Published in final edited form as:

J Air Waste Manag Assoc. 2009 October ; 59(10): 1173–1185.

Using a Continuous Time Lag to Determine the Associations Between Ambient PM_{2.5} Hourly Levels and Daily Mortality

Joan G. Staniswalis and Hongling Yang

Department of Mathematical Sciences University of Texas at El Paso

Wen-Whai Li and

Department of Civil Engineering University of Texas at El Paso

Kerry E. Kelly

Institute for Combustion & Energy Studies University of Utah

Abstract

We are interested in understanding the possible association between exposure to short-term PM_{2.5} peaks that have changing physical characteristics throughout the day and observable health outcomes (daily mortality). To this end, modern statistical methods are used here that allow for a continuous time lag between hourly PM_{2.5} mass concentration and daily mortality. The *functional* linear regression model was used to study how hourly PM_{2.5} mass of past days continuously influence the daily mortality count of the current day. Using a Poisson likelihood with the canonical link, we found that a 10 $\mu\text{g}/\text{m}^3$ increase in the hourly PM_{2.5} above the hourly average is associated with 1.7% (.1, 3.4), 2.4% (1.2, 3.7), 1.6% (.6, 2.7), .8% (–.2, 1.8) higher risk of mortality on the same day, next day, 2-days and 3-days later, respectively. The increase in relative risk is statistically significant for lags of 0–2 days, but not at lag 3. The highest association between PM_{2.5} mass concentration and daily mortality was found to occur in the morning when both mass and PM number concentrations peak at roughly 8 a.m. (lag of 15, 39, and 63 hours). This morning time interval corresponds to automobile traffic rush-hour that coincides with a morning atmospheric inversion that traps high concentrations of nanoparticles.

Keywords

nanoparticles; fine particles; functional linear model; smooth distributed lag model

About the Authors

Joan G. Staniswalis is Professor at The University of Texas at El Paso, Department of Mathematical Sciences. Hongling Yang is currently a lecturer at The University of Texas at El Paso, having recently completed her doctorate in statistics at Arizona State University, Department of Mathematics, Phoenix, AZ. Wen-Whai Li is Professor at The University of Texas at El Paso, Department of Civil Engineering. Kerry Kelly is Associate Director, Institute for Combustion & Energy Studies, University of Utah.

IMPLICATIONS

In this paper, the different sampling rates of mortality (daily) and PM_{2.5} (hourly) are accommodated by a historical functional linear model without any reduction of the hourly PM_{2.5} data. All lags of PM_{2.5} are simultaneously considered, thereby avoiding variable selection methods. Daily mortality was significantly associated with hourly PM_{2.5} (p-value=.0011) and most highly associated with a morning peak of PM_{2.5} mass concentration, which is the primary peak in the total number of particles and is dominated by ultrafine particles. The results of the analysis imply that short-term exposure to elevated PM_{2.5} mass and number concentrations might be the better predictor of daily mortality in El Paso.

INTRODUCTION

Particulate matter (PM) air pollution has been associated with adverse health effects, decreased heart-variability in the elderly, and respiratory and cardiopulmonary related morbidity and mortality. Among the various sizes of PM, fine PM (PM of less than or equal to $2.5 \mu\text{m}$ in aerodynamic diameter or $\text{PM}_{2.5}$) is of particular concern. Pope and Dockery¹ summarized the evaluations of health effects associated with long-term and short-term exposures to ambient PM conducted in recent years. They reported that thoracic PM (PM of less than or equal to $10 \mu\text{m}$ in aerodynamic diameter or PM_{10}) is associated with all-cause mortality, lung cancer, and nonmalignant respiratory mortality for males and coronary heart disease in females. Higher health risks may be associated with exposure to fine PM rather than to thoracic PM.

Most recently, the Womens Health Initiative Observatory Study reported by Miller et al² found a 76% increase in the risk of death from cardiovascular disease for every increase of $10 \mu\text{g}/\text{m}^3$ in long-term averaged $\text{PM}_{2.5}$ concentration. Dominici et al³ evaluated the short-term health effects of $\text{PM}_{2.5}$ and reported positive association between day-to-day variation in $\text{PM}_{2.5}$ and hospital admissions with the largest effect observed at 0 day lag (same day) for all cardiovascular health outcomes except ischemic heart disease (at 2 day lag time instead) using daily averaged $\text{PM}_{2.5}$ and hospital admission data from 204 counties in the U.S. In particular, the risk of heart failure increased by 1.28 % per $10 \mu\text{g}/\text{m}^3$ increase in same-day $\text{PM}_{2.5}$. Both studies echoed the earlier statement made by the U.S. Environmental Protection Agency⁴ that the new studies support previous conclusions that short-term exposure to fine PM is associated with both mortality and morbidity, where short-term is referring to 24-hour exposure. Other recent findings show changes in heart rate variability^{5,6} and triggering of myocardial infarction^{7,8} associated with 1–4 hour peaks in PM_{10} .

We are interested in understanding the possible association between exposure to diurnal episodic (hourly) $\text{PM}_{2.5}$ concentrations and short-term observable health outcomes (daily mortality). To this end, modern statistical methods are used that allow for a continuous time lag between hourly $\text{PM}_{2.5}$ and daily mortality. Moreover, the different sampling rates for mortality (daily) and $\text{PM}_{2.5}$ (hourly) can be accommodated with no reduction of the hourly $\text{PM}_{2.5}$ profile prior to its inclusion as a covariate in a regression model. We propose the use of the functional linear regression model⁹ to study how hourly $\text{PM}_{2.5}$ of past days continuously influence the daily mortality count of the current day. We wish to test the hypothesis that daily mortality rate in El Paso is associated with exposure to elevated hourly $\text{PM}_{2.5}$ concentrations.

Most studies of the association between $\text{PM}_{2.5}$ and daily mortality have two common elements in the statistical analysis: (i.) hourly $\text{PM}_{2.5}$ measurements are collapsed by summary statistics such as the daily mean, and (ii.) variable selection methods are used to determine the lag of daily average $\text{PM}_{2.5}$ most highly associated with daily mortality. Staniswalis et al¹⁰, an exception to (i.), demonstrated a significant association between PM_{10} , lag 3, and daily mortality in El Paso, TX using the first principal component score of hourly PM_{10} , an association that would have otherwise been missed with the daily average of PM_{10} . Notable exceptions to (ii.) involve the use of the distributed lag model (DLM), a regression model including time lagged variables as covariates. The standard unconstrained DLM includes the average daily PM (PM_{10} or $\text{PM}_{2.5}$) for all lags from say 0 to q days as covariates in the regression model. A polynomial DLM avoids the problem of multicollinearity among the lagged PM levels by constraining the coefficients to be a polynomial function of the lag number¹¹. The smooth DLM replaces the polynomial function with a smooth function of the lag number. The smooth DLM was first applied to daily summaries of pollution/mortality data from Milan, Italy by Zanobetti et al¹².

As an extension to the study of Staniswalis et al¹⁰, we examine the association between the newly available PM_{2.5} and mortality data with the intention to delineate the effects caused by naturally occurring wind events (the major source for elevated PM₁₀ concentration) and diurnal anthropogenic traffic-related pollution (the major source for elevated PM_{2.5} concentration). We use the smooth DLM with hourly PM_{2.5} measurements, rather than daily mean PM_{2.5}, thereby avoiding both (i.) and (ii.) above, namely, arbitrary collapsing of the hourly measurements, and variable selection for determining the lag of the pollutant most highly associated with daily mortality. The smooth DLM is an example of a historical functional linear model¹³; we review the historical functional linear model and fitting by penalized regression splines prior to its application.

