ENVIRONMENTAL HEALTH PERSPECTIVES

http://www.ehponline.org

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, Casey Olives, Lianne Sheppard, Paul D. Sampson, Timothy V. Larson, Joshua P. Keller, and Joel D. Kaufman

http://dx.doi.org/10.1289/EHP131

Received: 6 August 2015
Revised: 5 March 2016
Accepted: 18 May 2016
Published: 24 June 2016

Note to readers with disabilities: *EHP* will provide a 508-conformant version of this article upon final publication. If you require a 508-conformant version before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

NIH National Institute of Environmental Health Sciences

Environ Health Perspect DOI: 10.1289/EHP131
Advance Publication: Not Copyedited

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,[2], Lianne Sheppard,[2,3], Paul D. Sampson,[4], Timothy V. Larson,[2,5], Joshua P. Keller,[3], and Joel D. Kaufman,[2,6,7]

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

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

[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), the National Institute of Environmental Health Sciences (T32ES015459), and the National Research Foundation of Korea (Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education: 2013R1A6A3A04059017). We would like to thank Fred Lurmann and the Southern California Children's Health Study research team for providing $PM_{2.5}$ data collected in the Southern California Children's Health Study.

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 U.S. Environmental Protection Agency (EPA) Federal Reference Method (FRM) network for $PM_{2.5}$ was established in 1999.

Objectives: We evaluated a novel statistical approach to produce high quality exposure predictions from 1980-2010 in the continental U.S. 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. Temporal trends before 1999 were estimated by using a) extrapolation based on $PM_{2.5}$ data in FRM/IMPROVE, b) $PM_{2.5}$ sulfate data in the Clean Air Status and Trends Network, and c) visibility data across the Weather-Bureau-Army-Navy network. We validated the models using $PM_{2.5}$ data collected before 1999 from IMPROVE, California Air Resources Board dichotomous sampler monitoring (CARB dichot), the Children's Health Study (CHS), and the Inhalable Particulate Network (IPN).

Results: In our validation using pre-1999 data, the prediction model performed well across three trend estimation approaches when validated using IMPROVE and CHS data (R^2=0.84–0.91) with lower R^2s in early years. Model performance using CARB dichot and IPN data was worse (0.00–0.85) most likely due to smaller numbers of monitoring sites and inconsistent sampling methods.

Conclusions: 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., Eeftens 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 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 countries. 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-size particle data, or physical or chemical models complemented by visibility, emission, meteorology, and satellite data (Beelen et al. 2014; Brauer et al. 2012; Hogrefe et al. 2009; Hystad et al. 2012; Lall et al. 2004; Molnar et al. 2015; Ozkaynak et al. 1985; Paciorek et al. 2009; Yanosky et al. 2009). 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 $PM_{2.5}$ concentrations, IMPROVE sites were established to monitor visibility and located mostly in wilderness areas and national parks (Hand 2011; U.S. EPA 2004a). We downloaded all available data from 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 any 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 temporal trend estimation, whereas we reserved the IMPROVE data from 1990-1998 for model validation. We categorized all monitoring sites into three regions: East, Mountain West, and West Coast (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 (U.S. EPA 2005), 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 model evaluation in years 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 1979-1982 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). We applied the same inclusion criteria to compute annual averages to sites in the three model evaluation networks. These criteria reduced the number of IPN sites from 102 for 1979-1982 to 16 for 1980-1981, whereas the other three networks gave same or consistent numbers of sites.

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 $PM_{2.5}$ 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 land use imaginary 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) (Bild et al. 2002) and associated MESA Air project (Kaufman et al. 2012). These participants provided historical residential addresses dating back to 1980. In addition, we generated coordinates of 12,501 points on a 25 kilometer grid across the continental U.S.

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). To briefly summarize, the MESA Air spatio-temporal prediction model analyzed 2-week averages of $PM_{2.5}$ as a function of a spatially varying long-term mean, spatially varying temporal trends, and spatio-temporal residuals. The spatially varying temporal trends were composed of spatially-varying trend coefficients and trend basis functions. The trend basis functions were estimated from singular value decomposition of the data from sites with long time series (Fuentes et al. 2006). The spatially-varying long-term mean and trend coefficients were estimated using universal kriging, which integrates geographic predictors and spatial smoothing (Banerjee et al. 2003). Prior to regression modeling, we used partial least squares (PLS) to reduce the dimension of the hundreds of geographic variables to a limited number of 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 parameterized by three components: the range, partial sill, and nugget. The spatially-dependent and temporally-independent spatio-temporal residuals were modeled by using simple kriging. Whereas the MESA Air model was based on 2-week averages, in this work we modeled the log annual average $PM_{2.5}$ concentrations from 1999 through 2010. For the trend estimation, we considered only sites with more than six years of monitoring out of the twelve possible years. To avoid unnecessary complexity in the model, we assumed a single temporal trend, no spatial structure for the trend coefficient (zero range and partial sill), and two PLS predictors. 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 root mean square error (RMSE) and MSE-based R-square (R^2) statistics for annual averages (Keller et al. 2015). The MSE-based R^2 was computed by subtracting from 1 the ratio of the MSE to the variance of the data. This evaluates predictions compared to observations about the identity line. In contrast, traditional regression-based R^2, the squared correlation coefficient, compares predictions to observations about a regression line; this can result in overestimation of prediction ability. We presented cross-validation statistics for each year and all twelve years combined for all sites, and for all twelve years combined within each of the three regions. In addition to spatial performance, we examined temporal performance by using the median of cross-validation statistics at each site where there are more than 6 years of data. To aid in assessing bias, we have also provided slopes and intercepts from the regression of cross-validated predictions on observations.

