Relationship Between $PM_{2.5}$ Collected at Residential Outdoor Locations and a Central Site

Barbara Jane George, Donald A. Whitaker, Robert C. Gilliam, Jenise L. Swall, and Ronald W. Williams

National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC

Abstract

We develop regression models to describe the relationship between ambient $PM_{2.5}$ mass concentrations measured at a central-site monitor with those at residential outdoor monitors. Understanding the determinants and magnitude of variability and uncertainty in this relation ship is critical for understanding personal exposures in the evaluation of epidemiological data. Our repeated measures regression models address both temporal and spatial characteristics of data measured in the 2004-2007 Detroit Exposure and Aerosol Research Study, and they take into account missing data and other data features. The models incorporate turbulence kinetic energy and planetary boundary layer height, meteorological data that are not routinely considered in models that relate central-site concentrations to exposure to health effects. We found that turbulence kinetic energy was highly statistically significant in explaining the relationship of $PM_{2.5}$ measured at a particular stationary outdoor air monitoring site with $PM_{2.5}$ measured outside nearby residences for the temporal coverage of the data.

IMPLICATIONS

This work combines sophisticated statistical methods with meteorological data not previously considered in models relating central-site concentrations of $PM_{2.5}$ to exposure to health effects, and it addresses uncertainty of the representativeness of central site measurements for estimating personal exposure, key for the National Research Council's top priority for airborne particulate matter research. Turbulence kinetic energy is found to be an important explanatory variable for temporal and spatial characteristics of ambient

$PM_{2.5}$ concentrations from Detroit. This result will enable reducing or characterizing sources of uncertainty in models that link health effects such as respiratory problems to central-site measurements of $PM_{2.5}$.

INTRODUCTION

Quantitative source-to-exposure models hold great promise for air pollution source apportionment, risk assessment, and linking risk management actions to outcomes. Construction of these models requires an understanding of the determinants and magnitude of variability and uncertainty at all steps in the pathway leading to exposure. The National Research Council cites its top priority for airborne particulate matter research as determining the quantitative relationship between concentrations of particulate matter and gaseous copollutants measured at outdoor air monitoring sites and determining their contributions to actual personal exposures, especially for subpopulations and individuals.[1] In this paper, we develop regression models that describe the nature of relationships between ambient $PM_{2.5}$ mass concentrations measured at a central site monitor and at residential outdoor monitors using data from the Detroit Exposure and Aerosol Research Study (DEARS),[2] a critical step in understanding personal exposures for the evaluation of epidemiological data.

The primary goal of DEARS was to characterize and assess the impact of local sources on the relationship between concentrations measured at a central site monitor and concentrations measured at nearby residences in the Detroit area. DEARS builds on results from previous studies including the 1990 Particle Total Exposure Assessment Methodology (PTEAM) study in Riverside, California,[3-5] the 1995-1996 particulate matter and manganese studies in Toronto, Canada and Indianapolis, Indiana,[6-8] and the 1996-2001 particulate matter panel studies in, for example, Atlanta, Baltimore, Boston, Fresno, Research Triangle Park, and Seattle.[9-19] The study design and field implementation for DEARS are described in Williams et al.[2]

Previous research assessed the relationship between 12-hour or 24-hour integrated ambient $PM_{2.5}$ mass, i.e., particulate matter with particle aerodynamic diameters of 2.5 μm and smaller in ambient air, and residential outdoor $PM_{2.5}$ mass either directly or as part of a broader analysis effort. In the first large-scale probability-based study of personal exposure to particles, PTEAM participants were monitored for two consecutive, roughly 12-hour time periods with concurrent $PM_{2.5}$ samples collected indoors and outdoors at their homes. Clayton et al.[4] found that daytime and nighttime outdoor residential levels had similar medians but that the correlation between the fixed-site and the residential outdoor concentrations was higher for the nighttime data (r=0.96) than for the daytime data (r=0.83). For the particulate matter and manganese studies, Pellizzari et al.[8] reported correlations of outdoor to fixed-site log-scale $PM_{2.5}$ concentrations for Indianapolis (r=0.969) and for Toronto (r=0.818). For the 28-day, 38-household RTP panel study, Williams et al.[14] used regression for each home to relate the 24-hour ambient values to the corresponding outdoor values and estimated an average slope of 0.93 with an average coefficient of correlation of 0.94. In a companion paper, Williams et al.[15] fit a series of regression models for these data, again regressing the outdoor $PM_{2.5}$ data on the corresponding ambient data and estimated the slopes to range between 0.90 and 0.92.

There are studies in the literature that go beyond the simple regression and correlation analyses of the relationship between ambient fixed-site $PM_{2.5}$ mass and ambient residential outdoor $PM_{2.5}$ mass, including studies that relate personal exposure to central-site ambient concentrations. Noulett et al.[20] assessed the relationship between $PM_{2.5}$ and children's personal exposure in Prince George, British Columbia during the winter of 2001, evaluating the effect of spatial variation and meteorology in a complex airshed. Their study considered the meteorological influences of wind speed, wind direction, and inversion strength and found that levels of $PM_{2.5}$ during the 6-week study period were strongly influenced by the surface-based inversion or stability, that is, that the presence of a thermal inversion was significantly correlated with high ambient $PM_{2.5}$ levels and with personal exposures.

McBride et al.[21] developed a hierarchical Bayesian model for personal $PM_{2.5}$ exposure in 1998 in Baltimore as a function of outdoor $PM_{2.5}$, indoor $PM_{2.5}$, personal $PM_{2.5}$, wind speed, relative humidity, and time spent in microenvironments. Noting the absence of a spatial model relating ambient $PM_{2.5}$ to outdoor $PM_{2.5}$, their model included an assumption that the outdoor and ambient concentrations came from the same distribution. The model results showed evident bias in the posterior distribution for the white noise in the time series for mean outdoor $PM_{2.5}$ concentration, which the authors attributed to a likely underspecified submodel relating outdoor concentrations to meteorological conditions.

Holloman et al.[22] developed a hierarchical Bayesian model relating ambient $PM_{2.5}$ concentrations measured at Aerometric Information Retrieval System/Air Quality Subsystem (AIRS/AQS) monitors to model-simulated human exposure and to cardiovascular mortality in North Carolina in 1999-2000. They noted that the failure of approximating true exposure using ambient levels is well documented. They linked monitor readings to ambient levels over the study region using a spatial statistical model. The spatial model incorporated maximum temperature, average wind speed, and two sinusoidal terms to capture seasonal cycles. The authors noted uncertainty about the true ambient surface due to errors in the monitor data and the necessity of spatial interpolation. Calder et al.[23] constructed a hierarchical Bayesian model that related mortality in the same North Carolina data to exposure to $PM_{2.5}$ conditional on the ambient $PM_{2.5}$ concentrations. Their model incorporated an exposure simulator with a latent ambient $PM_{2.5}$ spatial field to account for the varying levels of uncertainty in ambient $PM_{2.5}$ concentrations over space and time. As in Holloman et al.,[22] this spatial component included maximum temperature, average wind speed, and sinusoidal terms for residual yearly cycles not explained by the other covariates. They found that replacing the spatial component of the model with an assumption of a constant $PM_{2.5}$ surface dramatically increased the uncertainty of the model results. Their work suggested that removing the spatial component resulted in poorer model fit and subsequent inflation of parameter uncertainty. They noted their model is simplistic in that it does not account for uncertainty in the variogram parameters, nonstationarity in space and time, or anisotropy in the spatial process.

Using a latent variable approach for data for 2000 for Ohio, Calder[24] derived information about $PM_{2.5}$ from PM_{10} monitor measurements and developed space-time interpolations that were improved over the interpolations using $PM_{2.5}$ monitor measurements alone. The Bayesian model used in this study incorporated temperature and wind speed as covariates, using wind speed and wind direction to specify the anisotropic properties of the latent processes. Empirical variograms indicated that the spatial correlation length for $PM_{2.5}$ depended on both wind speed and wind direction.