DATA

The study was restricted to the 2000–2005 six year period for which records of hourly PM_{2.5} and temperature were obtained from the Aerometric Information Retrieval System for the site located on the edge of the University of Texas at El Paso (CAMS12). CAMS12 and CAMS40 (both state-operated continuous ambient monitoring stations) are the only monitoring stations providing hourly PM_{2.5} data during the study period. CAMS40 is located near a bus/train terminal in an industrial section of El Paso, which is not heavily populated, and the PM_{2.5} levels at CAMS40 are not likely to be representative of El Paso. Only PM_{2.5} data collected at CAMS12 were used for the analysis. PM_{2.5} at this station may not be a perfect representation of exposure for the general population of El Paso, as it is well known that use of data from a central monitoring site can lead to exposure misclassifications under various circumstances. CAMS12 data are likely to underestimate PM_{2.5} for the region, but CAMS12 data were the only complete set of PM_{2.5} data for the region and the study time-period. PM₁₀ data at this site, however, were excluded from this study because of excessive missing values between July 2000 and October 2001 (> 20% of the study period).

B-splines were used to interpolate the hourly PM_{2.5} mass concentration data, to impute any missing observations spanning fewer than 5 hours. Data gaps of 5 or more hours in a given day were imputed with the mean for that day based on the available PM_{2.5} hourly observations. Twenty-five days were missing all hourly data, so the missing values were imputed with the mean of all available hourly PM_{2.5} data over the 6 year period.

Figure 1 is a graphical summary of the hourly PM_{2.5} data from this time period. Figure 2 displays the directions of maximum variation (first PC direction) of the hourly profiles for PM_{2.5} for each year of the study period. A pooled estimate of the first PC direction (as in Staniswalis et al¹⁰ for hourly PM₁₀) is not computed because the shape of the curves are not the same by year. The direction of maximum day-to-day variation in the hourly profiles of PM_{2.5} is in the size of an 8 pm evening peak (Years 2000, 2004, 2005), a late afternoon peak (Year 2003), and an early afternoon peak (Year 2001, 2002). Table 1 displays summary statistics of daily average PM_{2.5} and temperature, and hourly PM_{2.5} for for the study period.

Complete mortality data corresponding to all natural deaths in El Paso County were obtained from the Texas Department of Health for the six year interval 2000–2005. Natural deaths were identified with ICD-9 codes (International Classification of Diseases codes) less than 800. For the time period under study, the average number of daily deaths is 11.3, and 12.6 is the variance.

HISTORICAL FUNCTIONAL LINEAR MODEL

A historical functional linear model¹³ refers to the special case when the influence of the covariate on the response is of a feed-forward nature. This means that the behavior of daily mortality $y(t)$ depends only on the behavior of hourly PM_{2.5} $x(s)$ at past times $s \leq S(t)$. Here the transformation $S(t) = 24t$ converts the daily index $t \in \{1, \dots, T\}$ of mortality $y(t)$ to the

hourly index s of $PM_{2.5}$ $x(s)$. Daily mortality is modeled as a Poisson distributed variable with mean denoted by $\mu_t = E[y(t)]$. The canonical link for the Poisson distribution is used, meaning that $\eta_t = \ln \mu_t$ is the linear predictor, which we model as

$$\eta_t = \alpha(t) + \int_0^{S(t)} x(S(t) - u) \beta(u) du. \quad (1)$$

The *intercept* function $\alpha(t)$ is the overall summary of other factors influencing mortality y on the current day t . The *slope* function $\beta(u)$ reflects the strength of the association between pollution and mortality at a lag of u hours. For example, the value of $\beta(u)|_{u=12 \text{ hours lag}}$ describes the influence of the noon reading of hourly $PM_{2.5}$ on today's deaths, whereas $\beta(u)|_{u=36 \text{ hours lag}}$ describes the influence of yesterday's noon hour reading of $PM_{2.5}$ on today's deaths. The relative risk RR associated with a $10 \mu\text{g}/\text{m}^3$ uniform increase in the hourly PM across a lag of $[T_1, T_2]$ hours, while holding other variables fixed, can be computed according to

$$RR = \exp\left(10 \int_{T_1}^{T_2} \beta(u) du\right). \quad (2)$$

Parametric models that are linear in daily average $PM_{2.5}$ assume a functional form for $\beta(u)$. For example, a linear model in $\bar{x}_{t,0}$, daily average $PM_{2.5}$ with zero lag:

$$\eta_t = \alpha(t) + \beta_0 \bar{x}_{t,0}, \quad (3)$$

is formulated in terms of a historical functional linear model by specifying $\beta(\cdot)$ to be the boxcar function

$$\beta(u) = \begin{cases} \frac{1}{24} \beta_0 & u \in [0, 24) \text{ hours} \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

In which case, the relative risk RR according to equation (2) is $\exp(10\beta_0)$ for a $10 \mu\text{g}/\text{m}^3$ uniform increase in the hourly PM across $[0, 24)$ hours on the current day.

Consider instead the linear model

$$\eta_t = \ln \mu_t = \alpha(t) + \beta_2 \bar{x}_{t,2}, \quad (5)$$

where $\bar{x}_{t,2}$ is daily average $PM_{2.5}$ at a two day lag. In this case $\beta(u)$ is specified as

$$\beta(u) = \begin{cases} \frac{1}{24} \beta_2 & u \in [48, 72) \text{ hours} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The relative risk RR according to equation (2) is $\exp(10\beta_2)$ for a $10 \mu\text{g}/\text{m}^3$ uniform increase in the hourly PM across $[48, 72)$ lagged hours. As a final example, if daily average $PM_{2.5}$ is used to predict daily mortality with a 0-, 1-, and 2-day lag:

$$\eta_t = \ln \mu_t = \alpha(t) + \beta_0 \bar{x}_{t,0} + \beta_1 \bar{x}_{t,1} + \beta_2 \bar{x}_{t,2}, \quad (7)$$

then

$$\beta(u) = \begin{cases} \frac{1}{24}\beta_0 & u \in [0, 24) \text{ hours} \\ \frac{1}{24}\beta_1 & u \in [24, 48) \text{ hours} \\ \frac{1}{24}\beta_2 & u \in [48, 72) \text{ hours} \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

In each of the above parametric examples, the coefficients $\beta_0, \beta_1, \beta_2$ could be estimated from the data using the usual time-series methods for Poisson regression and would have the usual interpretation as the coefficient of daily average $\text{PM}_{2.5}$ with 0-, 1-, or 2-day lag, respectively.

In this paper, a functional form for $\beta(u)$ is not specified, instead it is estimated using nonparametric smoothing methods described in the next section. In the data analysis, the nonparametric estimate of $\beta(u)$ will be compared with the parametric estimate given by equation (8).

Additional Modeling Assumption

It is reasonable to suppose that only pollution exposures corresponding to time lags $u \leq \tau$ are likely to impact mortality on a given day, leading to the following truncated historical functional linear model for $\eta_t = \ln \mu_t$:

$$\eta_t = \alpha(t) + \int_0^{\min(S(t), S(\tau))} x(S(t) - u) \beta(u) \, du. \quad (9)$$

In any case, computer implementation of the historical functional linear model requires truncation of the lags used in prediction of the current day. Truncating the past influence of pollution to τ days has the effect of stabilizing the variance of the estimator of $\beta(\cdot)$, although, it comes at the cost of introducing one more unknown parameter.