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 RMSEs and MSE-based R^2s using these external data sources for all years and regions as well as by year and region. We also estimated intercepts and slopes of the best-fit lines.

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.

In addition, we 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/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) $\mu g/m^3$, respectively (Table 1). The number of monitoring sites was small in 1999 compared to 2000-2010 (Supplemental Material, Figure S1) and most sites for 1999-2010 were located in the East region (Figure 1). 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 described earlier. Whereas the extrapolated trend based on the $PM_{2.5}$ data was linear, the trends estimated using $PM_{2.5}$ sulfate and visibility measurements had different rates of decrease in different time periods with approximate linearity over time.

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 (Supplemental Material, Tables S2-S3). Temporally-characterized R^2s at each site over years were lower (0.55-0.58) than spatially-characterized R^2s in each year across sites, possibly because of relatively small temporal variability for twelve years compared to large spatial variability across U.S.. 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 S2 (data not shown).

Supplemental Material, Figure S3 shows estimated regression and variance parameters for the long-term mean, the temporal trend coefficient, and spatio-temporal residuals, whereas Figure S4 displays loadings of geographic variables for each PLS predictor. Regression coefficients of the two PLS predictors for both the long-term mean and trend coefficient were statistically significantly different from 0, reflecting that spatial variation in the long-term mean and temporal trend is explainable by the geographic variables used to create the PLS predictors. 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. Contribution of the temporal trend to cross-validated predictions was smaller than the long-term mean (Supplemental Material, Table S4).

Tables 2 and 3 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) (Table 3, Supplemental Material, Figure S5). CARB dichot data gave high R^2s over 0.5 except for some years, whereas IPN data consistently showed low R^2s (Table 3, Supplemental Material, Figures S6-S7). Variability of predicted $PM_{2.5}$ annual average concentrations tended to be smaller than the observations with slopes less than 1 in regressions on observations (Supplemental Material, Tables S5-S6). Supplemental Material, Figures S8 and S9 show that the differences between maximum and minimum of predicted $PM_{2.5}$ annual averages across three trend estimation approaches over years at IMPROVE sites. Median differences were small and mostly less than 2 ug/m^3. In addition, the differences were larger in early years than recent years, indicating increasing prediction uncertainty of trend estimation in early years.

Figure 4 shows predicted $PM_{2.5}$ concentrations dramatically decreased across decennial years from 1980 through 2010 with only a few areas that remained consistently high in the continental U.S. over all three decades. The decreasing trend was also clear over 31 years 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 and Supplemental Material, Figure S10). The two sets of predictions showed high correlations with 2000 annual averages (0.86-0.89) with slightly lower correlation and the slope more attenuated for movers 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. While the prediction model performed well for 1999-2010 as assessed by cross-validation, the pre-1999 external validation is a more important indicator for evaluating historical predictions. We found the pre-1999 predictions also generally performed well across three trend estimation approaches, particularly for the external 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/MESA Air participant residences tended to be higher than and somewhat unsystematically different from 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; U.S. 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 (Xing et al. 2013). The decreasing trend of annual sulfur dioxide emissions from power plants matches well with that of sulfate concentrations in the eastern half of the U.S. between 1990 and 2003 (U.S. EPA 2004b). 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 from the visibility data demonstrated a somewhat different shape from that of the $PM_{2.5}$ sulfate trend, which could possibly be driven by meteorological influence (Hand et al. 2014). In addition to a non-linear relationship between $PM_{2.5}$ concentrations and visibility depending on chemical composition and weather conditions, the change of sampling methods for visibility from the more subjective human eye to the more objective optical instruments beginning in 1992

(Hyslop et al. 2009; U.S. EPA 2005) coincides with the observed state of a marked downward trend.

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. investigators developed ratio models that predict monthly averages of $PM_{2.5}$ concentrations for 1988-1998 by multiplying the ratios 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 (Ozkaynak 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. One strength of using the IMPROVE data as a validation dataset is that it is national. It gave the highest R^2s among all external validation datasets, possibly due to its 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, gave lower R^2s, including values less than 0.5 for some years. These low R^2 estimates could be caused by the lower between-site variability in California (vs. the entire U.S.) as well as the small number of sites, a few of which had poor predictions. Another 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, possibly driven by the limited amount of IPN sites and the inconsistency of the IPN sampling protocol with that of FRM. 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. Furthermore, the IPN years of 1980-1981 are the earliest years of our prediction period and may reflect the most uncertainty in trend estimation.

