protected Visual Environment (IMPROVE) networks. Whereas FRM sites were located mostly in urban areas to monitor population-level of $PM_{2.5}$ concentrations, IMPROVE sites were established to monitor visibility and located mostly in wilderness areas and national parks (U.S. EPA 2004; Hand 2011). We downloaded the data from all FRM sites from 1999 through 2010 and IMPROVE sites from 1990 through 2010 from the EPA Air Quality database (U.S. EPA 2014). We computed annual averages of $PM_{2.5}$ for each site that met minimum inclusion criteria of at least two-thirds complete data points for a year (with exact numbers dependent on the sampling schedule) and less than 45 consecutive missing days of sampling. We used the $PM_{2.5}$ data collected in FRM and IMPROVE for 1999-2010 for model development including trend estimation, whereas we reserved the IMPROVE data from 1990-1998 for model validation. We identified all monitoring sites in three regions: the East, Mountain West, and West Coast regions (Figure 1).

In order to estimate temporal trends for the entire 1980 through 2010 time period, including all years without FRM $PM_{2.5}$ measurements, we obtained two additional sources of data: annual average concentrations of $PM_{2.5}$ sulfate measured in the Clean Air Status and Trends Network (CASTNet) from 1987 through 2010 (U.S. EPA 2013) and daily noon-time visual ranges, as a measure of visibility, monitored in the Weather-Bureau-Army-Navy (WBAN) network from 1980 through 2010. Because most visibility measurements collected by optical instruments had maximum of 16.093 km (10 miles) and these instruments replaced measurements taken by the human eye in 1990s, we truncated all measurements to a maximum 16.093 km distance. We computed annual averages of visibility after excluding days with heavy fog, dust, and precipitation, and after applying the same inclusion criteria as for $PM_{2.5}$ data.

For the model evaluation prior to 1999, we obtained $PM_{2.5}$ data from three different networks in addition to IMPROVE: the Southern California Children's Health Study (CHS) for 1988-2001 (Peters et al. 2004), the California Air Resources Board dichotomous sampler monitoring (CARB dichot) for 1994-2003 in California (Blanchard et al. 2011), and the Inhalable Particulate Network (IPN) for 1980-1981 over the continental U.S (U.S. EPA 1985). CHS $PM_{2.5}$ data collected using two-week samplers were converted to FRM-equivalent $PM_{2.5}$ for computing annual averages (Peters et al. 2004). Likewise, for the CARB dichot data we adopted a published conversion equation to estimate FRM-equivalent $PM_{2.5}$ (Blanchard et al. 2011).

Geographic variables and geocoding

We considered more than 800 variables representing geographic characteristics including traffic, land use, emission, elevation, and vegetation index (Supplemental Material, Table S1). Computation of these variables at each of all $PM_{2.5}$ monitoring sites was implemented in ArcGIS 10.2. For land use characteristics, we used data collected in different time periods to incorporate time-varying spatial features into the model: land cover data from the 1970s and 1980s, and satellite data generated in 2006. Our final list of geographic variables was pruned to about 300 variables after we eliminated the less informative variables with little variability. To illustrate our predictions over time, we geocoded residential addresses of 7,552 participants in the Multi-Ethnic Study of Atherosclerosis (MESA) and associated MESA Air project who consented to use of their addresses and provided historical residential addresses dating back to 1980, as well as 12,501 coordinates of points on a 25 kilometer grid in the continental U.S according to standardized procedures.

Development of the $PM_{2.5}$ model for 1980-2010

The $PM_{2.5}$ model for the period of 1980-2010 was developed based on the framework of the $PM_{2.5}$ spatio-temporal prediction model in MESA Air (Keller et al. 2015; Lindstrom et al. 2014; Sampson et al. 2011; Szpiro et al. 2010). We modeled the log annual average $PM_{2.5}$ concentrations from 1999 through 2010 as a function of a spatially varying long-term mean, a spatially varying temporal trend, and spatio-temporal residuals. The spatially varying temporal trend is composed of a spatially-varying trend coefficient and a trend basis function. The trend basis function is estimated from singular vector decomposition of the data (Fuentes et al. 2006). We restricted these data to sites with more than six years of monitoring out of the twelve possible years. We estimated the spatially-varying long-term mean and trend coefficient using universal kriging, which integrates geographic predictors and spatial smoothing (Banerjee et al. 2003). We used partial least squares (PLS) to reduce the dimension of the hundreds of geographic variables to two derived predictors that are the linear combinations that maximize their covariance with $PM_{2.5}$. The spatial dependence structure in the kriging model for the long-term mean was assumed to be exponential and was indexed by the range, partial sill, and nugget parameters. To avoid unnecessary complexity in the model, we did not allow a spatial structure for the trend coefficient (zero range and partial sill). We also used kriging to model the spatially-dependent and temporally-independent spatio-temporal residuals. We examined alternative modeling choices including a spatial structure for the trend coefficient and interaction terms by three regions.

We explored various approaches to estimate the temporal trend before 1999. These included the backward extrapolation of the temporal trend basis function estimated from the 1999-2010 FRM $PM_{2.5}$ data, and estimation of the temporal trend using other sources of data such as emission, meteorological variables, visibility, and $PM_{2.5}$ sulfate; all these other measurements have been shown to be associated with $PM_{2.5}$ in previous studies (Hand et al. 2014; Malm et al. 2002; Ozkanak et al. 1985). Ultimately we selected three approaches for in-depth evaluation of the historical trend estimation: 1) extrapolation of the linear trend estimated based on the $PM_{2.5}$ data in FRM and IMPROVE for 1999-2010, 2) estimation of the trend using the $PM_{2.5}$ sulfate data in CASTNet for 1987-2010 and extrapolation for 1980-1986, and 3) estimation of the trend using the visibility data in WBAN for 1980-2010. We also examined alternative approaches, including combining two data sources into one temporal trend, estimating two temporal trends, and replacing the trend by meteorological variables as spatio-temporal covariates.

To evaluate our model for 1999-2010, we performed 5-fold cross-validation and computed mean square error (MSE) and MSE-based R-square (R^2) statistics for annual averages (Keller et al. 2015). We presented cross-validation statistics yearly for these twelve years, in all regions for each year as well as all years combined and in each of the three regions for all years combined.

Model evaluation for the pre-1999 period

We externally validated the model using four distinct $PM_{2.5}$ datasets, all sampled before 1999: 1) IMPROVE data for 1990-1998, 2) CARB dichot data for 1988-2001, 3) CHS data for 1994-2003, and 4) IPN data for 1980-1981 (Table 1). We predicted annual averages of $PM_{2.5}$ concentrations at monitoring sites in each of the four monitoring networks and computed out-of-sample MSEs and MSE-based R^2s using these external data sources for all years and regions as well as by year and region.

Predictions

We created maps of $PM_{2.5}$ predictions on a 25 km grid over the contiguous U.S. in 1980, 1990, 2000, and 2010 to examine spatially-varying changes of $PM_{2.5}$ concentrations over time. We also selected 10 grid coordinates with the highest populations in each of the three regions and explored the trends of predictions over 31 years.

We also conducted some analyses to provide information on the degree to which exposure estimation based on data from the year 2000 reflects concentrations predicted by our approach in the earlier period. In order to investigate the sensitivity of temporally- and spatially-varying individual exposures that incorporate changes in people's residences over time, we predicted $PM_{2.5}$ concentrations at all home addresses from 1980 through 2000, the year of the baseline exam, among members of the MESA Air cohort and computed a 21-year average weighted by residence times across historical addresses for each participant. These predictions were compared to annual averages estimated for the same participants in 2000, the year of the baseline exam. We stratified this comparison by the 5,086 who did not move during 1980-2000 ("non-movers") and 2,466 people who moved at least once.

RESULTS

Means of $PM_{2.5}$ annual averages for 1999-2010 in FRM and IMPROVE were 12.03 (SD=3.23) and 5.44 (2.94) μg/m^3, respectively (Table 1). The number of monitoring sites was small in 1999 compared to 2000-2010 and most sites were located in the East region (Figure 1, Supplemental Material, Figure S1). Annual average concentrations of $PM_{2.5}$ decreased over time from 1999 through 2010, particularly in the East and West Coast regions (Supplemental Material, Figure S2). Figure 2 displays the estimated temporal trends from 1980 through 2010 using the three trend estimation approaches. Whereas the extrapolated trend based on the $PM_{2.5}$ data was linear, the trends estimated using $PM_{2.5}$ sulfate and visibility measurements were only generally linear and had different rates of decrease in different time periods.

