TECHNICAL PAPER

ISSN 1047-3289 *J. Air & Waste Manage. Assoc.* **55**:893–902

Temporal Analysis of Airborne Particulate Matter Reveals a Dose-Rate Effect on Mortality in El Paso: Indications of Differential Toxicity for Different Particle Mixtures

Joan G. Staniswalis
Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX

Norris J. Parks
Center for Environmental Resource Management, University of Texas at El Paso, El Paso, TX

Julia O. Bader
Statistical Consulting Laboratory, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX

Yolanda Muñoz Maldonado
Department of Statistics, Texas A&M University, College Station, TX

ABSTRACT

One of two topics explored is the limitations of the daily average in summarizing pollutant hourly profiles. The daily average of hourly measurements of air pollutant constituents provides continuity with previous studies using monitoring technology that only provided the daily average. However, other summary statistics are needed that make better use of all available information in 24-hr profiles. The daily average reflects the total daily dose, obscuring hourly resolution of the dose rate. Air pollutant exposures with comparable total daily doses may have very different effects when occurring at high levels over a few hours as opposed to low levels over a longer time. Alternative data-based choices for summary statistics are provided using principal component analysis to capture the exposure dose rate, while preserving ease of interpretation. This is demonstrated using the earliest hourly particle concentration data available for El Paso from archived records of particulate matter $(PM)_{10}$. In this way, a significant association between evening PM_{10} exposures and nonaccidental daily mortality is found in El Paso from 1992 to 1995, otherwise missed using the daily average. Secondly, the nature and, hence, effects of particles in the ambient aerosol during El Paso sandstorms is believed different from that of particles present during still-air conditions resulting from atmospheric temperature inversions. To investigate this, wind speed (ws) is used as a surrogate variable to label PM_{10} exposures as Low-ws (primarily fine particles), High-ws (primarily coarse particles), or Mid-ws (a mixture of fine and coarse particles). A High-ws evening is significantly associated with a 10% lower risk of mortality on the succeeding third day, as compared with comparable exposures at Low- or Mid-ws. Although this analysis cannot be used to form firm conclusions because it uses a very small data set, it demonstrates the limitations of the daily average and suggests differential toxicity for different particle compositions.

IMPLICATIONS

The limitation of the daily average in summarizing pollutant hourly profiles is one of two topics of this paper. We show that the effects of airborne PM on daily mortality can be underestimated, because the daily average describes chronic exposures but does not capture information about acute exposures, that is, large exposures occurring over short periods of time. A principal component data analysis is shown to be useful for characterizing hourly measurements of air pollution constituents. Secondly, it is shown that in El Paso, the risk of PM-induced mortality is higher during still-air inversions than it is during sandstorms.

INTRODUCTION

Particulate matter $(PM)_{10}$ refers to air pollutant particle concentrations in $\mu g/m^3$, particles with median aerodynamic diameter <10 μm. U.S. regional daily averages indicate that PM_{10} exposures in El Paso, Texas, are relatively low compared with the rest of the nation and that associations with daily mortality are not statistically significant (Figures 6, 7, and 9 of this reference).[1] But that picture is incomplete, not telling the whole story. This study addresses whether daily mortality is associated with the temporal dynamics of PM_{10} particle concentration occurring within a 24-hr period. Another contribution of our article is a study of the associations between daily mortality and exposures to particles inherent to still-air inversions versus sandstorms.

We and others[2–4] have noted a ubiquitous evening peak in hourly PM_{10} concentrations during still-air conditions (wind speed ≤6 mph, 2.7 m/sec) at various monitoring sites in El Paso over the last 12 years and more recently in respirable $PM_{2.5}$ (particles with median aerodynamic diameter <2.5 μm). We are concerned about the health effects of this dramatic change in atmospheric particle concentration over time, about a 4-fold increase over a few hours, regularly occurring during winter evenings. $PM_{2.5}$ is ~25% of the total PM_{10} concentration in El Paso, and peaks in PM_{10} are associated with peaks in $PM_{2.5}$,[3] which are believed to have greater adverse effects on health than the coarse particles. Elevated concentrations of fine particles over a 2-hr period are believed to transiently increase the risk of myocardial infarction for several hours, as well as for several days after exposure.[5] There is evidence that ultrafine particles diffuse rapidly into the systemic circulation and in this way perhaps exert direct effects on the heart and vessels.[6] As a consequence, the daily average might underestimate the association between air pollution and acute cardiovascular events.[5] Other studies have demonstrated that at high ambient concentrations, acute short-term (1-hr) exposure to diesel engine exhaust produces a well-defined and marked systemic and pulmonary inflammatory response in healthy human volunteers.[7,8]

A fundamental axiom of pharmacology and toxicology states that the rate of administration of a biologically active material is an important factor influencing the response. For a given dosage, an acute administration may result in death, whereas the same dosage administered chronically may only produce a transient toxic or limited response. Within 1 day, atmospheric PM_{10} levels in El Paso often display marked peaks that may lead to very high acute exposures. Summarizing the 24 PM_{10} observations within a day by the daily mean, however, throws this information away. In this study, large exposures over short time periods within a day are described by considering data-based choices of alternate summary statistics for the 24-hr PM_{10} profiles. These are included in a predictive model for daily mortality, thereby providing a missing piece of the story.

The second contribution of this article is in allowing for an interaction between wind speed and PM_{10} concentration in the predictive model for daily mortality. PM_{10} concentrations in El Paso have been shown to peak under both low and high wind speed conditions (Figure 6 of the reference),[2] and the chemical composition of the PM is believed to differ under low and high wind speed conditions.[2] Casual visual inspection of the filters in the PM_{10} monitors supports this in that the filters tend to be "gray" under conditions of still-air associated with atmospheric inversions that trap visible urban air pollution, whereas the filters tend to be "brown" under conditions of high wind speed that stir up loose sand in our desert community. Previously deposited contaminants, such as arsenic and lead, presumed to originate from area smelter and refinery activity, and the use of leaded gas (sold in the United States until 1985 and in Mexico until 1992), are believed to be resuspended and transported with high winds.[9]

Studies for other locations[1,10,11] have found that variations in concentrations of daily average PM_{10} in ambient air are associated with daily variations in nonaccidental mortality after accounting for meteorological variables and temporal trends. Schwartz et al.[12,13] reported that fine particles (<2.5 μm), not coarse particles, were associated with daily deaths. Smith et al.[14] concluded that the coarse particles (2.5–10 μm) effects can by no means be neglected. The associations between PM_{10} and daily mortality have been widely interpreted as reflecting the effects of particulate air pollution on persons who have heightened susceptibility because of chronic heart and lung diseases. Using a discrete Fourier transform of the time series given by the daily average of pollutants over an 8-yr period, Dominici et al.[15] found evidence that the associations between PM and daily mortality are not only because of an advance in the timing of the death of a few days for frail individuals.