This study includes some limitations and implications for future research. We used time-constant geographic variables which do not account for changes in spatial characteristics over time. However, among the approximately 300 geographic variables that we used for estimating PLS predictors were two sources of land use data: land cover data created in 1970s and 1980s and satellite land use imaginary data generated in 2007. These two data representing spatial differences in land use on two different time periods about 30 years apart, and our 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 collected 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. Some previous studies have used aerosol optical depth (AOD) data to improve prediction models for $PM_{2.5}$ (Beckerman et al. 2013; Hystad et al. 2012; Kloog et al. 2011). These models used short-term or long-term averages of AOD. Future studies should investigate how to incorporate AOD measurements into spatio-temporal prediction models for extended time periods and whether the addition of AOD improves the model performance.

As with application of any predicted exposure to health analyses, using predicted $PM_{2.5}$ concentrations from our historical prediction model may impact the estimates in subsequent health analyses due to exposure measurement error. As others have shown, we note that the high R^2s we obtained do not guarantee the accuracy or proper coverage of health effect estimates due to Berkson- and classical-like measurement error (Szpiro 2011a). Several simulation studies have shown that well-performing exposure models can still produce biased and/or imprecise health effect estimates (Alexeeff et al. 2014; Szipro et al. 2011b). One possible explanation for this feature is that the monitor locations do not represent the study population locations, resulting in monitored exposures spatially non-compatible with population's exposures (Szpiro & Paciorek 2013).

Our results suggest the importance of incorporating changes in air pollution concentrations in cohort studies. We showed that long-term $PM_{2.5}$ prediction averages for 31 years incorporating mobility were systematically higher than 2000 predictions among non-movers and were non-systematically different in movers. This pattern varied by cities, as suggested by the Supplemental Material, Figure S10, possibly depending on the extent of exposure contrast and the population's mobility between low and high exposure areas within a city. Using exposure predictions from a later period of follow-up in epidemiological study, as commonly done (Beelen et al. 2008; Cesaroni et al. 2013), may not adequately represent long-term exposures and might impact health effect findings.

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 point-wise prediction 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 accurately represent long-term average estimates of some people, and could impact subsequent health analyses.

REFERENCES

Alexeeff SE, Schwartz J, Kloog I, Chudnovsky A, Koutrakis P, Coull BA. 2015. Consequences of kriging and land use regression for PM2.5 predictions in epidemiologic analyses: insights into spatial variability using high-resolution satellite data. J Expo Sci Environ Epidemiol 25(2):138-44.

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

Beelen R, Hoek G, van den Brandt PA, Goldbohm RA, Fischer P, Schouten LJ, et al. 2008. Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). Environ Health Perspect 116(2):196-202.

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

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

Beckerman BS, Jerrett M, Serre M, Martin RV, Lee SJ, van Donkelaar A, et al. 2013. A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States. Environ Sci Technol 47(13):7233-7241.

Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. 2002. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 156(9):871-881.

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

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

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

Chen H, Goldberg MS, Crouse DL, Burnett RT, Jerrett M, Villeneuve PJ, et al. 2010. Back-extrapolation of estimates of exposure from current land-use regression models. Atmos Environ 44:4346-4354.

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

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

Gulliver J, de Hoogh K, Hansell A, Vienneau D. 2013. Development and back-extrapolation of NO2 land use regression models for historic exposure assessment in Great Britain. Environ Sci Technol. 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, et al. 2014. Widespread reductions in haze across the United States from the early 1990s through 2011. Atmos Environ 94:671-679.

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

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

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

Hystad P, Demers PA, Johnson KC, Brook J, van Donkelaar A, Lamsal L, et al. 2012. Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. Environ Health 11:22.

Kaufman JD, Adar SD, Allen RW, Barr RG, Budoff MJ, Burke GL, et al. 2012. Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). Am J Epidemiol 176(9):825-837.

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

Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos Environ 45:6267-6275.

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

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

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

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

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

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

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

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

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

Puett RC, Hart JE, Yanosky JD, Paciorek CJ, Schwartz J, Suh H, et al. 2009. Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. Environ Health Perspect 117(11):1697-701.

Sampson PD, Szpiro AA, Sheppard L, Lindström J, Kaufman JD. 2011. Pragmatic Estimation of a Spatio-Temporal Air Quality Model with Irregular Monitoring Data. Atmos Environ 45:6593-6606.

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

Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar S, Kaufman JD. 2010. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. Environmetrics. 21:606-631.

Szpiro AA, Sheppard L, Lumley T. 2011a. Efficient measurement error correction with spatially misaligned data. Biostatistics 12(4): 610–623.

Szipro AA, Paciorek CJ, Sheppard L. 2011b. Does more accurate exposure prediction necessarily improve health effect estimates? Epidemiology 22(5);680-685.

