Harvard University
Harvard University Biostatistics Working Paper Series

Year 2006 *Paper* 47

Mixed Multiplicative Factor Analysis Model for Air Pollution Exposure Assessment

Margaret C. Nikolov* Brent A. Coull†
Paul J. Catalano‡ John J. Godleski**

*Harvard School of Public Health, mkaramit@hsph.harvard.edu

†Harvard University, bcoull@hsph.harvard.edu

‡Harvard School of Public Health and Dana Farber Cancer Institute, catalano@hsph.harvard.edu

**Harvard School of Public Health, jgodlesk@hsph.harvard.edu

http://biostats.bepress.com/harvardbiostat/paper47

Mixed Multiplicative Factor Analysis Model for Air Pollution Exposure Assessment

Margaret C Nikolov, Brent A Coull, Paul J Catalano, John J Godleski

Department of Biostatistics
Harvard University

July 21, 2006

SUMMARY

A primary objective of current air pollution research is the assessment of health effects related to specific sources of air particles, or particulate matter (PM). Because most PM health studies do not observe the activity of the pollution sources directly, investigators must infer pollution source contributions based on a complex mixture of exposure. Methods such as source apportionment and multivariate receptor modeling use standard factor analytic techniques to estimate the source-specific contributions from a large number of observed chemical concentrations. In the interest of a more flexible source apportionment, we propose a multiplicative factor analysis with a mixed model on the latent source contributions. A factor analysis with multiplicative errors serves to maintain the non-negativity of the measured chemical concentrations. A mixed model on the latent source contributions provides for systematic effects on source activity as well as an adjustment for residual correlation in the source-specific exposures. In a simulation study, we examine the impact of (1) accounting for meteorological covariates and (2) adjusting for temporal correlation in the exposures on the estimation of the source profiles and the source activities. Finally, we expore the influence of meteorological conditions on source-specific exposures in an analysis of real PM exposure data.

Table 1: Major Sources of Boston Air Pollution

Source	Chemical Components
Road Dust	silicon and aluminum
Power Plants	sulfur and sulfate
Oil Combustion	nickel and vanadium
Motor Vehicles	black carbon, organic carbon, elemental carbon

1 Introduction

Evaluation of health effects associated with major sources of air pollution, such as power plants and motor vehicles, often relies on the characterization of complex air pollution exposures. In most health effects studies, investigators are unable to measure the activity of the pollution sources directly, and instead collect samples of ambient air, which reflect dynamic mixtures of source contributions. Methods such as source apportionment and multivariate receptor modeling use factor analytic techniques to estimate the contributions of a small number of pollution sources from the measured mixture components. While the exposure assessment literature contains a large amount of research that focuses on estimation of source-specific contributions (i.e., Koutrakis and Spengler 1987; Kavouras *et al.* 2001; and for review see Seigneur *et al.* 1999; Hopke 2003; Kim *et al.* 2004), little work has been done to explore the role of important factors that influence source activity in a formal way.

Over the past decade, researchers at the Harvard School of Public Health (HSPH) have been conducting animal toxicology studies to evaluate the mechanisms of morbidity and mortality associated with ambient air particulate matter (PM). As part of these studies, samples of Boston aerosol have been collected, concentrated, and analyzed for a series of elements and other chemical components. Source apportionment analyses of these exposure mixtures have indicated four major sources of Boston PM, often referred to as resuspended road dust, coal-fired power plants, oil combustion (primarily for home-heating), and motor vehicle exhaust based on their key chemical components as described in Table 1. Standard source apportionment methods used to analyze the Boston PM exposures characterize pollution sources strictly in terms of the source profiles and the source contributions.

Presently, researchers seek to better characterize source activity. A primary research objective is to allow the unobserved source contributions to depend on systematic effects, such as meteorological conditions. Consider, for instance, the potential for temperature to influence source activity. During periods of low temperature, home heating typically increases, which results in elevated contributions from oil combustion. Under this scenario, incorporating information on temperature in the source apportionment model may provide for better characterization of the oil combustion pollution source. Extending source apportionment methods to allow for systematic effects, such as meterology, on the source activities may lead to better estimation of the source profiles and, more importantly, the source contributions.

Another limitation of most existing source apportionment methods is failure to adequately adjust for temporal correlation in exposures measured on consecutive days. Most current source apportionments assume independent exposures, an assumption that is often unrealistic given the similarity of ambient air composition from one day to the next. In particular, the HSPH toxicology studies, by design, often collected air samples in blocks of three consecutive days. A second research objective aimed at improving source characterization is to account for residual correlation due to the clustered study design in the source apportionment model. Park, Guttorp, and Henry (2001) build on the multivariate receptor model using a time series approach to account for temporal dependence in the exposures. Christensen and Sain (2002) develop a nested block bootstrap method to incorporate a dependence structure in multivariate receptor modeling.