In places such as the Paso del Norte Airshed, where the level and type of PM_{10} exposure can vary greatly throughout a given day, reduction of the hourly PM_{10} measurements using the daily mean results in a loss of information to the extent that average PM_{10} is not a statistically significant predictor of mortality. In this article, the 24 hourly PM_{10} observations are viewed as the result of sampling a daily PM_{10} curve defined on the entire interval (0, 24) to which techniques from functional data analysis are applied. In our judgment, the daily patterns of variation of PM_{10} over time are best captured by the method of principal component analysis (PCA) of curves. A PCA of the curves is used to obtain better summary statistics than the daily mean that can be used as linear predictors in the log-linear model for daily mortality. The principal component (PC) scores allow for reduction of the 24-hr PM_{10} data while differentiating between days such as October 13, 1992, June 6, 1993, and October 22, 1995, which have similar daily averages (84.6, 99.5, and 75 $\mu g/m^3$, respectively) but very different 24-hr profiles (see Figure 1). This article reports on the first application of this technique to hourly air quality observations.

Chemical composition and size distribution of inhalable PM in El Paso county are not available for the time period of this study, so wind speed is used as a surrogate variable for the PM_{10} chemical makeup in El Paso. High wind speed is used to indicate the predominance of coarse particles (although it has been shown to have a $PM_{2.5}$ component[2]), whereas low wind speed indicates the predominance of urban air pollution from combustion. This use of wind speed as a surrogate variable is supported by observations of Noble et al.[4] This is additionally justified in the "Methods" section.

DATA

The study was restricted to the 1992–1995 4-year period, because it had the most complete records of hourly PM_{10}. The data consisting of PM_{10} levels measured using beta gauges were obtained from the U.S. Environmental Protection Agency Aerometric Information Retrieval System (U.S. EPA AIRES) for El Paso County's centrally located Chamizal National Memorial Site, a mixed residential, semi-industrial area. The Chamizal site is located in south El Paso, west of the Cordova International Bridge, east of downtown El Paso, and 1 km from the Bridge of the Americas into Ciudad Juarez. This choice is supported by Noble et al.:[4] "the Chamizal site was more representative of the total El Paso urban area because of the unimpeded airflow." Although several other locations had monitors collecting hourly PM_{10} measurements, only the Chamizal site had been validated by the Texas Natural Resource Conservation Commission and transferred to EPA AIRES at the time this work was initiated.

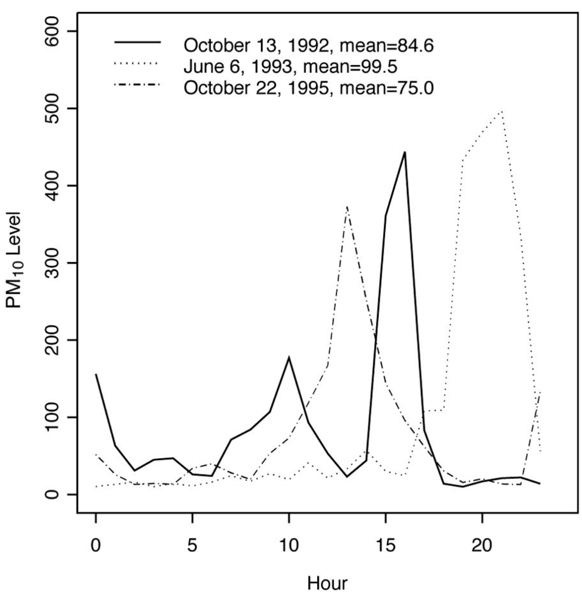

Figure 1. PM_{10} profiles for three days with similar daily averages.

Complete mortality data corresponding to all natural deaths in El Paso County were obtained from the Texas Department of Health for the 1461 days in the 4-yr interval 1992–1995. Natural deaths were identified with ICD-9 codes (International Classification of Diseases codes in effects for 1979–1998) <800. Twenty-four hourly measurements of wind speed, dew point, temperature and PM_{10} were obtained for use as predictors of daily mortality. These weather data were obtained from the National Climatic Data Center of the National Oceanic and Atmospheric Administration for the El Paso Airport station, located in northeast El Paso ~7 km from the Chamizal PM monitoring site.

Each day within the study period contributes one observation of total mortality and 24 hourly observations for each of several variables: wind speed, dew point, temperature, and PM_{10}. We view the 24 hourly observations collected on any given day as the result of sampling a stochastic process contaminated with white noise. The stochastic process is indexed by time (hours) on the interval (0, 24). The discrete Fourier series representation of the 24 hourly observations can be obtained using an orthonormal basis of 24 complex exponential functions that can be expressed in terms of sine and cosine functions. Trigonometric series consisting of sine and cosine terms for 23 frequencies and a constant term were fit to each of the daily 24 observations of wind speed, dew point, temperature, and PM_{10} to reconstruct a 24-hr profile on the entire interval (0, 24) from the discrete set of observations on the regular grid. The 24-hr profile on the entire interval (0, 24) interpolates the 24 observations for a given variable. This number of frequencies was chosen so that the interpolating functions closely agreed with the raw data at the sampling points. Information for a variable was considered missing on a given day whenever there existed a data gap of 5 or more hours. Data gaps of <5 hr were filled in by the trigonometric series. Missing data for PM_{10} reduced the number of days available for analysis from 1461 over the 4-yr interval to 1212 days. Missing data for dew point, temperature, and wind speed, together with the missing days for PM_{10}, additionally reduced the number of days with complete data to only 997 days, ~70% days with complete data. The daily average of PM_{10} ranged from 0.2–133.4 $\mu g/m^3$. The number of daily deaths averaged 8.5 with a min of 1 and max of 21. The population of El Paso County was ~620,000 in 1992. By 1995, the population had increased by roughly 8% to 670,000.