Szipro AA, Paciorek CJ. 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. Environmetrics 24;501-517.

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

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

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

U.S. EPA. 2004b. The particle pollution report: current understanding fo air quality and emissions through 2003 (Report No. EPA 454-R-04-002). U.S. Environmental Protection Agency. North Carolina, Research Triangle Park.

U.S. EPA. 2005. Review of the national ambient air quality standards for particle matter: policy assessment of scientific and technical information (Report No. EPA 452-R-05-005a) U.S. Environmental Protection Agency. North Carolina, Research Triangle Park

U.S. EPA. 2013. CASTNET Factsheet. 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 August, 3 2015].

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

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

Xing J, Pleim J, Mathur R, Pouliot G, Hogrefe C, Gan CM, et al. 2013. Historical gaseous and primary aerosol emissions in the United States from 1990 to 2010. Atmos Chem Phys 13:7531-7549.

Yanosky JD, Paciorek CJ, Suh HH. 2009. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the Northeastern and Midwestern United States. Environ Health Perspect 117(4):522-529.

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

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

Yu HL and Wang CH. 2010. Retrospective prediction of intraurban spatiotemporal distribution of PM2.5 in Taipei. Atmos Environ 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, number of observations, and sampling period for the monitoring sites that meet the minimum inclusion criteria for computing representative annual averages

Table 2. 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		FRM/IMPROVE[a] $PM_{2.5}$		CASTNet[a] $PM_{2.5}$ sulfate		WBAN[a] visibility	
Validation statistics		R^2	RMSE (μg/m³)	R^2	RMSE (μg/m³)	R^2	RMSE (μg/m³)
Year/region	N[b]						
All[c]	72 (423)	0.91	1.14	0.84	1.49	0.86	1.41
1990	30	0.85	1.04	0.78	1.26	0.70	1.48
1991	36	0.83	1.40	0.78	1.56	0.70	1.84
1992	37	0.91	1.19	0.84	1.59	0.85	1.57
1993	45	0.92	1.20	0.83	1.76	0.87	1.53
1994	50	0.92	1.03	0.84	1.45	0.89	1.20
1995	58	0.91	1.15	0.86	1.41	0.86	1.40
1996	56	0.93	0.93	0.88	1.26	0.91	1.10
1997	57	0.93	1.01	0.86	1.42	0.90	1.21
1998	54	0.90	1.28	0.83	1.70	0.87	1.46
East[c]	21 (120)	0.88	1.27	0.67	2.10	0.84	1.45
Mountain West[c]	34 (202)	0.25	0.93	0.04	1.06	0.00	1.39
West Coast[c]	17 (101)	0.69	1.33	0.67	1.37	0.66	1.39

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

	Estimated trend		FRM/IMPROVE[a] $PM_{2.5}$		CASTNet[a] $PM_{2.5}$ sulfate		WBAN[a] visibility	
	Validation statistics		R^2	RMSE (μg/m^3)	R^2	RMSE (μg/m^3)	R^2	RMSE (μg/m^3)
Validation data[a]	Year	N[b]						
CHS	All[c]	13 (120)	0.76	4.00	0.76	3.98	0.81	3.59
	1994	12	0.71	5.19	0.69	5.34	0.80	4.33
	1995	12	0.66	5.97	0.63	6.31	0.75	5.17
	1996	12	0.77	4.40	0.75	4.56	0.82	3.86
	1997	12	0.83	3.12	0.84	3.01	0.88	2.64
	1998	12	0.83	2.87	0.87	2.55	0.87	2.54
	1999	12	0.73	4.30	0.75	4.13	0.74	4.16
	2000	12	0.80	3.43	0.82	3.24	0.82	3.31
	2001	12	0.82	3.79	0.85	3.44	0.86	3.27
	2002	12	0.81	3.20	0.82	3.12	0.79	3.31
	2003	12	0.88	2.39	0.90	2.22	0.89	2.30
CARB dichot	All[c]	33 (162)	0.55	5.54	0.48	5.98	0.61	5.17
	1988	8	0.09	9.70	0.00	10.52	0.15	9.40
	1989	12	0.25	9.07	0.10	9.94	0.33	8.55
	1990	11	0.68	4.77	0.53	5.74	0.76	4.08
	1991	12	0.31	9.24	0.16	10.16	0.43	8.35
	1992	14	0.51	5.35	0.40	5.91	0.63	4.68
	1993	15	0.54	3.88	0.33	4.67	0.66	3.30
	1994	13	0.77	4.08	0.69	4.72	0.84	3.37
	1995	12	0.71	3.46	0.63	3.91	0.70	3.54
	1996	15	0.52	4.00	0.66	3.37	0.57	3.81
	1997	15	0.41	3.19	0.59	2.66	0.45	3.08
	1998	16	0.31	4.11	0.37	3.94	0.30	4.14
	1999	12	0.85	2.39	0.84	2.50	0.82	2.64

