Exploring relationships between outdoor air particulate-associated polycyclic aromatic hydrocarbon and PM2.5: A case study of benzo(a)pyrene in California metropolitan regions

Agnes B. Lobscheid[a] *, Thomas E. McKone[a,b], and Daniel A. Vallero[c]

[a] Indoor Environment Department, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory MS90R3058, Berkeley, CA 94720, USA

[b] Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA 94720, USA

[c] U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, North Carolina 27711, USA

*Corresponding author. Tel: +1-510-486-6945; Fax: +1-510-486-6658. Email address: ablobscheid@lbl.gov

Abstract

Polycyclic aromatic hydrocarbons (PAHs) and particulate matter (PM) are co-pollutants emitted as by-products of combustion processes. Convincing evidence exists for PAHs as a primary toxic component of fine PM ($PM_{2.5}$). Because $PM_{2.5}$ is listed by the US EPA as a "Criteria Pollutant", it is monitored regularly at sites nationwide. In contrast, very limited data is available on measured ambient air concentrations of PAHs. However, between 1999-2001, ambient air concentrations of $PM_{2.5}$ and benzo(a)pyrene (BaP) are available for California locations. We use multivariate linear regression models (MLRMs) to predict ambient air levels of BaP in four air basins based on reported $PM_{2.5}$ concentrations and spatial, temporal and meteorological variables as variates. We obtain an R^2 ranging from 0.57-0.72 among these basins. Significant variables ($p<0.05$) include the average daily $PM_{2.5}$ concentration, wind speed, temperature and relative humidity, and the coastal distance as well as season, and holiday or weekend. Combining the data from all sites and using only these variables to estimate ambient BaP levels, we obtain an R^2 of 0.55. These R^2-values, combined with analysis of the residual error and cross validation using the PRESS-statistic, demonstrate the potential of our method to estimate reported outdoor air PAH exposure levels in metropolitan regions. These MLRMs provide a first step towards relating outdoor ambient $PM_{2.5}$ and PAH concentrations for epidemiological studies when PAH measurements are unavailable, or limited in spatial coverage, based on publicly available meteorological and $PM_{2.5}$ data.

Keywords: toxic air pollutants, particulate matter, regression models, combustion sources

1. Introduction

Environmental decisions must frequently be made without sufficient data on exposure. In particular, data for toxic air pollutant exposures are usually less reliable than data for criteria pollutants such as ozone and particulate matter (PM). For example, the spatial coverage, intensity, and confidence about PM exposures are much greater than those for polycyclic aromatic hydrocarbons (PAHs). Since it is unlikely that toxic air pollutant data in general, and PAH data specifically, will substantially improve in the next decade, indirect means of estimating exposures to air toxics are needed. Our research demonstrates one such approach.

PAHs and PM are co-pollutants emitted as by-products of combustion processes (Fraser et al. 1998; Guo et al. 2003; Kaupp and McLachlan 2000; Kiss et al. 1998; Leotz-Gartziandia et al. 2000; Vardar and Noll 2003; Hou et al., 2006; Spencer et al., 2006). Convincing evidence exists that PAHs, such as benzo(a)pyrene (BaP), are significant toxic components of the fine fraction of PM with aerodynamic diameters less than 2.5 μm ($PM_{2.5}$) leading to adverse human health risks (Adonis and Gil 2000; Dejmek et al. 2000; Deng et al. 2006; Massolo et al. 2002; Pohjola et al. 2003; Strandell et al. 1994). Major sources of atmospheric emissions of PM2.5 and particulate-bound PAHs are gasoline and diesel powered motor vehicles and residential heating, primarily wood combustion (Bjorseth and Ramdahl 1985; Schauer and Cass 2000; Zheng et al. 2002). Around 7 million metric tons of $PM_{2.5}$ (US EPA 2000) and between 10,000-30,000 metric tons of associated particle-phase PAHs (Baek et al. 1991; US EPA 1998a) are emitted to the atmosphere annually in the US.

The important chemical reaction steps leading to the formation of PAHs and $PM_{2.5}$ involve the oxidation of benzene, the formation of cyclopentadienyl radicals, and reaction with C_3H_3 radicals (Richter and Howard 2000). PAHs can subsequently form soot via hydrogen abstraction and acetylene-addition (Richter and Howard 2000). Larger molecular-weight PAHs, such as BaP, are semivolatile organic compounds, with vapor pressures between 10^{-2} and 10^{-5} kilopascals. They are primarily particle-associated, and have been measured on archived PM2.5 filters (Pleil et al. 2004). BaP and other PAHs partition by adsorbing onto particle surfaces or absorbing into an organic material on the particle. This partitioning has been studied and characterized extensively (Harner and Bidleman 1998; Lohman et al. 2000; Mader and Pankow 2002; Odum et al. 1994; Offenberg and Baker 2002).

Because both PM_{10} and $PM_{2.5}$ are classified as criteria pollutants by the US EPA (1997), they are monitored regularly at sites nationwide. Outdoor air $PM_{2.5}$ concentrations are available through the US EPA's Air Quality Subsystem (AQS; formerly the Aerometric Information Retrieval System or AIRS (US EPA 2004). Since 1999, over a thousand US EPA and affiliated monitoring stations have collected data on ambient air $PM_{2.5}$. Of these, 93 stations are in California. In contrast however, measured outdoor air concentrations of PAHs are extremely limited. Relating outdoor air concentrations of primarily particulate-associated PAHs, such as BaP, to available $PM_{2.5}$ concentrations offers an important opportunity to characterize levels of human exposure where measured outdoor PAH concentrations are lacking. Of the limited studies reported in the literature relating particulate-associated PAHs with PM, the majority report poor ($r < 0.6$) Pearson product-moment correlation coefficients (r) between BaP and coarse PM, or PM_{10} (Lewis et al. 1995; Sakai et al. 2002).

The current literature identifies key temporal and meteorological factors that influence both PAH and $PM_{2.5}$ outdoor air concentrations. This suggests the need to further explore these factors for predictive relationships associating PAH concentrations and $PM_{2.5}$ levels. The photochemical degradation of PAHs is a function of particle composition and meteorological factors, such as humidity (Kamens et al. 1998) and temperature (Dimashki et al. 2001). Furthermore, temporal parameters, such and day of week (weekday/weekend/holiday), can impact PM and PAH levels, because of the prevalence of certain activities, such as residential wood combustion and motor vehicle activity.

Our objectives in this paper are to investigate and evaluate the reliability of daily (24-hour) average ambient air concentrations of $PM_{2.5}$, adjusted for spatial, temporal and daily averaged meteorological parameters, to characterize ambient air particulate-associated PAH, specifically BaP, concentrations.

2. Materials and Methods

Our analysis is based on co-located measurements of outdoor air BaP concentrations and $PM_{2.5}$ concentrations measured throughout California by the California Air Resources Board (CARB), and reported through the AQS. These reported BaP concentrations are measured by HPLC techniques from a portion of PM_{10} filters (CARB 1998), and will henceforth be referred to as BaP_{10}^{rpt}. $PM_{2.5}$ concentrations are determined by gravimetric analysis (CARB 2002). The AQS data set of concentrations reported between January 1999-December 2001 is available for

four air basins in California--the San Francisco Bay (SF Bay), San Joaquin Valley (SJ Valley), South Coast, and San Diego air basins. After 2001, PAHs are not reported for California. Although BaP_{10}^{rpt} concentrations are reported as early as January 1988 for some of these air basins, corresponding $PM_{2.5}$ levels are not recorded before 1999.

We obtain daily (24-hour average) meteorological data corresponding to the AQS sites in each basin from the National Climate Data Center (NCDC/NOAA 2003). Because NCDC meteorological data are not available at a sufficiently close distance for two of the four SF Bay AQS sites: Concord (37.94 latitude and -122.02 longitude), and Fremont (37.54 latitude and –121.96 longitude), we use meteorological data from the Bay Area Air Quality Management District (BAAQMD). **Figure 1** displays the co-located AQS BaP_{10}^{rpt} and $PM_{2.5}$ sites with corresponding publicly available meteorological data.

We first develop initial descriptive statistics of BaP_{10}^{rpt} and $PM_{2.5}$ concentrations in each of the four air basins individually at each site, and for all sites within a basin. These include the per cent non-detects (NDs) and simple Pearson product-moment correlation coefficients (r) between BaP_{10}^{rpt} and $PM_{2.5}$, and the relative mass concentrations of BaP_{10}^{rpt} to $PM_{2.5}$.