METHODS

Regression Analysis

Following Schwartz,[16] a log-linear model capturing the seasonal and weather dependence of daily mortality was constructed. The first step was to fit a baseline regression model in time that included 24 sine and 24 cosine terms to cover all cycles with periods from 1 month to 24 months to model seasonal variations in daily mortality. An annual dummy variable, linear and quadratic time trends were included to account for changes in the population size, as well as dummy variables for the day of week. The fitted baseline model was indistinguishable from a cubic smoothing spline function in time fit to the data using the S-PLUS software.

Next, the 24 hourly observations of the weather variables, dew point and temperature, were summarized with the daily average. A log-linear model was built from the baseline model by considering the following variables for the concurrent day and with lags up to 3 days: standardized average dew point; standardized average temperature; dummy variables for hot days, set at the ninety-fifth percentile of average temperature; cold days, set at the fifth percentile of average temperature; humid days, set at the ninety-fifth percentile of average dew point; and hot and humid days. The backward variable selection procedure was generous in that it kept weather variables in the model that were significant at the 0.2 level of significance, whereas forcing all terms from the baseline model. Poisson regression implemented in Statistical Analysis System's (SAS) Proc Genmod was used repeatedly for the backward variable selection. A plot of the smoothed squared residuals versus the predicted deaths suggested a homoscedastic variance structure. The lack of heteroscedasticity in these residuals could be because of the small range of the number of daily deaths in El Paso County; 90% of the daily death counts were between 4 and 14.

The final step in building the regression model considered summary measures of PM_{10} as predictors of daily mortality. The next section describes computation of the PC scores as summary statistics for the 24 hourly PM_{10} measurements.

Principal Component Analysis (PCA)

Let n denote the number of days with complete data for the time period of this study. The daily PM_{10} profiles $\{X_i(t)\}_{i=1}^{n}$ are modeled as a random sample, that is, n independent identically distributed realizations of $X(t)$ a smooth stochastic process contaminated with white noise and indexed by $t \in (0, 24)$. PCA of curves uses the sample $\{X_i(t)\}_{i=1}^{n}$ to estimate the principal modes of variation of $X(t)$ about the average hourly profile $\mu(t) = E[X(t)]$. It is based on the Karhunen-Loeve expansion for the stochastic process $X(t)$ and may be viewed as an extension of PCA of a random vector that uses a spectral analysis of the covariance. Castro et al.[17] explain well the connection between PCA of curve data and the more familiar PCA of vector data.[18] The approach of PCA for curve data is not new and has been successfully used in the context of biomedical, engineering, and oceanographic research investigations.[19–22] PCA may be used both for understanding $X(t)$, a smooth random curve, or $X(t)$, a smooth random curve contaminated with white noise.

PCA produces an approximation to $X(t)$ in terms of a few real-valued functions, say $\Psi_1(t), \ldots, \Psi_K(t)$:

$$X(t) \approx \mu(t) + \sum_{k=1}^{K} A_k \Psi_k(t) \tag{1}$$

where the A_k are uncorrelated random variables with zero means and variances $E[A_k^2] = \lambda_k$. PCA is a data-based method for choosing both the functions $\Psi_1(t), \ldots, \Psi_K(t)$, called the PC directions or PC harmonics, and for choosing K the number of terms in the approximation. The PC scores $A_1, \ldots, A_K$ are the random, real-valued coefficients in the linear approximation $\mu(t) + \sum_{k=1}^{K} A_k\Psi_k(t)$ that serve to provide the "best" approximation to the $X(t)$ in terms of $\Psi_1(t), \ldots, \Psi_K(t)$. The PC harmonics are assumed to be smooth functions in t and ordered so that $\lambda_1 \geq \ldots \geq \lambda_K$. Whenever $X(t)$ is contaminated with white noise, the PCA can incorporate nonparametric smoothing[19,23] to obtain smooth estimates of $\Psi_1(t), \ldots, \Psi_K(t)$.

The interpolating trigonometric series expansion used as a preprocessing step to fill in missing data provides an approximation to $X(t)$, but the trigonometric functions are not necessarily the most parsimonious for capturing the important features of $X(t)$. PCA in turn does provide the most parsimonious approximation of $X(t)$ in terms of a complete orthonormal collection of functions $\Psi_1(t), \ldots, \Psi_k(t)$.

The use of the PC scores $A_1, \ldots, A_K$ in place of the 24-hr average can bring greater power for detecting associations between pollutant and health effects. For example, the most prominent feature in the daily PM_{10} profiles is a local peak near 8:00 p.m., which varies in magnitude from day to day. We will see that the PC analysis reports this as the PC direction of greatest variation from day-to-day, accounting for 40% of the total variation in daily PM_{10} profiles, whereas the daily average only accounts for 28% of the total daily variation. A confidence interval for the slope parameter of PM_{10} as a predictor of daily mortality in an ordinary least squares regression would be 84% narrower ($\sqrt{.28/.40} = .84$) based on an analysis using the first PC direction rather than the daily mean.

Greater power is not realized whenever the first PC direction $\psi_1(t)$ of the 24-hr profile is simply a constant. The first PC direction of a pollutant is a constant whenever the primary source of variation from day to day is in the level of exposure and not in the shape of the profile. A constant $\psi_1(t)$ is interpreted to indicate that the greatest source of day-to-day variation in the daily profiles is described by an overall shift in level from the mean profile, $\mu(t) \pm$ constant, and that the shape of the profile is relatively stable from day to day, as compared with the overall level. Whenever the first PC direction of a pollutant profile is merely a constant, the first PC score corresponds to the daily mean (actually, daily mean deviation about $\mu[t]$).