	2000	6	0.53	2.41	0.46	2.59	0.41	2.69
	2001	3	0.00	9.41	0.00	9.34	0.00	9.19
IPN	All[c]	16 (18)	0.16	6.15	0.02	6.63	0.00	7.40
	1980	6	0.40	5.11	0.27	5.62	0.00	6.96
	1981	12	0.11	6.61	0.00	7.09	0.00	7.61

a. FRM = Federal Reference Method; IMPROVE = Interagency Monitoring of Protected Visual Environment; CASTNet = Clean Air Status and Trends Network; WBAN = Weather-Bureau-Army-Navy; 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 (blue, green, and red circles represent West, Mountain West, and East regions); Maps generated using locations of regulatory monitoring sites downloaded from the EPA website (http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily) and boundaries in the R package

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

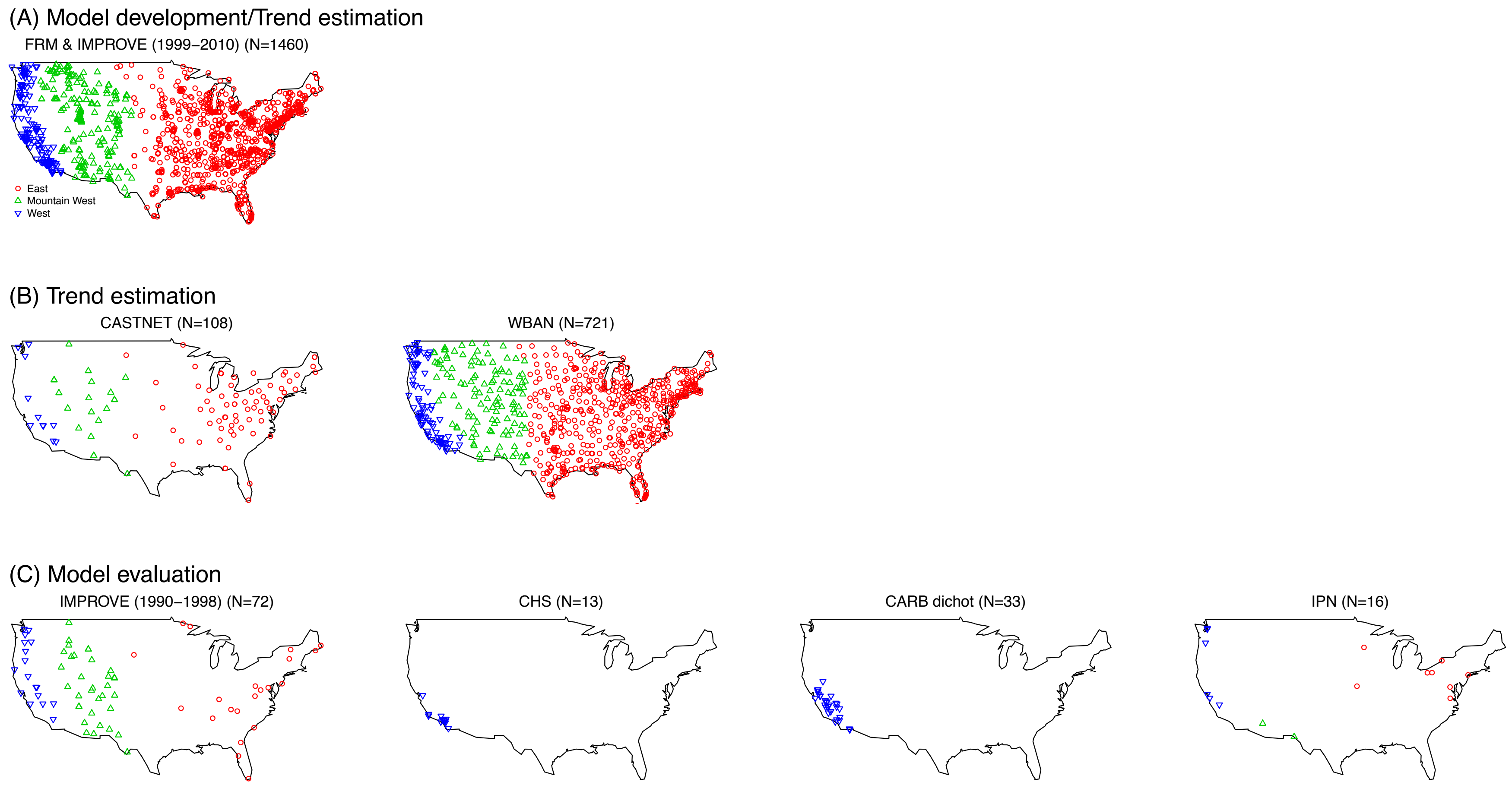
Figure 3. Scatter plots of observed and predicted $PM_{2.5}$ annual averages from the $PM_{2.5}$ historical model using the FRM/IMPROVE $PM_{2.5}$ trend across IMPROVE sites for 1990-1998

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; Maps generated using model outputs discussed in the "Development of the $PM_{2.5}$ model for 1980-2010" of the "Methods" section and boundaries obtained from the U.S. census