Next, we run multivariate linear regression models (MLRMs) in a stepwise fashion, using both Microsoft Excel and S-PLUS (Mathsoft Inc. 2000), to find the "best-fit" MLRM that adjusts the relationship between the BaP concentrations (BaP^{est}) and $PM_{2.5}$ concentrations for spatial, temporal and surface meteorological variables. The specific meteorological and temporal variables we adjust $PM_{2.5}$ concentrations in our MLRM are:

1) average daily (24-hour average) temperature (*T*) [°C and Kelvin, K];
2) average daily (24-hour average) wind speed (*WS*) [m/s];
3) relative humidity (24-hour average) (*RH*) [%];
4) weekend or holiday (binary);
5) season (winter is December-February; spring is March-May; summer is June-August; and autumn is September-November);
6) rain (binary).

We run separate analysis to test for a season binary variable for each of four seasons, and another set using a binary variable for 'winter or fall' only. Additionally, because RH is not reported in the NCDC unedited climate data, we calculate it (Eq.S-1 in Supplementary Material).

Because population densities tend to increase around coastal areas, PM emissions in California tend to rise sharply at the coastline where prevailing wind patterns bring clean ocean air against dense traffic. Thus, we expect some 'aging' with respect to the accumulation of BaP on particles as one moves east in a California air basin (Miguel et al. 2004). We therefore include the westward distance [km] to the coast to account for this potential effect. Lastly, we explore whether there are any site-specific effects on particle composition or other meteorological phenomenon that are not characterized by our other explanatory variables by including a dummy variable for each site within a basin to explain the variation in BaP_{10}^{rpt} levels.

We investigate the relationship between BaP_{10}^{rpt} and $PM_{2.5}$ adjusted for the meteorological and temporal variables based on the following MLRM procedure:

$y_i = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \text{error}$ **(Eq. 1)**

where y_i are the observed BaP_{10}^{rpt} [ng/m3]; β_o is the intercept; β_1 to β_n are the coefficients associated with the continuous and any/all of the binary and categorical variables $x_1, x_2, \ldots, x_n$. Our "best-fit" MLRM has:

1) Normal distribution of residuals with constant variance; and
2) residual errors are independent of the explanatory variables.

To assess the precision of the final "best-fit" MLRM to estimate BaP concentrations in each of the four air basins, we compare the standard error of the estimate (SEE) and validate our models using the Predictive Residual Sum of Squares (PRESS). These statistics allow us to evaluate the reliability and the validity of our MLRMs. The SEE is computed as (Hamburg 1970):

$$SEE\ of\ \hat{y}_i = \sum_{i=1}^{n} \sqrt{\frac{(y_i - \hat{y}_i)^2}{n-2}} \qquad \textbf{(Eq. 2)}$$

where y_i are the observed (BaP_{10}^{rpt}) and $\hat{y}_i$ are the predicted outcomes (BaP^{est} $[ng/m^3]$) , respectively, based on our "best-fit" MLRM; and n is the sample size of corresponding co-located $PM_{2.5}$ and BaP_{10}^{rpt} measurements.

Although a small SEE and an R^2 near 1 are necessary conditions for a valid MLRM, these statistics sometimes fail to fully account for chance correlations. Therefore, we apply the PRESS statistic to assess the validity of the regression models. With cross validation we can simulate a validation set by creating multiple modified observations from the existing data set (taking away each observation only once) and developing a model for each reduced data set. The response

values, BaP_{10}^{rpt}, of the deleted observations are then predicted from the model and the squared differences between predicted and 'actual' values are summed into the PRESS statistic, calculated as (Wold, 1991):

$$PRESS = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(1 - h_{ii})^2} \qquad \textbf{(Eq. 3)}$$

where y_i and $\hat{y}_i$ are defined above (Eq. 2) and h_{ii} are the diagonal elements of the "hat" matrix, calculated as:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad \textbf{(Eq. 4)}$$

Where **X** is a two-column and n-row matrix in which the first column is filled with 1's corresponding to the intercept (b) and the second column are the n predictors, corresponding to values of the sum of the coefficients on the x_n variables; **X**' refers to the transpose of **X**. According to Wold (1991), if PRESS<SSY (the sum of the squared observed responses, y_i, or BaP_{10}^{rpt}), the model predicts better than chance.

3. Results and Discussion

3.1 BaP_{10}^{rpt} Concentrations

Overall, more than 60% of the BaP_{10}^{rpt} concentrations reported in each basin are below the LOD (0.05 ng/m^3), resulting in a bimodal distribution of BaP_{10}^{rpt}. As shown in Table 1, the majority of these non-detects (NDs) are in spring and summer. Therefore, BaP^{est} values from our MLRM analyses best characterize autumn and winter activities, such as residential wood combustion. BaP_{10}^{rpt} values ≥LOD in the four air basins are similarly distributed (see *Supplementary Material*), with peak concentrations between 0.1-0.5 ng/m^3.

In autumn, ANOVA of the BaP_{10}^{rpt}≥LOD indicates that the differences in mean BaP_{10}^{rpt} among sites within each basin are not significant ($p>0.05$). When NDs are included, there is a significant ($p<0.05$) difference between the mean BaP_{10}^{rpt} at sites in the South Coast Basin in the autumn. This is most likely due to the wider range in the proportion of NDs (i.e., 33% at Burbank, and 73% at Simi Valley). However, in the winter, BaP_{10}^{rpt} are significantly ($p<0.05$) different in each basin, regardless of whether or not NDs are included. This difference in the mean concentrations within each basin is most likely due to varying source-activity levels, e.g., residential wood combustion. In the *Supplementary Material* we present the BaP_{10}^{rpt} distributions (boxplots) for each site in each basin for autumn and winter seasons.

3.2 $PM_{2.5}$ Concentrations

All reported $PM_{2.5}$ concentrations are > LOD. We present distributions (boxplots) of autumn and winter $PM_{2.5}$ concentrations in the *Supplementary Material*. Similar to winter BaP_{10}^{rpt}, ANOVA indicates that the mean $PM_{2.5}$ concentrations are significantly different ($p<0.05$) between sites within each respective basin for the winter. However, unlike BaP_{10}^{rpt} values, the autumn mean $PM_{2.5}$ concentrations are also significantly different ($p<0.05$) within each basin, except in San Diego. The distributions (histograms) of $PM_{2.5}$ concentrations within a basin are presented in the Supplementary Material.

3.3 Relative mass of BaP_{10}^{rpt} to $PM_{2.5}$

The logarithm of the relative mass ratios of BaP_{10}^{rpt} (≥ LOD) to $PM_{2.5}$, or $\log(BaP_{10}^{rpt}/PM_{2.5})$, in the four basins are similarly distributed (**Figure 2),** indicating similar relationships between BaP_{10}^{rpt} and $PM_{2.5}$ levels. Further, the $\log(BaP_{10}^{rpt}/PM_{2.5})$ at sites within the SF Bay and SJ Valley basins are similarly distributed (**Figure 2).** At each site within these two basins, the mean and medians of the logarithm of the $\log(BaP_{10}^{rpt}/PM_{2.5})$ are on the order of -4.9. However, the means and medians of the $\log(BaP_{10}^{rpt}/PM_{2.5})$ at each site within the South Coast basin, are slightly lower. **Figure 2** shows that there are significant differences in the site-specific $\log(BaP_{10}^{rpt}/PM_{2.5})$ in the South Coast basin. For the four westernmost sites in the South Coast, the mean and median are close to -5.0. Additionally, the mean and median $\log(BaP_{10}^{rpt}/PM_{2.5})$ at the two easternmost sites in the South Coast, i.e., Fontana and Rubideux, are the lowest of all the sites in California (mean and median of -5.4 at Fontana and -5.2 at Rubideux). Based on these concentrations , the lower $\log(BaP_{10}^{rpt}/PM_{2.5})$ at these two sites appear to be due to the relatively higher $PM_{2.5}$ concentrations in the autumn, instead of lower BaP_{10}^{rpt} concentrations (see *Supplementary Material*). In contrast, the lower $\log(BaP_{10}^{rpt}/PM_{2.5})$ in the winter at these sites is due to lower BaP_{10}^{rpt} values relative to other South Coast sites (except Simi Valley) and not higher $PM_{2.5}$ concentrations.

At the two San Diego sites, the median (and mean) $\log(BaP_{10}^{rpt}/PM_{2.5})$ mass ratios are the highest of all the sites in California, i.e., -4.8 (-4.9) at Chula Vista and -4.5 (-4.6) at El Cajon. When compared to other sites, this appears due to relatively higher BaP_{10}^{rpt} values in the autumn and winter and correspondingly lower $PM_{2.5}$ concentrations.