In this paper, we consider source apportionment methods to better characterize source activity. Following Wolbers and Stahel (2005; and Billheimer 2001), we impose a multiplicative error structure on the factor analysis model. A multiplicative factor analysis with log-normal source contributions and log-normal errors respects the non-negativity of the observed elemental concentrations. The typical factor analysis model with additive error, in constrast, is not conducive to this property. In the additive error formulation, normally distributed errors allow for negative elemental concentrations. Alternatively, log-normally distributed errors perserve non-negativity; however, these distributional assumptions violate the mean zero assumption in that log-normal errors are always positive.

We build on Wolbers and Shahel (2005; and Billheimer 2001) in two important ways. First, we propose a mixed model on the latent source contributions to allow source activity to depend on systematic effects, such as meteorological covariates, while adjusting for temporal correlation through random block effects. Second, we take a Bayesian approach to estimation. Our ultimate goal in developing such an elaborate source apportionment is two-fold; (1) to assess factors that are associated with higher source contributions, and (2) to gain better estimates of the latent source countributions in the interest of improving estimation of source-specific health effects. Since studies investigating the health effects of air pollution typically consist of a small number of unique exposures, Bayesian methods provide a distinct advantage over classical methods given the ability to leverage historical exposure information (Nikolov *et al.* 2006). In keeping with this paradigm, we fit the mixed multiplicative factor analysis model using Bayesian methods.

The remainder of this paper is arranged as follows. Section 2.2 describes the exposure and covariate data in detail. Section 2.3 presents the mixed multiplicative factor analysis model, discusses the implications of this modeling framework, and describes a Bayesian approach to estimation. Section 2.4 provides a simulation study to evaluate the impact of adjusting for covariates and temporal correlation in the source apportionment, and Section 2.5 implements the proposed methods to analyze the HSPH exposure data. Finally, we summarize our findings and conclusions in Section 2.6.

2 Data

Our exposure data consists of detailed information on the chemical composition of concentrated ambient particles (CAPs) collected in Boston between September 1996 and March 2003. Located on Huntington Avenue, a major road in Boston, the Harvard Ambient Particle Concentrator (HAPC) takes in ambient air and concentrates the samples approximately 30 times without altering the physical and chemical properties of the mixture. The CAPs exposures are then measured for sulfate (SULF) via ion chromatography, black carbon (BC) using an aethalometer, elemental carbon (EC) and organic carbon (OC) determined with a thermal and optical reflectance method, and elemental

Table 2: Summary of CAPs measured in ug/m^3 ($N = 139$)

Element	Mininum	25th Perc	Median	75th Perc	Maximum
Si	1.434	5.612	9.013	13.794	47.856
S	3.453	15.438	29.510	53.651	378.171
Ni	0.002	0.032	0.053	0.094	0.589
OC	9.100	44.700	76.144	119.231	617.552
Al	0.055	1.279	2.906	4.998	21.726
Ti	0.079	0.281	0.441	0.651	1.780
Ca	0.780	2.507	3.526	5.349	15.540
SULF	9.172	43.127	83.100	152.650	1237.400
Se	0.001	0.010	0.022	0.048	0.255
V	0.003	0.058	0.091	0.148	0.681
Br	0.010	0.050	0.092	0.152	0.442
BC	1.746	5.836	10.545	15.318	42.009
EC	2.400	12.701	22.535	31.150	91.445

concentrations (in ug/m^3) collected via X-ray fluorescence (XRF), specifically: aluminum (Al), arsenic (As), barium (Ba), bromine (Br), calcium (Ca), cadmium (Cd), chlorine (Cl), chromium (Cr), copper (Cu), iron (Fe), potassium (K), manganese (Mn), nickel (Ni), sodium (Na), lead (Pb), sulfur (S), selenium (Se), silicon (Si), titanium (Ti), vanadium (V), and zinc (Zn).

As noted by Park, Guttorp, and Henry (2001), an important first step in source apportionment is to select a subset of species that are contributed by major pollution sources. In this paper, we focus on a subset of $P = 13$ elements deemed to be major components of the four known sources of Boston PM; Si, Al, Ti, Ca, S, SULF, Se, Br, Ni, V, OC, BC, and EC. Table 2 summarizes the complete CAPs exposure data ($N = 139$).