An additional modeling assumption is made that further distinguishes our model from that of Zanobetti et al¹². In a final step taken to reduce the dimensionality of the regression problem, it is assumed that the intercept function $\alpha(t)$ in equation (9) is a periodic function

$$\alpha(t) = \alpha(t + 365), \quad t \in \{1, \dots, T\}. \quad (10)$$

This intercept function $\alpha(t)$ is accounting for seasonal changes over time t that influence daily mortality y_t . However, assuming periodicity of $\alpha(\cdot)$ necessitates augmenting (9) to include a slowly varying nonparametric function $s_1(t)$ to model non-periodic changes over time t , for example, changes in y attributable to population growth over time:

$$\eta_t = \alpha(t) + s_1(t) + \int_0^{\min(S(t), S(\tau))} x(S(t) - u) \beta(u) \, du + s_2(\text{temperature}_t) + \text{terms for day-of-week}_t. \quad (11)$$

Equation (11) now also includes nonparametric additive terms $s_2(\cdot)$ for modeling the dependence of mortality on the daily mean of temperature, and indicator variables for day-of-week. The nonparametric functions in the additive model given by equation (11) are all modeled as linear combinations of a flexible cubic B-spline basis. Detailed treatment of these functions is included in the Appendix for interested readers. The data analysis was written in the S programming language using S-PLUS.

Similarly, other variables, either co-pollutants (such as, NO₂, ozone, PM_{10-2.5}) or atmospheric conditions (mixing height, humidity, wind speed) can be incorporated into the model to evaluate the dependence of mortality on them. In this study, we have limited the nonparametric additive term to one for ambient temperature. Dew point was not included in the model because it is highly correlated with temperature, $r = .6$ sample correlation.

RESULTS

The truncated historical functional linear model (11) was fit to the El Paso 2000–2005 pollution/mortality data. Hourly PM_{2.5} was a significant predictor of daily mortality adjusted for temperature and day-of-week. Figure 3 overlays the fitted component functions in the additive model (11) that includes the covariate hourly PM_{2.5} for $\tau=4,6,8,10$ and 12 days in the prediction of daily mortality. The parameter τ determines the number of days for which hourly PM_{2.5} measurements are included as a predictor of daily mortality, for example $\tau = 12$ includes the hourly PM_{2.5} measurements for the current day and the previous eleven days. The values $\tau=4,6,8,10$ and 12 days were chosen arbitrarily for visual inspection. The knots for the periodic intercept function $\alpha(t)$ were placed at 2-month (60-days) intervals, whereas, the knots for the slope function $\beta(\cdot)$ were placed at 2-day intervals; the knots for the smooth dependence of time $s_1(t)$ were placed at 2-year intervals, and for the smooth dependence on temperature $s_2(\text{temperature})$ at 10-degree intervals. For each component in Figure 3, the placement of the knots are indicated by the tic-marks. The values of the smoothing parameter (λ) were obtained by trial and error and set to $\lambda_\alpha = 10^4$, $\lambda_\beta = 10^7$, and the smoothing parameters for the other smooth functions were set to 10^6 . A larger value of λ results in a smoother fit of the regression, as compared with a smaller λ that provides a much rougher fit. The shapes of all of the component functions in the full model are fairly consistent across all values of truncation τ . However, because the dimension of the B-spline basis needed for estimation of the slope $\beta(\cdot)$ in the full model increases in τ , the power of the scaled deviance test statistic goes down as τ increases. Thus, it was found that the p-value for significance of PM_{2.5} ranges from .0003 to .003 as τ increases from 4 to 12.

Computer simulations¹⁴ suggested that late truncation was preferred over early truncation to reduce bias. Thus, those computer simulations suggest that results shown in Figure 3b are preferred with truncation at $\tau = 8$ days for prediction of mortality. See Figure 4 for the fitted components of equation (11) along with ± 2 standard error bars with $\tau = 8$.

The fit in Figure 4a for the intercept function $\alpha(t)$ indicates that daily mortality tends to be highest in January, decreasing in time until a minimum at about 180 days (June/July), then increasing for the rest of the year without attaining the global maximum seen in January. The fit in Figure 4b for the slope function $\beta(\cdot)$ is convex with a global maxima between .5–3 days, suggesting that the association between daily mortality and PM_{2.5} is highest for a .5–3 day lag. The pointwise confidence intervals for the estimate of the slope function $\beta(\cdot)$ do not contain 0 for lags within roughly .5–3 days, signifying that PM_{2.5} is a significant predictor of daily mortality. Furthermore, Table 2 shows that an increase of 10 $\mu\text{g}/\text{m}^3$ in the hourly PM_{2.5} across a day would be associated with a relative risk of 1.017 (1.001, 1.034), 1.024 (1.012, 1.037), 1.016 (1.006, 1.027), and 1.008 (.998, 1.018) deaths on the same day and following three days, respectively; see equation (2) for RR . The increases in RR for lags 0–2 days are statistically significant, but not for the 3-day lag. These relative risk estimates are consistent with those reported based on daily average PM_{2.5}, namely, relative risks between .9 to 1.17, with the exception of one study reporting a nonsignificant finding of 1.32 (Environmental Protection Agency⁴ 2006, Table 1).

The fit in Figure 4c for $s_1(t)$ shows that daily mortality was steadily increasing in time over the six year period, although the rate of increase was lower over 2002–2004. Figure 4d displays

the fit for $s_2(\text{temperature})$, showing that mortality tends to decrease as temperature increases. Although this seems counterintuitive at first, it is consistent with Figure 4a, which shows that June/July (months with the highest temperatures) on average have fewer deaths; whereas December/January (months with the lowest temperatures) on average have the most deaths. So it seems plausible that the estimated smooth function $s_2(\text{temperature})$ is an adjustment of the smooth intercept $\alpha(t)$ shown in Figure 4a. Lastly, Figure 4e displays the coefficients for the indicator variables for day of week (Saturday through Friday). Only the coefficient for the indicator function for Friday is significantly different from zero, and because the coefficient is greater than zero, on average there are more deaths on Friday than the other days of the week.

So far, because this analysis has placed the knots for estimation of the slope $\beta(\cdot)$ at 2-day intervals, fine details of the effects of hourly $\text{PM}_{2.5}$ on daily mortality are obscured by the rigidity of the B-spline basis. In order to develop a better understanding of the association between $\text{PM}_{2.5}$ and daily mortality at specific hours, the regression analysis is executed with regularly spaced knots at two-hour intervals, $\lambda_\beta = 10^8$ and $\tau = 3$ for the estimation of the slope function $\beta(\cdot)$. The knots and smoothing parameters for the other additive components $\alpha(\cdot)$, $s_1(t)$, and $s_2(\text{temperature})$ remain as before. Figure 5 displays the function estimates with ± 2 standard error bars in solid lines, from which it can be seen that morning hourly- $\text{PM}_{2.5}$ levels around 8 a.m. (lags of 15, 39, 63 hours) are most highly associated with mortality. The fits for the other terms in the model were similar to those given in Figure 4, and are not displayed.