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 5,086

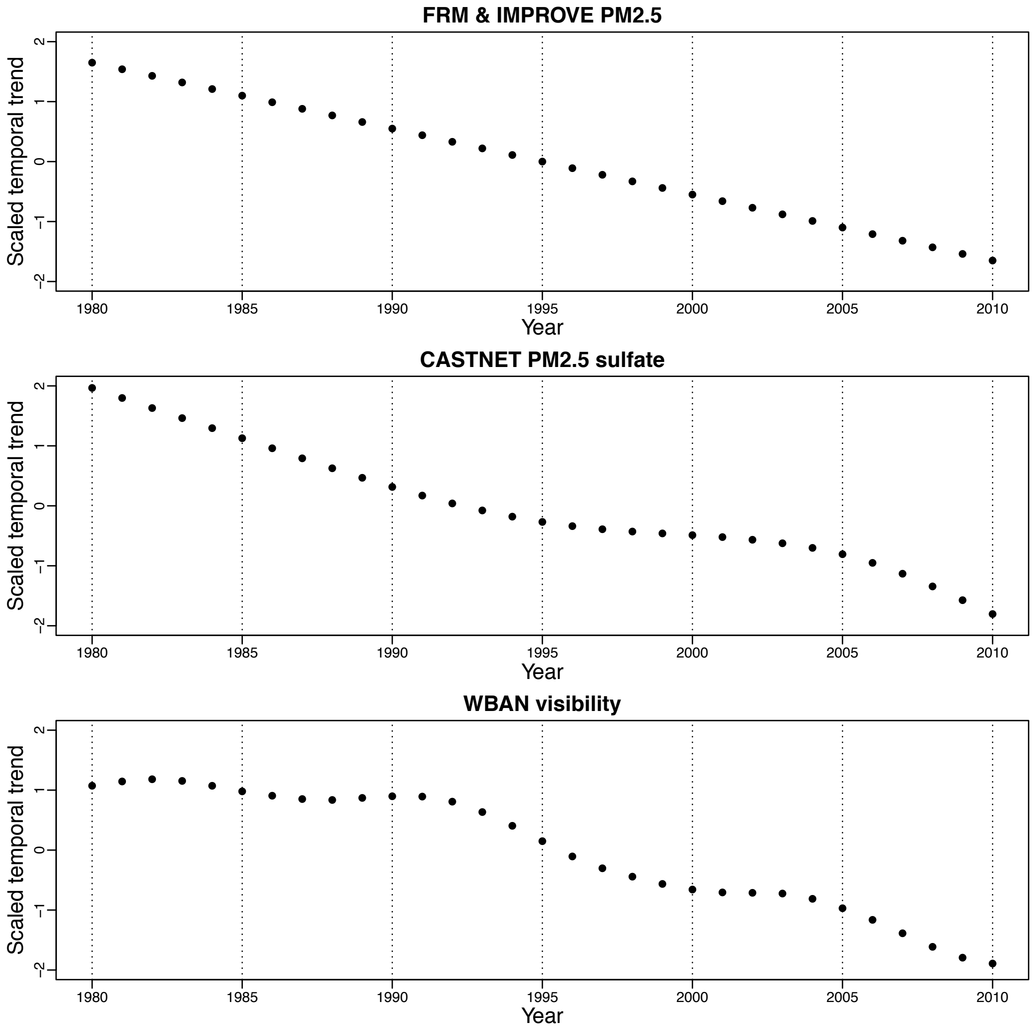
participants who never moved for 1980-2000 and 2,466 MESA/MESA Air participants who moved at least once

Figure 1.


(A) Model development/Trend estimation
FRM & IMPROVE (1999–2010) (N=1460)
East
Mountain West
West
(B) Trend estimation
CASTNET (N=108)
WBAN (N=721)
(C) Model evaluation
IMPROVE (1990–1998) (N=72)
CHS (N=13)
CARB dichot (N=33)
IPN (N=16)

Environ Health Perspect DOI: 10.1289/EHP131
Advance Publication: Not Copyedited

Figure 2.



Environ Health Perspect DOI: 10.1289/EHP131
Advance Publication: Not Copyedited

Figure 3.