In our paper, we develop regression models that describe the nature of relationships between ambient $PM_{2.5}$ mass concentrations measured at a central site monitor and at residential outdoor monitors, taking into account repeated measurements over time. These models address the uncertainty of the representativeness of central site measurements for estimating personal exposure cited by National Research Council (NRC)[1], with the goal of explaining the relationship between residential outdoor $PM_{2.5}$ and $PM_{2.5}$ measured at a central site. Pope and Dockery[25] gave a thorough review of research since 1997 on the effects of exposure to particulate air pollution on human health, and they discussed gaps in scientific knowledge and reasons for skepticism concerning what is known about the health effects. Reducing or characterizing the sources of uncertainty for the representativeness of central site measurements is a necessary step in being able to link health effects to central site measurements. Our work matches repeated measures regression model assumptions and methods with data measured in DEARS, taking into account missing data and other data features. Regression models are fitted, and the results are interpreted in light of their uncertainties and relevance to explaining temporal and spatial variability of ambient $PM_{2.5}$.

MATERIALS AND METHODS

Design

DEARS was designed to sample non-smoking households in residential neighborhoods based on proximity to point or line sources of $PM_{2.5}$ and $PM_{10-2.5}$, their components, and selected air toxics (see www.epa.gov/dears/). Measurements were taken over three summers and three winters in the Detroit area at a central-site monitor, indoors and outdoors at residences, and on vests worn by participating individuals. Measurements were collected from each participating household for five consecutive days, by design, always Tuesday through Saturday, for a pair of sequential summer and winter seasons. A season was made up of seven weeks with a different set of households being measured each week. The goal was to take measurements from 40 households for five days per season for each of two seasons, thus totaling 120 households over the three summers and three winters represented by the study. $PM_{2.5}$ data were collected as 24-hour integrated average mass concentrations using the Personal Environmental Monitor as described in Williams et al.[2] The households were selected from residential neighborhoods in six Exposure Monitoring Areas (EMAs) representing various sources of $PM_{2.5}$ and other pollutants. A map of the Detroit area is provided as Figure 1 and shows locations of the EMAs and the Michigan Department of Environmental Quality's (MDEQ's) Allen Park site, the central-site monitor location used by DEARS. MDEQ's Allen Park site was selected as the fixed "central" site, in part, because it is a permanent, state facility in the $PM_{2.5}$ Speciation Trends Network[26] not overly influenced by nearby sources. No measurements were taken at EMA 2 during the DEARS study and none were taken at EMA 5 during summer 2004 and winter 2005 by design. EMA 7 was in Belleville, southwest of Detroit.

Exploratory Analysis

Data collection for DEARS began in the summer of 2004 and ended in the winter of 2007. Figure 2 gives the 24-hour integrated average $PM_{2.5}$ mass concentrations ($\mu g/m^3$) measured by the Personal Environmental Monitor outside residences for up to five days and at the central-site monitor for each day of the first two seasons of the study. The data analyzed in this paper are limited to the first summer and winter seasons of DEARS, July 13 through August 28, 2004 and February 1 through March 19, 2005, respectively. Both seasons from DEARS are modeled in consideration of seasonality; subsequent years are not modeled in this work. The figure shows that the temporal variability of $PM_{2.5}$ appears to follow large scale weather patterns. For example, around July 29, the Great Lakes region had a high pressure system with light winds, and the $PM_{2.5}$ mass concentrations were higher during that time. Also, in the August 4 to 10 period, a very strong Canadian high pressure system moved southeast into the eastern U.S., advecting clean air by way of strong northerly winds. $PM_{2.5}$ concentrations were low during that period. The variability, in this manner, possibly reflects regional meteorologically driven components as well as local emissions, chemistry, and transport, and it possibly reflects seasonality. The spatial variability is less pronounced with no apparent simple relationship between residential outdoor and central-site concentrations, although EMA 7 in Belleville was seen to have consistently low concentrations. The concentration data considered in this work are blank corrected; they are not censored in relation to a minimum detection limit. Details on the study design, the Personal Environmental Monitor, and the measurement procedures, including blank correction, are discussed in Williams et al.[2]

DEARS data are not balanced in the statistical sense—data collection start dates for participating households were staggered over each season, and there are <10 data points for many households. The counts of $PM_{2.5}$ measurements taken by week within each EMA and at the central site are shown in Table 1. Measurements were seldom taken from more than one household in any particular EMA on any particular day, by design, thus household is confounded with week. Given the large day-to-day variability, the effect of week may dominate the relationship of the outdoor measurements to the central-site values in a spatial sense. Measurements were taken outside 39 households in summer 2004 and 35 households in winter 2005. Several participating households in summer 2004 were replaced by new households in winter 2005 in the same EMA.

A total of 25 households had data collected on all 10 days as designed, while 22 households had data collected on 5 or fewer days. In addition to days not scheduled for sampling in accordance with the study design, there were also $PM_{2.5}$ measurements that were unavailable for particular days when the $PM_{2.5}$ sample was lost due to, for example, a torn filter or pump failure. Central-site concentration measurements were unavailable for 3 of the 35 days in summer 2004, and residential outdoor measurements were unavailable on 5 other days. In winter 2005, 16 residential outdoor measurements were unavailable. The number, mean, and other descriptive statistics for the central-site and residential outdoor measurements are given by season in Table 2. In comparing summer 2004 and winter 2005 results, note that roughly half of the outdoor measurements were taken at the same houses across the seasons. Descriptive statistics are also given for the logarithm of the residential outdoor concentration and for the ratio of residential outdoor to central site concentrations. All these measured $PM_{2.5}$ data were above the detection limit, but all were field and laboratory blank corrected to adjust for background mass on the filter due to handling. As a consequence, one residential outdoor value was negative with the corresponding log-transformed value treated as a missing value; this observation was omitted from our analysis.

Meteorological Data

Table 3 provides outdoor $PM_{2.5}$ and its log-transformed value when residential outdoor concentrations and central-site concentrations are considered together as simply "outdoor" concentrations. This approach handles the central site as an additional EMA. Table 3 also gives descriptive statistics for North American Regional Reanalysis (NARR) meteorological data for the dates corresponding to those for DEARS. The NARR data do not represent a daily average but rather are a snapshot at 2:00pm local daylight-savings time. Mid-afternoon is an ideal time to capture turbulent mixing near the surface each day.[27] Furthermore, the NARR data are considered representative of the meteorology 32km around the analysis grid point. These data were downloaded from the National Oceanic and Atmospheric Administration, National Climatic Data Center, and the ncdf package in R was used to open and read these meteorological data sets from netCDF format.[28-30] The NARR data described in Table 3 are turbulence kinetic energy (m^2/s^2), relative humidity at 2 meters (%), temperature at 2 meters, U and V wind components at 10 meters above the surface, and planetary boundary layer height (m). The U and V are the east-west and north-south Cartesian wind components, respectively, and so wind speed is given by $(U^2+V^2)^{1/2}$.[27]

The NARR meteorological data include planetary boundary layer height and turbulence kinetic energy. These two important, interrelated boundary layer measures of atmospheric stability are not commonly considered in models that relate ambient concentration to exposure and then relate exposure to health effects. The planetary boundary layer is the layer of the atmosphere directly affected by contact with the earth's surface. The planetary boundary layer as a whole tends to be compressed or thinner in areas of high pressure than in areas of low pressure as a result of the subsidence or the low-level horizontal divergence associated with synoptic high pressure. The subsidence and divergence tend to warm the air aloft, effectively capping the planetary boundary layer. The planetary boundary layer tends to be deeper near areas of low pressure because of rising air, stronger winds, and cooler temperatures aloft. Turbulence is the direct measure of the mixing rigor of the planetary boundary layer. It follows a diurnal cycle typically peaking early to mid-afternoon.

Regression Models

The goal of the regression models in this work is to characterize the relationship between ambient $PM_{2.5}$ mass concentrations measured at a central-site monitor and those measured at residential outdoor monitors. We want to explain the differences between these pairs of values on average, and a key component of this work is the use of appropriate statistical methodology. Data features that are taken into account include measurement error in the concentrations, non-normality of the concentration data, correlated errors resulting from the repeated measures structure, and lack of balance in the data due to both the study design and lost samples.