Wind Speed as a Surrogate Variable for the Type of PM_{10}

Still-air conditions (wind speed ≤6 mph, 2.7 m/sec) often occur during the winter as a result of atmospheric temperature inversions, which in turn trap urban air pollution. Under still-air conditions, particles of the respirable ambient aerosol ($PM_{2.5}$ or less) have recently been demonstrated to be a mix of crystalline and anthropogenic organic materials.[24,25] Extremely high wind speeds (sandstorms; wind speed >17 mph, 7.6 m/sec) occur periodically, typically in the fall and spring, blowing out urban pollution and introducing high amounts of larger entrained crustal particles from unprotected surfaces. These include geologic materials from the surrounding desert and unpaved roads, as well as trace elements from resuspension of deposited metals previously emitted from several regional point sources.[9]

Wind speed is used as a surrogate variable for the PM_{10} chemical makeup in El Paso. This use of wind speed as a surrogate variable for the type of PM is supported by observations of Noble et al.[4] during a 21-day period in El Paso, winter 1999. They observed that PM_{10} tended to peak in the afternoon and evenings, and this was likely because of the increased wind speed suspending dust in the air. Li et al.[2] report that hours with low or extremely high wind speeds tend to yield higher PM concentrations than hours with light to moderate wind speeds. Low wind speed conditions are believed to correspond to exposures that are primarily made up of a fine PM of urban type, whereas high wind speed is believed to correspond to coarse PM and fine PM from resuspended fugitive dust. Middle levels of wind speed ought to correspond to a mixture of these. It is a limitation of this study that this conjecture has not been substantiated with data.

RESULTS

Principal Component Analysis of PM_{10}

PCA was applied to the PM_{10} interpolating functions by year without smoothing to obtain an ordered set of PC harmonics for each year.[26] Consistency of the patterns in the PC harmonics across all of the years was verified by visual inspection. However, for the PC scores to be commensurate across years, it is necessary that the PC scores for all years ultimately be computed using the same PC harmonics and hourly mean PM_{10} profile. This was accomplished by pooling the PC harmonics across the years by pointwise averaging, normalizing so that the squared integral was one, and then recomputing the PC scores using the pooled PC harmonics. This pooling of the PC harmonics by pointwise averaging across the study years removed much of the roughness present in the individual PC harmonics. The hourly mean PM_{10} profile of the interpolating functions is plotted in Figure 2.

The first two PC harmonics of PM_{10} together explained ~55% of the day-to-day variation in the daily profiles (40, 15, 10, and 7%, respectively, for each of the first four directions). The first harmonic (Figure 3) captures the shape of the evening local peak between 19:00 and 23:00 hr; the first PC score (PC1) reflects the magnitude of the local peak. Large positive PC1 scores correspond to days with large local peaks of PM_{10} exposure around 8 p.m., whereas negative PC1 scores correspond to days with the 8 p.m. peak lower than the yearly average of this evening peak. The first harmonic also captures a minor local peak in PM_{10} occurring between 5:00 and 10:00 hr. In a study of the Paso del Norte airshed air quality between August 1999 and March 2000, Li et al.[2] attributed the evening peak in PM_{10} hourly measurements to the formation of radiation-inversions and to times when wood burning and home cooking prevailed in the air basin. They attribute the minor morning peak to ground-based inversions combined with the effects of morning traffic.

The second harmonic (Figure 3) provides a contrast between daytime and the evening exposure levels. If the

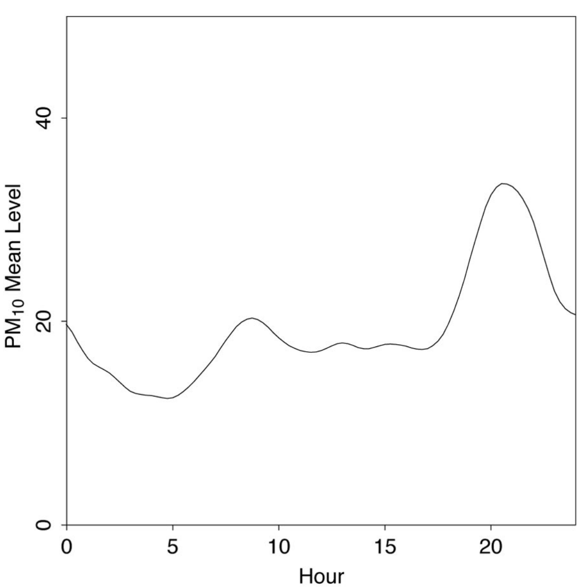

Figure 2. PM_{10} hourly mean 1992–1995.

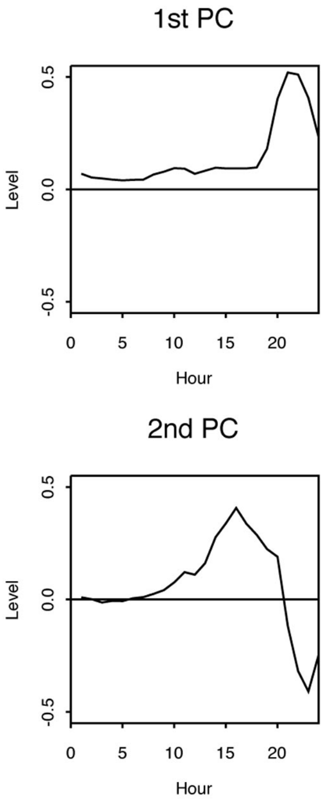

Figure 3. Pooled PC harmonics 1992–1995.

second PC score (PC2) is positive (negative), then the cumulative deviation-from-the-mean daytime PM_{10} exposure between 10:00 and 18:00 hr is higher (lower) than that for the evening exposure between 19:00 and 23:00 hr.

Return to Figure 1, which displays the PM_{10} profile for 3 days with similar daily PM_{10} averages. For these days, the PC scores based on the pooled PC harmonics are provided in Table 1. PC1 is largest for June 6, the day with the largest evening peak around 8:00 p.m. PC2 contrasts PM_{10} deviations-from-the-mean between 10:00 a.m. and 6:00 p.m. with those in the evening. PC2 is positive for the 2 days in October with large PM_{10} exposures between 10:00 a.m. and 6:00 p.m. and low exposures around 8:00 p.m. PC2 is negative for June 6, which had a low exposure between 10:00 a.m. and 6:00 p.m. but a very high PM_{10} level around 8:00 p.m.

Table 1. PC scores for three days with similar daily PM_{10} averages.

Date	PC1	PC2	Daily mean PM_{10} (μg/m^3)
October 13, 1992	81.688	358.243	84.6
June 6, 1993	772.813	−181.797	99.5
October 22, 1995	129.488	301.796	75.0

Model I: Regression Analysis with PM_{10} PCA Scores

The first two PC scores for the concurrent day and with lags 1:3 were included in the log-linear model using backward variable selection while forcing inclusion of the variables from the baseline and weather model. PC1 with lag 3 was the only significant term at the 0.05 level of significance; the scaled deviance did not indicate lack-of-fit.