3.4 Unadjusted correlations (r) and R^2 between BaP_{10}^{rpt} and $PM_{2.5}$

Overall, we obtain the highest correlation coefficients (r) using natural-logarathmic (ln) transformed BaP_{10}^{rpt} and $PM_{2.5}$ concentrations, as opposed to the non-transformed

concentrations. This relationship generally applies on a per-site basis and for all sites combined within a basin (see *Supplementary Material*). Only four sites: San Jose (SF Bay basin), Simi Valley (South Coast basin), and Bakersfield and Fresno (SJ Valley basin) have r>0.7, indicating a relatively strong linear correlation between the two pollutant concentrations. Small negative correlations at Fontana (South Coast basin; in both the ln-transformed and non-transformed concentrations), and Chula Vista (San Diego basin; in only the non-transformed concentrations) suggests unique features in either $PM_{2.5}$ composition and/or BaP and/or $PM_{2.5}$ sources in these areas.

3.5 The "best-fit" MLRMS

Figure 3(a-d) displays our "best-fit" MLRM for each of the four air basins. We exclude BaP_{10}^{rpt}< LOD because if they are included, our MLRM violates the assumptions of (i) linearity, (ii) constant variance of the residuals, and (iii) errors that are independent of the explanatory variables. We also found that a natural-logarithmic transformation of the continuous variables did not violate these assumptions and produced a higher R^2 than leaving the continuous variables untransformed. Further, we report model results based on absolute temperature [K], because this consistently provides higher R^2-values and a significant intercept ($p<0.01$), than [^oC]. Additionally, because most basins have few, if any, spring measurements, and because autumn is not significant ($p>0.05$) in all basins for explaining BaP concentrations (but winter and summer are), we rely on the binary 'fall or winter' season variable. Compared with the MLRMs including dummy variables for each season with measurements available, this "fall or winter" variable does not decrease the R^2 nor increase the SEE appreciably.

As shown in **Table 2**, in each of the four basins we obtain an R^2 of at least 0.57, or explain at least 57% of the variation in $\ln(BaP_{10}^{rpt})$ concentrations by the MLRMs. The linear expressions of the independent variables in our "best-fit" MLRMs for the four air basins are listed in **Table 2**. The $\ln(PM_{2.5})$ is positively related to $\ln(BaP_{10}^{rpt})$ and is significant at $p<0.01$ in each basin, except in San Diego ($p<0.05$). ln(T) [K] is also significant ($p<0.01$) and inversely related to $\ln(BaP_{10}^{rpt})$ concentrations in all four basins. This effect of temperature on particle associated PAH levels has been reported elsewhere and is likely due to increased partitioning or sorption with decreasing temperature (Miguel et al. 2004; Motelay Massei et al. 2003). The natural-logarithm of the other continuous meteorological variables, RH and WS are also inversely related to $\ln(BaP_{10}^{rpt})$ concentration.

Interestingly, the binary variable indicating a weekend or weekday measurement is significant in our "best-fit" MLRMs at $p<0.01$ for the South Coast and SJ Valley air basin, but is not significant ($p>> 0.10$) for the SF Bay and the San Diego basins. This may be due partly to the higher number of weekend measurements reported in the South Coast (47 out of 140, or 34%), and the SJ Valley (22 out of 77, or 29%) basins, as compared to the SF Bay basin (only 9 out of 50, or 18%). However, because 43% of the 28 measurements were taken on the weekend in San Diego, another reason is needed to explain why weekend is not significant in the San Diego basin. Perhaps seasonal activities, such as residential wood combustion, are less prevalent there than in the other basins.

Further evaluation is needed to determine why WS is a significant variable ($p<0.05$) only in the MLRM for the South Coast basin. This might be due to the fact that ln(WS) is also significantly ($p<0.01$) correlated with the 'fall or winter' binary variable ($r= -0.4$) in the South Coast Basin. Correlation analysis reveals that WS is significantly correlated ($p<0.01$) with BaP_{10}^{rpt} levels and PM2.5 in all basins, with the exception of San Diego basin for WS and BaP_{10}^{rpt}. We can rule out a correlation between wind speed and temperature because we found they are not correlated ($R^2= 0.007$) using daily meteorological data from the NCDC. In our "best-fit" MLRMs for each basin, ln(WS) is significant at $p=0.05$ only in the South Coast model and "fall or winter" is significant at $p<0.01$ in the SF Bay and at $p<0.10$ in the South Coast and San Diego basins.

Because this is the first attempt to characterize outdoor BaP levels based on our set of explanatory variables we retain all variables, with the exception of the site-specific dummy variable, that are significant (at $p\leq0.05$) in at least one of the air basin MLRMs. Rain was the only variable that did not fulfill this requirement and this is corroborated by previously published research by Panther et al. (1999) who report no significant correlation between rainfall levels and TSP or PAH concentrations.

Although the site-specific dummy variable for the southernmost site (Bakersfield) in the SJ Valley basin is a borderline significant ($p=0.05$) estimator of BaP_{10}^{rpt} concentrations in the final MLRM, we exclude it from the final model because our objective is to ideally formulate general relationships that can be applied throughout a particular region or air basin, not at particular sites. By doing so, the multiple R^2s decreases from 0.60 to 0.57. Despite the large distance between the three sites in the SJ Valley, relative to distances between sites in the South

Coast and SF Bay region, we obtain a multiple R^2 in SJ Valley comparable to those obtained for South Coast and San Diego basins.

Additionally, the dummy variable for the Burbank site in the South Coast Basin is significant ($p<0.05$) in the MLRM and explains an additional 5% of the variation in $\ln(BaP_{10}^{rpt})$ when added to our "best-fit" MLRM, suggesting there are unique source-activity or meteorological events in sub-regions of the South Coast basin. We recommend further research, outside the scope of this exploratory study, to determine any site-specific sources influencing this finding.

When we combine all the corresponding BaP_{10}^{rpt} and $PM_{2.5}$ concentrations from the four basins to evaluate a MLRM to estimate BaP levels in metropolitan regions of California, we obtain an R^2 slightly lower (0.55) than the R^2 value for any of the individual basins. In the rightmost column of **Table 2** we present the linear expression for the combined data set. **Figure 4** provides the corresponding graphical presentation of the "best-fit" MLRM based on the seven significant variables from our basin MLRM analysis applied to the data from all sites in the four basins.

3.6 The reliability of our 'best-fit' MLRMs

The SEE of $\ln(BaP^{est})$ and the geometric spread implied, that is $\exp[SEE(\ln(BaP^{est})]$, allows us to characterize the accuracy of our BaP^{est} relative to BaP_{10}^{rpt}. The SEE of $\ln(BaP^{est})$ is fairly consistent among the four air basins (**Table 3)**, ranging from 0.48 to 0.68. When we interpret exp(SEE) as the geometric error, we estimate the BaP_{10}^{est} are within a factor less than two among the four metropolitan air basins. That is, a factor of two captures one standard deviation of the normal distribution of the residuals in $\ln(BaP^{est})$. Among the specific locations, the one geometric standard deviation of the residuals is 1.6 for SF Bay and the South Coast, 1.7 for San Diego, and 2.0 for the SJ Valley basin. Additionally, the exp(SEE) for our MLRM combining data from all basins, is 1.9. As shown by **Figures 3(a-d)** and **Figure 4**, all the BaP^{est} values are within the 95th percentile confidence interval of each of our MLRMs.

3.7 Cross Validation of our MLRMs

Because an independent set of data that has undergone strict quality assurance measures is not available at this time, we perform a cross validation of our MLRMs using a subset of our data and the PRESS statistic. First, we apply the MLRM based on combined data from all the sites, to each respective basin. This results in R^2-values ranging from 0.41 (San Diego and SF Bay) to

0.56 (SJ Valley), as summarized in **Table 3**. Next, we apply the MLRM based on the highest R^2 (0.72) and lowest SEE (0.48), using data from the SF Bay basin. Based on this application, we obtain R^2-values ranging from 0.47 (San Diego) to 0.72 (SF Bay). These results, summarized in **Table 3**, suggest that the MLRMs based on the coefficients of the seven variables summarized in **Table 2** from all sites are slightly better (lower SEE with comparable R^2) in explaining the variation in $\ln(BaP_{10}^{rpt})$ levels, than the SF Bay MLRM. However, when the MLRM based on data from all sites is applied to SF Bay it performs rather poorly compared with the best-fit model for that region. This may be due to particular meteorology or source characteristics that affect the SF Bay region, and requires further research.