We developed regression models for $PM_{2.5}$ concentration data where the regression assumptions and specifications are matched closely to the data. These models adopted the linear mixed model formulation to address the within-subject repeated measures covariability. We chose the SAS MIXED procedure[31] since it is relatively robust against lack of balance. The SAS GLM procedure for general linear models, an alternate choice, accounts for within-subject covariability; however, it ignores data from any subject with any missing repeated measures.[32] We used the SAS MIXED procedure to analyze the non-missing repeated measures (variance component) data for participating household identifier (PID) nested within week, see Figure 3. These variance components resulting from multiple observations for participating households were modeled in the compound symmetry (exchangeable) covariance structure of the repeated statement shown in Figure 3. Compound symmetry is a reasonable choice for unbalanced data and gave a better fit to these data than an autoregressive covariance structure.

Residuals diagnostics from SAS MIXED favored the logarithmic scale for the dependent variable, $PM_{2.5}$, for these data from DEARS. That is, the studentized residuals for the overall mean appeared noticeably closer to normally distributed than their original-scale counterparts. Measurement error was captured in the log-transformed $PM_{2.5}$ concentration dependent variable by treating the central site as another EMA. The regression models used the log-transformed $PM_{2.5}$ concentration data measured at the central site as the reference cell, thus the expected value of the ratio of residential outdoor $PM_{2.5}$ to central-site $PM_{2.5}$ is estimated by exponentiating the EMA-specific single fixed effect regression coefficient. The form of the single fixed effect regression models is $\log(PM_{2.5}) = \beta_0 + (\beta_1 \times \text{EMA}) + (b \times \text{Time}) + e$, where b is the within-subject repeated-measures random effect of time. This setup is consistent with the measurement error assumptions of commonly-used regression software such as SAS MIXED and SAS GLM procedures.

RESULTS

Models with Spatial and Temporal Information

Our first model regressed $\log(PM_{2.5})$ onto a single explanatory categorical variable representing EMA. In summer 2004 the overall EMA fixed effect was not significant (p-value = 0.8525), and the winter 2005 finding was similar (p-value = 0.8712). We concluded from this single fixed effect regression that EMA alone does not significantly explain variability in the dependent variable. We next used multiple fixed effects regression with fixed effects for categorical EMA and a date indicator variable. The indicator variable for date was highly significant (p-value < 0.0001 for each season) and so partially explained the variability of the dependent variable. In this multiple fixed effects regression, EMA was also highly significant in each season (p-value = 0.0055 for summer 2004; p-value = 0.0001 for winter 2005), consistent with this form of the model being more appropriate than the single fixed effect regression. The multiple fixed effects regression model accounted for day-to-day temporal variability as well as spatial variability expressed in EMA. From Figure 2, we see that the day-to-day variability of $PM_{2.5}$ appears noticeably more pronounced than day-specific variability for these EMAs.

The Bayesian Information Criterion (BIC) is a penalized log-likelihood statistic favoring parsimony, and we used it to compare these two models. The multiple fixed effects regression model had smaller BIC values and is therefore preferred.[31, 33-35] The model with categorical EMA and the date indicator had BICs of 114.7 and 65.8 for summer 2004 and winter 2005, respectively. The single fixed effect regression model with categorical EMA alone had BICs of 411.1 and 416.0, respectively. While the model including a date indicator is preferred over the single fixed effect regression for these data from Detroit, the date indicator is a surrogate for substantive explanatory variables. We wanted to identify explanatory variables for the relationship between $PM_{2.5}$ outside residences and at central-site monitors, so our model development next considered meteorological variables even though the model with the date indicator fit the data well.

Models with Meteorological Data

We next specified regression models with a suite of meteorological variables: turbulence kinetic energy (m^2/s^2), relative humidity at 2 meters, temperature at 2 meters, U and V wind components at 10 meters, and height of the planetary boundary layer (m). These NARR data are available for a grid of latitudes and longitudes, which measures approximately 0.3 degrees (32km) resolution at the lowest latitude (see Spatial Coverage at http://www.cdc.noaa.gov/ data/gridded/data.narr.html). For the first of the next two models, we used meteorological data from the grid point closest to each residence and to the central site. Distances were calculated as great circle distances, and EMA was grouped into 3 classes based on distance to the closest meteorological grid point, as shown in Figure 4, where latitudes and longitudes are intentionally withheld. This approach effectively paired $PM_{2.5}$ concentrations measured in DEARS with meteorological data at the closest grid point in the NARR data. Since the residences and central site in DEARS were not closely associated with the meteorological grid latitudes and longitudes, the $PM_{2.5}$ data were also paired with meteorological data that average nearby NARR values. These averaged meteorological values were calculated as the inverse-distance weighted, composite meteorological values based on the four grid points surrounding the central site. These four grid points are corners of a rectangle, three of which were singly closest to the locations where $PM_{2.5}$ was measured. While it is arguably naïve to use these four grid points to interpolate data on a meteorology surface, the focus of this work is model development, not estimation. The meteorological variables listed above and EMA class were considered jointly as main effects in a closest-grid-point (unweighted) multiple fixed effects regression model that included the interaction of turbulence kinetic energy and EMA class. The interaction term gives a test of equal slopes for turbulence kinetic energy for the EMA classes, and a significant p-value would add support to the idea that turbulence kinetic energy is valuable in spatially relating outdoor residential $PM_{2.5}$ concentrations to those at a central site. Main effect model terms with coefficient p-values< 0.1 were retained initially as explanatory in unweighted and distance weighted models that also included EMA class and the interaction of turbulence kinetic energy and EMA class. Backward elimination was used to remove additional main effect terms when the corresponding BIC was improved (lower). The initial models included both turbulence kinetic energy and height of the planetary boundary layer, but, for these $PM_{2.5}$ data, turbulence kinetic energy explained more variability.

For the summer 2004 unweighted model, turbulence kinetic energy, temperature, and the V (north-south Cartesian) wind component were highly statistically significant for explaining the variability of $\log(PM_{2.5})$. The interaction of turbulence kinetic energy and the EMA class was not significant (p-value = 0.1313), so turbulence kinetic energy was not found to be differentially influential by EMA class. There was, however, a suggestion that the slope of turbulence kinetic energy for EMA class 1 differed from that for the central site (p-value = 0.0870) and that the slope for EMA class 2 did not differ from that for the central site (p-value = 0.9879). The distance-weighted version of the summer 2004 model gave consistent results, but the interaction term was further from significant (p-value = 0.2085). For the winter 2005 unweighted model, turbulence kinetic energy, and both the U and V wind components were highly statistically significant for explaining the variability of log($PM_{2.5}$). The interaction of turbulence kinetic energy and EMA class was not significant (p-value = 0.1116). As seen for summer 2004, there was a suggestion that the slope of the turbulence kinetic energy was not exactly the same for each EMA class, but here the slope of turbulence kinetic energy for EMA class 2 seemed to differ from that of the central site (p-value = 0.0808) while the slope for EMA class 1 did not (p-value = 0.1743). The distance-weighted version of the winter 2005 model was consistent with its unweighted counterpart. The BICs for these summer 2004 models were 318.6 and 318.9 for the unweighted and weighted versions and for the winter 2005 models were 192.5 and 188.5, respectively. These unweighted and weighted multiple fixed effects regression models included the interaction term to test differences in turbulence kinetic energy among the EMA classes. They took into account the repeated measurements at each residence, nesting the household identifier (PID) within EMA class. The next models expand on this idea by including a time effect in addition to the EMA class spatial effect.