In the model evaluation for 1999-2010, cross-validated R^2s for all twelve years combined and each single year were high, varying between 0.77 and 0.87 across the three trend estimation approaches (Table 2). The East and West Coast regions gave higher R^2s (0.80-0.86) than those in the Mountain West region (0.59-0.60). Supplemental Material, Figure S3 shows estimated regression and variance parameters for the long-term mean, the temporal trend coefficient, and spatio-temporal residuals. Regression coefficients of two PLS predictors for both the long-term mean and trend coefficient were statistically significantly different from 0, reflecting the large contributions of PLS predictors to these two components. Significant range and partial sill parameters for the long-term mean show an additional important contribution of the spatial correlation structure to the long-term mean. The cross-validation statistics of alternative modeling approaches in the sensitivity analyses were consistent with (and no better than) or poorer than those of our primary approach shown in Table 2 (data not shown).

Tables 3 and 4 show the external validation statistics for the pre-1999 period using IMPROVE data and the CHS, CARB dichot, and IPN data, respectively. Using IMPROVE data, the R^2s were consistently high for all years and each year separately (0.70-0.91) across the three trend estimation approaches (Table 2, Figure 3). The R^2s were slightly higher for the model using the extrapolated linear trend based on $PM_{2.5}$ data than estimated trends from

$PM_{2.5}$ sulfate and visibility data. In addition, the earliest years (1990 and 1991) gave lower R^2s (0.70-0.85) than the other years (0.83-0.93). The East region produced higher R^2s (0.67-0.88) than the Mountain West region. When the model was validated using the CHS data, the R^2s were also generally high (0.71-0.90). CARB dichot data gave high R^2s over 0.5 except for some years whereas IPN data consistently showed low R^2s.

Figure 4 shows predicted $PM_{2.5}$ concentrations dramatically decreased over 31 years with only a few areas that remained consistently high in the continental U.S. over the entire time period. The decreasing trend over time was also clear across the 10 most populated grid coordinates in each region (data not shown). Thirty-one year residence-weighted average $PM_{2.5}$ predictions for MESA Air participants were generally higher than the corresponding annual averages at their residence in 2000 (Figure 5). The two sets of predictions showed more inconsistency for movers with slightly lower correlation than for non-movers.

DISCUSSION

We developed a 31-year prediction model to estimate fine-scale ambient $PM_{2.5}$ concentrations in the continental U.S., including the time period prior to 1999 when extensive monitoring data became available. Key aspects of our approach to historical (pre-1999) prediction were our consideration of various trend estimation approaches and our model validation with multiple external validation datasets. The prediction model performed well for 1999-2010 as assessed by cross-validation. The pre-1999 predictions also generally performed well across three trend estimation approaches when validated based on external $PM_{2.5}$ monitoring network data, particularly IMPROVE and CHS data. The model performance was better in the more highly populated East region. Twenty one-year average $PM_{2.5}$ concentrations for 1980-2000 at MESA Air participant residences tended to be higher than and somewhat inconsistent with annual averages in 2000, though the correlation was higher among those with stable residence locations.

Developing a prediction model for estimating long-term $PM_{2.5}$ concentrations for the time period when there is little available $PM_{2.5}$ monitoring data requires using external information to estimate a temporal trend. Our three approaches for trend estimation gave consistently good model performance as assessed by R^2s, with a slight edge to the linearly extrapolated trend for predictions before 1990. This could be because the three trends we considered, while based on three different data sources, all showed similarly decreasing patterns with only slightly different shapes. We considered $PM_{2.5}$ sulfate data useful for trend estimation as a large reduction of $PM_{2.5}$ in 1990s and early 2000s was likely to be due to a large reduction of sulfate, particularly in the East region (Malm et al. 2002; US EPA 2003). The non-linear decrease of the estimated trend from $PM_{2.5}$ sulfate data could be due to the timing of implementation of policies regulating sulfur dioxide emissions. The CASTNet sites were located mostly in rural areas which may not represent $PM_{2.5}$ concentrations from urban sources or affecting population centers. However, as sulfate is an important regional pollutant that exhibits homogenous concentrations on a large spatial scale due to long-range transport, the rural sites still allow us to assess large regional trends over time as intended by the CASTNet monitoring design. The trend estimated based on the visibility data showed a somewhat different shape from that of the $PM_{2.5}$ sulfate trend. In addition to a non-linear relationship between $PM_{2.5}$ concentrations and visibility, the change of sampling methods for visibility from human eye to optical instruments in late 1990s may impact interpretation of the pattern (Hyslop et al. 2009; US EPA 2003).

Our historical model was based on a spatio-temporal framework using annual averages of $PM_{2.5}$ concentrations for multiple years. Other studies in Europe and Canada predicted annual averages of NO_2, NO_X, and $PM_{2.5}$ by back-extrapolation (Beelen et al. 2014; Chen et al. 2010; Gulliver et al. 2013; Meng et al. 2015). The back-extrapolation approach computed the difference of spatial averages between the two time periods or the ratio of a short-term average to an annual average based on a few fixed site measurements and then added to or multiplied by predictions in recent years in order to obtain estimates in early years. In contrast with the back-extrapolation approach, our spatio-temporal approach allows prediction for an extended time period when there are no measurements.

Like other authors, we considered various alternative approaches to historical prediction. Most previous studies used ratios of $PM_{2.5}$ to PM_{10} to leverage PM_{10} data collected before $PM_{2.5}$ monitoring began, as opposed to our approach directly using $PM_{2.5}$ along with an estimated temporal trend. Some U.S. studies developed ratio models that predict monthly averages of $PM_{2.5}$ concentrations for 1988-1998 by multiplying by PM_{10} for Nurse's Health Study participants residing in Northeastern and Midwestern regions (Paciorek et al. 2009; Yanosky et al. 2009) and expanded to the continental U.S. (Yanosky et al. 2014). In Taipei, Taiwan, another study developed a ratio model for predicting historical monthly averages of $PM_{2.5}$ (Yu et al. 2010). In separate analyses to mimic this approach, we also applied our model to annual average ratios. Our cross-validated R^2s were high between 1999 and 2010 (R^2=0.84-0.90) consistent with those in our original model. However, R^2s in the out-of-sample validation using IMPROVE data were lower, particularly in early years such as 1990 and 1991 (R^2=0.13 and 0). This poor model performance could be due to relatively poor prediction performance of PM_{10} rather than $PM_{2.5}$. A spatio-temporal prediction model for PM_{10} annual averages in the continental U.S. achieved a cross-validated R^2 of 0.55 (Hart et al. 2009), much lower than the cross-validated R^2 of 0.88 in a spatial prediction model for

$PM_{2.5}$ annual averages in 2000 (Sampson et al. 2013). It is also possible that temporal and spatial patterns of PM_{10} vary rather differently from those of $PM_{2.5}$.

In addition to ratios, we also explored modeling approaches that incorporated visibility or meteorology to predict historical $PM_{2.5}$ concentrations. A group of studies used the extinction coefficient, the inverse visual range multiplied by a constant, solely or jointly with $PM_{2.5}$ and PM_{10} data based on their high correlation with $PM_{2.5}$ concentrations (Ozkanak et al. 1985; Paciorek et al. 2009; Yanosky et al. 2009). The good model performance using the visibility trend in our model confirms the usefulness of visibility data for predicting $PM_{2.5}$. However, our results showed slightly better model performance using $PM_{2.5}$ data than visibility data when validated on the national scale using IMPROVE data. We examined our models after adding meteorological measurements as spatio-temporal covariates and found worse model performance than our preferred approach.

We evaluated our historical prediction model using four available external validation datasets; together these covered 13 years of the 19 year period for 1980-1998 in much of the United States. Previous studies for historical $PM_{2.5}$ prediction models either presented cross-validated results using data before 1999 but without any external validation datasets (Paciorek et al. 2009; Yanosky et al. 2009; Yanosky et al. 2014), or reported external validation results based on a limited dataset for a short time period (Hogrefe et al. 2009; Lall et al. 2004; Ozkaynak et al. 1985; Yu et al. 2010). Our model performed particularly well when evaluated against IMPROVE and CHS data. The IMPROVE data, as a national network with a strength as a validation dataset, gave the highest R^2s among all external validation datasets, possibly due to the advantage of validating for the 1990-1998 time period when the estimated trend is less uncertain.