Next, the evening exposure associated with PC1 was characterized as primarily due to either urban pollution or blowing sand, by examining the surrogate variable of wind speed (ws) between 19:00 and 23:00 hr. If the ws hourly profile had no more than 1 hourly reading >6 mph (2.7 m/sec), then evening was labeled a "Low-ws" exposure. If the the ws hourly profile had at least 2 hourly readings ≥17 mph (7.6 m/sec), then evening was labeled a "High-ws" exposure. If the evening exposure was neither Low or High, it was labeled a "Mid-ws" exposure. In this way, 110 High-ws, 114 Low-ws, and 773 Mid-ws evening exposures were identified.

PC1 for the concurrent day and with lags 1:3 were included in the Poisson regression model using backward variable selection, while forcing inclusion of the variables from the baseline and weather model. Indicator variables of High-ws and Low-ws for the concurrent day and with lags 1:3 were also included in the backward variable selection both as additive effects and as interactions with the PC1 scores. PC1 with lag 3, as well as the indicator of High-ws at a lag of 3 were the only significant terms at the 0.05 level of significance. Table 2 (Model I) provides the final parameter estimates, standard errors, and *P* values for the test of significance based on a Poisson regression using SAS Proc Genmod. The test of significance using the Wald and the likelihood ratio tests yielded the same rounded *P* values.

Table 2. Model I significant parameter estimates from a Poisson regression using the raw PC1 scores (n = 997).

Model I Parameters	Genmod Estimate (SE)	*P* Value	Relative Risk Three Days Later
Dew point, lag 1	−0.04 (0.02)	0.04	
Cold days, lag 3	0.19 (0.07)	0.005	
Humid days, lag 3	0.15 (0.06)	0.02	
PC1, lag 3	0.0003 (0.0002)	0.04	1.0206
Indicator of evening-high-ws, lag 3	−0.104 (0.04)	0.01	0.90

Note: Reported here is the relative risk three days later associated with a 68 unit change in the PC1 score, which corresponds to a 10 μg/m^3 change in the daily mean.

The parameter estimates and standard errors for PC1 with a lag of 3 were comparable with and without the High-ws indicator in the fitted regression model. The top 1, 5, and 10% largest PC1 scores were not influential in that upon their removal, the parameter estimates and standard errors were stable. However, upon removing 10% of the largest PC1 scores, PC1 with lag 3 was no longer significant ($P = 0.27$), but the indicator for High-ws was still significant ($P = 0.01$).

The interaction between PC1 and the indicator for High-ws was not statistically significant. This implies that the association between PM_{10} and mortality is not significantly different on Low-ws/Mid-ws as compared with High-ws days. However, the indicator for High-ws was significant when entered into the model as an additive effect. The High-ws evening exposure is associated with a 10% lower risk of mortality as compared with comparable levels of PM_{10} exposure during conditions of Low-ws or Mid-ws. This 10% lower risk of mortality 3 days later associated with High-ws conditions is constant across all PC1 scores.

The change in mortality associated with a 10 $\mu g/m^3$ in the magnitude of the daily average PM_{10} is typically reported in the literature. Because the mean of the first PC direction is 0.147, this would correspond to a change in the PC1 score of ~10/0.147 = 68 units. An increase of 68 units in the PC1 score corresponds to roughly a change of 34 $\mu g/m^3$ per hour in the magnitude of the evening PM_{10} peak (Figure 3). This change corresponds to a 2.06% increase in mortality 3 days later.

Model II: Regression Analysis with Transformed PC1 Scores

The distribution of the PC1 score has a long positive tail and a few outliers. Shifted-ln PC1 = ln(PC1 + 81) served to transform the PC1 scores to a more symmetric distribution without outliers. The shifted-ln PC1 score with a lag of 3 days, baseline, and weather variables were all forced into the regression model for mortality while using backward variable selection on the indicator variable: High-ws and Low-ws, with lag 3, included both as an additive effect and as an interaction with the shifted-ln PC1 score, lag 3. The indicator for High-ws with a lag of 3 days was significant as an interaction with shifted-ln PC1 but not as an additive effect. Table 3 (Model II) provides the final parameter estimates. There was no indication of lack-of-fit using the scaled deviance.

This analysis using the shifted-ln PC1 scores agrees with the analysis reported above using the raw PC1 scores in that High-ws evenings are associated with a lower risk of mortality 3 days later as compared with Low-ws or Mid-ws evenings. However, High-ws indicator is an additive effect in Model I with PC1 but interacts with shifted-ln PC1 in Model II. Recall that an increase of 10 $\mu g/m^3$ in the magnitude of the daily average PM_{10} corresponds to an increase of 68 units in the PC1 score. The relative risk of mortality 3 days later associated with a 68-unit increase in the PC1 score is:

Table 3. Model II significant parameter estimates from a Poisson regression using the shifted-ln PC1 scores ($n = 997$).

Model II Parameters	Genmod Estimate (SE)	*P* Value
Dew point, lag 1	−0.05 (0.02)	0.04
Cold days, lag 3	0.19 (0.07)	0.004
Humid days, lag 3	0.16 (0.06)	0.02
Shifted-ln PC1, lag 3	0.033 (0.015)	0.03
Interaction, lag 3 (shifted-ln PC1 with indicator of high-ws)	−0.026 (0.01)	0.01

$$\frac{\exp(\hat{\beta}\ln(PC1+81+68))}{\exp(\hat{\beta}\ln(PC1+81))}=$$

$$[\exp\hat{\beta}]^{\ln(PC1\,+\,81\,+\,68)\,-\,\ln(PC1\,+\,81)} \qquad (2)$$

and will now depend upon: (1) the ws conditions, because of the significance of an interaction with ln-shifted PC1, and (2) the shifted-ln PC1 score before the 68-unit increase. This yields a 0.36–15% increase in mortality 3 days later under Low-ws or Mid-ws conditions ($\hat{\beta} = .033$) at the max (PC1 = 510.1) and min (PC1 = −80) values of the PC1 scores, respectively. Under High-ws conditions ($\hat{\beta} = .033 - .026$), the same range of PC1 scores (510.1 to −80) results in only a 0.076–3.01% increase in mortality 3 days later.