Models with Alternate Repeated Measures Set-up

The DEARS design took repeated measurements at each particular household during a 5-day period corresponding to a particular week. In this sense, the household identifier (PID) is nested within week in addition to being nested within EMA class. Our next models accounted for the repeated measurements at each residence by nesting the household identifier (PID) within week. These models included the categorical variable for week as a main effect and also in interaction terms with turbulence kinetic energy and EMA class, where the three-way interaction term was used to examine the slope for turbulence kinetic energy at spatial-temporal combinations. The first model for each season jointly considered all the meteorological variables, EMA class, and week as main effects, and it included the three-way interaction and corresponding lower order interactions for turbulence kinetic energy, EMA class, and week. As before, main effect model terms with coefficient p-values< 0.1 were retained initially as explanatory in unweighted and distance-weighted models that also included the interaction terms.

The summer 2004 unweighted model with turbulence kinetic energy, temperature, and V (north-south Cartesian) wind component, EMA class, and week as main effects plus all interactions for turbulence kinetic energy, EMA class, and week gave the best fit with BIC 217.5. A reduced model that included these same main effects and the single interaction of turbulence kinetic energy and week gave a slightly worse fit with BIC 218.5, suggesting interactions with EMA class help explain the variability in $\log(PM_{2.5})$ despite their lack of significance reflected by p-values. This result is interesting in part because the BIC imposes a penalty for additional model terms. The distance-weighted summer 2004 version of the best fitting unweighted model (with all three-way and two-way interactions for turbulence kinetic energy, EMA class, and week) gave an improved BIC of 214.4. The winter 2005 unweighted model with turbulence kinetic energy, U (east-west Cartesian) wind component, and V (north-south Cartesian) wind component as main effects and with main effects for EMA class and week and all interactions for turbulence kinetic energy, EMA class, and week gave a good fit with BIC of 15.4. A reduced model that included these same main effects and the single interaction of turbulence kinetic energy and week gave the best fit with BIC of 3.1. The distance-weighted version of this reduced model also gave BIC of 3.1. The p-values, variance components, and BICs for models with household identifier (PID) nested within EMA class are given in Table 4; those for models with PID nested within week are given in Table 5. The V (north-south Cartesian) wind component was highly significant in all of the weighted and unweighted models that included meteorology terms. Turbulence kinetic energy was highly significant in all those models that did not include week, and the interaction of turbulence kinetic energy with week was highly significant in all those models that included week. Temperature was significant in all of the summer 2004 models. The U (east-west Cartesian) wind component was significant in all of the winter 2005 models. Planetary boundary layer height was significant in the weighted model without week for winter 2005, but perhaps collinearity in the explanatory meteorological variables or spatial features of the residences relative to the central site suppresses the importance of the boundary layer height in the other models. It is noteworthy that turbulence kinetic energy is highly significant as a main effect or as an interaction term in both seasons for these data but has not typically been considered in spatial and other models for $PM_{2.5}$. As indicated by the BICs for the summer 2004 models, further research is needed to identify explanatory variables that are jointly as effective as date indicator.

DISCUSSION AND CONCLUSIONS

This work establishes that turbulence kinetic energy is highly statistically significant in explaining the relationship in DEARS data of $PM_{2.5}$ measured at a particular stationary outdoor air monitoring site with $PM_{2.5}$ measured outside nearby residences for the temporal coverage of the data. Meteorological data such as turbulence kinetic energy and planetary boundary layer height are not routinely considered in models that relate central-site concentrations to health effects, and our findings support their importance. The statistical methodology of this analysis addresses the non-normality of the concentration data and the correlated error structure resulting from repeated measures. The set-up avoids introducing measurement error for $PM_{2.5}$ into regression explanatory variables and uses software well-suited to unbalanced data. Our models are straightforward to interpret in that they did not combine data across seasons, and our findings support that the temporal variability of $PM_{2.5}$ indicated in Figure 2 is important. These Detroit-area $PM_{2.5}$ data are more sensitive to time than space for the meteorological data used in our work. The variability likely reflects regional meteorology as well as local emissions, chemistry, and transport, and possibly reflects seasonality.

We conclude that models for these DEARS data that use the household identifier (PID) nested within week for the repeated measures covariance structure (Table 5) are preferred over the models with PID nested within EMA (Table 4). From a model development perspective, the reduced model with EMA class, turbulence kinetic energy, temperature, V (north-south Cartesian) wind component, week, and the interaction of turbulence kinetic energy with week is favored for summer 2004season. Similarly, the reduced model with

EMA class, turbulence kinetic energy, U (east-west Cartesian) wind component, V (north-south Cartesian) wind component, week, and the interaction of turbulence kinetic energy with week is favored for winter 2005. Their BIC values are the lowest, or nearly so, for models without an indicator variable for date, and these models are parsimonious.

There are many uncertainties remaining in this analysis, and these affect the measurement data as well as the meteorology data. Diurnal patterns for the correlation between fixed-site and residential outdoor concentration seen by Clayton et al.[4] in PTEAM suggest the 24-hour data collection in DEARS may contribute to the uncertainty of the findings from this analysis. As indicated by Table 1, household is confounded with week and EMA within a season for most of the $PM_{2.5}$ data since, by design, data were collected at a participating DEARS household for at most 5 days in one week of a season. Also, local $PM_{2.5}$ sources were not explicitly considered in this work. The proximity of participating residences to known or suspected point or line sources was considered in the design of DEARS; however, we chose to include source contributions, rather than their distances, parsimoniously through the U and V wind components. There are uncertainties in the NARR meteorological data, but it is not clear how to characterize these. The estimation and other uncertainties of the NARR data may well be dwarfed by the other uncertainties of this analysis. An added, and perhaps important, uncertainty of this analysis comes from the facts that the meteorological grid is large relative to the range of DEARS residences and that the grid cell centers are not located close to the DEARS residences.

While boundary layer meteorological data are not routinely considered in regression models for particulate matter, Thornburg et al.[36] included Monin-Obukhov length in addition to wind direction, wind speed, distance between EMAs, and weekday/weekend in spatial-temporal modeling for $PM_{10-2.5}$. Their research was part of the evaluation of a newly developed coarse particle exposure monitor used in DEARS; the $PM_{10-2.5}$ data used in their modeling were collected summer 2006 and winter 2007. Monin-Obukhov length, like turbulence kinetic energy, is a measure of atmospheric stability. Monin-Obukhov length was the only explanatory variable in their regression found to be statistically significant for both summer and winter $PM_{10-2.5}$ concentrations. This Thornburg et al.[34] finding is consistent with our findings.

Further modeling using $PM_{2.5}$ measurements from other localities is needed to evaluate the importance of planetary boundary layer height, turbulence kinetic energy, and perhaps other measures of atmospheric stability such as Monin-Obukhov length, in explaining uncertainty of the representativeness of central site measurements for estimating personal exposure. In this work turbulence kinetic energy was found to be significant, but its interrelationship with planetary boundary layer height may promote the latter's significance elsewhere. When available, turbulence kinetic energy is one of the most important measures used to characterize the turbulence in the planetary boundary layer. It is a more straightforward measure of mixing or turbulence than Monin-Obukhov length, a scaling parameter related to turbulence kinetic energy. Turbulence kinetic energy values are nonnegative; they are directly proportional to the kinetic energy of the mixing eddies. Monin-Obukhov length values are not restricted to being positive; they are proportional to the height above the surface at which buoyant factors first dominate over mechanical production of turbulence[27]. Given these characteristics, in our judgment turbulence kinetic energy is preferred over Monin-Obukhov length as a correlate to pollution concentrations and health outcomes. Further model development for the $PM_{2.5}$ data from DEARS could investigate inclusion of time-lagged explanatory variables or sinusoidal terms for residual synoptic cycles not explained by the other covariates. The latter are similar to the terms included for residual yearly cycles in the works by Calder et al.[23] and Holloman et al.[22]

DISCLAIMER

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here under contract EP-D-05-065 to Alion Science &Technology. It has been subjected to Agency review and approved for publication.

ACKNOWLEDGEMENTS

We thank Linda Sheldon of EPA's National Exposure Research Laboratory for her question that sparked this work and Fred Dimmick of EPA's Process Modeling Research Branch, Human Exposure and Atmospheric Sciences Division for his comments about the mixing layer of the atmosphere. His comments led us to meteorologists Brian Eder and Robert Gilliam of EPA's Atmospheric Modeling and Analysis Division, who recommended turbulence kinetic energy and planetary boundary layer height as important boundary layer measures. We also thank James Chromy of RTI International, Fred Dimmick, Brian Eder, and the anonymous peer reviewers for helpful discussions and constructive comments during the course of preparing the final manuscript.