We also observed consistently high R^2s when validating against the data in CHS which deployed monitoring sites in urban and residential areas. All CHS monitoring sites were in Southern California and thus may not be generalizable across the U.S.. The CARB dichot data, also restricted to California locations, however gave lower R^2s, including values less than 0.5 for some years. One possible reason for this poor performance is that the CARB dichot network used a different sampling protocol than FRM. Our simplified data-driven calibration method may not have performed well compared to an approach incorporating site-specific meteorological conditions (Blanchard et al. 2011). Model performance could have also been impacted by a set of CARB dichot sites in the highest $PM_{2.5}$ concentration areas (Figure 4). The IPN data gave the lowest R^2s overall. In addition to the inconsistency of the IPN sampling protocol with that of FRM, the limited amount of IPN data might have been influential. With 6 and 12 sites for 1980 and 1981, respectively, a few sites with poor predictions had a large impact on the R^2 estimates (data not shown). Furthermore, the IPN years of 1980-1981 are the earliest years of our prediction period and may reflect the most uncertainty in trend estimation.

One limitation of this study is our use of time-constant geographic variables which do not account for changes in spatial characteristics over time. However, our estimated PLS predictors from the geographic variables included two sources of land use data: land cover data created in 1970s and satellite data generated in 2007. These two data representing spatial differences on two different time periods about 30 years apart, and our explicit modeling of the temporal trend with these covariates incorporated gave us the ability to capture changes of land use features over time in our model. In addition, a study in Vancouver, Canada, found the model performance for predicting NO and NO_2 in 2003 was consistent with geographic variables between 2003 and 2010 (Wang et al. 2013). Although this time period is only 7 years and much shorter than our 31 years, these findings suggest that spatial patterns in urban areas with stable physical environments can be characterized by geographic variables from one of many time periods.

Our results suggest the importance of incorporating a temporal trend of air pollution concentrations changing over time and varying over space in cohort studies. Using exposure predictions from a later period of follow-up in epidemiological study, as commonly used in studies (Beelen et al. 2008; Cesaroni et al. 2013), may not precisely represent long-term exposures and might impact health effect findings even for people who never moved.

CONCLUSIONS

Our 31-year national $PM_{2.5}$ prediction model can be widely applicable to epidemiological studies, particularly for assessing the association of long-term air pollution exposure and health outcomes in cohort studies. While there remains unavoidable uncertainty about the quality of the predictions for the earliest time periods, the overall strong performance of our model assures that we can provide good $PM_{2.5}$ estimates that are temporally well aligned with health data, including health outcomes collected before extensive monitoring data exist. In addition, application of this model will allow estimation of individual-level concentrations across historical addresses over time and thus will improve assessment of the impact of air pollution on progression of disease conditions over the life course. Our findings also suggest that long-term average $PM_{2.5}$ estimates obtained from single addresses or restricted time periods after health observation may not always accurately represent long-term average estimates, and could impact subsequent health analyses.

REFERENCES

Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M, et al. Exposure Assessment for Estimation of the Global Burden of Disease Attributable to Outdoor Air Pollution. Environ Sci & Technol 2012;46(2), 652-660.

Banerjee S, Carlin BP, Gelfand AE, Hierarchical Modeling and Analysis for Spatial Data. Boca Raton, FL: Chapman & Hall/CRC Press. 2004:21-68.

Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. Lancet. 2014:1;383(9919):785-795.

Bellander T, Berglind N, Gustavsson P, Jonson T, Nyberg F, Pershagen G, Järup L. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. Environ Health Perspect. 2001;109(6):633-639.

Blanchard CL, Tanenbaum S, Motallebi N. Spatial and temporal characterization of PM2.5 mass concentrations in California, 1980-2007. J Air Waste Manag Assoc. 2011;61(3):339-351.

Cesaroni G, Badaloni C, Gariazzo C, Stafoggia M, Sozzi R, Davoli M, Forastiere F. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. Environ Health Perspect. 2013;121(3):324-231.

Chen H, Goldberg MS, Crouse DL, Burnett RT, Jerrett M, Villeneuve PJ, Wheeler AJ, Labreche F, Ross NA. Back-extrapolation of estimates of exposure from current land-use regression models. Atmos Environ 2010;44:4346-4354.

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. 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. Environ Sci Technol. 2012:16;46(20):11195-1205.

Fuentes M, Guttorp P, Sampson P D. Using transforms to analyze space-time processes: Statistical methods for spatio-temporal system (Chapman & Hall/CRC monographs on statistics & applied probability 107). Boca Raton, FL: Chapman & Hall/CRC Press. 2006:77-149.

Gulliver J, de Hoogh K, Hansell A, Vienneau D. Development and back-extrapolation of NO2 land use regression models for historic exposure assessment in Great Britain. Environ Sci Technol. 2013:16;47(14):7804-7811.

Hand JL, Copeland SA, Day DE, Dillner AM, Indresand H, Malm WC, et al. 2011. Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States: Report V. June (http://vista.cira.colostate.edu/improve/publications/Reports/2011/PDF/Cover_TOC.pdf).

Hand JL, Schichtel BA, Malm WC, Copeland S, Molenar JV, Frank N, Pitchford M. Widespread reductions in haze across the United States from the early 1990s through 2011. Atmos Environ 2014;94:671-679.

Hart JE, Yanosky JD, Puett RC, Ryan L, Dockery DW, Smith TJ, Garshick E, Laden F. Spatial modeling of PM10 and NO2 in the continental United States, 1985-2000. Environ Health Perspect. 2009;117(11):1690-1696.

Hogrefe C, Lynn B, Goldberg R, Rosenzweig C, Zalewsky E, Hao W, et al. A combined model–observation approach to estimate historic gridded fields of PM2.5 mass and species concentrations. Atmos Environ 2009:43;2561-2570.

Hyslop NP. Impaired visibility: the air pollution people see. Atmos Environ 2009;43:182-195.

Keller JP, Olives C, Kim SY, Sheppard L, Sampson PD, Szpiro AA, et al. A unified spatiotemporal modeling approach for prediction of multiple air pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. Environ Health Perspect 2015:123(4):301-309.

Lall R, Kendall M, Ito K, Thurston GD. Estimation of historical annual PM2.5 exposures for health effects assessment Atmos Envion 2004;38(31):5217-5226.

Lindstrom J, Szpiro AA, Sampson PD, Oron AP, Richards M, Larson TV, Sheppard L. A flexible spatio-temporal model for air pollution with spatial and spatio-temporal covariates. Environ Ecol Stat. 2014;21:411–433.

Malm WC, Schichtel BA, Ames RB, Gebhard KA. A 10-year spatial and temporal trend of sulfate cross the Unite States. J Geophy Res. 2002;107:4627-4646.

Meng X, Chen L, Cai J, Zou B, Wu CF, Fu Q, Zhang Y, Liu Y, Kan H. A land use regression model for estimating the NO2 concentration in Shanghai, China. Environ Res. 2015;137:308-315.

Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, Kaufman JD. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med. 2007;356:447-458.

Molnar P, Stockfelt L, Barregard B, Sallsten G. Residential NO_X exposure in a 35-year cohort study. Changes of exposure, and comparison with back extrapolation for historical exposure assessment. Atmos Environ 2015;115:62-69.

Ozkaynak H, Schatz AD, Thurston GD, Isaacs RG, Husar RB. Relationships between Aerosol Extinction Coefficients Derived from Airport Visual Range Observations and Alternative Measures of Airborne Particle Mass. JAPCA 1985;35:1176-1185.

Paciorek CJ, Yanosky JD, Puett RC, Laden F, Suh HH. Practical large-scale spatio-temporal modeling of particulate matter concentrations. Ann Appl Stat 2009;3(1):370-397.

Peters JM, Avol E, Perhane K, Gauderman WJ, Gilliland F, Jerrett M, et al. Epidemiologic investigation to identify chronic effects of ambient pollutants in southern California. 2004. California Air Resources Board and the California Environmental Protection Agency.

Puett RC, Hart JE, Yanosky JD, et al. Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environ Health Perspect*. 2009;117(11):1697-701.
Sampson PD, Szpiro AA, Sheppard L, Lindström J, Kaufman JD. Pragmatic Estimation of a Spatio-Temporal Air Quality Model with Irregular Monitoring Data. Atmos Environ. 2011;45:6593-6606.

Sampson PD, Richards M, Szpiro, AA, Bergen S, Sheppard L, Larson TV, Kaufman JD. A regionalized national universal kriging model using partial least squares regression for estimating annual PM2.5 concentrations in epidemiology. Atmos Environ. 2013;75;383-392.

Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar S, Kaufman JD. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. Environmetrics. 2010;21:666-631.

U.S. EPA. Inhalable particulate network report: Operation and data summary (mass concentrations only (Report No. EPA 600/S4-84-088). 1985. U.S. Environmental Protection Agency. Washington, DC.