When we randomly select a subset of data (n= 64) from all of the sites (N=296), the resulting ratio of the PRESS statistic to the SSY is 0.11. According to Wold (1991), this indicates a nearly ”excellent model” (PRESS/SSY<0.1), leading us to believe the reliability of the “all-sites” MLRM is quite high. Presently, we consider the all-sites MLRM as a useful surrogate measure to estimate outdoor exposure levels of BaP in metropolitan regions of California where measurements are not available. Whether this MLRM is representative of metropolitan regions outside of California, requires further investigation.

4. Conclusions

Our research demonstrates that MLRMs can be useful when measurements are not available or are limited in spatial coverage, particularly in epidemiological studies relating outdoor ambient $PM_{2.5}$ and PAH concentrations to disease surveillance. Our models provide a first step in developing reliable proxy measures of PAH concentration using widely available $PM_{2.5}$ measurements and available meteorological data. There are uncertainties, however, both in the reliability of this method for regions outside of California and whether this approach is sufficiently robust for exposure studies. In the following paragraphs we describe and evaluate sources of uncertainty in our MLRM, including: model formulation, precision and accuracy of inputs and interpretation of results. We also summarize the relevance of our findings in the context of these uncertainties.

4.1 Model Formulation

We select a parsimonious set of variables to take into account both temporal, meteorological, and spatial factors that we believe do not “over-fit” the model nor result in variables that are randomly correlated with BaP concentrations. However, additional variables may increase the

precision of our MLRMs. Examples of such factors include solar radiation [W/m^2] or hours of sunlight at a certain site to capture the rapid reaction and degradation of particle-associated BaP. However, these variables are not available for all sites and are likely explained by the ln(T) variable. Similarly, atmospheric stability (reported by the BAAQMD, but not in the NCDC data) may better explain the variation in the BaP_{10}^{rpt} levels. This may be particularly relevant in the South Coast basin due to the impact of its low and varying inversion layer. The extent to which WS may serve as a proxy for atmospheric stability needs further examination, especially because WS is significant ($p<0.05$) only in the South Coast basin.

Including other chemical constituents to adjust ln($PM_{2.5}$)concentrations may also increase the model precision. Hughes et al. (2000) demonstrated how particulate matter composition changes in the South Coast basin as one travels from Long Beach (coast) to Riverside (inland) due to the addition/accumulation of secondary ammonium nitrate and organics to the original sea-salt and primary carbon constituents. Kleeman et al. (1999) found that this phenomenon was the "single largest contribution to ambient PM2.5 concentrations" at the Riverside site. Unfortunately, sea-salt and background sulfate concentrations are not reported at the AQS sites and therefore cannot be tested in our model framework at this time.

4.2 Accuracy and quality assurance issues related to input data

Undoubtedly, model reliability increases considerably when input data pass quality assurance measures. The AQS BaP_{10}^{rpt} and $PM_{2.5}$ concentrations were determined based on standard operating procedures intended to minimize the effects of sampling artifacts due to:

1) loss of PAH from particles on filters by volatilization;
2) sorption of PAHs onto filter; and
3) reaction of compounds with reagents during sampling.

In addition, AQS sites are part of the US EPA administered State and Local Air Monitoring Station (SLAMS) network. As such, the EPA's Office of Air Quality Planning and Standards requires annual monitoring summaries of each SLAMS monitor, and additional details upon request. These quality assurance criteria strengthen the reliability of our MLRMs.

In contrast however, the NCDC unedited local climatological data were not yet quality assured. These data are simply collected by automated weather stations operated by the US government. In three instances, we had to limit the calculated RH to 100%, because we calculated over 100% based on **Eq.S-1** (Supplementary Material) and the reported station

pressure and wet bulb temperatures. However, because we use lognormally-transformed meteorological variables, the degree to which the lack of knowledge about the 'true' value of the meteorological data decreases the reliability of our "best-fit" MLRMs, is likely minimal.

4.3 Alternative modeling efforts to characterize outdoor air concentrations

To assess levels of HAPs in outdoor air, the US EPA has applied process-based transport models such as the Assessment System for Population Exposure Nationwide (ASPEN) in exposure studies (e.g., the Cumulative Exposure Project (US EPA 1998b; Woodruff et al. 1998) and the National Air Toxic Assessment for the Integrated Urban Air Toxics Strategy (US EPA 1999). Although ASPEN has a tendency to underestimate monitored concentrations (Pratt et al. 2000; US EPA 1998b), our "best-fit" per-basin, and "all sites" MLRMs have R^2-values (0.57-0.72, and 0.55, respectively) are comparable with the ASPEN. R^2-values of 0.59, 0.53, and 0.23 have been reported for ASPEN when it is applied to the Northeast, West and Central regions of the US (Rosenbaum et al. 1999).

Although our R^2-values are comparable, ASPEN and other process-based models tend to incorporate additional uncertainties associated with problem specification and the algorithms (such as reactive decay, deposition, etc) used to formulate the model. Process-based models also rely on estimated emissions, which have considerable uncertainty due to a lack of knowledge arising from missing, incomplete, or erroneous data. Therefore, we believe that the additional uncertainties generally associated with process-based models result in less reliable outdoor air concentration estimates than those we achieve with our "best-fit" MLRMs to estimate outdoor air BaP concentrations.

4.4 Relevance of our findings

Our results are encouraging in light of the scarcity of monitoring data available to characterize human exposure levels to toxic air pollutants, such as PAHs. Our "best-fit" MLRMs explain approximately the same amount of variation in $\ln(BaP^{est})$ concentrations (multiple $R^2 > 55$ %), in each of the four air basins. This is a considerable increase from the R^2-values of $\ln(BaP_{10}^{rpt})$ against $\ln(PM_{2.5})$, which explain approximately 35% of the variation in both the SJ Valley and SF Bay basin, only 7% in the South Coast, and only 3% in the San Diego basin (see Supplementary Material).

There are limits to the representativeness of our results. For example, certain activities, such as coal combustion, are not carried out in California, and therefore our results may not

represent metropolitan regions with these types of emissions. Additionally, generalizing our MLRM results is uncertain due to unknown contributions from other combustion sources, such as automobiles without catalytic converters. Because the data we rely on for BaP_{10}^{rpt} is concentrated in the autumn and winter months, our MLRM derived estimates may tend to reflect ambient air concentrations due to residential wood combustion.

Our MLRM methods to estimate outdoor-air BaP levels based on surrogate $PM_{2.5}$ concentrations can likely be applied toward characterizing levels of other particulate-associated PAHs, such as benzo(k)fluoranthene, benzo(ghi)perylene, indeno(1,2,3 cd)pyrene, and dibenz(a,h)anthracene, as well as other semivolatile organic compounds, such as many pesticides, PCBs, and dioxins. Our approach may therefore complement the research and exposure studies proposed as part of the human health tracking surveillance system in California (CPRC 2004). Because of the relatively similar features of the MLRMs of the respective metropolitan regions (R^2, significant variables) the authors lay the foundation for future studies and validation.

Acknowledgements

We thank Arpad Horvath, Robert Spear, and S. Katherine Hammond of University of California, Berkeley (UC-Berkeley), for their early reviews and Coe Owen of the US EPA, for air concentration data and Kurt Malone of the Bay Area Air Quality Management District for meteorological data. This research was supported in part by the US EPA National Exposure Research Laboratory through Interagency Agreement No. DW-988-38190-01-0, with the Lawrence Berkeley National Laboratory operated for the US Department of Energy (DOE) under Contract Grant No. DE-AC02-05CH11231 and by Cooperative Agreement Number U50/CCU922409-01 from the US Centers for Disease Control and Prevention (CDC). Additional funding assistance was provided by the University of California Toxic Substance and Research Teaching Program and the Russell M. Grossman Endowment from the School of Public Health, UC-Berkeley. Although the research described in this article has been funded and reviewed by the US EPA, it does not necessarily reflect the views of the federal government and no official endorsement should be inferred.

Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

References

Adonis M., Gil L., 2000. Polycyclic aromatic hydrocarbon levels and mutagenicity of inhalable particulate matter in Santiago, Chile. Inhalation Toxicology 12:1173-1183.

Baek S., Field R.A., Goldstone M.E., Kirk P., Lester J., Perry R., 1991. A review of atmospheric polycyclic aromatic hydrocarbons: Sources, fate and behavior. Water, Air, and Soil Pollution 60:279-300.