Using a mixed model (Diggle *et al.* 2002) with random effects for blocks of consecutive days, we estimated the temporal correlation in the measured concentrations for each of the chemical species. Table 3 presents the estimated correlations, where the four columns contain the elements mainly contributed by the four major sources of Boston PM described in Table 1. Notice that, in general, the estimates are quite similar among elements contributed by the same pollution source. This finding suggests that the temporal correlation in the measured chemical components may be explained

Table 3: Estimates of Temporal Correlation in CAPs

Elem	$\hat{\rho}$	Elem	$\hat{\rho}$	Elem	$\hat{\rho}$	Elem	$\hat{\rho}$
Si	0.54	S	0.32	Ni	0.26	OC	0.76
Al	0.60	SULF	0.29	V	0.17	BC	0.35
Ti	0.54	Se	0.46			EC	0.41
Ca	0.49	Br	0.45				

Table 4: Summary of Meteorological Data

Covariate	Mininum	25th Perc	Median	75th Perc	Maximum
Wind Speed	0.000	6.043	8.695	12.000	21.000
Pressure	29.230	29.870	30.020	30.160	30.590
Relative Humidity	30.000	48.000	60.000	75.000	100.000
Temperature (^{o}C)	-8.889	2.778	14.444	22.511	32.222

by the temporal correlation in the underlying sources.

In the interest of examining the impact of meteorology on source activity, we obtained data on wind speed, pressure, relative humidity, and temperature in degrees Celcius in the Boston area for the dates corresponding to our exposures from the National Climatic Data Center (available online at *http://www.ncdc.noaa.gov/oa/ncdc.html*). Table 4 summarizes the meteorological data.

3 Model and Notation

3.1 Modeling Framework

Following Wolbers and Stahel (2005; and Billheimer 2001), we specify a multiplicative factor analysis model to describe the relationship between the unobserved source activity and the observed elemental concentrations. Let $\mathbf{Y}_{ij}$ be the $(P \times 1)$ vector of *non-negative* elemental concentrations and let $\boldsymbol{\eta}_{ij}$ be the $(K \times 1)$ *non-negative* source contributions $(K < P)$ for day j in block i $(j = 1, ..., n_i$ and $i = 1, ..., N_B)$. We assume

$$\mathbf{Y}_{ij} = (\boldsymbol{\Lambda}\boldsymbol{\eta}_{ij}) \circ \boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}, \tag{1}$$

where $\mathbf{\Lambda}$ is the factor pattern, the columns of which $\{\boldsymbol{\lambda}_k^{(c)}\}$ represent the source profiles, $\boldsymbol{\eta}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}$ follow log-normal distributions, and $\circ$ denotes elementwise multiplication, such that

$$log(\mathbf{Y}_{ij}) = log(\mathbf{\Lambda}\boldsymbol{\eta}_{ij}) + log(\boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}). \tag{2}$$

where $log(\boldsymbol{\eta}_{ij}) \overset{iid}{\sim} MVN_K(\boldsymbol{\mu}_{\boldsymbol{\eta}}, \mathbf{\Sigma}_{\boldsymbol{\eta}})$, $log(\boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}) \overset{iid}{\sim} MVN_P(\mathbf{0}, \mathbf{\Psi})$, $\mathbf{\Sigma}_{\boldsymbol{\eta}}$ and $\mathbf{\Psi}$ are diagonal, and $\boldsymbol{\eta}_{ij} \perp \boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}$.

We propose a mixed model (Diggle *et al.* 2002) on the log of the source contributions; $\forall k = 1, ..., K$,

$$log(\eta_{ijk}) = \mathbf{X}_{ij}^T\boldsymbol{\alpha}_k + \mathbf{Z}_{ij}^T\mathbf{b}_{ik} + \varepsilon_{ijk}^{\eta}, \tag{3}$$

where $\mathbf{b}_{ik} \overset{iid}{\sim} MVN(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{b}_k})$, $\varepsilon_{ijk}^{\eta} \overset{iid}{\sim} N(0, \sigma_{\eta_k}^2)$, and $\varepsilon_{ijk}^{\eta} \perp \mathbf{b}_{ik}$. Also, $\forall k \neq k'$, $\mathbf{b}_{ik} \perp \mathbf{b}_{ik'}$, and $\varepsilon_{ijk}^{\eta} \perp \varepsilon_{ijk'}^{\eta}$. This formulation is very flexible and allows for systematic effects as well as random effects on the the source contributions.

Using this framework, we can explore the influence of meteorology on source activity through the fixed effects, $\boldsymbol{\alpha}$. In addition, we can adjust for residual correlation in the source contributions due to the clustered study design by specifying a random effect b_i for each block i of consecutive days. Finally, we specify unique fixed effects, $\boldsymbol{\alpha}_k$, and unique random effects $\mathbf{b}_{ik}$ for each source k, so that we can estimate separate covariate effects and correlations for the different underlying pollution sources.

3.2 Model Identifiability

The source apportionment model specified in (1) and (2) is not identifiable without further assumptions. Because the source profiles are unknown *and* the source contributions are unobserved, the factor analysis model does not have a unique solution. However, the model may be made identifiable by constraining parameters in $\mathbf{\Lambda}$. We consider the following set of identifiability conditions, which result in a confirmatory, rather than exploratory, factor analysis (Park, Spiegelman, and Henry 2002).