ABOUT THE AUTHORS

Barbara Jane George is a mathematical statistician in the Exposure Modeling Research Branch, Human Exposure and Atmospheric Sciences Division of NERL EPA. Donald A. Whitaker is a research physical scientist in the Exposure Measurements and Analysis Branch, Human Exposure and Atmospheric Sciences Division of NERL EPA. Robert C. Gilliam is a physical scientist in the Atmospheric Model Development Branch, Atmospheric Modeling and Analysis Division of NERL EPA. Jenise L. Swall is a statistician in the Applied Modeling Branch, Atmospheric Modeling and Analysis Division of NERL EPA. Ronald W. Williams is a research chemist for the Exposure Measurements and Analysis Branch, Human Exposure and Atmospheric Sciences Division of NERL EPA. Please address correspondence to: Dr. B. J. George, U.S. Environmental Protection Agency, National Exposure Research Laboratory, Human Exposure and Atmospheric Sciences Division, 109 T.W. Alexander Drive, Mail Drop E205-02, Research Triangle Park, NC 27711, USA; phone: 919-541-4551; fax: 919-541-4787; e-mail: george.bj@epa.gov.

REFERENCES

1. National Research Council. *Research Priorities for Airborne Particulate Matter: IV. Continuing Research Progress*; The National Academies Press: Washington, DC, 2004.

2. Williams, R.; Rea, A.; Vette, A.; Croghan, C.; Whitaker, D.; Stevens, C.; McDow, S.; Fortmann, R.; Sheldon, L.; Wilson, H.; Thornburg, J.; Phillips, M.; Lawless, P.; Rodes, C.; Daughtrey, H. The Design and Field Implementation of the Detroit Exposure and Aerosol Research Study (DEARS); *J. Expos. Sci. Environ. Epidemiol.* **2009**, 19, 643–659.

3. Thomas, K. W.; Pellizzari, E. D.; Clayton, C. A.; Whitaker, D. A.; Shores, R. C.; Spengler, J.; Ozkaynak, H.; Froehlich, S. E.; Wallace, L. A. Particle Total Exposure Assessment Methodology (PTEAM) 1990 Study: Method Performance and Data Quality for Personal, Indoor, and Outdoor Monitoring; *J. Expos. Anal. Environ. Epidemiol.* **1993**, 3, 203–226.

4. Clayton, C. A.; Perritt, R. L.; Pellizzari, E. D.; Thomas, K. W.; Whitmore, R. W.; Wallace, L. A.; Ozkaynak, H.; Spengler, J. D. Particle Total Exposure Assessment Methodology (PTEAM) Study: Distributions of Aerosol and Elemental Concentrations in Personal, Indoor, and Outdoor Air Samples in a Southern California Community; *J. Expos. Anal. Environ. Epidemiol.* **1993**, 3, 227–250.

5. Ozkaynak, H.; Xue, J.; Spengler, J.;Wallace, L.; Pellizzari, E.; Jenkins, P. Personal Exposure to Airborne Particles and Metals: Results from the Particle TEAM Study in Riverside, California; *J. Expos. Anal. Environ. Epidemiol.* **1996**, 6, 57–78.

6. Clayton, C.; Pellizzari, E.; Rodes, C.; Mason, R.; Piper, L. Estimating distributions of longterm particulate matter and manganese exposures for residents of Toronto, Canada; *Atmos. Environ.* **1999**, 33, 2515–2526.

7. Crump, K. S. Manganese exposures in Toronto during use of the gasoline additive, methylcyclopentadienyl manganese tricarbonyl; *J. Expos. Anal. Environ. Epidemiol.* **2000**, 10, 227–239.

8. Pellizzari, E.; Clayton, C.; Rodes, C.; Mason, R.; Piper, L.; Fort, B.; Pfeifer, G.; Lynam, D. Particulate matter and manganese exposures in Indianapolis, Indiana; *J. Expos. Anal. Environ. Epidemiol.* **2001**, 11, 423–440.

9. Wheeler, A.; Zanobetti, A.; Gold, D. R.; Schwartz, J.; Stone, P.; Suh, H. H. The Relationship between Ambient Air Pollution and Heart Rate Variability Differs for Individuals with Heart and Pulmonary Disease; *Environ. Health Perspect.* **2006**, 114, 560–566.

10. Landis, M. S.; Norris, G. A.; Williams, R. W.; Weinstein, J. P. Personal exposures to $PM_{2.5}$ mass and trace elements in Baltimore, MD, USA; *Atmos. Environ.* **2001**, 35, 6511–6524.

11. Rea, A. W.; Zufall, M. J.; Williams, R. W.; Sheldon, L.; Howard-Reed, C. The Influence of Human Activity Patterns on Personal PM Exposure: A Comparative Analysis of Filter-based and Continuous Particle Measurements; *J. Air & Waste Manage. Assoc.* **2001**, 51, 1271–1279.

12. Rojas-Bracho, L.; Suh, H. H.; Koutrakis, P. Relationships among personal, indoor, and outdoor fine and coarse particle concentrations for individuals with COPD; *J. Expos. Anal. Environ. Epidemiol.* **2000**, 10, 294–306.

13. Evans, G.; Highsmith, R.; Sheldon, L.; Suggs, J.; Williams, R.; Zweidinger, R.; Creason, J.; Walsh, D.; Rodes, C.; Lawless, P. The 1999 Fresno particulate matter Exposure studies: comparison of community, outdoor, and residential PM mass measurements; *J. Air & Waste Manage. Assoc.* **2000**, 50, 1887–1896.

14. Williams, R.; Suggs, J.; Rea, A.; Leovic, K.; Vette, A.; Croghan, C.; Sheldon, L.; Rodes, C.; Thornburg, J.; Ejire, A.; Herbst, M.; Jr, W. S. The Research Triangle Park particulate matter panel study: PM mass concentration relationships; *Atmos. Environ.* **2003**, 37, 5349–5363.

15. Williams, R.; Suggs, J.; Rea, A.; Sheldon, L.; Rodes, C.; Thornburg, J. The Research Triangle Park particulate matter panel study: modeling ambient source contribution to personal and residential PM mass concentrations; *Atmos. Environ.* **2003**, 37, 5365–5378.

16. Mar, T. F.; Koenig, J. Q.; Jansen, K.; Sullivan, J.; Kaufman, J.; Trenga, C. A.; Siahpush, S. H.; Liu, L.-J. S.; Neas, L. Fine Particle Air Pollution and Cardiorespiratory Effects in the Elderly; *Epidemiol.* **2005**, 16, 681–687.

17. Sheppard, L.; Slaughter, J. C.; Schildcrout, J.; Liu, L.-J. S.; Lumley, T. Exposure and measurement contributions to estimates of acute air pollution effects; *J. Expos. Anal. Environ. Epidemiol.* **2005**, 15, 366–376.

18. Liu, L.-J. S.; Box, M.; Kalman, D.; Kaufman, J.; Koenig, J.; Larson, T.; Lumley, T.; Sheppard, L.; Wallace, L. Exposure Assessment of Particulate Matter for Susceptible Populations in Seattle; *Environ. Health Perspect.* **2003**, 111, 909–918.

19. Goswami, E.; Larson, T.; Lumley, T.; Liu, L.-J. S. Spatial Characteristics of Fine Particulate Matter: Identifying Representative Monitoring Locations in Seattle, Washington; *J. Air & Waste Manage. Assoc.* **2002**, 52, 324–333.

20. Noulett, M.; Jackson, P. L.; Brauer, M. Winter measurements of children's personal exposure and ambient fine particle mass, sulphate and light absorbing components in a northern community; *Atmos. Environ.* **2006**, 40, 1971–1990.