Model III: Regression Analysis with Daily Mean PM_{10}

The daily mean PM_{10} for the concurrent day and with lags 1:3 were included in the log-linear model using backward variable selection while forcing inclusion of the variables from the baseline and weather model. No pollutant terms were significant. Nevertheless, average PM_{10} with a lag of 3 days was forced into the Poisson regression using SAS Proc Genmod (Table 4). An increase of 10 $\mu g/m^3$ in the magnitude of the daily average PM_{10} corresponds to an increase in mortality of ~1.7% 3 days later.

DISCUSSION

For the study of 90 U.S. cities, a 10 $\mu g/m^3$ increase in the daily average PM_{10} corresponds to anywhere from no significant change to ~3% increase in death with a lag of

Table 4. Model III parameter estimates from a Poisson regression using the daily average (n = 997).

Model III Parameters	Genmod Estimate (SE)	*P* Value	Relative Risk Three Days Later
Dew point, lag 1	−0.03 (0.02)	0.12	
Cold days, lag 3	0.19 (0.07)	0.004	
Humid days, lag 3	0.15 (0.06)	0.03	
Average PM_{10}, lag 3	0.0017 (0.001)	0.10	1.017

Note: Reported here is the relative risk three days later associated with a 10 $\mu g/m^3$ change in the daily mean.

1 day (see Figure 7 of the reference).[1] We have demonstrated the limitations of the daily average of PM_{10} as a predictor of daily mortality. A log-linear regression analysis based on the average PM_{10} reports a smaller health effect as compared with a log-linear regression analysis based on the first-PC score. Using a PCA of the 24 hourly measurements, an equivalent 10 $\mu g/m^3$ increase in the daily average PM_{10} corresponds with a 2.06% increase in mortality 3 days later (Model I), whereas an analysis using the daily average predicts an increase of 1.7% 3 days later (Model III). The estimate of the change in the mortality based on the daily mean is ~20% lower than that provided by the analysis based on the raw PC scores. The differences are even greater between the predictions based on ln-shifted PC1 (Model II) versus daily average PM_{10} (Model III).

Figure 4 displays PM_{10} profiles of 3 days for which the difference among the profiles is only in the size of the local evening peak around 8 p.m. These 3 days are now used to compare the relative risk of mortality 3 days later based on the regression results incorporating the daily

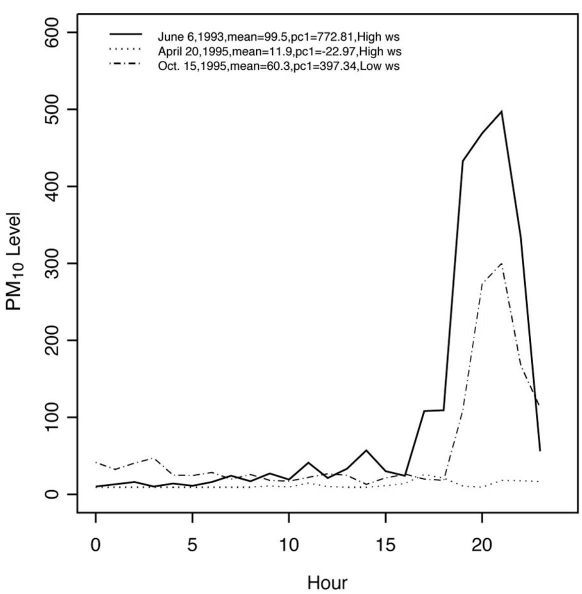

Figure 4. Used for comparison of the relative risk.

Table 5. Relative risk of mortality three days later computed under four models.

Model: Parameter Estimate (SE, *P* Value)	Relative Risk Three Days Later: June 6, 1993 PC1 = 772.81 high-ws average PM_{10} = 99.5	Relative Risk Three Days Later: Oct. 15, 1995 PC1 = 397.34 low-ws average PM_{10} = 60.3
Model III: average PM_{10}, lag 3: 0.0017 (0.001, 0.10)	1.16	1.09
Model with PC1, lag 3: 0.0003 (0.0002, 0.03)	1.27	1.13
Model I: PC1, lag 3 and additive indicator of evening high-ws, lag 3: 0.0003 (0.0002, 0.04); −.104 (0.04, 0.01)	1.27	1.26
Model II: ln(PC1 + 81), lag 3 and multiplicative indicator of evening high-ws, lag 3: 0.033 (0.015, 0.03); −0.026 (0.01, 0.01)	1.02	1.19

Note: The table reports the relative risk of mortality three days later for each of October 15, 1995, and June 6, 1993, relative to April 20, 1995 (average PM_{10} = 11.9, PC1 = −22.97, high-ws).

mean versus the first PC direction (Table 5, first two rows), as if all weather conditions were comparable. The relative risk of mortality 3 days later using the first-PC direction is 1.27 (June 6, 1993) and 1.13 (October 15, 1995), as compared with 1.16 (June 6, 1993) and 1.09 (October 15, 1995) using the PM_{10} daily average. The relative risk is computed for each of the 2 days relative to the third (April 20, 1995). This example serves to demonstrate how studies using average PM_{10} can be underestimating the health effects of PM_{10}.

Note that the reference day, April 20, 1995, had High-ws but low atmospheric PM_{10}, most likely because of some precipitation (0.02 in, 0.05 cm) on the day before. If the indicator of evening High-ws is now taken into consideration as an additive effect, then the relative risk of mortality 3 days later using the first-PC direction remains 1.27 for June 6, 1993, because the reference day is also an evening High-ws day. However, the relative risk of mortality 3 days later adjusted for wind speed is 1.26 for October 15, 1995, because it is an evening Low-ws day (Table 5, the row labeled Model I). Lastly, if the indicator of evening High-ws is entered into the model with shifted-ln PC1 scores as a multiplicative effect, the relative risk of mortality 3 days later is only 1.02 for June 6, 1993, and 1.19 for October 15, 1995 (Table 5, the row labeled Model II) relative to the reference day April 20, 1995. Although the evening PM_{10} peak for June 6, 1993, is ~1.7 times as high as that of October 15, 1995 (Figure 4), about the same or lower risk for mortality 3 days later is reported when adjusting for evening wind speed as an additive or multiplicative effect.