C1: There are at least $K-1$ zero elements in each column of $\mathbf{\Lambda}$

C2: The rank of $\mathbf{\Lambda}^{(k)}$ is $K-1$, where $\mathbf{\Lambda}^{(k)}$ is the matrix composed of the rows containing the assigned 0s in the kth column with those assigned 0s deleted.

C3: $\lambda_{pk} = 1$ for some p $(p = 1, 2, .., P)$ for each $k = 1, 2, .., K$

The C1-C3 conditions are sufficient but not necessary to establish identifiability. While there exist alternative conditions, other commonly used proposals are also sufficient but not necessary. For instance, Park, Spiegelman, and Henry (2002) proposed sufficient conditions which, instead of placing constraints on the factor loadings, assume that some sources are absent on some days. These authors argued that in some settings, this alternative set of constraints may be plausible if one knows that a particular source, such as a power plant in the region, has been shut down for some period of time. In the same vein, Bandeen-Roche (1994) considered situations in which a subset of the source contributions is known. In our setting, however, we do not have information on the presence or absence of a particular source on a particular day. Thus, given the existing literature on the pollution mixture in the Boston area (Oh *et al.* 1997), it seems safer to assume that certain elements are not markers for certain sources. One important result of the C1-C3 identifiability constraints is that the scale of each factor is now on the scale of the element whose factor loading is constrained to one.

3.3 Model Implications

When compared to the standard additive factor analysis formulation, the modeling framework described above leads to different interpretation of the model parameters. First, the multiplicative error structure on the factor analytic model, in conjuction with our distributional assumptions, implies that, conditional on the source contributions, the elemental concentrations are log-normally distributed. In our application, this result is an advantage of our modeling framework in that the HSPH data suggest that the observed elements and compounds more closely follow a log-normal distribution as opposed to a normal distribution (Figure 1). The degree of skew in the empirical distributions is reflected in the specific variances, with larger values of ψ_p indicating a higher degree of skew in the corresponding element, Y_p. One drawback, however, is that the log-normal distribution does not directly allow for null concentrations.

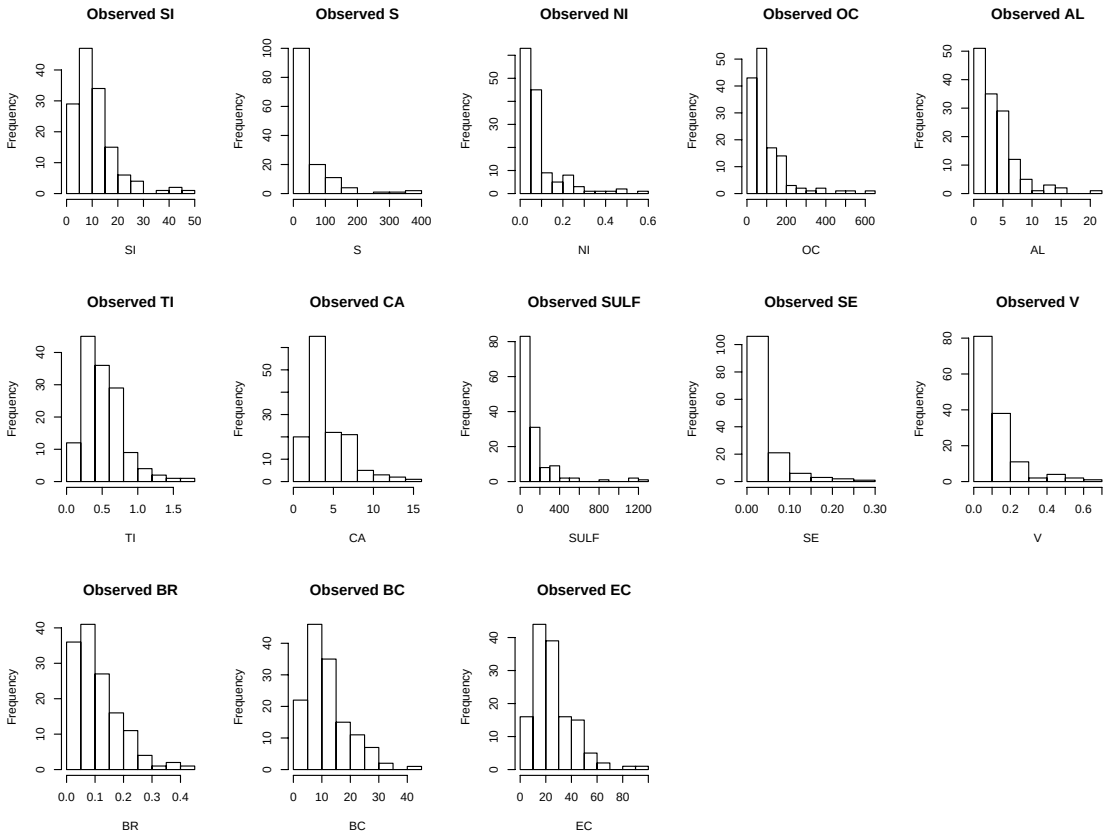

Figure 1: Histograms of the Measured Chemical Components

The assumptions of multiplicative error and log-normality also have important implications on the covariance structure of our measured outcomes. In the usual factor analysis model with additive normal error,