As a basis for comparison with other published results, relative risk is also calculated based on the parametric Poisson linear model (8) that uses the daily average $\text{PM}_{2.5}$ at 0-, 1-, and 2-day lags; see Table 2. The changes in relative risk for lags 1–2 are statistically significant, but not for lag 0. The nonparametric model gave statistically significant changes in relative risk for all lags 0–2, but the relative risk estimates themselves at lag 0 are similar. The parametric estimates of the slope function $\beta(\cdot)$ under equation (8) together with ± 2 standard error bars are graphed in Figure 5 in dashed lines. The parametric estimate seems to reflect a mean value of the nonparametric smooth estimates over each of the 24-hour periods, obscuring the relationship between hourly $\text{PM}_{2.5}$ and daily mortality. Our findings using hourly and daily average $\text{PM}_{2.5}$ (see Table 2) are in line with those of Franklin et al.¹⁵ who reported a relative risk of 1.012% for all-cause mortality associated with a $10 \mu\text{g}/\text{m}^3$ increase in the previous day's $\text{PM}_{2.5}$ average concentration in a study of 1.3 million deaths in 27 US communities.

We next examined the physical characteristics of $\text{PM}_{2.5}$ at this site. Temporal variations of $\text{PM}_{2.5}$ mass and size-resolved number concentrations were determined in two consecutive seasons (October 26–November 12, 2006 for the fall season and March 29 to April 15, 2007 for the spring season) using an Aerodynamic Particle Sizer, a Scanning Mobility Particle Sizer, and 2 Tapered Element Oscillating Microbalance monitors.^{16,17} A $\text{PM}_{2.5}$ mass peak of less magnitude occurring in the morning hours between 6 to 9 am was observed for 50% of the time during the exploratory study. It was observed that the number concentration peaked in the morning hours while another number concentration peak of less magnitude was observed in the evening (Figures 6 and 7) and the mode changes from ~ 20 nm to ~ 40 nm at night (Figure 8). The total number of particles peaked in the morning as well as in the evening, while the mode of the particle size changed from ~ 20 nm to ~ 40 nm, indicating different PM sources may be responsible for the mass and number concentrations and/or agglomeration of particles in the atmosphere during the day. We thus believe that mortality is associated with exposure to high number of nanoparticles in the morning hours, most likely generated by traffic-related activities, rather than the ambiguously defined mass concentration and hypothesize a better predictor of number concentration for mortality.

DISCUSSION

The goal in this study is to use $PM_{2.5}$ as an indicator to shed light on the associations between mortality and pollutants. Our analysis uses the raw hourly $PM_{2.5}$ mass concentrations in a linear model for prediction of daily mortality avoiding the arbitrary use of the daily mean, which inadvertently obscures the significance of acute (hourly) PM exposure on various health outcomes¹⁰. At high ambient concentrations, acute short-term (1-hr) exposure to diesel engine exhaust has been shown to produce a well-defined and marked systemic and pulmonary inflammatory response in healthy human volunteers^{18,19}. Studies have begun to link short-term increases in hourly $PM_{2.5}$ concentrations with adverse cardiac effects²⁰. However, few studies have evaluated the effect of short-term increases in $PM_{2.5}$ concentration on daily mortality.

Our fitted model differs from that of Zanobetti et al¹² in that the smooth function in time included two additive components: a periodic piece of period 365 days $\alpha(t)$, and a slowly varying function in time $s_1(t)$. See equation (11). This has the advantage of reducing the dimensionality of the model. Further, Zanobetti et al¹² used daily average $PM_{2.5}$ in a functional linear model, and we used hourly measurements.

The association over time of $PM_{2.5}$ and mortality were studied by varying selection of the knots and truncation point τ in the historical functional linear model; see equation (11). This is reviewed below. Figure 4, shows that $PM_{2.5}$ lags of up to .5–3 days are most highly associated with daily mortality. This is consistent with the published literature reviewed by the Environmental Protection Agency⁴. Such a global viewpoint over the effects of time was achieved by: (1) Placing the knots for estimation of $\beta(\cdot)$ at regular 2 day intervals, and (2) Allowing $PM_{2.5}$ hourly levels for day t and the previous 7 days ($\tau = 8$) to enter the linear model for prediction of mortality on day t .

Figure 5 allows for a more detailed view in time of the relationship between $PM_{2.5}$ and mortality that is not possible in studies that use the daily average PM. Such a detailed view was obtained by: (1) Placing the knots for estimation of $\beta(\cdot)$ at regular 2 hour intervals, and (2) Choosing $\tau = 3$, to zoom in on the association between daily mortality and hourly $PM_{2.5}$ on the current day and previous two days. The highest association between $PM_{2.5}$ mass concentration and mortality occurs with morning $PM_{2.5}$ levels around roughly 8 a.m. (lag of 15 hours) or earlier.

The irrelevance of the even higher diurnal evening $PM_{2.5}$ mass peak (Figure 1) to mortality and the association between PM_{10} mass and mortality previously reported for El Paso¹⁰ deserve some investigation. Staniswalis et al¹⁰ found that diurnal evening peak (likely caused by traffic related emissions) and occasional afternoon peak (caused by high winds) explain 40% and 15%, respectively, of the day-to-day variation in the daily PM_{10} profiles in El Paso. However, the obvious morning peak (Figure 2 of Staniswalis et al¹⁰) was obscured mathematically in their principal component analysis of a 5-yr record of PM_{10} and mortality data. They further observed that mortality in El Paso may be underestimated by 20% if daily average PM_{10} instead of hourly PM_{10} is used and that relative risk is approximately 20% higher when the evening PM_{10} peak occurred under low-wind condition (wind speed less than 2.5 m/sec).

It is well known in the literature that while the direction of maximum variation (evening peak) may be significantly associated with the response (mortality), the direction of maximum variation (evening peak) is not necessarily the most highly correlated covariate with a given response (mortality)²¹. Power to detect statistical significance of a covariate is driven by two features of the covariate: the variance of the covariate and the correlation with the response. A covariate may be highly correlated with a response, but there may not be enough power to detect this when the variance of the covariate is small (a PC direction of lesser variation). Figure 2 shows that indeed the evening peak is still the primary direction of maximum variation in

this study explaining the day-to-day variation of PM_{2.5} mass, although mathematically it is not as dominant as that observed in the PM₁₀ analysis of Staniswalis et al¹⁰.

The conclusion that mortality risk associated with either diurnal evening or occasional afternoon PM₁₀ peaks under high-wind condition is lower than that under low-wind conditions¹⁰ and the fact that PM₁₀ mass in the region is composed of mostly coarse PM (or PM_{10-2.5}, at 75 to 90%)^{22,23,24} imply that the coarse fraction of PM₁₀ may be less potent than PM_{2.5} and that PM_{2.5} may be a better predictor than PM₁₀ for daily mortality. Unfortunately, we were unable to fit the functional linear model on PM₁₀ data at this CAMS 12 site to further validate our premise due to excessive missing data.

The fact that the morning PM_{2.5} peak is highly correlated with daily mortality, as shown in Figure 5, is interesting, particularly when the evening peak is not implicated in the analysis (Figure 1). We thus looked into the physical and chemical difference between morning and evening peaks. Chemical speciation data for the morning and afternoon peaks during this period of time is not available. The most relevant data available is a set of 4 week time-resolved PM_{2.5} chemical speciation data collected at a station approximately 3.5 miles northwest of the CAMS12 site²³ in the winter of 2002. Concentrations of anions, EC, OC, and elemental composition of PM_{2.5} for the 3-hr samples collected in the evening (6–9 pm) appear to be greater than that collected at other 3-hr intervals. Concentrations of chemical composition of morning PM_{2.5} (6–9 am) is consistently less than that observed in the evening (6–9 pm) and show little difference from that observed throughout the day. Thus, species and magnitudes of chemicals absorbed onto PM_{2.5} do not appear to be a factor in the dominance of morning peak on mortality.