U.S. EPA. The particle pollution report: Current understanding of air quality and emissions through 2003 (Report No. EPA 454-R-04-002) 2003. U.S. Environmental Protection Agency. Washington, DC.

U.S. EPA. Air quality criteria for particulate matter (Report No. EPA 600/P-99/002aF-bF): Volume 1. 2004. U.S. Environmental Protection Agency. Washington, DC.

U.S. EPA. CASTNET Factsheet. 2013. U.S. Environmental Protection Agency. Washington, DC. (http://epa.gov/castnet/javaweb/docs/CASTNET_Factsheet_2013.pdf).

U.S. EPA. 2014. Air Quality SyStem data for downloading. Available: http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm [accessed 3 November 2014].

Wang R, Handerson SB, Sbihi H, Allen RW, Brauer M. Temporal stability of land use regression models for traffic-related air pollution. Atmos Environ 2013;64:312-319.

Weichenthal S, Villeneuve PJ, Burnett RT, van Donkelaar A, Martin RV, Jones RR, DellaValle CT, Sandler DP, Ward MH, Hoppin JA. Long-term exposure to fine particulate matter: association with nonaccidental and cardiovascular mortality in the agricultural health study cohort. Environ Health Perspect. 2014;122(6):609-15.

Yanosky JD, Paciorek CJ, Suh HH. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the Northeastern and Midwestern United States. Environ

Health Perspect. 2009;117(4):522-529.

Yanosky JD, Paciorek CJ, Laden F, Hart JE, Puett RC, Liao D, Suh HH. Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. Environ Health. 2014;13:63.

Young MT, Sandler DP, DeRoo LA, Vedal S, Kaufman JD, London SJ. Ambient air pollution exposure and incident adult asthma in a nationwide cohort of U.S. women. Am J Respir Crit Care Med. 2014;15;190(8):914-921.

Yu HL and W CH. Retrospective prediction of intraurban spatiotemporal distribution of PM2.5 in Taipei. Atmos Environ 2010;44:3053-3065.

Table 1. Summary of $PM_{2.5}$ monitoring data used for $PM_{2.5}$ historical model development and validation

Network[a]	Spatial coverage	Regulatory monitoring network	Number of sites[b]	Number of observations[b]	Sampling period[b]	Annual average of $PM_{2.5}$ ($\mu g/m^3$)	
						Mean	SD
FRM	National (urban)	Yes	1,282	9,233	1999-2010	12.03	3.23
IMPROVE	National (rural)	Yes	178	1,567	1999-2010	5.44	2.94
			72	423	1990-1998	6.05	3.75
CASTNet	National (rural)	Yes	108	1,485	1987-2010	3.15	1.91
IPN	National (urban/rural)	Yes	16	18	1980-1981	21.31	6.69
CARB dichot	California (urban/rural)	Yes	33	247	1988-2001	19.35	7.78
CHS	Southern California (urban)	No	13	120	1994-2003	16.12	8.17

a. FRM = Federal Reference Method; IMPROVE = Interagency Monitoring of Protected Visual Environment; CASTNet = Clean Air Status and Trends Network; IPN = Inhalable Particulate Network; CARB dichot = California Air Resources Board dichotomous sampler monitoring; CHS = Children's Health Study

b. Number of sites and observations, and sampling period for the monitoring sites that meet the minimum inclusion criteria for computing representative annual averages

Table 2. Cross-validation statistics of the historical $PM_{2.5}$ models for 1999-2010 by year and region

Estimated trend		Linear trend from FRM/IMPROVE[a] $PM_{2.5}$		Generally-linear trend from CASTNet[a] $PM_{2.5}$ sulfate		Generally-linear trend from WBAN[a] visibility	
Cross-validation statistics		R^2	MSE	R^2	MSE	R^2	MSE
Year/region	N[b]						
All[c]	1,460 (10,800)	0.87	2.08	0.86	2.15	0.86	2.25
1999	523	0.86	3.29	0.86	3.29	0.85	3.46
2000	865	0.85	2.38	0.85	2.39	0.85	2.44
2001	988	0.86	2.32	0.86	2.32	0.86	2.35
2002	1,054	0.84	2.39	0.84	2.44	0.84	2.47
2003	969	0.85	2.13	0.84	2.20	0.84	2.26
2004	980	0.86	1.99	0.85	2.06	0.85	2.12
2005	940	0.88	2.08	0.88	2.14	0.87	2.24
2006	898	0.86	1.87	0.85	1.93	0.84	2.05
2007	937	0.86	1.85	0.86	1.94	0.85	2.07
2008	902	0.82	1.82	0.81	1.90	0.79	2.10
2009	884	0.80	1.61	0.79	1.73	0.77	1.89
2010	860	0.83	1.63	0.81	1.82	0.80	1.97
East[c]	1,056 (7,956)	0.86	1.19	0.86	1.26	0.86	1.26
Mountain West[c]	239 (1,594)	0.59	3.84	0.59	3.91	0.60	3.77
West Coast[c]	165 (1,250)	0.84	5.50	0.84	5.52	0.80	6.61

a. FRM = Federal Reference Method; IMPROVE = Interagency Monitoring of Protected Visual Environment; CASTNet = Clean Air Status and Trends Network; WBAN = Weather-Bureau-Army-Navy
b. Number of sites (Number of observations when different from the number of sites)
c. Annual averages from 1999 through 2010

Table 3. External validation statistics of the historical $PM_{2.5}$ models using $PM_{2.5}$ IMPROVE data for 1990-1998 by year and region

Estimated trend		Linear trend from FRM/IMPROVE[a] $PM_{2.5}$		Generally-linear trend from CASTNet[a] $PM_{2.5}$ sulfate		Generally-linear trend from WBAN[a] visibility	
Validation statistics		R^2	MSE	R^2	MSE	R^2	MSE
Year/region	N[b]						
All[c]	72 (423)	0.91	1.29	0.84	2.23	0.86	1.98
1990	30	0.85	1.07	0.78	1.59	0.70	2.20
1991	36	0.83	1.96	0.78	2.43	0.70	3.40
1992	37	0.91	1.42	0.84	2.52	0.85	2.47
1993	45	0.92	1.44	0.83	3.08	0.87	2.33
1994	50	0.92	1.07	0.84	2.09	0.89	1.44
1995	58	0.91	1.32	0.86	1.97	0.86	1.96
1996	56	0.93	0.87	0.88	1.58	0.91	1.21
1997	57	0.93	1.03	0.86	2.01	0.90	1.47
1998	54	0.90	1.64	0.83	2.88	0.87	2.13
East[c]	21 (120)	0.88	1.60	0.67	4.42	0.84	2.10
Mountain West[c]	34 (202)	0.25	0.87	0.04	1.11	0.00	1.94
West Coast[c]	17 (101)	0.69	1.76	0.67	1.88	0.66	1.93

a. FRM = Federal Reference Method; IMPROVE = Interagency Monitoring of Protected Visual Environment; CASTNet = Clean Air Status and Trends Network; WBAN = Weather-Bureau-Army-Navy
b. Number of sites (Number of observations when different from the number of sites)
c. Annual averages from 1990 through 1998

Table 4. External validation statistics of the historical $PM_{2.5}$ models using CHS, CARB dichot, and IPN data by year

	Estimated trend		Linear trend from FRM/IMPROVE[a] $PM_{2.5}$		Generally-linear trend from CASTNet[a] $PM_{2.5}$ sulfate		Generally-linear trend from WBAN[a] visibility	
	Validation statistics		R^2	MSE	R^2	MSE	R^2	MSE
Validation data[a]	Year	N[b]						
CHS	All[c]	13 (120)	0.76	16.04	0.76	15.87	0.81	12.91
	1994	12	0.71	26.90	0.69	28.55	0.80	18.79
	1995	12	0.66	35.60	0.63	39.78	0.75	26.76
	1996	12	0.77	19.33	0.75	20.82	0.82	14.94
	1997	12	0.83	9.71	0.84	9.04	0.88	6.96
	1998	12	0.83	8.26	0.87	6.48	0.87	6.44
	1999	12	0.73	18.48	0.75	17.07	0.74	17.31
	2000	12	0.80	11.77	0.82	10.51	0.82	10.95
	2001	12	0.82	14.38	0.85	11.84	0.86	10.70
	2002	12	0.81	10.27	0.82	9.73	0.79	10.96
	2003	12	0.88	5.69	0.90	4.92	0.89	5.28
CARB dichot	All[c]	33 (162)	0.55	30.69	0.48	35.75	0.61	26.75
	1988	8	0.09	94.11	0.00	110.73	0.15	88.36
	1989	12	0.25	82.20	0.10	98.75	0.33	73.11
	1990	11	0.68	22.74	0.53	32.95	0.76	16.64
	1991	12	0.31	85.45	0.16	103.26	0.43	69.74
	1992	14	0.51	28.60	0.40	34.92	0.63	21.90
	1993	15	0.54	15.06	0.33	21.85	0.66	10.92
	1994	13	0.77	16.66	0.69	22.24	0.84	11.37
	1995	12	0.71	11.99	0.63	15.30	0.70	12.50
	1996	15	0.52	16.03	0.66	11.36	0.57	14.54
	1997	15	0.41	10.14	0.59	7.07	0.45	9.51
	1998	16	0.31	16.91	0.37	15.52	0.30	17.10
	1999	12	0.85	5.72	0.84	6.26	0.82	6.98