$$\mathbf{Y}_{ij} = \mathbf{\Lambda}\boldsymbol{\eta}_{ij} + \boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}},$$

where $\boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}} \overset{iid}{\sim} MVN_P(\mathbf{0}, \mathbf{\Psi})$, for diagonal $\mathbf{\Psi}$. In this framework, the variances and covariances of the measured components are simple functions of the factor loadings, $\{\lambda_{pk}\}$, and the variance-covariance of the latent sources, $V(\boldsymbol{\eta}_{ij})$.

$$Var(Y_{ijp}) = \boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_p^{(r)} + \psi_p \tag{4}$$

$$Cov(Y_{ijp}, Y_{ijp'}) = \boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_{p'}^{(r)} \tag{5}$$

Expressions (4) and (5) imply that the covariance (or correlation) between the measured components is explained by the latent sources and any additional variation is residual and unique to each element.

In the new modeling framework with multiplicative log-normal errors, the corresponding variances and covariances have more complicated forms (see Appendix A for derivation of the following expressions), with

$$Var(Y_{ijp}) = (\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}))^2 (e^{2\psi_p} - e^{\psi_p}) + (\boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_p^{(r)})(e^{2\psi_p}) \tag{6}$$

$$Cov(Y_{ijp}, Y_{ijp'}) = (e^{\frac{\psi_p + \psi_{p'}}{2}})(\boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_{p'}^{(r)}). \tag{7}$$

As demonstrated in (6), the variance of each element p is now a complicated function of its specific variance, ψ_p, and depends on the mean of the latent sources, $E(\boldsymbol{\eta}_{ij})$, as well as $V(\boldsymbol{\eta}_{ij})$. According to (7), the covariance between two measured elements now depends on the specific variances, such that the correlation between the measured components is no longer entirely explained by the latent sources. Finally, (6) and (7) indicate that both $V(Y_{ijp})$ and $Cov(Y_{ijp}, Y_{ijp'})$ are increasing functions of ψ. The variance, however, increases at a faster rate than the covariance, such that $Cor(Y_{ijp}, Y_{ijp'})$ is a decreasing function in ψ.

From a practical standpoint, the additive and multiplicative factor analysis models have the potential to yield quite different results. The methods deviate in terms of the estimated factor pattern, depending on the degree of skew in the measured components. Recall that elements that are highly skewed have relatively large ψ's. In the additive framework, the ψ's do not impact the covariance between any two elements; however, in the multiplicative framework, the ψ's play an important role.

Expression (7) indicates an interplay of the ψ's and the λ's in the multiplicative model. Specifically, as ψ_p increases, the λ's corresponding to pth element will decrease. Therefore, for elements that are highly skewed, and thus have large ψ's, the factor loadings corresponding to these elements will be reduced in the multiplicative framework as compared to the additive framework.

In contrast, as $\psi_p \to 0$,

$$Var(Y_{ijp}) \to \boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_p^{(r)},$$

and as $\psi_{p'} \to 0$,

$$Cov(Y_{ijp}, Y_{ijp'}) \to \boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_{p'}^{(r)}$$

in both the additive and multiplicative models. Hence, when the specific variances are all zero, the variances and covariances are identical in the two modeling frameworks.

3.4 Estimation

We take a Bayesian approach to estimation, and fit the multiplicative factor analysis with underlying mixed model using Markov Chain Monte Carlo (MCMC) methods in WinBUGS (Spiegelhalter, Thomas, and Best 2000). We specify normal priors on the fixed effects $\boldsymbol{\alpha}_k$ in the mixed model for $k = 1, 2, ..., K$, and inverse gamma priors on all variance parameters in $\boldsymbol{\Sigma}_{\mathbf{b}_k}$ (for $k = 1, 2, ..., K$), $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$, and $\boldsymbol{\Psi}$. To ensure non-negativity, and because negative components of source profiles are not interpretable, we define log-normal priors on the unconstrained parameters in $\boldsymbol{\Lambda}$.

In setting the hyperparameters of the prior distributions, we urge caution when working with the log-normal distribution. A log-normally distributed variable, say Z, is parameterized in terms of the underlying normal distribution, i.e.

$$log(Z) \sim N(\mu_Z, \sigma_Z^2),$$

where the moments of Z are

$$E(Z) = e^{\mu_Z + \frac{\sigma_Z^2}{2}}$$

$$V(Z) = e^{2(\mu_Z + \sigma_Z^2)} - e^{2\mu_Z + \sigma_Z^2}$$

(Casella and Berger 1990). Likewise, the hyperparameters for a log-normal prior distribution are generally specified in terms of the underlying normal moments. Therefore, when specifying a vague log-normal prior, it may seem natural to choose the typical settings for a vague normal prior. However, a vague normal prior distribution does not translate to a vague log-normal prior distribution. Consider, for instance, the typical vague normal prior with mean zero and large variance. Setting a large variance on the underlying normal distribution impacts the mean of the corresponding log-normal on the exponential scale, inducing a huge degree of skew and effectively leading to a very informative log-normal prior. Given that the log-normal is a skewed distribution whose moments both depend on the mean and variance of the underlying normal distribution, defining a vague log-normal prior is not straightforward.