CONCLUSION

We used a smooth distributed lag model to analyze 6 years of continuous record of hourly PM_{2.5} and daily mortality data in El Paso, Texas. We discovered a statistically significant association between the morning PM_{2.5} peak at a central monitoring site at .5–3 days lagged and daily mortality in the general population. By examining the available physical and chemical properties of PM_{2.5} for the site, we believe that the total number of nanoparticles, particularly those in the range of ~ 20 nm, likely generated by traffic-related activities is the principal factor of the association and that total number of particle may be a better and simplistic indicator for mortality. A limitation of this study is that in El Paso only one PM_{2.5} monitoring site was available with complete data for the study period. Future studies are needed to confirm the findings for El Paso, to determine the generalizability of the findings to other areas of the United States, and to evaluate the association between daily mortality and other pollutant components such as NO, CO, and NO₂.

ACKNOWLEDGEMENTS

Joan Staniswalis was partially supported by NIH-SCORE 2S06 GM008012. Wen-Whai Li and Kerry E. Kelly were partially supported by a grant from the Health Effects Institute (HEI, RFA05-1B). The authors thank JoAnn Lighty, University of Utah for instrumental support, and Dr. J.O. Ramsay, McGill University, for his functional analysis software written in S-PLUS that was used to impute missing hourly data. This publication was made possible by grant number S11 ES013339 from the National Institute of Environmental Health Sciences (NIEHS), NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIEHS, NIH, or HEI.

APPENDIX: Using a Continuous Time Lag to Determine the Associations Between Ambient PM_{2.5} Hourly Levels and Daily Mortality

This appendix lays out the matrix expression for the linear predictor $\eta_t = \ln \mu_t$ and the equations for estimation of the parameters.

NONPARAMETRIC CURVE FITTING WITH P-SPLINES

For model identifiability, the usual constraints

$$\sum_{i=1}^T s_1(t_i) = 0, \quad (12)$$

$$\text{and } E[s_2(\text{temperature})] = 0 \quad (13)$$

are imposed on the estimates in equation (11). The nonparametric functions in the additive model are all modeled as linear combinations of a cubic B-spline basis^{25,26}, meaning that the fitted functions are piecewise cubic polynomials smoothly joined together at prespecified points in time called knots. Cubic B-splines provide a convenient flexible basis for the representation of smooth curves, see for example, Ostro et al²⁷ who modeled the dependence of daily mortality on time and weather. B-splines have compact local support that can be exploited to obtain stable and efficient algorithms for nonparametric curve fitting. Adjacent B-splines have overlapping support that is exploited in the P-spline curve fitting technology²⁸. This will be described next.

Let the vector β denote the coefficients in the B-spline representation of the unknown regression functions, and the coefficients for the indicator variables of day-of-week. The vector β is estimated from the data by maximization with respect to β of the penalized Poisson log-likelihood^{29,30} with the canonical link function:

$$\mathcal{L}(\beta; \text{data}) - \mathbf{P}(\lambda, \beta), \quad (14)$$

the Poisson log-likelihood is conditional on the covariates. The second term in the penalized log-likelihood is a difference penalty that is explicitly given later in this Appendix. It has the effect of imposing smoothness on the regression function by requiring that the heights of neighboring B-splines with overlapping support “hold hands” to withstand erratic fluctuations in the data. This difference penalty is similar in flavor to the roughness penalty based on the squared integral of the second derivative of the regression function. A large value of the smoothing parameter λ results in a smoother fit of the regression function, as compared with small values of λ that provide a much rougher fit.

The estimate of the parameter β that maximizes equation (14) was computed by iteratively reweighted least squares³⁰. This is explained later in this Appendix. The scaled Poisson deviance³⁰ was used to test for significance of the covariate x as a predictor of y in equation (11) for the Poisson mean.

Estimating β Alone

Here the method of estimation is developed step by step, beginning with a simplified model for η_t , $t \in \{1, \dots, T\}$ that has no intercept. Set the truncation point for the historical functional linear model at $\tau \leq 365$ days. Later, the equations are developed for simultaneous estimation of both the intercept and the slope functions. So for now the model is

$$\eta_t = \int_0^{\min(S(t), S(\tau))} x(S(t) - u) \beta(u) \, du. \quad (15)$$

Let $\{N_j(\cdot)\}_{j=1}^K$ denote the cubic B-spline basis for a given knot sequence on $[0, S(\tau)]$. The dependence of K on τ is suppressed from the notation in order to simplify the notation in the presentation that follows. The slope function is approximated by

$$\beta(u) = \begin{cases} \sum_{j=1}^K b_j N_j(u) & u \in [0, S(\tau)] \\ 0 & u > S(\tau). \end{cases} \quad (16)$$

An equivalent formulation to (15) :

$$\eta_t = \sum_{j=1}^K b_j \phi_j(t; x), \quad (17)$$

is obtained by defining

$$x_+(S(t) - u) = \begin{cases} x(S(t) - u) & \text{when } u \leq S(t) \\ 0 & \text{otherwise,} \end{cases} \quad (18)$$

$$\phi_j(t, x) = \int_0^{S(\tau)} x_+(S(t) - u) N_j(u) \, du. \quad (19)$$

Approximate the integrals defining $\phi_j(t; x)$ with a quadrature rule using the Q equally spaced time points $\{g_q\}_{q=1}^Q$ in $[0, S(\tau)]$:

$$\phi_j(t, x) \approx \frac{S(\tau)}{Q} \sum_{q=1}^Q x_+(S(t) - g_q) N_j(g_q). \quad (20)$$

Let $\mathbf{X}'(t) = (x_+(S(t) - g_1), \dots, x_+(S(t) - g_Q))$, $\mathbf{b}' = (b_1, \dots, b_K)$, and

$$\mathbf{N}_{\mathbf{b}} = \begin{bmatrix} N_1(g_1) & N_2(g_1) & \cdots & \cdots & N_K(g_1) \\ N_1(g_2) & N_2(g_2) & \cdots & \cdots & N_K(g_2) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ N_1(g_Q) & N_2(g_Q) & \cdots & \cdots & N_K(g_Q) \end{bmatrix}_{Q \times K}. \quad (21)$$

Set $h = \frac{S(\tau)}{Q}$. Then equation (17) can be written in matrix form:

$$\eta_t = h \cdot \mathbf{X}'(t) \mathbf{N}_{\mathbf{b}} \mathbf{b}. \quad (22)$$

Setting $h = 1$ hour, $\mathbf{X}(t)$ can be evaluated from the observed raw pollution data. Define η and $\mathbf{X}$:

$$\eta = \begin{bmatrix} \eta_{t_1} \\ \eta_{t_2} \\ \vdots \\ \vdots \\ \eta_{t_T} \end{bmatrix}_{T \times 1} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}'(t_1) \\ \mathbf{X}'(t_2) \\ \vdots \\ \vdots \\ \mathbf{X}'(t_T) \end{bmatrix}_{T \times Q}, \quad (23)$$

so that the model (17) for the entire T days can be written as

$$\eta = h \cdot \mathbf{X} \mathbf{N}_b \mathbf{b}. \quad (24)$$

Estimating $\alpha(\cdot)$ and $\beta(\cdot)$ Simultaneously

Now let us include the intercept function $\alpha(t)$ with period $P = 365$ days

$$\eta_t = \alpha(t) + \int_0^{\min(S(t), S(\tau))} x(S(t) - u) \beta(u) \, du, \quad t \in \{1, \dots, T\} \quad (25)$$

and simultaneously estimate $\alpha(\cdot)$ and $\beta(\cdot)$. It is assumed that T is a multiple of P . Choose a cubic B-spline basis $\{M_j(t)\}_{j=1}^d$ of dimension d on $[0, P]$ for representation of $\alpha(t)$:

$$\alpha(t) = \sum_{i=1}^d a_i M_i(t \bmod P), \quad t \in \{1, \dots, T\}. \quad (26)$$