Bjorseth, A., Ramdahl, H., 1985. Source and emissions of PAH. In: Bjorseth A., Ramdahl, T., (Eds.), Handbook of Polycyclic Aromatic Hydrocarbons. Volume 2: Emission Sources and Recent Progress in Analytical Chemistry. Marcel Dekker Inc., New York, pp. 1-20.

CARB (California Air Resources Board), 1998. Standard operating procedure for the determination of selected polycyclic aromatic hydrocarbons (PAH) in ambient air (SOP MLD 028).

CARB (California Air Resources Board), 2002. Standard operating procedure for the determination of PM2.5 mass in ambient air by gravimetric analysis (SOP MLD 055): **2002** Available: http://www.arb.ca.gov/aaqm/sop/summary/sop055.htm [Accessed January 2007].

CPRC (California Policy Research Center), 2004. Report on the steering committee for implementing SB 702. Available: www.ucop.edu/cprc/ehssrpt.pdf [Accessed September, 2006].

Dejmek J., Solansky I., Benes I., Lenicek J., Sram R.J., 2000. The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome. Environmental Health Perspectives 108(12):1159-64.

Deng, W.J., Louie, P.K.K., Liu, W.K., Bi, X.H., Fu, J.M., Wong, M.H., 2006. Atmospheric levels and cytotoxicity of PAHs and heavy metals in TSP and $PM_{2.5}$ at an electronic waste recycling site in southeast China. Atmospheric Environment 40: 6945-6955.

Dimashki M., Lim L.H., Harrison R.M., Harrad S., 2001. Temporal trends, temperature dependence, and relative reactivity of atmospheric polycylic aromatic hydrocarbons. Environmental Science and Technology 35(11):2264-67.

Fraser M.P., Cass G.R., Simoneit B.R.T., Rasmussen R.A., 1998. Air quality model evaluation data for organics. 5. C6-C22 nonpolar and semipolar aromatic compounds. Environmental Science and Technology 32(12):1760-70.

Guo H., Lee S.C., Ho K.F., Wang X.M., Zou S.C., 2003. Particle-associated polycyclic aromatic hydrocarbons in urban air of Hong Kong. Atmospheric Environment 37:5307-5317.

Hamburg M., 1970. Statistical analysis for decision making. San Diego, CA: Harcourt Brace Jovanovich, Inc.

Harner T., Bidleman F., 1998. Octanol-air partition coefficient for describing particle/ gas partitioning of aromatic compounds in urban air. Environmental Science and Technology 32(10):1494-1502.

Hou, X., Zhuang, G., Sun, Y., Zhisheng, A., 2006. Characteristics and sources of polycyclic aromatic hydrocarbons and fatty acids in $PM_{2.5}$ aerosols in dust season in China. Atmospheric Environment 40: 3251-3262.

Hughes L.S., Allen J.O., Bhave P., Kleeman M.J., Cass G.R., Liu D.Y., Fergenson D.P., Morrical B.D., Prather K.A., 2000. Evolution of atmospheric particles along trajectories crossing the Los Angeles Basin. Environmental Science and Technology 34(15):3058-68.

Kamens R.M., Guo Z., Fulcher J.N., 1998. Influence of humidity, sunlight, and temperature on the daytime decay of polyaromatic hydrocarbons on atmospheric soot particles. Environmental Science and Technology 22(1):103-8.

Kaupp H., McLachlan M.S., 2000. Distribution of polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) and polycyclic aromatic hydrocarbons (PAHs) within the full size range of atmospheric particles. Atmospheric Environment 34:73-83.

Kiss G., Varga-Puchony Z., Rohrbacher G., Hlavey J., 1998. Distribution of polycyclic aromatic hydrocarbons on atmospheric aerosol particles of different sizes. Atmospheric Research 46:253-61.

Kleeman M.J., Hughes L.S., Allen J.O., Cass G.R., 1999. Source Contributions to the Size and Composition Distribution of Atmospheric Particles: Southern California in September 1996. Environmental Science and Technology 33(23):4331-4341.

Leotz-Gartziandia E., Tatry V., Carlier P., 2000. Sampling and analysis of polycyclic aromatic hydrocarbons (PAH) and oxygenated PAH in diesel exhaust and ambient air. Polycyclic Aromatic Compounds 20:245-58.

Lewis A., Kupiszewska D., Bartle K., Pilling M., 1995. City centre concentrations of polycyclic aromatic hydrocarbons using supercritical fluid extraction. Atmospheric Environment 29(13):1531-42.

Lohman R., Harner T., Thomas G.O., Jones K.C., 2000. A comparative study of the gas-particle partitioning of PCDD/Fs, PCBs, and PAHs. Environmental Science and Technology 34(23):4943-51.

Mader, B.T., Pankow, J.F., 2002. Study of the effects of particle-phase carbon on the gas/particle partitioning of semivolatile organic compounds in the atmosphere using controlled field experiments. Environmental Science and Technology 36(23):5218-28.

Massolo L., Muller A., Tueros A., Rehwagen M., Franck U., Ronco A., Herbarth O., 2002. Assessment of mutagenicity and toxicity of different-size fractions of air particulates from La Plata, Argentina, and Leipzig, Germany. Environmental Toxicology 17:219-31.

MathSoft, Inc. 2000. SPLUS 2000 Professional Release 3; Copyright 1988-2000.

Miguel A.H., Eiguren-Fernandez A., Jaques P.A., Froines J.R., Grant B.L., Mayo P.R., Siouta C., 2004. Seasonal variation of the particle size distribution of polycyclic aromatic hydrocarbons and of major aerosol species in Claremont, California. Atmospheric Environment 38:3241-3251.

Motelay-Massei A., Ollivon D., Garban B., Chevreuil M., 2003. Polycyclic aromatic hydrocarbons in bulk deposition at suburban site: assessment by principal component analysis of the influence of meteorological parameters. Atmospheric Environment 37:3135-3146.

NCDC/NOAA (National Climate Data Center/ National Oceanic and Atmospheric Association). 2003. Unedited Daily Meteorological Data. Available: http://www.ncdc.noaa.gov/servlets/DLY [Accessed September, 2004].

Offenberg J.H., Baker J.E., 2002. The influence of aerosol size and organic carbon content on gas/particle partitioning of polycyclic aromatic hydrocarbons. Atmospheric Environment 36:1205-1220.

Odum J.R., Yu J., Kamens R.M., 1994. Modeling the mass transfer of semivolatile organics in combustion aerosols. Environmental Science and Technology 28(13):2278-2285.

Panther B.C., Hooper M.A., Tapper N.J., 1999. A comparison of air particulate matter and associated polycylic aromatic hydrocarbons in some tropical and temperate urban environments. Atmospheric Environment 33:4087-99.

Pleil J.D., Vette A.F., Rappaport, S.M. Assaying particle-bound polycyclic aromatic hydrocarbons from archived PM2.5 filters. Journal of Chromatography A 1033:9-17.

Pohjola S.K., Lappi M., Honkanen M., Rantanen L., Savela K., 2003. DNA binding of polycyclic aromatic hydrocarbons in a human bronchial epithelial cell line treated with diesel and gasoline particulate extracts and benzo(a)pyrene. Mutagenesis 18(5):429-38.

Pratt G.C., Palmer K., Wu C.Y., Oliaei F., Hollerbach C., Fenske M.J., 2000. An assessment of air toxics in Minnesota. Environmental Health Perspectives 108(9):815-825.

Richter H., Howard J.B., 2000. Formation of polycylic aromatic hydrocarbons and their growth to soot- a review of chemical reaction pathways. Progress in Energy and Combustion Science 26:565-608.

Rosenbaum A.S., Axelrad D.A., Woodruff T.J., Wei Y.H., Ligocki M.P., Cohen J.P., 1999. National estimates of outdoor air toxics concentrations. Journal of Air and Waste Management Association 49(10):1138-1152.

Sakai R., Siegmann H.C., Sato H., Voorhees A.S., 2002. Particulate matter and particle-attached polycyclic aromatic hydrocarbons in the indoor and outdoor air of Tokyo measured with personal monitors. Environmental Research 89(1):66-71.

Schauer, J.J., Cass, G.R., 2000. Source apportionment of wintertime gas-phase and particle-phase air pollutants using organic compounds as tracers. Environmental Science and Technology 34(9), 1821-1832.

Spencer, M.T., Shields, L.G., Sodeman, D.A., Toner, S.M., Prather, K.A., 2006. Comparison of oil and fuel particle chemical signatures with particle emissions from heavy and light duty vehicles. Atmospheric Environment 40:5224-5235.