	2000	6	0.53	5.81	0.46	6.69	0.41	7.25
	2001	3	0.00	88.49	0.00	87.15	0.00	84.42
IPN	All[c]	16 (18)	0.16	37.82	0.02	44.00	0.00	54.80
	1980	6	0.40	26.12	0.27	31.57	0.00	48.45
	1981	12	0.11	43.68	0.00	50.21	0.00	57.97

a. FRM = Federal Reference Method; IMPROVE = Interagency Monitoring of Protected Visual Environment; CASTNet = Clean Air Status and Trends Network; WBAN = Weather-Bureau-Army-Navy; CHS = Children's Health Study; CARB dichot = California Air Resources Board dichotomous sampler monitoring; IPN = Inhalable Particulate Network

b. Number of sites (Number of observations when different from the number of sites)

c. Annual averages from 1990 through 1998

FIGURE LEGENDS

Figure 1. Maps of A) FRM and IMPROVE sites for 1999-2010 used in model development and trend estimation, B) CASTNet and WBAN sites used for trend estimation, and C) IMPROVE sites for 1990-1998, CHS, CARB dichot, and IPN sites used in model evaluation

Figure 2. Estimated temporal trends based on $PM_{2.5}$ annual averages in FRM and IMPROVE, $PM_{2.5}$ sulfate annual averages in CASTNet, and visibility annual averages in WBAN

Figure 3. Scatter plots of observed and predicted $PM_{2.5}$ annual averages from the $PM_{2.5}$ historical model using the FRM/IMPROVE $PM_{2.5}$ trend across IMPROVE sites for 1990-1998

Figure 4. Predicted $PM_{2.5}$ annual averages in 1980, 1990, 2000, and 2010 from the 31-year $PM_{2.5}$ model using the extrapolated temporal trend based on $PM_{2.5}$ data for 1999-2010

Figure 5. Scatter plots of predicted $PM_{2.5}$ annual averages from the 31-year $PM_{2.5}$ model using the extrapolated temporal trend based on $PM_{2.5}$ data for 1999-2010 for 2000 vs. long-term averages for 1980-2000 weighted by times of residences across home addresses of 2,466 participants who never moved for 1980-2000 and 5,086 MESA/MESA Air participants who moved at least once in each of the six MESA city areas

A) Model development/
Trend estimation

B) Trend estimation

C) Model evaluation

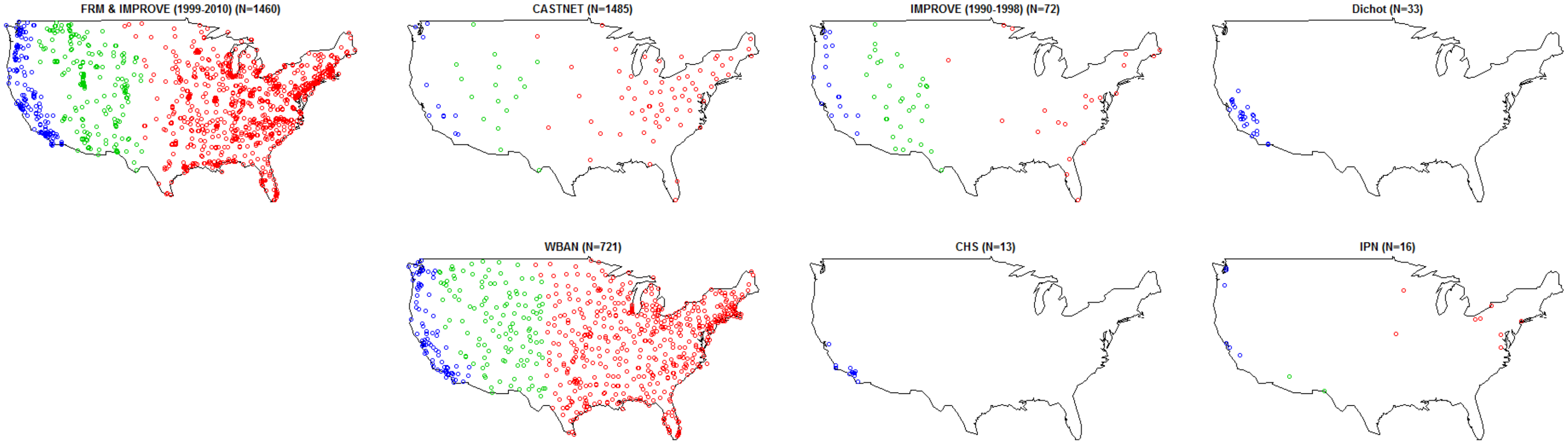

Figure 1. Maps of A) FRM and IMPROVE sites for 1999-2010 used in model development and trend estimation, B) CASTNet and WBAN sites used for trend estimation, and C) IMPROVE sites for 1990-1998, CHS, CARB dichot, and IPN sites used in model evaluation

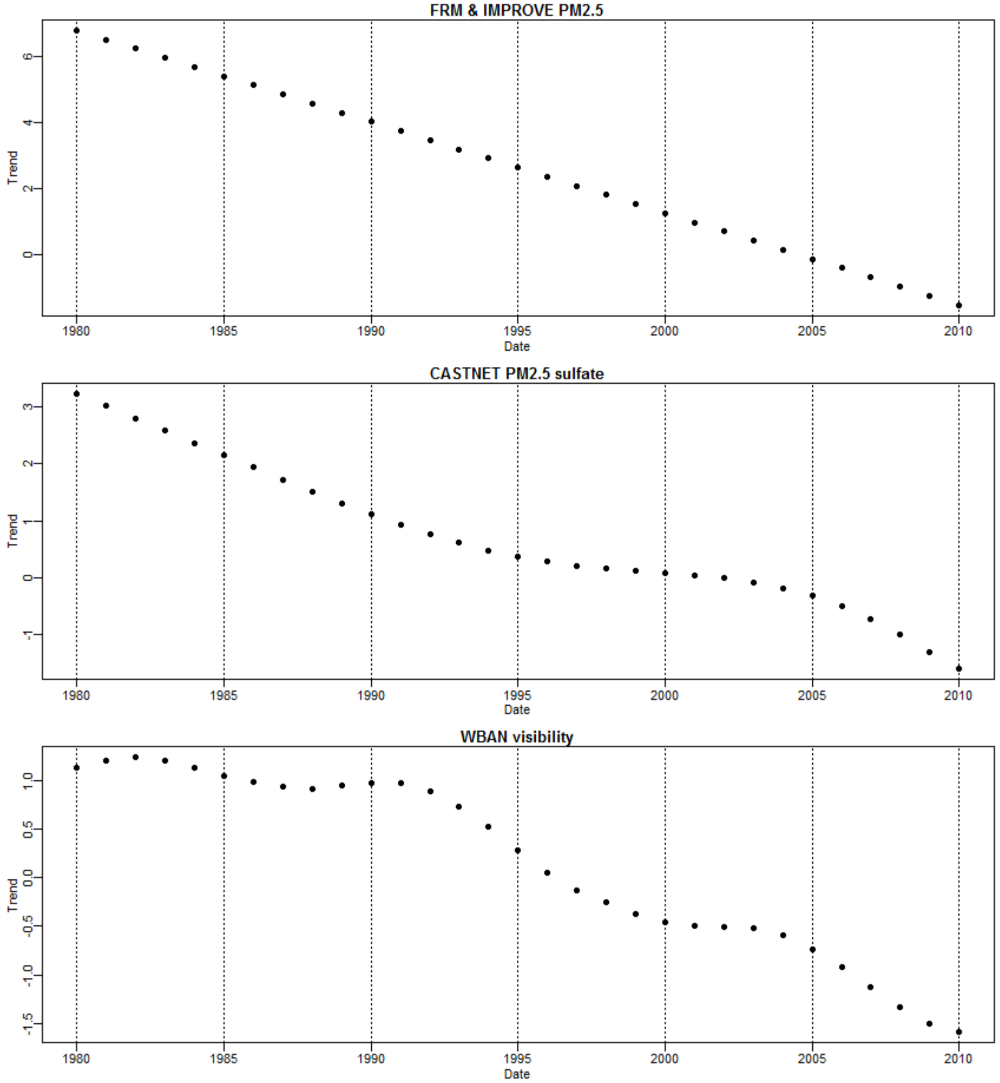

Figure 2. Estimated temporal trends based on $PM_{2.5}$ annual averages in FRM and IMPROVE, $PM_{2.5}$ sulfate annual averages in CASTNet, and visibility annual averages in WBAN