A similar issue in prior specification arises in the context of logistic regression. Bedrick *et al.* (1997) noted that for a logistic model, a normal prior for $\boldsymbol{\beta}$ is convenient in large sample situations in which the posterior is approximately normal. If the sample size is not large, however, one should be cautious about using a normal prior with large covariances, as the induced prior distributions for $Pr(y_i = 1)$ can have point masses at zero and one. In such cases, Bedrick *et al.* (1996) proposed using a conditional means prior, which specifies the prior distribution on the success probabilities directly, and sugggested that it may be preferable to use an analogous strategy in the lognormal setting as well.

Fortunately, for the purposes of our research, the issues pertaining to the log-normal distribution are not prohibitive, since we can specify a log-normal with high density in the reasonable range of our parameter space. For instance, we specify a log-normal prior distribution on the unconstrained parameters in $\boldsymbol{\Lambda}$.

Previous analyses have consistently estimated the majority of factor loadings within the range $(0, 1)$

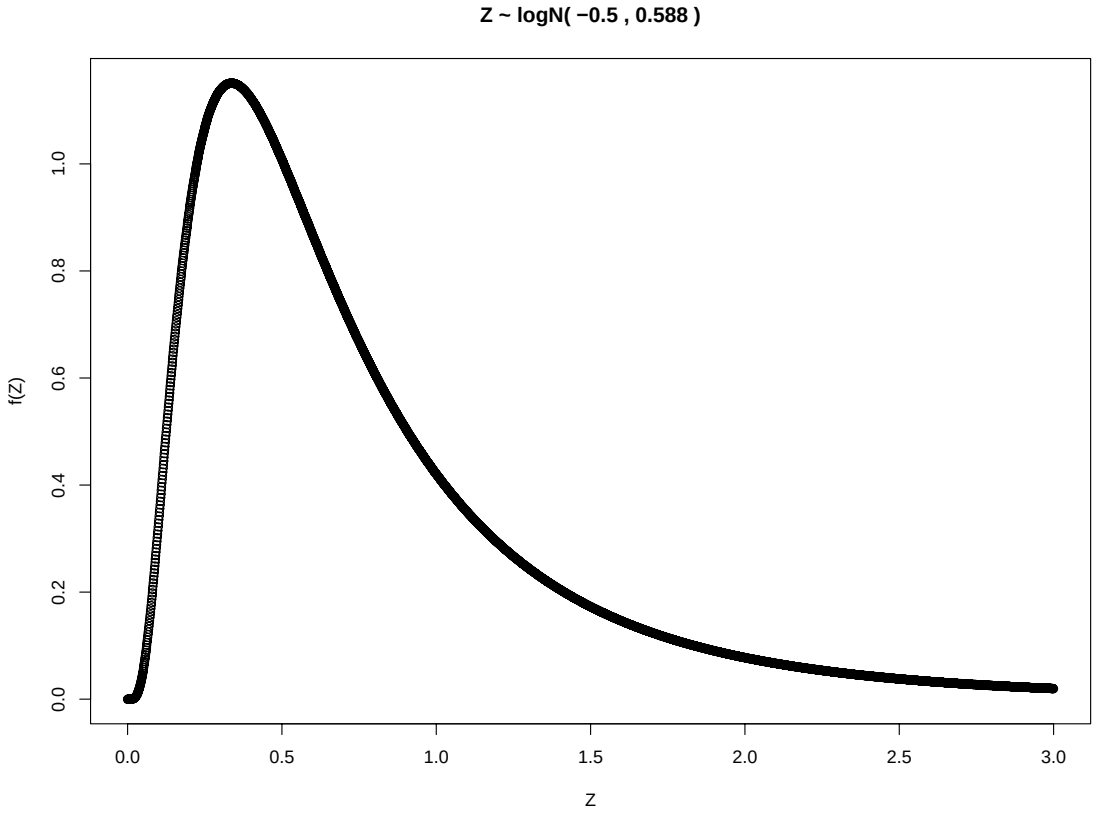

Figure 2: Log-Normal Prior on the Unconstrained Factor Loadings

and rarely do the estimates exceed 2. Given these prior results, our strategy for specifying the log-normal priors on the $\{\lambda_{pk}\}$ is to choose hyperparameters that yield a density that effectively assigns prior weight in the interval $(0, 1)$. For example, Figure 2 provides an example of such a log-normal distribution that would be a reasonable prior for the $\{\lambda_{pk}\}$ in our analysis. One may criticize that the log-normal prior is not particularly vague in the sense that it is not flat even in the reasonable range of our parameter space, and may instead argue for a uniform distribution. However, we prefer the log-normal prior distribution over a uniform prior distribution, because the log-normal prior allows the parameters to move beyond the reasonable range, while a uniform prior imposes a strict boundary on the parameter space.