Strandell M., Zakrisson S., Alsberg T., Westerholm R., Winquist L., Rannug U., 1994. Chemical analysis and biological testing of a polar fraction of ambient air, diesel engine, and gasoline engine particulate extracts. Environmental Health Perspectives 102(Suppl 4):85-92.

US EPA. 1997. National Ambient Air Quality Standards for Particulate Matter; Final Rule. 40 CFR Part 50. Available: http://www.epa.gov/ttncaaa1/t1/fr_notices/pmnaaqs.pdf [Accessed September 2006].

US EPA. 1998a. 1990 Emissions Inventory of Section 112(c)(6) Pollutants: POM, 2,3,7,8-TCDD/ 2,3,7,8-TCDF, PCBs, Hexachlorobenzene, Mercury, and Alkylated Lead. Final Report. United States Environmental Protection Agency, Emission Factor and Inventory Group (MD-14) and Visibility and Ecosystem Protection group (MD-15).

US EPA. 1998b. Urban Air Toxics Report: Appendix B. Modeled outdoor concentrations of HAPs: Analysis of data from the Cumulative Exposure Project for the Urban Area Source Program. Office of Policy, Planning and Evaluation, Office of Air Quality Planning and Standards: May 1998.

US EPA. 1999. National Air Toxics Program: The Integrated Urban Strategy; Final Notice. Fed Reg 64:38705-38740.

US EPA. 2000. National Air Pollutant Emissions Trends: 1900-1998. Office of Air Quality Planning and Standards, US EPA. EPA 454/R-00-002. Available: http://www.epa.gov/ttn/chief/trends/trends98/ [Accessed September 2006].

US EPA. 2004. AirData. Available: http://www.epa.gov/air/data.aqsdb.html [Accessed December 2004].

Vardar N., Noll K.E., 2003. Atmospheric PAH concentrations in fine and coarse particles. Environmental Monitoring and Assessment 87:81-92.

Wold S., 1991. Validation of QSARs. Quantitative Structure Activity Relationships 10:191-193.

Woodruff T.J., Axelrad D.A., Caldwell J., Morello-Frosch R., Rosenbaum A., 1998. Public health implications of 1990 Air Toxics Concentrations across the United States. Environmental Health Perspectives 106(5):245-251.

Zheng, M., Cass, G.R., Schauer, J.J., Edgerton, E.S., 2002. Source apportionment of PM2.5 in the southeastern United States using solvent-extractable organic compounds as tracers. Environmental Science and Technology 36(11), 2361-2371.

Appendix

List of Nomenclature

AIRS: Aerometric Information Retrieval System

AQS: US EPA's Air Quality Subsystem

ASPEN: Assessment System for Population Exposure Nationwide

BAAQMD: Bay Area Air Quality Management District

BaP: benzo(a)pyrene

BaP_{10}^{rpt}: reported particle-associated (PM_{10}) BaP levels

BaP^{est}: estimated BaP levels

CARB: California Air Resources Board

EPA: US Environmental Protection Agency

ln: natural logarithm

LOD: limit of detection

MLD: Monitoring and Laboratory Division

MLRM: multivariate linear regression model

NCDC: National Climate Data Center

ND: non-detect

PAHs: polycyclic aromatic hydrocarbons

PM: particulate matter

$PM_{2.5}$: fine PM with aerodynamic diameter less than 2.5 μm

PM_{10}: fine PM with aerodynamic diameter less than 10 μm

PRESS: Predictive Residual Sum of Squares

r : Pearson product-moment correlation coefficient

R^2: multiple correlation coefficient

RH: relative humidity

SEE: standard error of the estimate

SF Bay: San Francisco Bay Area Air Basin

SJ Valley: San Joaquin Valley Air Basin

SLAMS: State and Local Air Monitoring Station

SOP: Standard Operating Procedure

SSY: sum of square of responses

T: average daily (24-hour) temperature

WS: wind speed

Figure Captions

Figure 1. Location of US EPA AQS sites with corresponding meteorological sites (BAAQMD,2003 and NCDC, 2003) used in our MLRM analysis. Co-located $PM_{2.5}$ and BaP_{10}^{rpt} concentrations and meteorological data are available for four AQS sites in the SF Bay air basin, all six sites in the South Coast basin, and all three sites in the SJ Valley basin. The air basin boundaries are roughly defined, but otherwise the figure is drawn to scale.

Figure 2. The logarithm of the relative mass of BaP_{10}^{rpt} to $PM_{2.5}$ in the four air basins. Outliers are presented as "○".

Figure 3. The "best-fit" MLRMs and 95th% confidence intervals shown for the a) SF Bay, b) SJ Valley, c) South Coast, and d) San Diego air basins. The form for the MLRMs is: ln($PM_{2.5}$ [μg/m^3]) + ln(WS [m/s]) + ln(T [K]) + ln(RH [%])+ weekend/holiday + season + ln(coast distance [km]). Coefficients are provided in **Table 2**. These intervals are calculated by adding and subtracting the following from the BaP^{est} : $2 \times \exp\left(\frac{stdev(\ln(BaP^{est}))}{\sqrt{N}}\right)$

where *stdev* is the standard deviation and *N* is the size of the MLRM data set, given in **Table 2**.

Figure 4. "Best-fit" MLRM combining data from all four air basins in California, with 95th% confidence intervals. The form for the MLRMs is: ln($PM_{2.5}$ [μg/m3]) + ln(WS [m/s]) + ln(T [K]) + ln(RH [%])+ weekend/holiday + season + ln(coast distance [km]). Coefficients are provided in **Table 2.**

Table 1: The total sample size, n, from each air basin (not including the BaP_{10}^{rpt}<LOD) and the percent of BaP_{10}^{rpt}<LOD.

Basin		**%<LOD**			
Site	n	Fall	Winter	Spring	Summer
SF Bay	**51**	**46-60%**	**8-16%**	**50-83%**	**89-100%**
Fremont	17	60%	11%	75%	89%
Concord	10	50%	8%	50%	100%
San Francisco	13	46%	16%	83%	100%
San Jose	11	53%	11%	75%	100%
SJ Valley	**80**	**40-57%**	**0-17%**	**63-79%**	**88-100%**
Fresno	28	40%	0%	63%	100%
Bakersfield	36	53%	5.6%	79%	100%
Modesto	16	57%	17%	63%	88%
South Coast	**140**	**33-73%**	**0-40%**	**58-96%**	**76-100%**
Burbank	26	33%	5%	79%	76%
Los Angeles	27	53%	5%	67%	94%
Long Beach	27	46%	0%	79%	94%
Rubidoux	25	38%	28%	89%	97%
Fontana	29	40%	40%	58%	82%
Simi Valley	6	73%	33%	96%	100%
San Diego	**28**	**40%, 46%**	**0, 5.6%**	**88%, 89%**	**100%, 100%**
Chula Vista	23	46%	5.6%	89%	100%
El Cajon	5	40%	0%	88%	100%

Table 2
Outdoor air ln-BaP^{est} [ng/m^3] stepwise regression results (95th percentile confidence limits). Unless otherwise indicated, coefficients (βs) are significant at p<0.01.

Coefficient	Variable	SF Bay	SJ Valley	South Coast	San Diego	All Sites
β_o	Intercept	195 (123, 268)	184 (111, 255)	137 (98, 175)	229 (81, 377)	182 (151, 212)
β_1	ln($PM_{2.5}$ [μg/m^3])	0.5 (0.1,0.8)	0.5 (0.2, 0.7)	0.4 (0.2, 0.6)	0.8 (0.1, 1.5)[h]	0.5 (0.3, 0.6)
β_2	ln(T [K])	-34.1 (-46.5, -21.7)	-32.5 (-44.9, -20.2)	-23.9 (-30.6, -17.2)	-40.9 (-14.8, -66.9)	-32.1 (-37.4, -26.8)
β_3	ln(WS [m/s])	-0.1 (-0.6,0.5)[a]	-0.2 (-0.7, 0.2)[d]	-0.3 (-0.6, -0.1)[h]	1.0 (-0.2, 2.1)[e]	-0.2 (-0.4, -0.06)[j]
β_4	ln(RH [%])	-1.4 (-2.4, -0.5)	-0.7 (-1.6, 0.1)[e]	-1.0 (-1.2, -0.7)	-1.1 (-1.8, -0.4)	-0.8 (-1.0, -0.5)
β_5	Weekend or holiday	-0.02 (-0.4, 0.4)[b]	-0.4 (-0.8, -0.01)	-0.3 (-0.5, -0.1)	0.1 (-0.4, 0.6)[i]	-0.3 (-0.5, -0.1)
β_6	Fall or winter	0.8 (0.4, 1.2)	0.2 (-0.3, 0.7)[f]	0.2 (-0.01, 0.5)[e]	0.8 (-0.2,1.2)[e]	0.4 (-0.2, 0.6)
β_7	ln(coast distance [km])	0.1 (-0.1, 0.2)[c]	0.2 (-0.6, 0.9)[h]	-0.2 (-0.3, -0.04)	0.5 (0.05, 0.9)[h]	0.02 (-0.04, 0.1)[k]
	***N* =**	50 [p]	78 [p]	140	28	296
	R^2=	0.72	0.57	0.57	0.65	0.55
	SEE =	0.48	0.68	0.50	0.51	0.62

The regression model is: $\beta_o + \beta_1 \times \ln(PM_{2.5}$ [μg/m3]) + $\beta_2 \times \ln$ (T [K]) + $\beta_3 \times \ln$(WS [m/s]) + $\beta_4 \times \ln$ (RH [%]) +$\beta_5 \times$ (Weekend or holiday) + $\beta_6 \times$ season + $\beta_7 \times \ln$(coast distance [km]) + error.