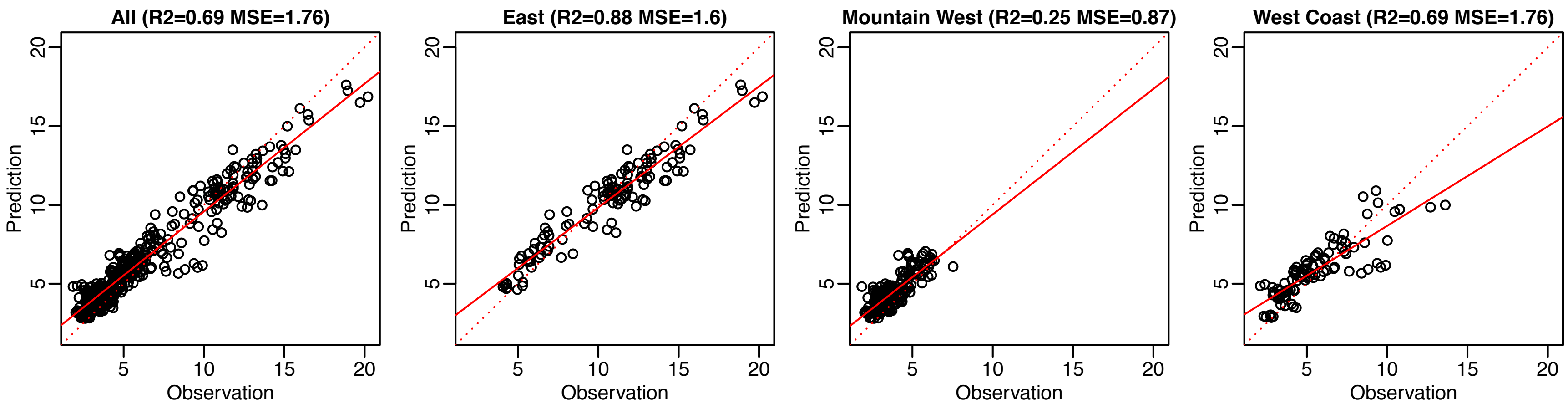


Figure 4.