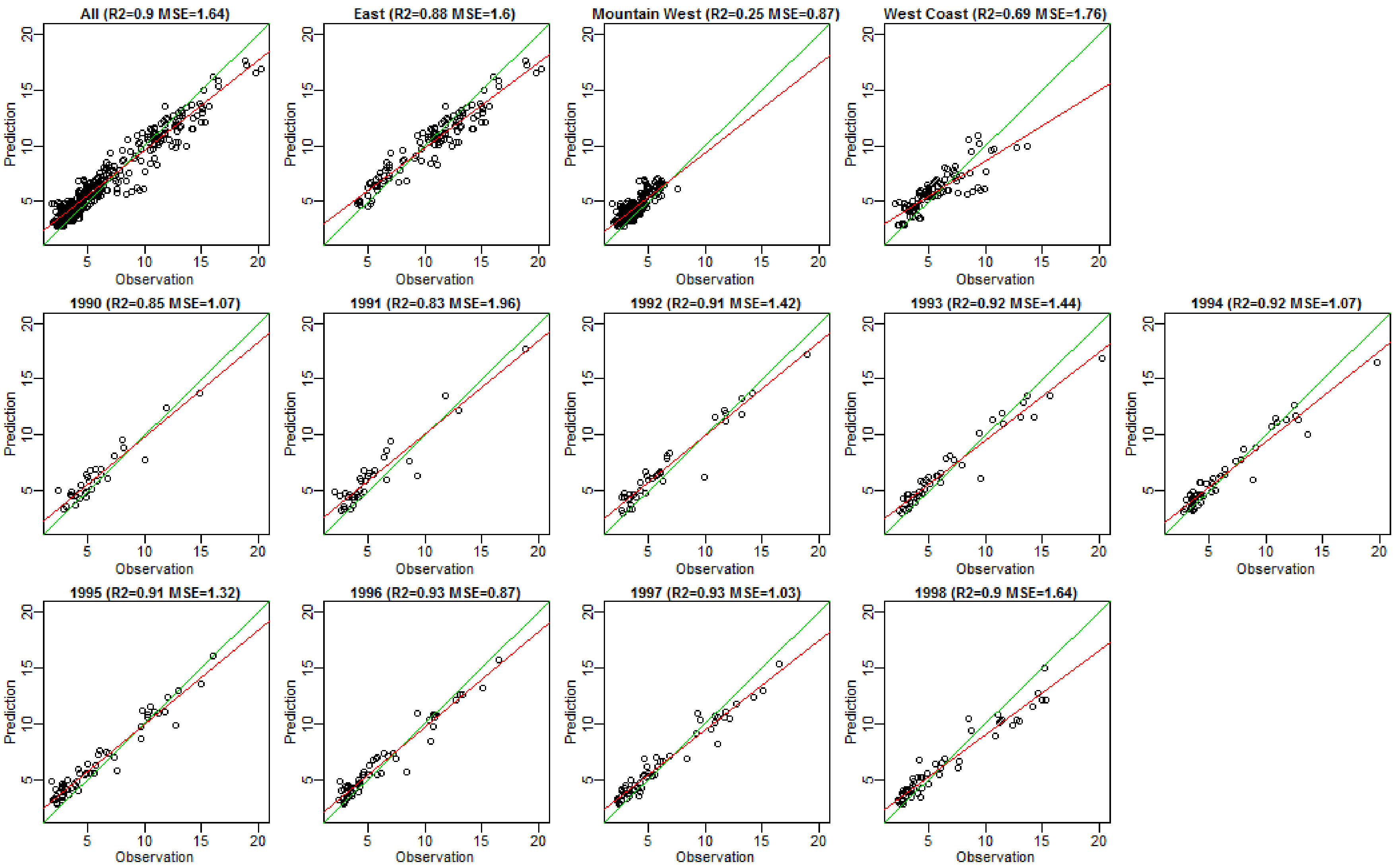

Figure 3. Scatter plots of observed and predicted $PM_{2.5}$ annual averages from the $PM_{2.5}$ historical model using the FRM/IMPROVE $PM_{2.5}$ trend across IMPROVE sites for 1990-1998

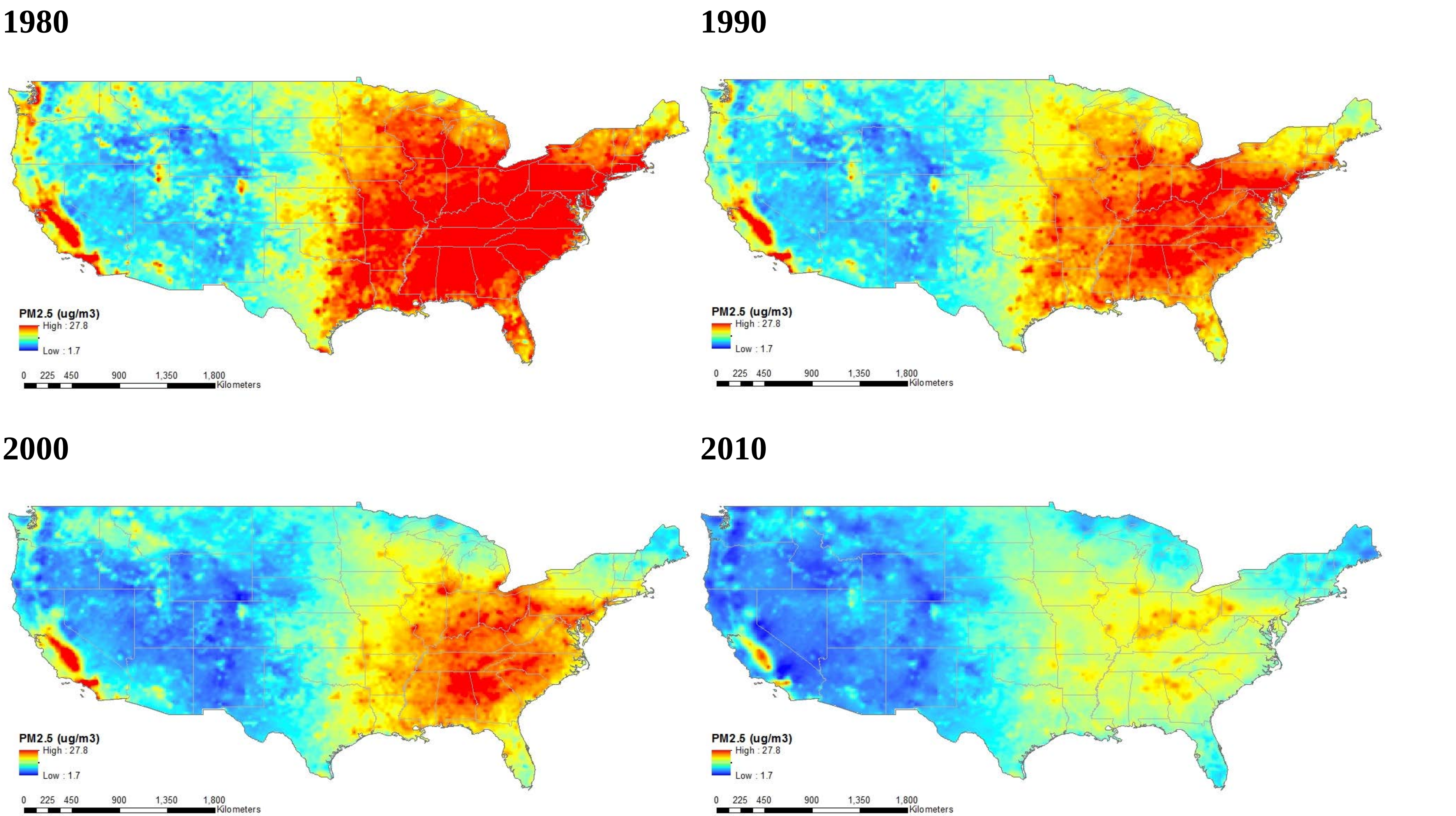

Figure 4. Predicted $PM_{2.5}$ annual averages in 1980, 1990, 2000, and 2010 from the 31-year $PM_{2.5}$ model using the extrapolated temporal trend based on $PM_{2.5}$ data for 1999-2010