[a-n] p-value: [a] 0.74; [b] 0.94; [c] 0.27; [d] 0.29; [e] p<0.10; [f] 0.15; [g] 0.14; [h] p<0.05; [i] 0.0.61 ; [j] 0.15; [k] 0.48
[p]: the final sample size has been decreased from that shown in Table 1 because of missing meteorological data.

Table 3
R^2 and SEE-values resulting from applying the "best-fit" MLRMs from three regions (with highest R^2, lowest SEE and from all sites combined MLRM; based on coefficients listed in Table 1) to other regions.

Region MLRM	**Resulting R^2 and SEE-values**				
	SF Bay	**SJ Valley**	**South Coast**	**San Diego**	**All Sites**
All Sites					
R^2	0.41	0.56	0.47	0.41	0.55
SEE	0.90	1.1	0.71	0.68	0.62
SF Bay (highest R^2)					
R^2	0.72	0.50	0.47	0.47	0.48
SEE	0.48	0.91	1.2	0.87	0.79

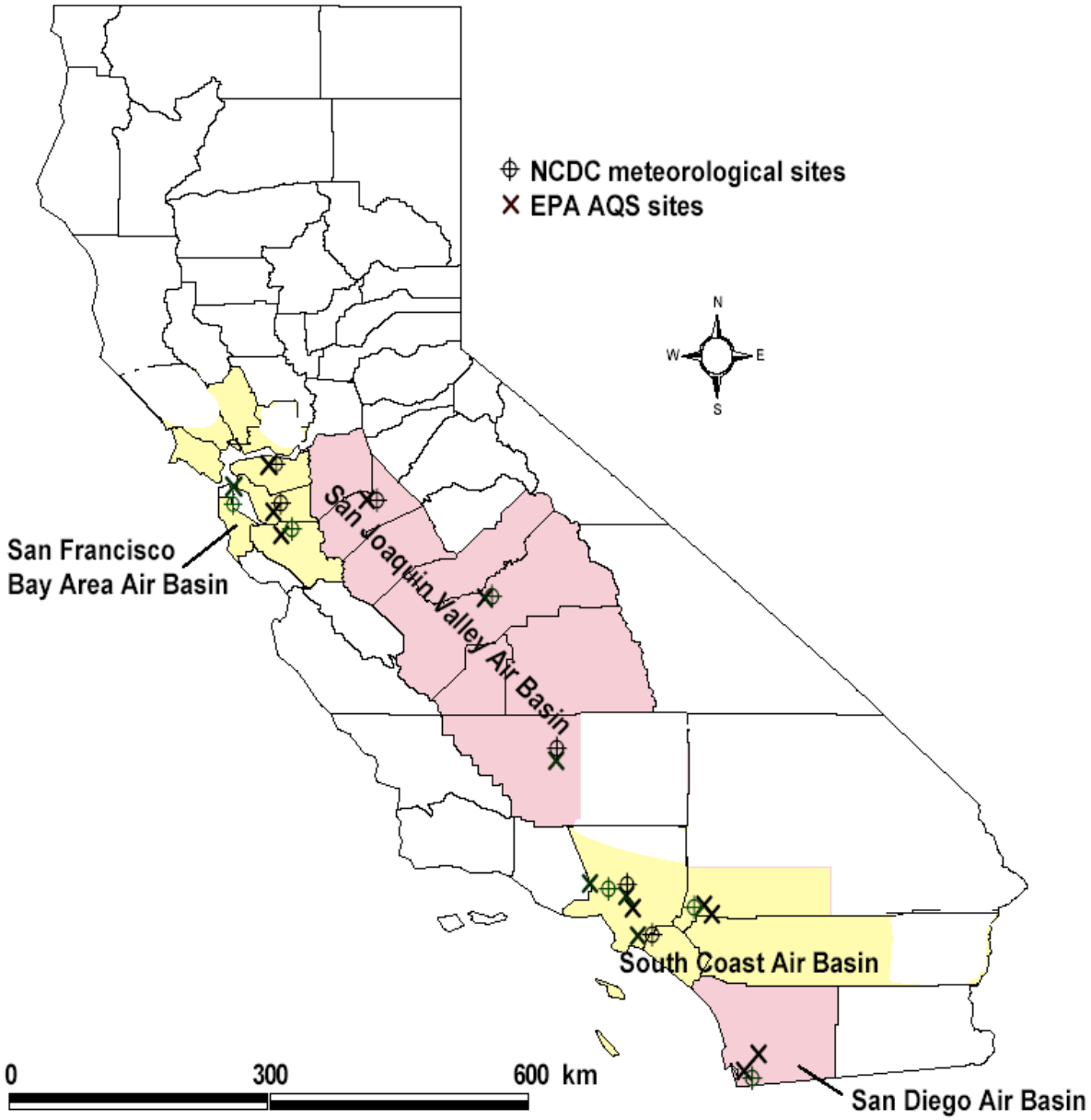
NCDC meteorological sites
EPA AQS sites
N
W
E
S
San Francisco
Bay Area Air Basin
San Joaquin Valley Air Basin
South Coast Air Basin
San Diego Air Basin
0
300
600 km

Figure 1

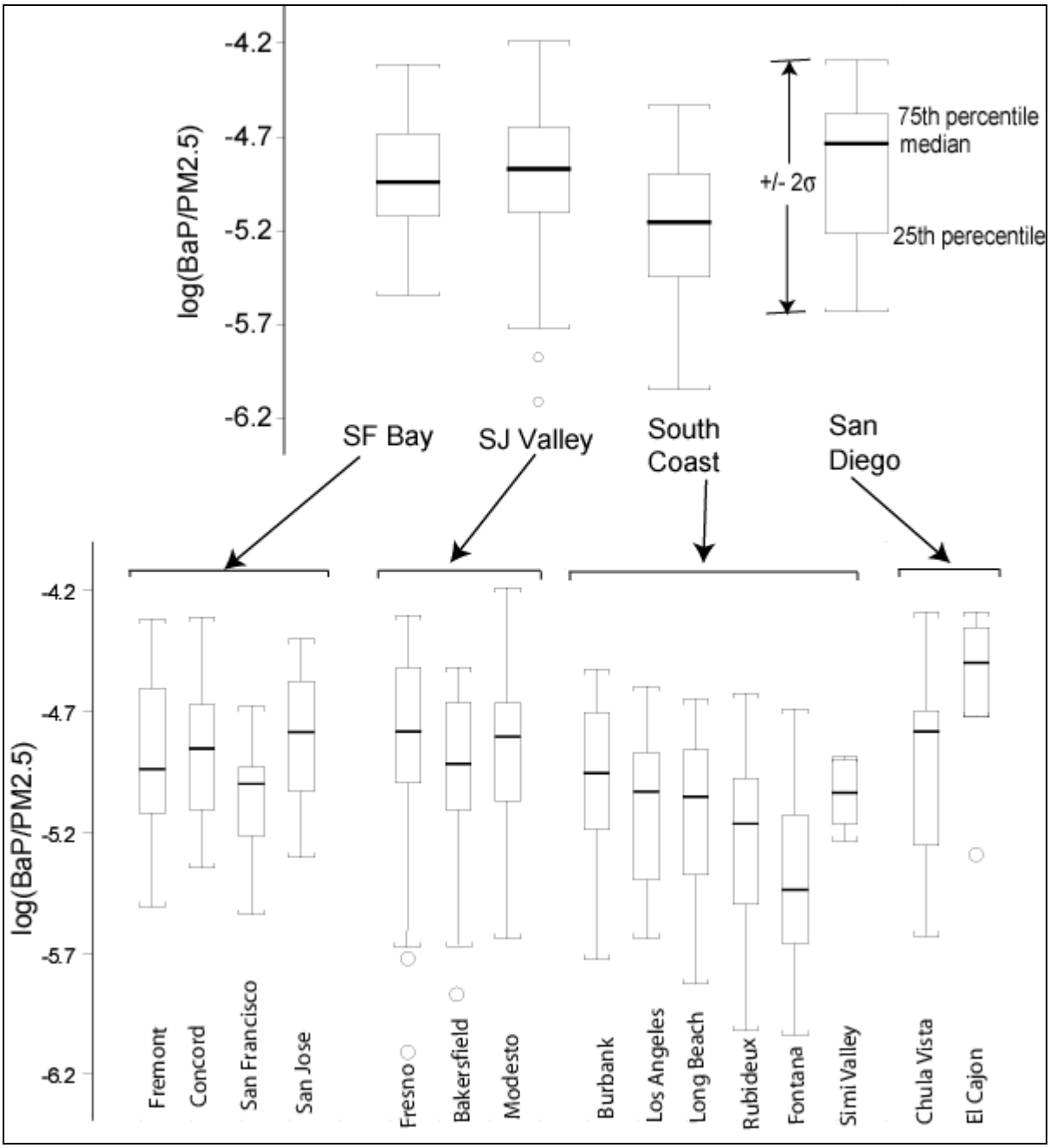
log(BaP/PM2.5)
-4.2
-4.7
-5.2
-5.7
-6.2
75th percentile
median
+/- 2σ
25th perecentile
SF Bay
SJ Valley
South Coast
San Diego
Fremont
Concord
San Francisco
San Jose
Fresno
Bakersfield
Modesto
Burbank
Los Angeles
Long Beach
Rubideux
Fontana
Simi Valley
Chula Vista
El Cajon