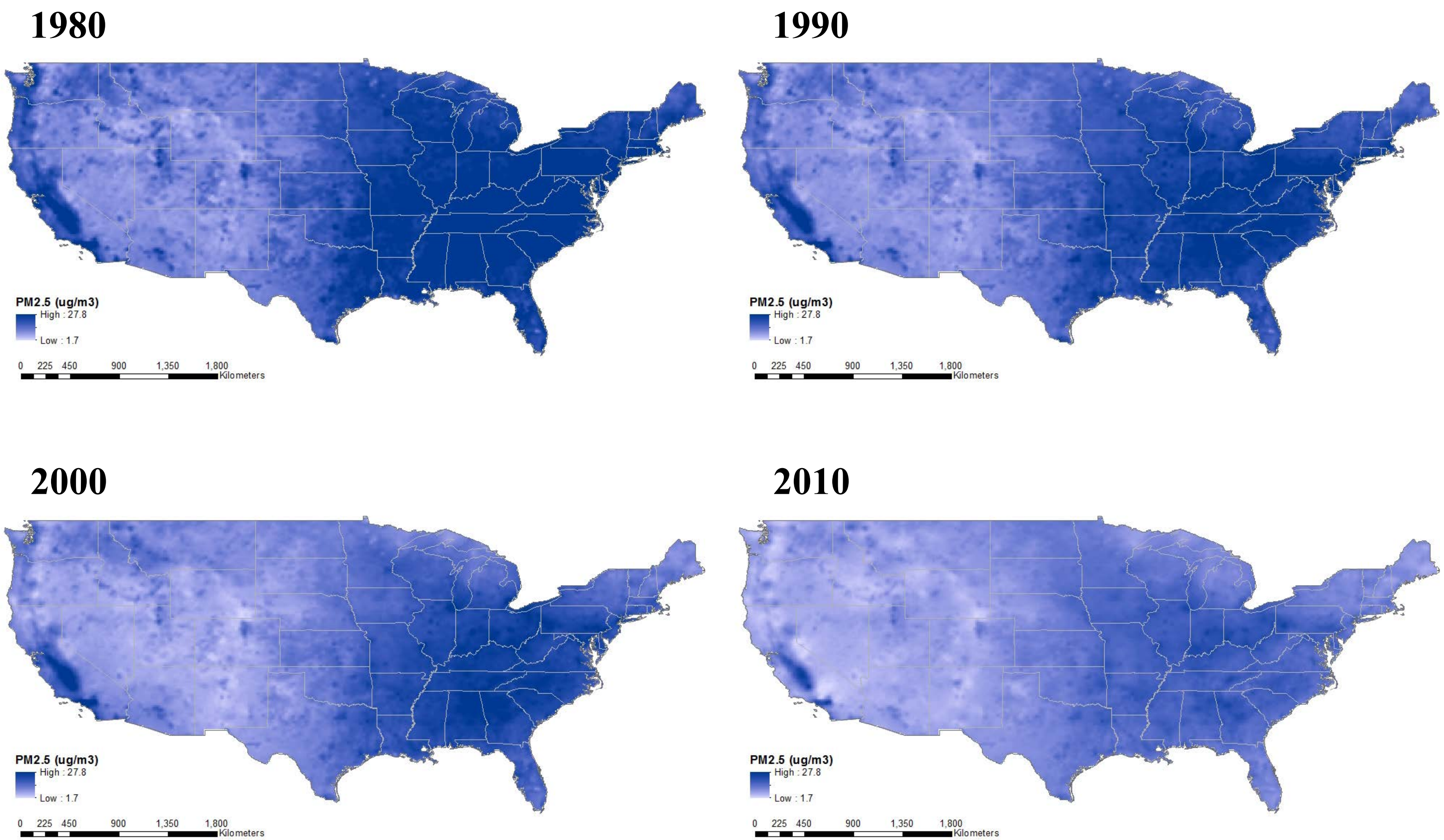
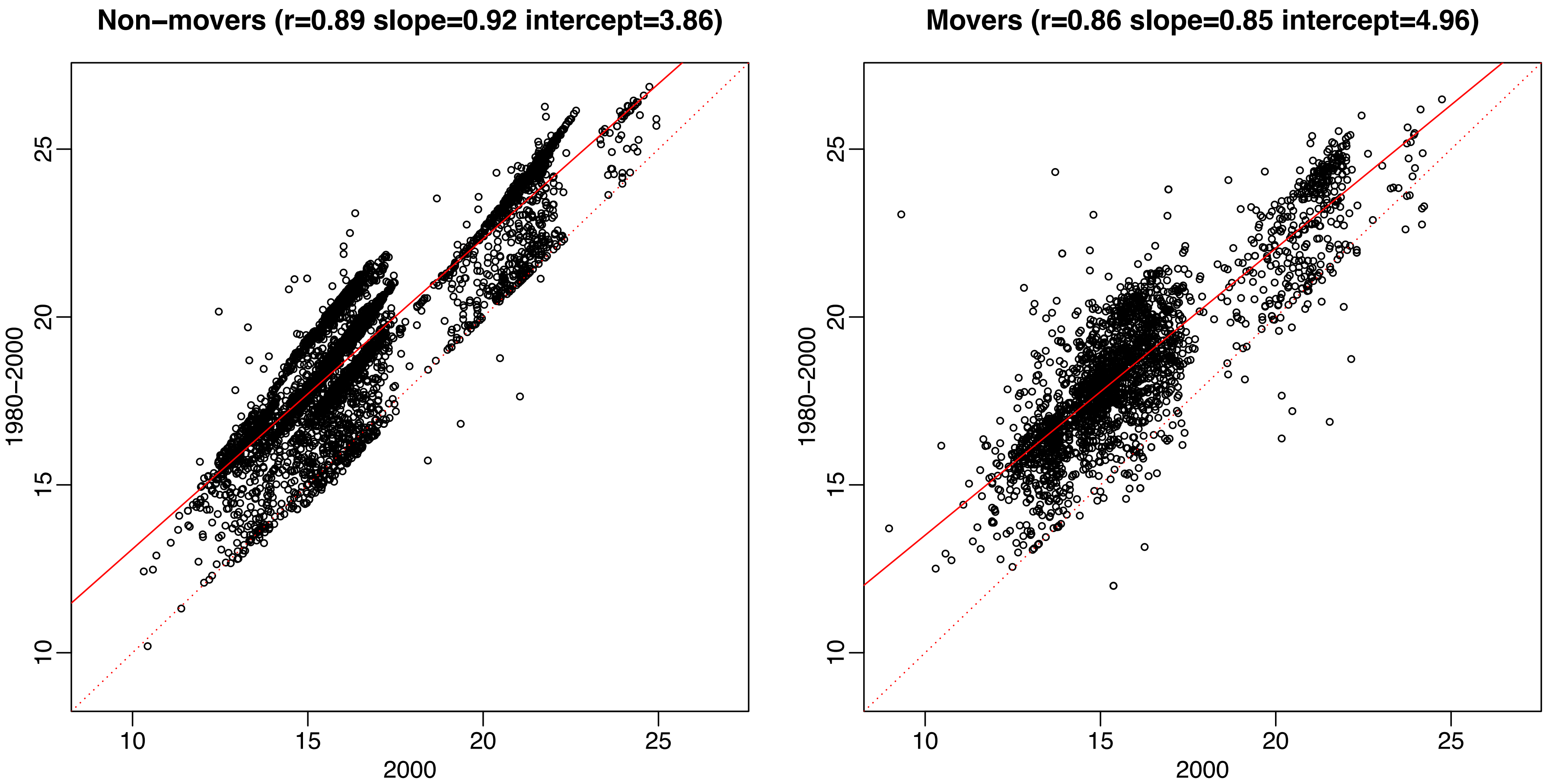

1980
PM2.5 (ug/m3)
High : 27.8
Low : 1.7
0 225 450 900 1,350 1,800
Kilometers
1990
PM2.5 (ug/m3)
High : 27.8
Low : 1.7
0 225 450 900 1,350 1,800
Kilometers
2000
PM2.5 (ug/m3)
High : 27.8
Low : 1.7
0 225 450 900 1,350 1,800
Kilometers
2010
PM2.5 (ug/m3)
High : 27.8
Low : 1.7
0 225 450 900 1,350 1,800
Kilometers

Figure 5.


Non–movers (r=0.89 slope=0.92 intercept=3.86)
Movers (r=0.86 slope=0.85 intercept=4.96)
1980–2000
2000
10
15
20
25