The matrix form of $\alpha(t)$ in terms of the B-spline basis is :

$$\begin{bmatrix} \alpha(t_1) \\ \alpha(t_2) \\ \vdots \\ \alpha(t_T) \end{bmatrix} = \begin{bmatrix} \mathbf{N}_a \\ \mathbf{N}_a \\ \vdots \\ \mathbf{N}_a \end{bmatrix}_{T \times d} \mathbf{a}, \quad (27)$$

where $\mathbf{a}' = (a_1, \dots, a_d)$ and

$$\mathbf{N}_a = \begin{bmatrix} M_1(t_1) & M_2(t_1) & M_3(t_1) & \cdots & M_d(t_1) \\ M_1(t_2) & M_2(t_2) & M_3(t_2) & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ M_1(t_p) & \cdots & \cdots & \cdots & M_d(t_p) \end{bmatrix}_{P \times d}. \quad (28)$$

Incorporating equation (27) into (24),

$$\eta = [\mathbf{J} \mathbf{X}] \begin{bmatrix} \mathbf{N}_a & 0 \\ 0 & h \mathbf{N}_b \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \quad (29)$$

with $\mathbf{J}_{T \times P}$ composed by stacking identity matrices of dimension $P \times P$, that is,

$$\mathbf{J}' = [\mathbf{I}_{P \times P}, \quad \dots, \quad \mathbf{I}_{P \times P}]. \quad (30)$$

The number of identity matrices in the stack must match the number of years of data. Set

$$\tilde{\mathbf{X}} = [\mathbf{J}\mathbf{X}], \quad \tilde{\mathbf{N}} = \begin{bmatrix} \mathbf{N}_a & 0 \\ 0 & h\mathbf{N}_b \end{bmatrix}, \quad \beta = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \quad (31)$$

then

$$\eta = \tilde{\mathbf{X}}\tilde{\mathbf{N}}\beta \quad (32)$$

$$= \mathbf{X}\beta \quad (33)$$

$$= [\mathbf{X}_a \mathbf{X}_b] \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}. \quad (34)$$

Regularized Estimates of $\alpha(t)$ and $\beta(t)$

The Poisson log-likelihood of the data is given by

$$\mathcal{L}(\beta) = \sum_{i=1}^T (-\mu_i + y_i \ln \mu_i), \quad (35)$$

where μ_i is the parameter for the Poisson distribution that generated the mortality count y_i .

Assume the canonical link function $\mu_i = \exp(\mathbf{X}'_i \beta)$ with $\mathbf{X}'_i$ denoting the i^{th} row of $\mathbf{X}$ given in equation (33). Thus,

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^T (y_i - \mu_i) \mathbf{X}_i \quad (36)$$

$$= \mathbf{X}' (\mathbf{Y} - \boldsymbol{\mu}), \quad (37)$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$.

The roughness penalty²⁸ needed to constrain variation of the functions $\alpha(\cdot)$ and $\beta(\cdot)$ is described next. Recall that the intercept $\alpha(\cdot)$ has been written as a linear combination of B-splines defined on the interval $[0, P]$ with coefficients denoted by $\mathbf{a} \in \mathbb{R}^d$. The intercept function is periodically extended to the interval $[0, T]$. Similarly, the slope function $\beta(\cdot)$ has been written as a linear combination of B-splines defined on the interval $[0, T]$ with coefficients denoted by $\mathbf{b} \in \mathbb{R}^K$. The penalty needed in equation (14) is the sum $P(\lambda_a, \mathbf{a}) + P(\lambda_b, \mathbf{b})$ of the penalties for each function alone:

$$P(\lambda_a, \mathbf{a}) = \frac{1}{2} \lambda_a \|D_a \mathbf{a}\|_2^2 \quad (38)$$

$$P(\lambda_b, \mathbf{b}) = \frac{1}{2} \lambda_b \|D_b \mathbf{b}\|_2^2, \quad (39)$$

where $\|\cdot\|_2$ denotes the usual Euclidean norm of a vector. Here D_a and D_b are matrix representations of the second-order difference operator acting on the coefficients $\mathbf{a}$ and $\mathbf{b}$ in the B-spline representation of $\alpha(\cdot)$ and $\beta(\cdot)$, respectively. The general form of D_a and D_b is given as

$$\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & -2 & -1 \end{bmatrix}, \quad (40)$$

where the dimension of the matrix is $(d-2) \times d$ for D_a and $(K-2) \times K$ for D_b .

Define the D to be a block diagonal matrix of dimension $(d+K) \times (d+K)$:

$$D = \begin{bmatrix} \lambda_a D_a^t D_a & 0_{d \times K} \\ 0_{K \times d} & \lambda_b D_b^t D_b \end{bmatrix}. \quad (41)$$

The penalty on the Poisson log-likelihood is

$$P(\lambda_a, \mathbf{a}) + P(\lambda_b, \mathbf{b}) = \frac{1}{2} \begin{bmatrix} \lambda_a \mathbf{a}^t D_a^t D_a \mathbf{a} & 0_{d \times K} \\ 0_{K \times d} & \lambda_b \mathbf{b}^t D_b^t D_b \mathbf{b} \end{bmatrix} \quad (42)$$

$$= \frac{1}{2} \beta^t D \beta. \quad (43)$$

The partial derivative of this penalty with respect to β is given by

$$\frac{\partial}{\partial \beta} (P(\lambda_a, \mathbf{a}) + P(\lambda_b, \mathbf{b})) = D \beta, \quad (44)$$

Finally, putting together (37) and (44), the maximizer $\hat{\beta}$ of (14) solves the nonlinear equation

$$\mathcal{X}'(Y - \mu) - D \beta = 0. \quad (45)$$

The estimate of β is computed using iteratively reweighted least squares as follows. Set

$$V = \text{diagonal}(\mu_1, \dots, \mu_n), \quad (46)$$

$$Z = V^{-1}(Y - \mu) + X\beta. \quad (47)$$

Here V is the covariance of the vector of independent Poisson observations Y , and Z are the pseudo-observations. Both V and Z depend on β . The covariance of Z conditional on X and β is V^{-1} .

Initialize $\widehat{\beta} = \widehat{\beta}_o$, where the components of a are $\ln(\bar{Y})$ and $\mathbf{b} = \mathbf{0}$.

1. Let $\widehat{V}_o$, $\widehat{Z}_o$ denote the covariance matrix and pseudo-observations evaluated at the current value of $\widehat{\beta}$.
2. Update the current fitted value according to

$$\widehat{\beta} = [X' \widehat{V}_o X + D]^{-1} X' \widehat{V}_o \widehat{Z}_o. \quad (48)$$

When the inverse does not exist, use the generalized inverse.