21. McBride, S. J.; Williams, R. W.; Creason, J. Bayesian hierarchical modeling of personal exposure to particulate matter; *Atmos. Environ.* **2007**, 41, 6143–6155.

22. Holloman, C. H.; Bortnick, S. M.; Morara, M.; Stauss, W. J.; Calder, C. A. A Bayesian Hierarchical Approach for Relating $PM_{2.5}$ Exposure to Cardiovascular Mortality in North Carolina; *Environ. Health Perspec.* **2004**, 112, 1282–1288.

23. Calder, C. A.; Holloman, C. H.; Bortnick, S. M.; Strauss, W.; Morara, M. Relating ambient particulate matter concentration levels to mortality using an exposure simulator; *J. Am. Stat. Assoc.* **2008**, 103, 137–148.

24. Calder, C. A. A dynamic process convolution approach to modeling ambient particulate matter concentrations; *Environmetrics* **2008**, 19, 39–48.

25. Pope, C. A.; Dockery, D. W. 2006 Critical Review: Health Effects of Fine Particulate Air Pollution: Lines that Connect; *J. Air & Waste Manage. Assoc.* **2006**, 56, 709-742.

26. Flanagan, J. B.; Jayanty, R.; Rickman , Jr., E. E.; Peterson, M. R. $PM_{2.5}$ Speciation Trends Network: Evaluation of Whole-System Uncertainties Using Data from Sites with Collocated Samplers; *J. Air & Waste Manage. Assoc.J. Air & Waste Manage. Assoc.* **2006**, 56, 492–499.

27. Stull, R. B. *An Introduction to Boundary Layer Meteorology*; Kluwer Academic Publishers: Dordrecht, 1997.

28. Pierce, D.W. *ncdf: Interface to Unidata netCDF Data Files*; R Foundation for Statistical Computing: Vienna, Austria, 2006; available at http://cirrus.ucsd.edu/~pierce/ncdf (accessed May 2010).

29. R Development Core Team, *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2007.

30. Rew, R.; Davis, G.; Emmerson, S.; Davies, H.; Hartnett, E.; Heimbigner, D. *The NetCDF Users Guide*; University Corporation for Atmospheric Research, Boulder Colorado, 2010; available at http://www.unidata.ucar.edu/software/netcdf/docs/ (accessed May 2010).

31. SAS Institute Inc., SAS/STAT® 9.1 User's Guide; SAS Institute Inc.: Cary, NC, 2004.

32. Wolfinger, R.; Chang, M. Comparing the SAS GLM and MIXED procedures for Repeated Measures; Technical Report, 1995.

33. Schwarz, G. Estimating the Dimension of a Model; *An. Stat.* **1978**, 6, 461–464.

34. Davidian, M.; Giltinan, D. M. *Nonlinear Models for Repeated Measurement Data*; Chapman & Hall/CRC: Boca Raton, 1995.

35. Kadane, J. B.; Lazar, N. A. Methods and Criteria for Model Selection; *J. Am. Stat. Assoc.* **2004**, 99, 279–290.

36. Thornburg, J.; Rodes, C. E.; Lawless, P. A.; Williams, R. Spatial and Temporal Variability of Outdoor Coarse Particulate Matter Mass Concentrations Measured with a New Coarse Particle Sampler During the Detroit Exposure and Aerosol Research Study; *Atmos. Environ.* **2009**, 43, 4251–4258.

LIST OF FIGURES

1. DEARS Exposure Monitoring Areas 1 to 7. Allen Park was the fixed "central" site for DEARS.

2. 24-hour integrated average $PM_{2.5}$ mass concentrations in $\mu g/m^3$ by date and EMA locations: (a) summer 2004 and (b) winter 2005. Note that one value in summer 2004 is negative as a result of blank correction.

3. SAS regression model code example with repeated measures for household identifier (PID) nested within week.

4. DEARS residences (both summer 2004 and winter 2005), central site monitor, and grid points for meteorological data. Note that latitudes and longitudes are intentionally suppressed.

Table 1. Counts of $PM_{2.5}$ measurements by EMA and week. Underlined counts indicate measurements are from more than one household.

	Week						
EMA	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Summer 2004							
1	5	5	5	5	5	4	10
3	5	5	5	5	10	5	5
4	5	5	8	5	5	0	4
6	5	4	4	9	5	9	5
7	0	5	5	5	4	5	5
Central site	4	5	4	5	5	5	4
Winter 2005							
1	0	5	0	8	9	2	0
3	5	5	5	10	14	5	5
4	0	0	5	0	0	4	10
6	4	0	10	0	0	10	10
7	3	5	4	5	5	5	5
Central site	5	5	5	5	5	5	5

Table 2. Descriptive statistics for outdoor $PM_{2.5}$ in $\mu g/m^3$ for Summer 2004 and Winter 2005.

	n	mean	std	min	med	max
Summer 2004						
Central Site $PM_{2.5}$	32	17.8	11.9	3.3	13.2	42.7
Residential Outdoor $PM_{2.5}$	181	16.0	10.3	-1.1	11.6	55.6
log(Residential Outdoor $PM_{2.5}$)	180	2.6	0.6	1.1	2.5	4.0
ResOutdoor/CentralSite $PM_{2.5}$	166	1.0	0.6	0.4	0.9	6.3
Winter 2005						
Central Site $PM_{2.5}$	35	18.8	15.9	3.7	13.5	66.4
Residential Outdoor $PM_{2.5}$	158	17.1	14.6	3.1	13.5	85.6
log(Residential Outdoor $PM_{2.5}$)	158	2.6	0.7	1.1	2.6	4.4
ResOutdoor/CentralSite $PM_{2.5}$	158	1.0	0.2	0.4	1.0	1.7

Table 3. Descriptive statistics for outdoor $PM_{2.5}$ and meteorology data for Summer 2004 and Winter 2005 with combined residential outdoor and central-site data.

	n	mean	std	min	med	max
Summer 2004						
Outdoor $PM_{2.5}$ ($\mu g/m^3$)	212	16.4	10.5	3.0	12.6	55.6
log(Outdoor $PM_{2.5}$) ($\mu g/m^3$)	212	2.6	0.6	1.1	2.5	4.0
Turbulence Kin. Energy (m^2/s^2)	212	1.4	0.7	0.2	1.2	3.4
Relative Humidity 2m (%)	212	61.6	10.4	39.1	61.0	90.3
Temperature 2m (K)	212	297.7	4.1	289.4	297.8	305.2
U Wind Component 10m	212	1.8	2.7	-4.3	2.0	8.3
V Wind Component 10m	212	0.7	3.9	-5.8	0.3	7.4
Planetary Bound. Layer Ht (m)	212	1508.3	355.0	498.8	1552.7	2155.6
Winter 2005						
Outdoor $PM_{2.5}$ ($\mu g/m^3$)	193	17.4	14.8	3.1	13.5	85.6
log(Outdoor $PM_{2.5}$) ($\mu g/m^3 3$)	193	2.6	0.7	1.1	2.6	4.4
Turbulence Kin. Energy (m^2/s^2)	193	1.1	0.7	0.2	1.0	2.8
Relative Humidity 2m (%)	193	74.5	14.1	31.9	77.0	93.6
Temperature 2m (K)	193	271.7	2.8	266.7	272.0	278.3
U Wind Component 10m	193	2.1	3.2	-5.7	2.4	7.6
V Wind Component 10m	193	-0.9	2.9	-6.5	-1.3	4.9
Planetary Bound. Layer Ht (m)	193	1029.3	506.8	181.2	1107.7	2017.3

Table 4. Regression model results for log($PM_{2.5}$) without week fixed effects.