CONCLUSIONS

In summary, an alternate analysis of the health effects of PM_{10} is reported here in which the PM_{10} profiles are data better summarized using principal component scores. In this way, a significant association between PM_{10} and daily mortality is found in El Paso that is otherwise overlooked using the daily average PM_{10}. This analysis cannot be used to form firm conclusions, because it uses a very small data set (one location, one monitoring site, small mortality counts, and only 4 years of data). From information gained in this limited study, when dealing with multiple monitors, it seems prudent to perform a PCA of the daily profiles separately for each monitor. Pooling information across monitors would be warranted only if the principal component directions agree. On a larger geographical scale, inspection of principal component directions may provide insight on how to best combine PM_{10} data from different cities.

It was found that the High-ws evening exposure is associated with a 10% lower risk of mortality as compared with comparable levels of PM_{10} exposure during conditions of Mid-ws or Low-ws. This is consistent with the belief that ultrafine particle exposures have the greater negative impact on health outcomes,[12,13] when considering only a time frame of a few days after exposure. A possible explanation for this is that fine particles have a larger surface area-to-mass ratio than coarse particles, and, therefore, comparable exposures of fine and coarse PM as measured by mass are not comparable when viewed in terms of surface area. It may be that the surface area of the PM exposure and not the mass is the better predictor of daily mortality, because a larger surface area provides a greater opportunity for the pollutant to damage biological tissue.[27,28]

Our findings are also in line with Smith et al.[14] who concluded that the coarse particles effects can by no means be neglected. On the other hand, Schwartz et al.[13] concluded that coarse particles associated with wind-blown dust are not associated with mortality risk by comparing the number of deaths on and after a day of a dust storm with deaths on control days in Spokane, WA. Differences between the reported effects of coarse particles in El Paso and Spokane may be because of differences in PM_{10} chemical composition. The coarse particles in Spokane are mostly natural dust[14] and, thus, ought to be far less toxic than the coarse particles in El Paso which contain resuspension of deposited metals previously emitted from several regional point sources.[9]

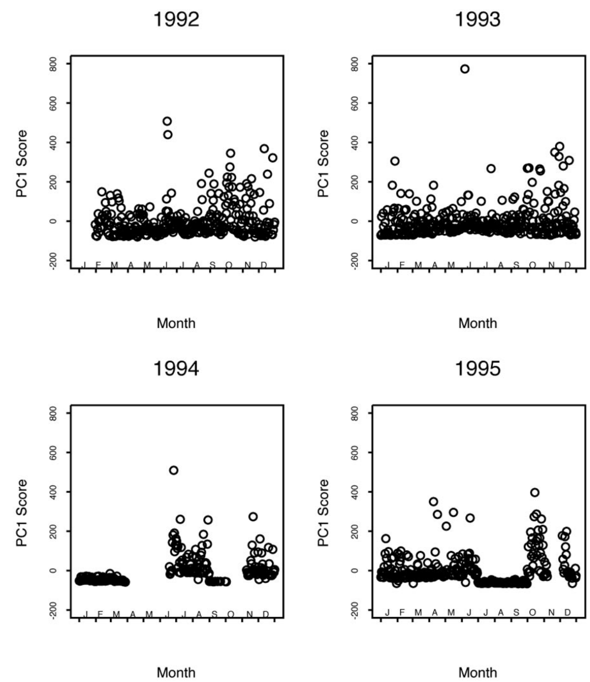

Figure 5. PC1 scores vs. time by year.

A possible next step in this study would be to combine the PC approach with that of Dominici et al.[15] This would involve a discrete Fourier transform of the time series given by the daily PC scores of pollutants (Figure 5, $n = 1212$), rather than the daily average, to estimate the long-term effects of pollution levels on daily mortality in El Paso. It would also be interesting to explore functional log-linear models[22] in which the 24 hourly PM_{10} observations without additional reduction are used as the independent variable.

ACKNOWLEDGMENTS

The authors thank Robert W. Gray, University of Texas at El Paso, Dr. James A. VanDerslice, University of Texas, Houston, Health Science Center (formerly), Houston, TX; Dr. James O. Ramsay, McGill University, Montreal, Quebec, Canada for his PCA software written in S-PLUS; and Dr. Timothy Ramsay, McLaughlin Centre for Population Health Risk Assessment, Ottawa, Ontario, Canada, for comments on an earlier draft. Dr. Staniswalis was partially supported by NIH-SCORE 2S06 GM008012; Dr. Bader is partially funded by NIH RCMI 2G12-RR008124. This work would not have been possible without the support of The Center for Border Health Research, El Paso, Texas (CBHR 131/1999).