Finally, we use the Bayesian *Deviance Information Criterion* (DIC; Spiegelhalter *et al.* 2002) to compare models. Like other information criterion, the DIC compares likelihoods from two compet-

ing models after adjusting for the "effective" number of parameters in the model, with this number depending on the model parameter priors. Let $\Theta = \{\boldsymbol{\Lambda}, \boldsymbol{\mu_\eta}, \boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_K, \boldsymbol{\Psi}, \boldsymbol{\Sigma_\eta}, \boldsymbol{\Sigma}_{\mathbf{b}_1}, ..., \boldsymbol{\Sigma}_{\mathbf{b}_K}\}$ denote the fixed parameters, and let $\mathbf{u} = \{\mathbf{b}_1, ..., \mathbf{b}_K\}$ denote the random effects. This criterion takes the form

$$DIC = D(\overline{\Theta}, \overline{\mathbf{u}}) + 2 * p_D,$$

where $D(\Theta, \mathbf{u}) = -2\log\left[p\left(\mathbf{y}|\boldsymbol{\mu}_\mathbf{y}\right)\right] + 2\log\left[f\left(\mathbf{y}|\boldsymbol{\mu}_\mathbf{y} = \mathbf{y}\right)\right]$ is the usual deviance, $\mathbf{y}$ represents the observed chemical concentrations, $\boldsymbol{\mu}_\mathbf{y}$ is the mean vector for $\mathbf{y}$, $\overline{\Theta}$, $\overline{\mathbf{u}}$, and $\overline{D(\Theta, \mathbf{u})}$ denote the posterior means of the model parameters and deviance, respectively, and $p_D = \overline{D(\Theta, \mathbf{u})} - D(\overline{\Theta}, \overline{\mathbf{u}})$. All necessary quantities can be calculated from the posterior samples generated by the MCMC model fitting.

4 Simulation Study

We conducted a simulation study to evaluate whether adjusting for covariates and temporal correlation improved estimation of the source profiles and the source activities. We also examined the impact of ignoring temporal correlation in assessing covariate effects on the latent source contributions.

In order to make our findings most relevant to the HSPH PM studies, we based our simulations on the $K = 4$ known sources of Boston pollution described in Table 1. To obtain realistic settings for the parameters in our model, we conducted a preliminary analysis of the complete exposure data ($N = 139$) described in Section 2.2. Since convergence problems are common when elemental concentrations are on widely different scales, each element was scaled by its sample standard deviation, which is equivalent to conducting a factor analysis on the sample correlation matrix, as opposed to the sample covariance matrix.

$$\mathbf{\Lambda} = \begin{pmatrix} & RoadDust & PowerPl & OilComb & Vehicles \\ Si & \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} & \lambda_{1,4} \\ S & \lambda_{2,1} & 1 & 0 & \lambda_{2,4} \\ Ni & 0 & 0 & 1 & 0 \\ OC & \lambda_{4,1} & \lambda_{4,2} & 0 & 1 \\ Al & 1 & 0 & \lambda_{5,3} & 0 \\ Ti & \lambda_{6,1} & \lambda_{6,2} & \lambda_{6,3} & \lambda_{6,4} \\ Ca & \lambda_{7,1} & \lambda_{7,2} & \lambda_{7,3} & \lambda_{7,4} \\ SULF & \lambda_{8,1} & \lambda_{8,2} & 0 & 0 \\ Se & 0 & \lambda_{9,2} & \lambda_{9,3} & \lambda_{9,4} \\ V & \lambda_{10,1} & \lambda_{10,2} & \lambda_{10,3} & \lambda_{10,4} \\ Br & \lambda_{11,1} & \lambda_{11,2} & \lambda_{11,3} & \lambda_{11,4} \\ BC & \lambda_{12,1} & \lambda_{12,2} & \lambda_{12,3} & \lambda_{12,4} \\ EC & 0 & 0 & \lambda_{13,3} & \lambda_{13,4} \end{pmatrix} \tag{8}$$

Our simulation parameters were based on the following multiplicative factor analysis of the standarized exposure data. First, to identify the model, we constrained parameters in $\mathbf{\Lambda}$ according to the C1-C3 identifiability conditions (Park, Spiegelman, and Henry 2002) as defined in (8). Here, aluminum, sulfur, nickel, and organic carbon identify road dust, power plants, oil combustion, and motor vehicles, respectively.