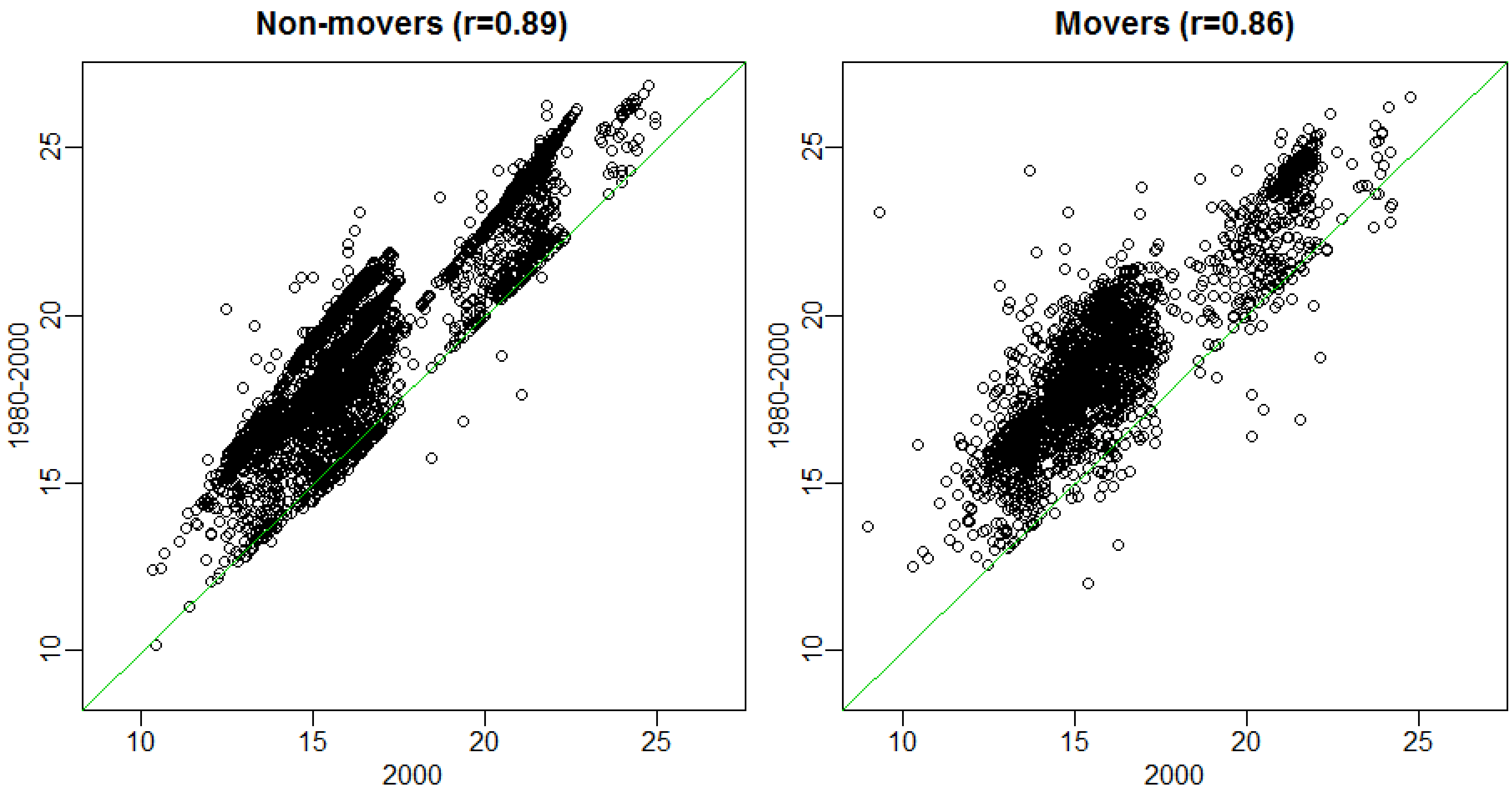

Figure 5. Scatter plots of predicted $PM_{2.5}$ annual averages from the 31-year $PM_{2.5}$ model using the extrapolated temporal trend based on $PM_{2.5}$ data for 1999-2010 for 2000 vs. long-term averages for 1980-2000 weighted by times of residences across home addresses of 2,466 participants who never moved for 1980-2000 and 5,086 MESA Air participants who moved at least once

Supplemental Material

Table S1. List of geographic variables

Category	Measure	Variable description
Traffic	Distance to the nearest road	Any road, A1, intersection
	Sum within buffers of 0.05-15 km	A1, A2+A3, truck route, intersections
Population	Sum within buffers of 0.5-3 km	Population in block groups
Land use (Urban)	Percent within buffers of 0.05-15 km	Urban or Built-Up land
		(residential, commercial, industrial, transportation, urban)
		Developed low, medium, and high density
		Developed open space
Land use (Rural)	Percent within buffers of 0.05-15 km	Agricultural land (cropland, groves, feeding)
		Rangeland (herbaceous, shrub)
		Forest land (deciduous, evergreen, mixed)
		Water (streams, lakes, reservoirs, bays)
		Wetland
		Barren land (beaches, dry salt flats, sand, mines, rock)
		Tundra
		Perennial snow or Ice
Position	Coordinates	Longitude, latitude
Source	Distance to the nearest source	Coastline, Coastline (rough)
		Commercial area
		Railroad, Railyard
		Airport
		Major airport
		Large port
		City hall
Emission	Sum within buffers of 3-30 km	$PM_{2.5}$
		PM_{10}

		CO SO_2 NO_x
Vegetation	Quantiles within buffers of 0.5-10 km	Normalized Difference Vegetation Index (NDVI)
Imperviousness	Percent within buffers of 0.05-5 km	Impervious surface value
Elevation	Elevation above sea levels Counts of points above or below a threshold within buffers of 1-5 km	Elevation value
Residual oil	Distance to the nearest boiler Sum within buffers of 0.1-3 km	Residual oil grade 4 or 6 Total residual oil active heating capacity