3. Iterate steps (1) and (2) to convergence.

Upon convergence, the estimates of the component curves of $\ln$ -relative risk $\widehat{\alpha}(t)$ and $\widehat{\beta}(t)$ are given by $\left[\begin{matrix} JN_a & 0 * \end{matrix} \right] \begin{bmatrix} N_b \\ 0_{(T-Q) \times Q} \end{bmatrix} \widehat{\beta}$ and $\left[0 * JN_a \right] \begin{bmatrix} N_b \\ 0_{(T-Q) \times Q} \end{bmatrix} \widehat{\beta}$, which are of the form $H_a \widehat{Z}_o$ and $H_b \widehat{Z}_o$, respectively. The pointwise standard errors of the estimated component curves are the square root of the diagonal entries of $H_a \widehat{V}_o^{-1} H_a'$ (intercept) and $H_b \widehat{V}_o^{-1} H_b'$ (slope) at the current value of $\widehat{\beta}$.

Now $X\widehat{\beta}$ is of the form HY with $H = X(X'V_oX + D)^{-1}X'V_o$. For testing hypothesis about the significance of $\beta(\cdot)$ using the scaled deviance, the degrees of freedom associated with a particular fit are taken to be $\text{trace}(H)$ from the first step of the iteratively reweighted least squares, when V_o is a multiple of the identity matrix. Standard errors are computed for the $\ln$ -relative risk estimates by exploiting the linearity of $\widehat{\beta}(u)$ in Y .

Obtaining Regularized Estimates of $\alpha(t)$, $\beta(t)$, $s_1(t)$, s_s (temperature) day – of – week Simultaneously

Augmenting the model for the mean of the daily mortality just augments the design matrix X and the block diagonal matrix D as indicated below. For ease of presentation, a common smoothing parameter λ_γ is assumed for estimation of $s_1(t)$, and $s_2(\text{temperature})$. Allowing for different smoothing parameters is a straight forward extension. Represent $s_1(t)$, $s_2(\text{temperature})$ in terms of a B-spline basis with coefficients γ so that now

$$\eta = [XZ] \begin{bmatrix} \beta \\ \gamma \end{bmatrix}. \quad (49)$$

Define D to be the block diagonal matrix:

$$D = \text{diagonal} \left(\lambda_a D_a^t D_a, \lambda_b D_b^t D_b, \lambda_\gamma D_\gamma^t D_\gamma \right), \quad (50)$$

where D_γ is the matrix representation of the second-order difference operator acting on the coefficients γ . The estimates of β and γ are the solutions to

$$(X \ Z)^t (Y - \lambda) - D \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = 0 \quad (51)$$

and are obtained as before by iteratively reweighted least squares with the starting value $\gamma = \mathbf{0}$ for the additional parameters in the augmented model.

Indicator variables for day-of-week with the corresponding coefficients are included in the model for the linear predictor η_t by augmenting the design matrix $(X \ Z)$, the parameter vector, and the penalty matrix D as described before for including the temperature variable. However, because day-of-week is not a smooth function, no smoothing is used in the estimation of the day-of-week effect, that is, the smoothing parameter for day-of-week is set to zero.

References

25. de Boor, C. A Practical Guide to Splines. Springer; New York: 1978.
26. Eubank, RL. Spline Smoothing and Nonparametric Regression. Marcel Dekker; New York: 1988.
27. Ostro B, Broadwin R, Green S, Feng WY, Lipsett M. Fine Particulate Air Pollution and Mortality in Nine California Counties: Results from CALFINE. *Environmental Health Perspectives* 2006;114(1): 29–33. [PubMed: 16393654]
28. Eilers PH, Mark BD. Flexible Smoothing with B-splines and Penalties. *Statistical Science* 1996;11 (2):89–121.
29. Green, PJ.; Silverman, BW. Nonparametric Regression and Generalized Linear Models: A roughness penalty approach. Chapman & Hall/CRC; London: 1994.
30. McCullagh, P.; Nelder, JA. Generalized Linear Models. Chapman & Hall/CRC; London: 1989.

REFERENCES

1. Pope AC, Dockery DW. Health Effects of Fine Air Pollution: Lines that Connect. *J. Air & Waste Manage. Assoc* 2006;56:709–742.
2. Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, Kaufman JD. Long-Term Exposure to Air Pollution and Incidence of Cardiovascular Events in Women. *New England J. Med* 2007;356:447–458. [PubMed: 17267905]
3. Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, Sanat JM. Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. *JAMA* 2006;295(10): 1127–1134. [PubMed: 16522832]
4. Environmental Protection Agency. Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure. National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency; Research Triangle Park: Jul. 2006 NC 27711EPA/600/R-06/063
5. Gold DR, Litonjua A, Schwartz J, Lovett E, Larson A, Nearing B, Allen G, Verrier M, Cherry R, Verrier R. Ambient Pollution and Heart Rate Variability. *Circulation* 2000;101(11):1267–1273. [PubMed: 10725286]
6. Magari SR, Schwartz J, Williams PL, Hauser R, Smith TJ, Christiani DC. The Association Between Personal Measurements of Environmental Exposure to Particulates and Heart Rate Variability. *Epidemiology* 2002;13(3):305–310. [PubMed: 11964932]

7. Peters A, Dockery DW, Muller JE, Mittleman MA. Increased Particulate Air Pollution and the Triggering of Myocardial Infarction. *Circulation* 2001;103:2810–2815. [PubMed: 11401937]
8. Peters A, von Klot S, Heier M, Trentinaglia I, Hormann A, Wichmann HE, Lowel H. Exposure to Traffic and the Onset of Myocardial Infarction. *New England J. Medicine* 2004;351(17):1721–1730.
9. Ramsay, JO.; Silverman, BW. *Functional Data Analysis*. Springer; New York: 1997.
10. Staniswalis JG, Parks NJ, Bader JO, Muñoz Maldonado Y. Temporal Analysis of Airborne Particulate Matter Reveals a Dose-Rate Effect on Mortality in El Paso: Indications of a Differential Toxicity for Different Particle Mixtures. *J. Air & Waste Manage. Assoc* 2005;55:893–902.
11. Schwartz J. The Distributed Lag Between Air Pollution and Daily Deaths. *Epidemiology* 2000;11:320–326. [PubMed: 10784251]
12. Zanobetti A, Wand MP, Schwartz J, Ryan LM. Generalized Additive Distributed Lag Models: Quantifying Mortality Displacement. *Biostatistics* 2000;1(3):279–292. [PubMed: 12933509]
13. Malfait N, Ramsay JO. The Historical Functional linear Model *The Canadian Journal of Statistics* 2003;31(2):115–128.
14. Yang, H. M.S. Thesis. University of Texas; El Paso: 2005. A Further Study of the Relationship Between PM10 Level and Daily Mortality in El Paso, Texas Using a Historical Functional Linear Model. AAT 1430952
15. Franklin M, Zeka A, Schwartz J. Association Between PM2.5 and All-Cause and Specific-Cause Mortality in 27 Communities. *Journal of Exposure Science and Environmental Epidemiology* 2007;17(3):279–287. [PubMed: 17006435]
16. Gamez, J. M.S. Thesis. Department of Civil Engineering, the University of Texas; El Paso: Jul. 2007 Diurnal Variation in Ambient Fine and Ultrafine Particle Concentrations near a Major Highway in El Paso, Texas..
17. Li, WW.; Gamez, J.; Baca, DJ.; Olvera, HA.; Garcia, JH.; Staniswalis, JG.; Garcia, N.; Garcia, M.; Pingitore, NE., Jr.; Amaya, M.; Kelly, K.; Lighty, J. Investigation of the Number Concentrations of Ultrafine Particles in the Nocturnal PM Peaks. the 3rd International Symposium on Nanotechnology, Occupational and Environmental Health; Academia Sinica, Taipei, Taiwan. August 29-September 1, 2007;
18. Nemmar A, Hoet PHM, Vanquickenborne B, Dinsdale D, Thomeer M, Hoylaerts MF, Vanbilloen H, Mortelmans L, Nemery B. Passage of Inhaled Particles Into the Blood Circulation in Humans. *Circulation* 2002;105:411–414. [PubMed: 11815420]
19. Salvi S, Blomberg A, Rudell B, Kelly F, Sandstrom T, Holgate ST, Frew A. Acute Inflammatory Responses in the Airways and Peripheral Blood After Short-Term Exposure to Diesel Exhaust in Healthy Human Volunteers. *Am. J. Respir. Crit. Care Med* 1999;159:702–709. [PubMed: 10051240]
20. Rosenthal FS, Carney FS, Olinger ML. Out-of-Hospital Cardiac Arrest and Airborne Fine Particulate Matter: A Case-Crossover Analysis of Emergency Medical Services Data in Indianapolis, Indiana. *Environmental Health Perspectives* 2008;116(5):631–636. [PubMed: 18470283]
21. Massy WF. Principal Components Regression in Exploratory Statistical Research. *J. of the American Statistical Assoc* 1965;60:234–254.
22. Li WW, Orquiz R, Pingitore NE Jr. Garcia JE, Espino TT, Gardea-Torresdey J, Chow J, Watson JG. Analysis of Temporal and Spatial Dichotomous PM Air Samples in the El Paso-Cd. Juarez Air Quality Basin. *J. Air & Waste Manage. Assoc* 2001;51:1511–1560.
23. Li WW, Cardenas N, Walton J, Trujillo D, Morales H, Arimoto R. PM Source Identification at Sunland Park, New Mexico Using a Simple Heuristic Meteorological and Chemical Analysis. *J. Air & Waste Manage. Assoc* 2005;55:352–364.
24. U.S. Environmental Protection Agency (U.S.EPA). Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency; Research Triangle Park: Dec. 2005 NC 27711EPA-452/R-05-005a