Figure 2

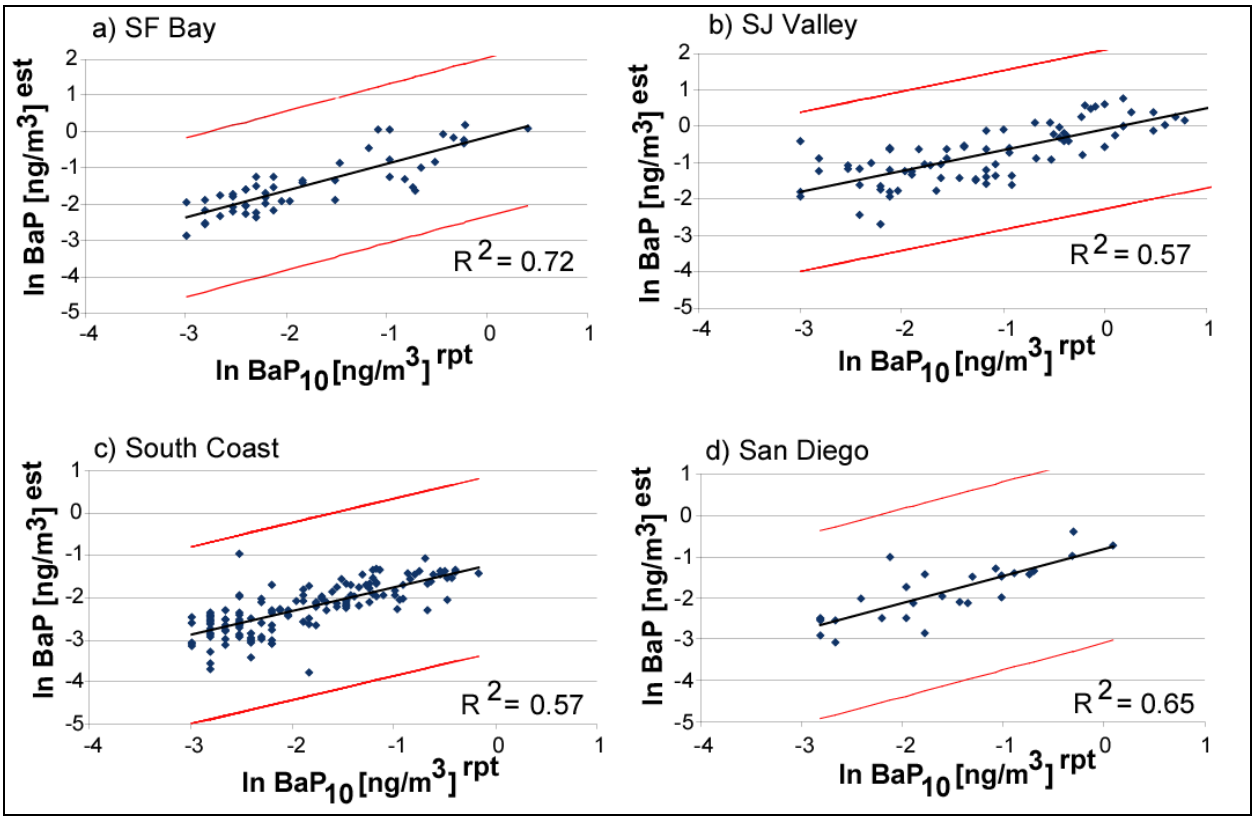
a) SF Bay
b) SJ Valley
c) South Coast
d) San Diego
ln BaP [ng/m3] est
ln BaP10 [ng/m3] rpt
R2 = 0.72
R2 = 0.57
R2 = 0.57
R2 = 0.65

Figure 3

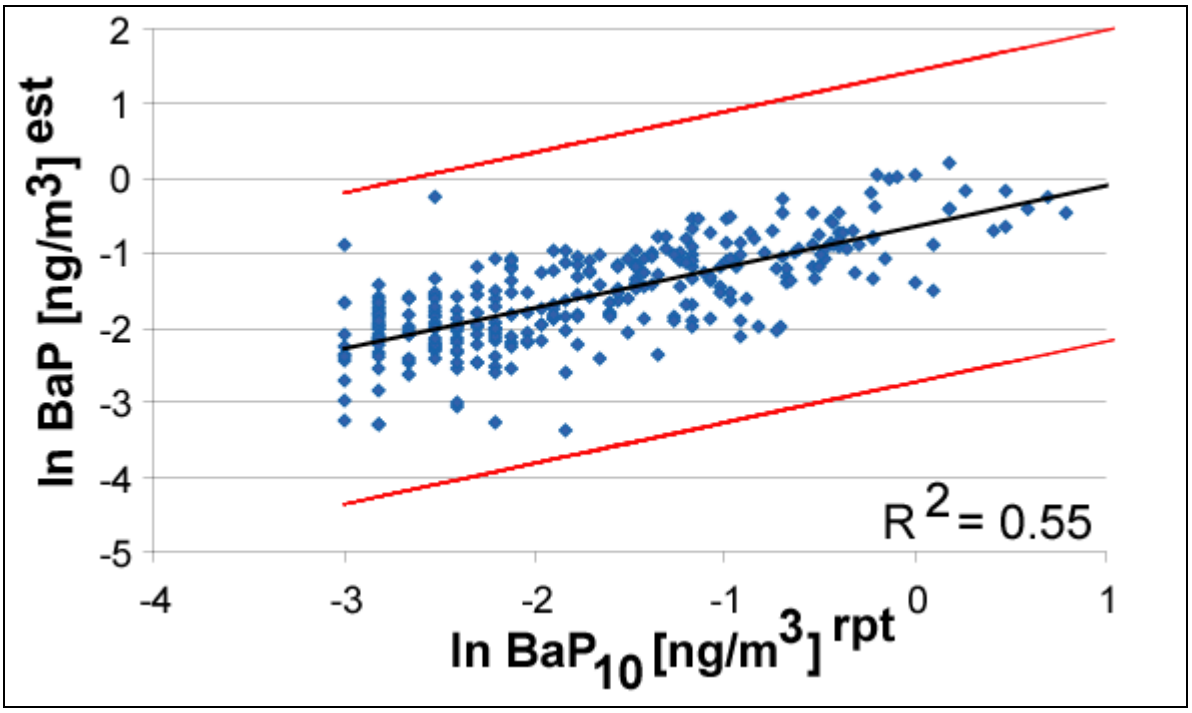

Figure 4