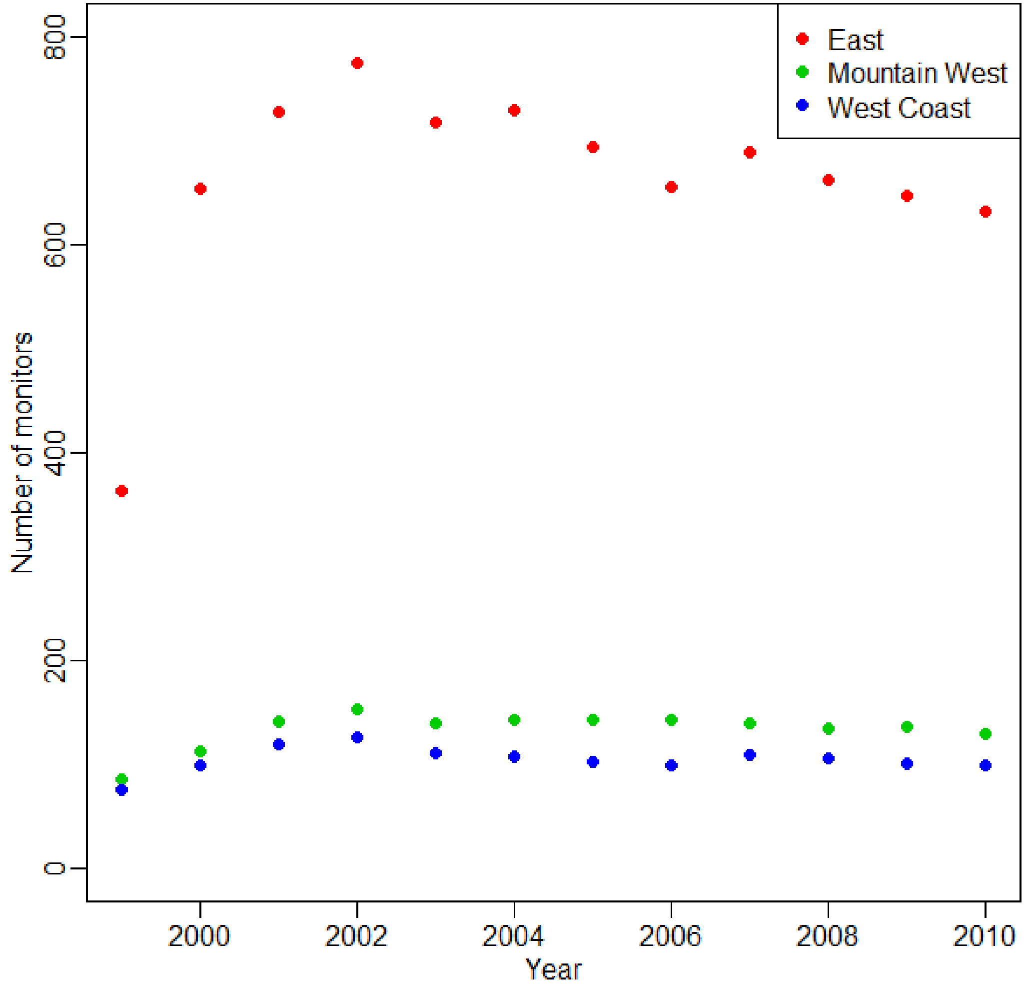

Figure S1. Number of monitoring sites for $PM_{2.5}$ in FRM and IMPROVE from 1999 through 2010

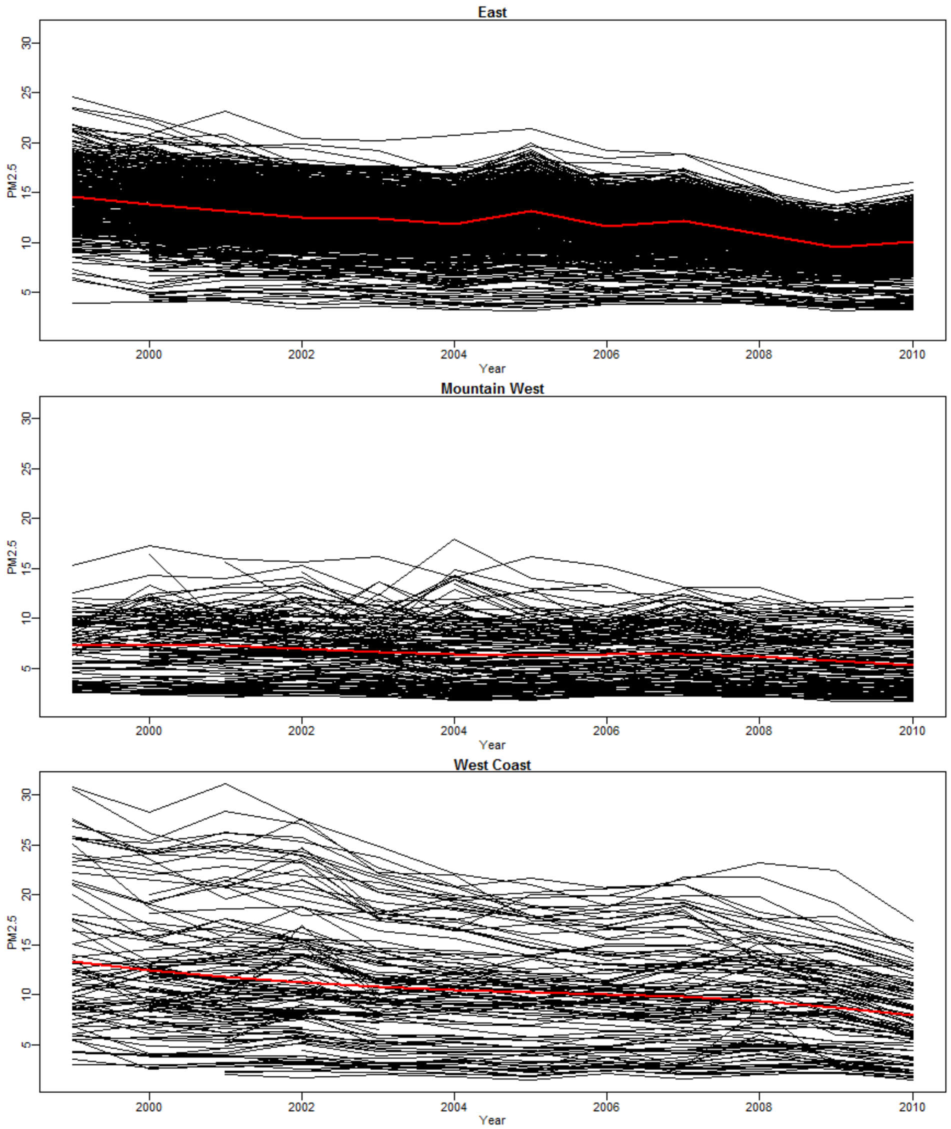

Figure S2. Time-series plots of annual averages for $PM_{2.5}$ across FRM and IMPROVE sites for 1999-2010 by region

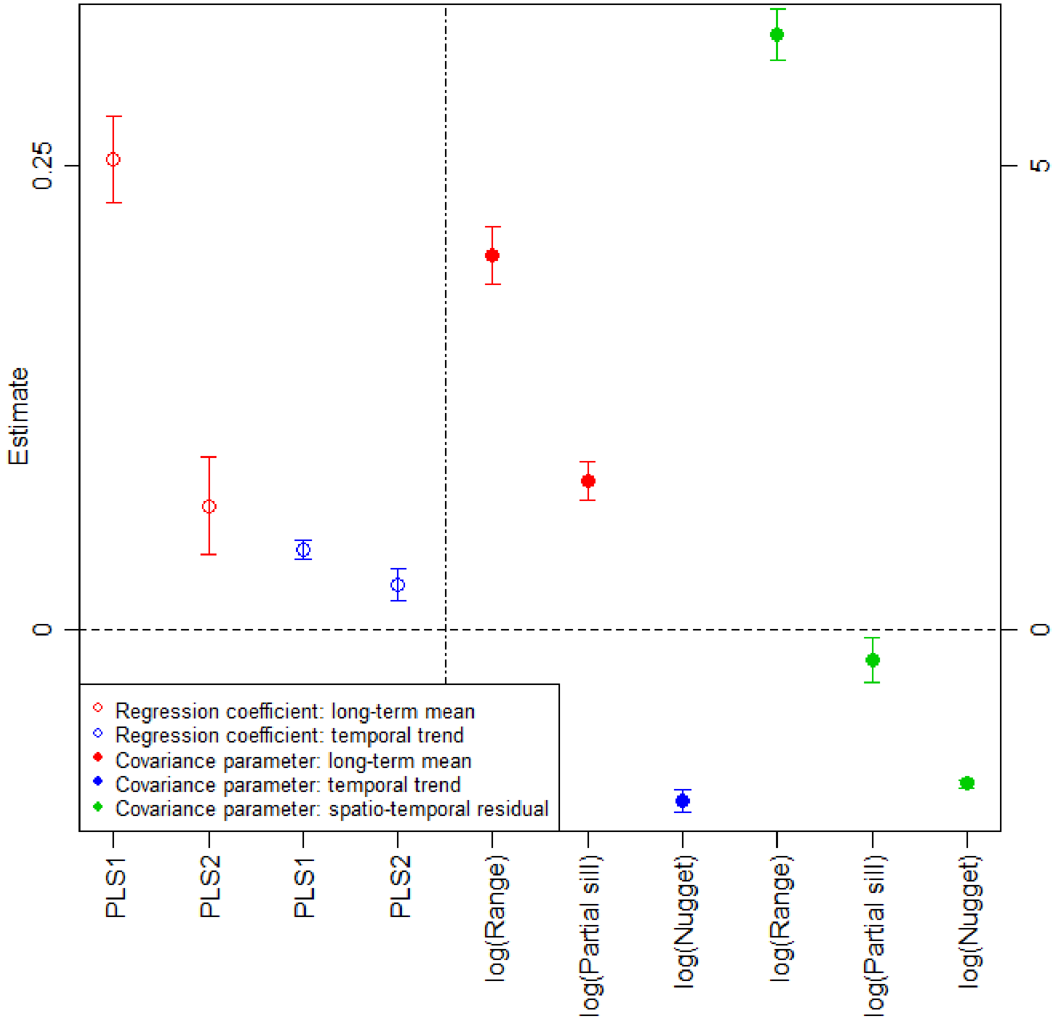

Figure S3. Estimated regression and variance parameters of the $PM_{2.5}$ prediction model for 1980-2010