REFERENCES

1. Dominici, F.; McDermott, A.; Daniels, M.; Zeger, S.L.; Samet, J.M. *Mortality Among Residents of 90 Cities; Revised Analysis of Time-Series Studies of Air Pollution and Health;* Health Effects Institute Research Report, 2003, Special Report, Part II, pp 9–24.
2. Li, W.W.; Orquiz, R.; Garcia, J.H.; Espino, T.T.; Pingitore, N.E.; Gardea-Torresdey, J.; Chow, J.; Watson, J.G. 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,* 1551–1560.
3. Jeon, S.J.; Meuzelaar, H.L.C.; Sheya, S.A.N.; Lighty, J.A.S.; Jarman, W.M.; Kasteler, C.; Sarofim, A.F. Exploratory Studies of PM_{10} Receptor and Source Profiling by gas chromatography/mass spectrometry and Principal Component Analysis of Temporally and Spatially Resolved Ambient Samples; *J. Air & Waste Manage. Assoc.* **2001**, *51,* 766–784.
4. Noble, C.A.; Mukerjee, S.; Gonzales, M.; Rodes, C.E.; Lawless, P.A.; Natarajan, S.; Myers, E.A.; Norris, G.A.; Smith, L.; Ozkaynak, H.; Neas, L.M. Continuous Measurement of Fine and Ultrafine Particulate Matter, Criteria Pollutants and Meteorological Conditions in Urban El Paso, Texas; *Atmos. Environ.* **2003**, *37,* 827–840.
5. Peters, A.; Dockery, D.W.; Muller, J.E.; Mittleman, M.A. Increased Particulate Air Pollution and the Triggering of Myocardial Infarction; *Circulation.* **2001**, *103*, 2810–2815.
6. Nemmar, A.; Hoet, P.H.M.; Vanquickenborne, B.; Dinsdale, D.; Thomeer, M.; Hoylaerts, M.F.; Vanbilloen, H.; Mortelmans, L.; Nemery, B. Passage of Inhaled Particles Into the Blood Circulation in Humans; *Circulation* **2002**, *105,* 411–414.
7. Salvi, S.; Blomberg, A.; Rudell, B.; Kelly, F.; Sandstrom, T.; Holgate, S.T.; 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.
8. Salvi, S.; Nordenhall, C.; Blomberg, A.; Rudell, B.; Pourazar, J.; Kelly, F.J.; Wilson, S.; Sandstrom, T.; Holgate, S.T.; Frew, A.J. Acute Exposure to Diesel Exhaust Increases IL-8 and gasoline range organics-α Production in Healthy Human Airways; *Am. J. Respir. Crit. Care Med.* **2000**, *161*, 550–557.
9. Garcia, J.H.; Li, W.W.; Arimoto, R.; Okrasinski, R.; Greenlee, J.; Walton, J.; Schloesslin, C.; Sage, S. Characterization of Potential Fugitive Dust Sources in the Paso del Norte Region; *Science Total Environ.* **2004**, *325*, 95–112.
10. Samet, J.M.; Zeger, S.L.; Dominici, F.; Curriero, F.; Coursac, I.; Dockery, D.W.; Schwartz, J.; Zanobetti, A. *The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity, Mortality, and Air Pollution in the United States*; Health Effects Institute Research Report 2000, Vol. 94, Part II: pp 5–79 .
11. PM2000: Particulate Matter and Health- The Scientific Basis for Regulatory Decision-Making; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 1073–1268.
12. Schwartz, J.; Dockery D.W.; Neas, L.M. Is Daily Mortality Associated Specifically with Fine Particles?; *J. Air & Waste Manage. Assoc.* **1996**, *46*, 2–14.
13. Schwartz, J.; Norris, G.; Larson, T.; Sheppard, L.; Claiborne, C.; and Koenig, J. Episodes of High Coarse Particle Concentrations Are Not Associated with Increased Mortality; *Environ. Health Perspect.* **1999**, *107*, 339–342.
14. Smith, R.L.; Spitzner, D.; Kim, Y.; Fuentes, M. Threshold Dependence of Mortality Effects for Fine and Coarse Particles in Phoenix, Arizona; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 1367–1379.
15. Dominici, F.; McDermott, A.; Zeger, S.L.; Samet, J.M. Airborne Particulate Matter and Mortality: Timescale Effects in Four US Cities; *Am. J. Epidemiol.* **2003**, *157*, 1055–1065.
16. Schwartz, J. Nonparametric Smoothing in the Analysis of Air Pollution and Respiratory Illness; *Canadian J. Stat.* **1994**, *22*, 471–487.
17. Castro, P.E.; Lawton, W.H.; Sylvestre, E.A. Principal Modes of Variation for Processes with Continuous Sample Curves; *Technometrics.* **1986**, *28*, 329–337.
18. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice Hall: Englewood Cliffs, NJ, 1998.
19. Rice, J.A.; Silverman, B.W. Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves*; J. R. Statist. Soc. B.* **1991**, *53*, 233–243.
20. Service, S.K.; Rice, J.A.; Chavez, F.P. Interactions Among Physical and Biological Variables During the Upwelling Period in Monterey Bay, CA; *Deep Sea Research* **1997**, *45*, 1669–1685.
21. Jones, M.C.; Rice, J.A. Displaying the Important Features of Large Collections of Similar Curves; *Am. Stat.* **1992**, *46*, 140–145.
22. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*. New York: Springer Series in Statistics, 1997.
23. Silverman, B.W. Smoothed Functional Principal Components Analysis by Choice of Norm; *Ann. Stat.* **1996**, *24*, 1–24.
24. Murr, L.E.; Esquivel, E.V.; Bang, J.J. Characterization of Nanostructure Phenomena in Airborne Particulate Aggregates and their Potential for Respiratory Health Effects; *J. Mater. Sci.: Mater. Med.* **2004**, *15*, 1–11.
25. Bang, J.J.; Trillo, E.A.; Murr, L.E. Utilization of Selected Area Electron Diffraction Patterns for Characterization of Air Submicron Particulate Matter Collected by a Thermophoretic Precipitator; *J. Air & Waste Manage. Assoc.* **2003**, *53*, 227–236.
26. Parks, N.J.; Staniswalis, J.G.; Bader J.; Mu~noz, Y. *Association of Particulate Matter Levels with Daily Mortality After Windstorms, Still-Air Inversions and Closely-Timed Combinations.* Report to the Center for Border Health Research, El Paso, Texas 2003. http://www.statlab.utep.edu/sample_projects.htm.
27. Maynard, A.D.; Maynard, R.L. A Derived Association Between Ambient Aerosol Surface Area and Excess Mortality Using Historic Time Series Data; *Atmos. Environ.* **2002**, *36*, 5561–5567.
28. Moshammer, H.; Neuberger, M. The Active Surface of Suspended Particles as a Predictor of Lung Function and Pulmonary Symptoms in Australian School Children; *Atmos. Environ.* **2003**, *37*, 1737–1744.

About the Authors

Joan G. Staniswalis is Professor at The University of Texas at El Paso, Department of Mathematical Sciences. Dr. Norris J. Parks is Adjunct Professor at The University of Texas at El Paso, Department of Civil Engineering. Dr. Julia O. Bader is Staff Statistician of The Statistical Consulting Laboratory, a core facility of the The University of Texas at El Paso, Border Biomedical Research Center. Yolanda Muñoz Maldonado is currently enrolled in the doctoral program at Texas A&M University, Department of Statistics, College Station, Texas. Address correspondence to: Joan G. Staniswalis, University of Texas at El Paso, Department of Mathematical Sciences, 500 West University Avenue, Bell Hall 124, El Paso, TX 79902; phone: +1-915-747-6992; fax: 1+-915-747-5856; e-mail: joan@math.utep.edu.