Figure 1. Boxplot (quartiles and 95th percentile) of PM_{2.5} on log₁₀ Scale by Hour of Day for 2000–2005, El Paso.

PM_{2.5} 1st PC Yrs 00-05

Figure 2.
First Harmonic from Principal Component Analysis of PM_{2.5}; Years 2000–2005 for El Paso, TX.

Figure 3.

Estimates of the Component Curves in Equation (11), Including Lags of Up to $\tau=4,6,8,10$, and 12; Years 2000–2005 for El Paso, TX. The placement of the knots are indicated by the tick-marks. The shapes of all of the component functions are consistent across all values of τ . Simulations¹⁴ suggest that late truncation is preferred over early truncation to reduce bias, the value $\tau = 8$ was selected for prediction of mortality.

p-value= 0.0011 df= 5 for significance of PM_{2.5}

Figure 4.

Estimates of the Component Curves in Equation (11). The nonparametric estimates (solid lines) including lags of up to $\tau=8$, displayed with ± 2 pointwise standard errors (dashed lines). (a) Mortality tends to be highest in January, decreasing in time until a minimum at June/July, then increasing for the rest of the year without attaining the global maximum attained in January; (b) The fitted slope function has a global maxima between .5–3 days, suggesting that the association between daily mortality and PM_{2.5} is highest for a .5–3 day lag. The pointwise confidence intervals do not contain 0 for lags within roughly .5–3 days, signifying that PM_{2.5} is a significant predictor of daily mortality; (c) On average, daily mortality is increasing in time, the rate of increase was lower over the years 2002–2004; (d) Mortality tends to decrease as temperature increases. This is consistent with (a), which shows that June/July on average have fewer deaths, whereas December/January have the most deaths. The estimate of smooth dependence on temperature may be viewed as an adjustment of the smooth intercept function in (a); (e) The coefficients of day-of-week suggest that deaths on average significantly increase on Friday.

Figure 5.

Estimates of the Component Curves in Equation (11). The solid lines are the nonparametric estimates of $\beta(u)$, $\tau=3$, with ± 2 pointwise standard errors. The x-axis tic-marks display the location of the knots in time. The dashed lines are the parametric estimates of the coefficients of $\beta(u)$ corresponding to the daily average $PM_{2.5}$ at lags 0, 1, and 2 days with ± 2 pointwise standard errors.

Figure 6.
Summary of hourly total number concentrations at CAMS12.

(a) Fall 2006 (10 nm – 30 nm view)

(b) Spring 2007 (5 nm – 30 nm view)

Figure 7.
Normalized diurnal distributions and modes of nanoparticles of sizes 30nm and below at CAMS12: : (a) Fall 2006 season; (b) Spring 2007 season.

(a) Fall 2006

(b) Spring 2007

Figure 8.
Normalized average particle size distributions: (a) Fall 2006 season; (b) Spring 2007 season.

Table 1
Summary Statistics for Temperature and PM_{2.5}: mean, sd, min, quartiles, max

Variable	mean	sd	min	P ₂₅	median (P ₅₀)	P ₇₅ max
Daily Average Temp (°F)	67	15	26	54	69	8094
Daily Average PM _{2.5} (µg/m ³)	9	6	.5	6	8	1199
Hourly PM _{2.5} (µg/m ³): time interval						
[0, 1)	10	17	0	4	6	11580
[1, 2)	8	10	0	3	6	10155
[2, 3)	8	8	0	3	6	9121
[3, 4)	7	7	0	3	5	991
[4, 5)	7	7	0	4	6	9161
[5, 6)	8	7	0	4	7	10137
[6, 7)	9	7	.1	5	8	1293
[7, 8)	10	6	.3	6	8	1277
[8, 9)	9	6	0	5	8	1157
[9, 10)	8	6	0	4	7	1048
[10, 11)	7	7	0	3	6	9101
[11, 12)	6	11	0	2	5	8399
[12, 13)	6	13	0	2	4	7353
[13, 14)	6	14	0	2	4	7265
[14, 15)	7	17	0	2	4	7396
[15, 16)	7	21	0	2	4	7636
[16, 17)	8	19	0	3	4	8405
[17, 18)	9	17	0	3	6	10417
[18, 19)	12	16	0	5	8	13270
[19, 20)	15	17	0	6	10	18180
[20, 21)	15	15	0	6	10	20189
[21, 22)	13	16	0	5	9	17347
[22, 23)	12	16	0	4	8	14417
[23, 24)	10	14	0	4	7	12287

Table 2

Predicted Effect of a $10\mu\text{g}/\text{m}^3$ Increase in Hourly $\text{PM}_{2.5}$ Across One Day

Model terms for $\text{PM}_{2.5}$	Relative Risk (Approximate 95% Confidence Interval) for same day	Relative Risk (Approximate 95% Confidence Interval)		
		1 day lag	2 days lag	3 days lag
hourly; equation (2) $\tau = 8$	1.017 (1.001,1.034)	1.024 (1.012,1.037)	1.016 (1.006, 1.027)	1.008 (.998,1.018)
daily average; equation (8)	1.018 (.997,1.039)	1.024 (1.004,1.046)	1.023 (1.003,1.044)	