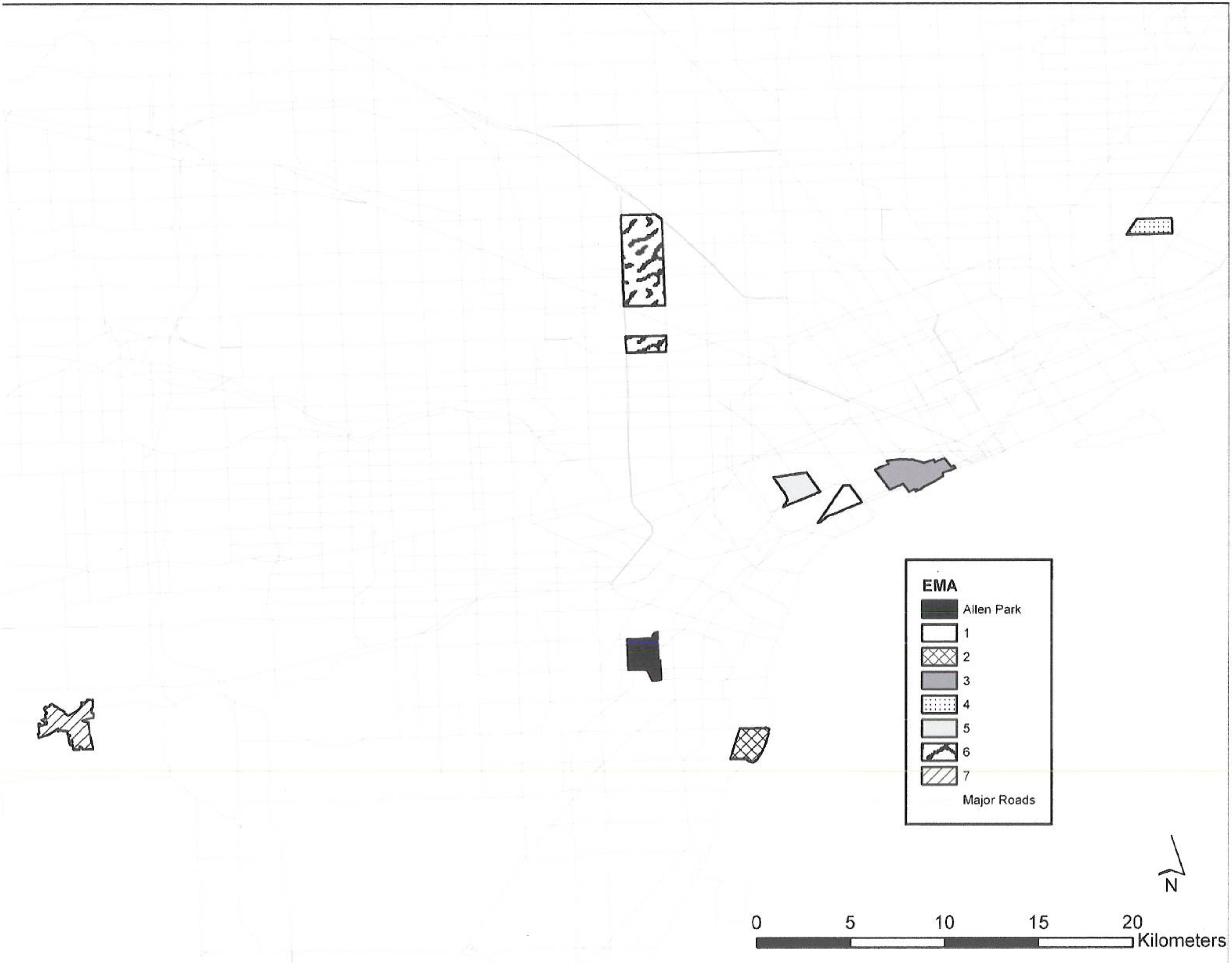
Models for log($PM_{2.5}$)	Single Fixed Effects Regression	Multiple Fixed Effects Regression	Unweighted Meteorology	Weighted Meteorology
Summer 2004				
p-values				
EMA	0.8525	0.0055		
EMA class			0.7490	0.6901
Date indicator		< 0.0001		
TKE[a]			0.0004	0.0005
Temperature			< 0.0001	< 0.0001
V wind component			< 0.0001	< 0.0001
TKE × EMA class			0.1313	0.2085
Variance components				
PID(EMA)	0.1304	-0.0002	0.0849	0.0832
Residual	0.3145	0.0694	0.1833	0.1843
Model comparison				
BIC	411.1	114.7	318.6	318.9
Winter 2005				
p-values				
EMA	0.8712	0.0001		
EMA class			0.2487	0.2839
Date indicator		< 0.0001		
TKE[a]			< 0.0001	< 0.0001
Relative humidity			0.0179	0.0281
U wind component			< 0.0001	0.0002
V wind component			< 0.0001	< 0.0001
PBL[b] height			0.1035	0.0496
TKE × EMA class			0.1394	0.2198
Variance components				
PID(EMA)	0.2951	0.0056	0.1020	0.1017
Residual	0.3660	0.0197	0.0944	0.0922
Model comparison				
BIC	416.0	-65.8	213.6	209.9

Note: All p-values are shown; p-values ≤ 0.05 are considered statistically significant; omitted model terms are indicated by the absence of corresponding p-values; [a]Turbulence kinetic energy, [b]Planetary boundary layer.

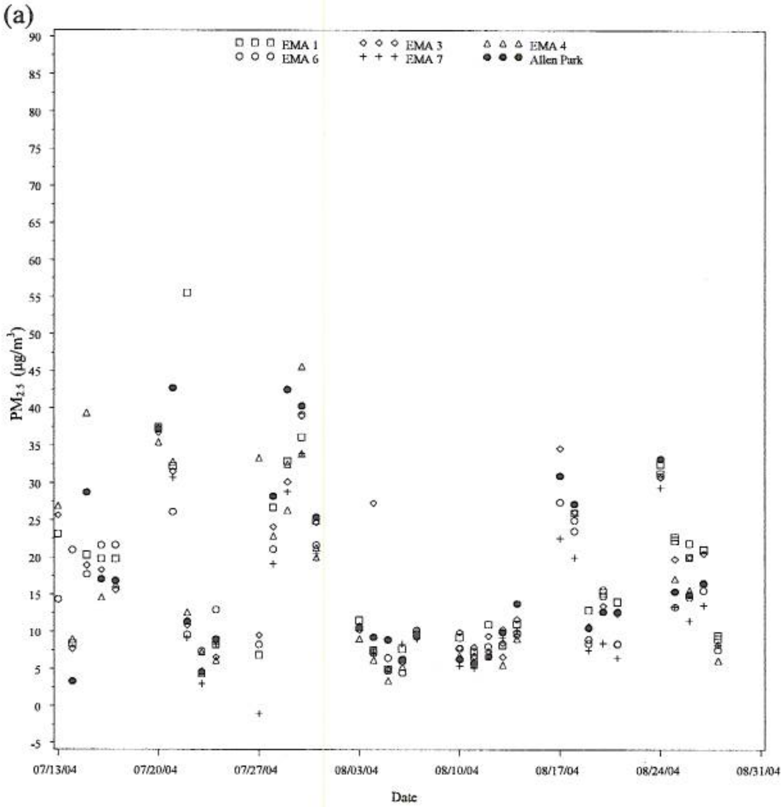
Table 5. Regression model results for log($PM_{2.5}$) including week.

Models for log($PM_{2.5}$)	Week & Interactions Regression	Reduced Regression With Week	Weighted Regression With Week
Summer 2004			
p-values			
EMA class	0.2802	0.0020	0.3268
TKE[a]	0.0236	0.0154	0.0101
Temperature	0.0011	0.0009	0.0014
V wind component	< 0.0001	< 0.0001	< 0.0001
Week	0.0735	0.0002	0.0506
TKE × EMA class	0.6861		0.7390
TKE × week	0.0009	< 0.0001	0.0008
EMA class × week	0.6597		0.5757
TKE × EMA class × week	0.6712		0.5948
Variance components			
PID(week)	-0.0126	-0.0140	-0.0127
Residual	0.1452	0.1433	0.1435
Model comparison			
BIC	217.5	218.5	214.4
Winter 2005			
p-values			
EMA class	0.9888	0.0006	0.0007
TKE[a]	0.6948	0.6329	0.3659
U wind component	< 0.0001	< 0.0001	< 0.0001
V wind component	< 0.0001	< 0.0001	< 0.0001
Week	< 0.0001	< 0.0001	< 0.0001
TKE × EMA class	0.3868		
TKE × week	< 0.0001	< 0.0001	< 0.0001
EMA class × week	0.7938		
TKE × EMA class × week	0.7168		
Variance components			
PID(week)	0.0073	0.0040	0.0038
Residual	0.0398	0.0391	0.0393
Model comparison			
BIC	15.4	3.1	3.1

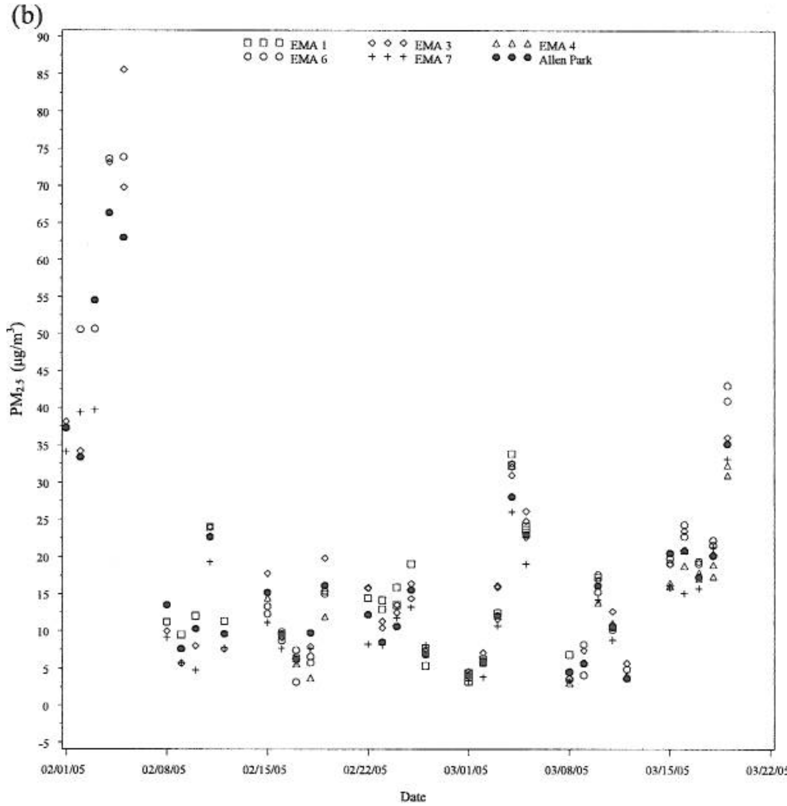
Note: All p-values are shown; p-values ≤ 0.05 are considered statistically significant; omitted model terms are indicated by the absence of corresponding p-values; [a]Turbulence kinetic energy.


EMA
Allen Park
1
2
3
4
5
6
7
Major Roads
N
0
5
10
15
20
Kilometers

(a)


EMA 1
EMA 3
EMA 4
EMA 6
EMA 7
Allen Park
PM2.5 (μg/m3)
Date
07/13/04
07/20/04
07/27/04
08/03/04
08/10/04
08/17/04
08/24/04
08/31/04

(b)


EMA 1
EMA 3
EMA 4
EMA 6
EMA 7
Allen Park
PM2.5 (μg/m3)
Date
02/01/05
02/08/05
02/15/05
02/22/05
03/01/05
03/08/05
03/15/05
03/22/05

Season 2 Reduced Model with Week

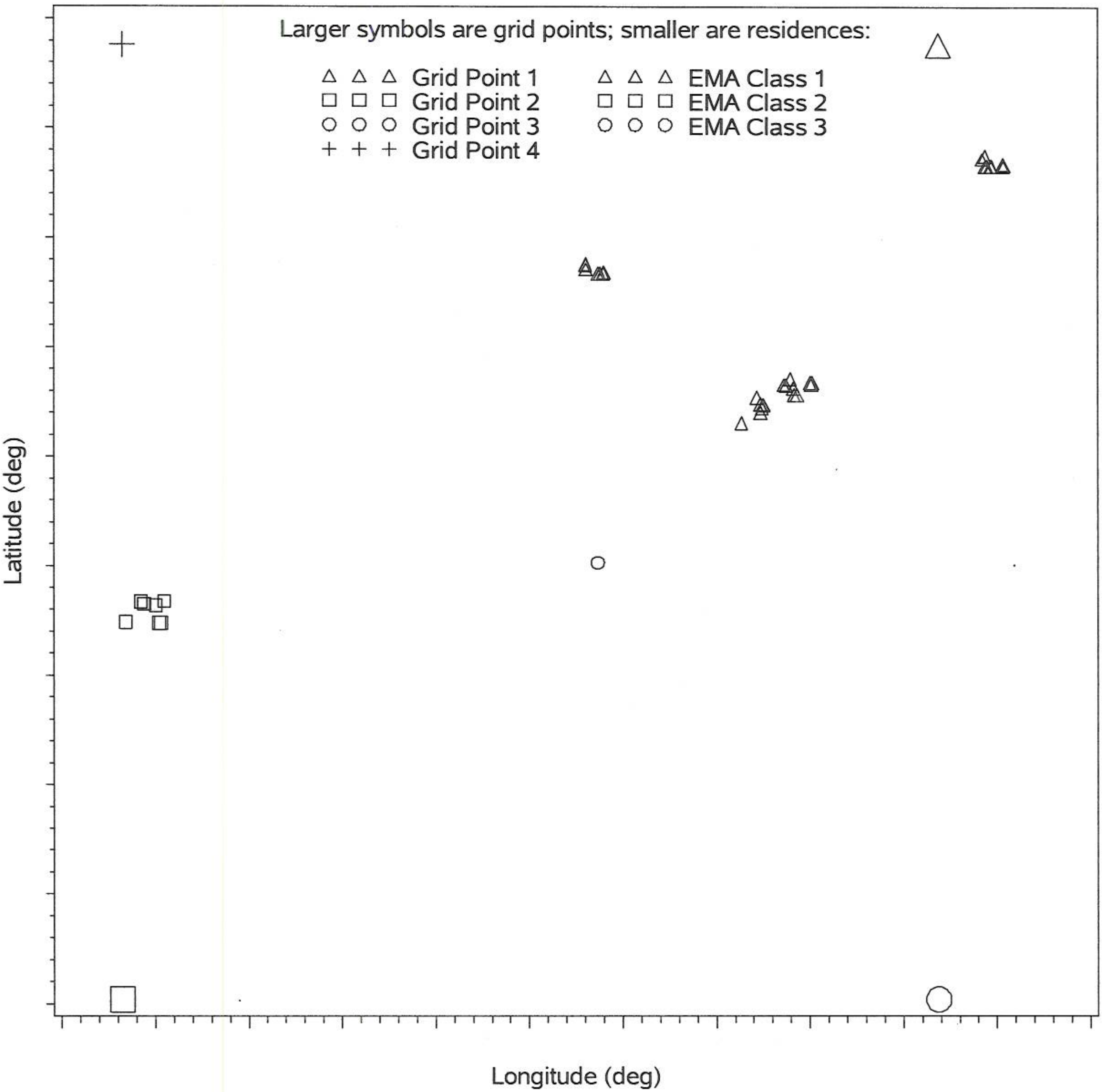
```
proc mixed;
class pid time emagrid week;
model logO_HMTW = emagrid week tke uu10m vv10m week*tke / solution;
repeated time / subject=pid(week) type=cs;
title 'Season 2 Reduced Model with week';
run;


/* comments on model and repeated statements:

   model  log(PM2.5) = EMA_class  week  turbulence_kinetic_energy
          U_wind_component  V_wind_component
          week_×_turbulence_kinetic_energy / solution;

   repeated  time / subject=PID_nested_within_week
          variance_type=compound_symmetry;

end of comments */
```


Larger symbols are grid points; smaller are residences:
Grid Point 1
Grid Point 2
Grid Point 3
Grid Point 4
EMA Class 1
EMA Class 2
EMA Class 3
Latitude (deg)
Longitude (deg)