As for the mixed model on source activity, we simulated the effect of a single covariate, X (temperature in degrees Celcius), on all sources except road dust as well as the source-specific correlations, i.e. $\forall\ k \neq 1$,

$$log(\eta_{ijk}) = \mu_{\eta_k} + \alpha_k * X_{ij} + b_{ik} + \varepsilon^{\eta}_{ijk}$$

and, for $k = 1$,

$$log(\eta_{ij1}) = \mu_{\eta_1} + b_{i1} + \varepsilon^{\eta}_{ij1}.$$

The parameters settings for the simulation study may be found in Appendix B. Covariate data for the simulation study was pulled in blocks of three consecutive days from the meteorological data described in Section 2.2.

Each simulated data set consisted of 50 blocks of three days for a total of 150 unique exposures. For each source k, we generated 50 random block effects and 150 errors from univariate normals, $b_{ik} \sim N(0, \sigma^2_{b_k})$ and $\varepsilon^{\eta}_{ijk} \sim N(0, \sigma^2_{\eta_k})$, respectively. The log contribution for source k, block i, day j was computed as

$$log(\eta_{ijk}) = \mu_{\eta_k} + \alpha_k X_{ij} + b_{ik} + \varepsilon^{\eta}_{ijk}.$$

To simulate observed exposures, we generated log errors on the elemental concentrations from a multivariate normal distribution, $log(\boldsymbol{\varepsilon}^{\mathbf{Y}}_{ij}) \sim MVN(\mathbf{0}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is $diag(\boldsymbol{\psi})$. The vector of elemental concentrations for day i, block j was computed as

$$\mathbf{Y}_{ij} = (\boldsymbol{\Lambda}\boldsymbol{\eta}_{ij}) \circ \boldsymbol{\varepsilon}^{\mathbf{Y}}_{ij}$$

where $\circ$ denotes elementwise multiplication.

For each simulated data set, we fit three different models. All three models assumed the multiplicative factor analytic structure,

$$\mathbf{Y}_{ij} = (\boldsymbol{\Lambda}\boldsymbol{\eta}_{ij}) \circ \boldsymbol{\varepsilon}^{\mathbf{Y}}_{ij}$$

where $log(\boldsymbol{\varepsilon}^{\mathbf{Y}}_{ij}) \overset{iid}{\sim} MVN_P(\mathbf{0}, \boldsymbol{\Psi})$ for diagonal $\boldsymbol{\Psi}$. The models differed with respect to the mixed model on the latent source contributions.

1. ***Simple Multiplicative Factor Analysis:*** For each source k $(k = 1, 2, .., K)$,

   $$log(\eta_{ijk}) = \mu_{\eta_k} + \varepsilon^{\eta}_{ijk}$$

   where $\varepsilon^{\eta}_{ijk} \overset{iid}{\sim} N(0, \sigma^2_{\eta_k})$ and, $\forall k \neq k'$, $\varepsilon^{\eta}_{ijk} \perp \varepsilon^{\eta}_{ijk'}$.

2. ***Multiplicative FA with Temperature Effects but Ignoring Temporal Correlation:*** For each source k $(k = 1, 2, .., K)$,

   $$log(\eta_{ijk}) = \mu_{\eta_k} + \alpha_k X_{ij} + \varepsilon^{\eta}_{ijk}$$

where $\varepsilon^{\eta}_{ijk} \overset{iid}{\sim} N(0, \sigma^2_{\eta_k})$ and, $\forall k \neq k'$, $\varepsilon^{\eta}_{ijk} \perp \varepsilon^{\eta}_{ijk'}$.

3. ***Multiplicative FA with Temperature Effects and Adjusting for Temporal Correlation:*** For each source k $(k = 1, 2, .., K)$,

$$log(\eta_{ijk}) = \mu_{\eta_k} + \alpha_k X_{ij} + b_{ik} + \varepsilon^{\eta}_{ijk}$$

where $b_{ik} \overset{iid}{\sim} N(0, \sigma^2_{b_k})$, $\varepsilon^{\eta}_{ijk} \overset{iid}{\sim} N(0, \sigma^2_{\eta_k})$, $b_{ik} \perp \varepsilon^{\eta}_{ijk}$, and, $\forall k \neq k'$, $b_{ik} \perp b_{ik'}$ and $\varepsilon^{\eta}_{ijk} \perp \varepsilon^{\eta}_{ijk'}$.

In all models, the prior distributions were defined as follows. We specified $logN(-0.5, 0.588)$ prior distributions (shown in Figure 2) on the unconstrained $\{\lambda_{pk}\}$. We specified $N(0, 10)$ prior distributions on the $\{\mu_{\eta_k}\}$ and the $\{\alpha_k\}$. Recall that while these settings may not seem vague on the normal scale, these parameters define the mean of the η_{ijk}'s which are log-normally distributed. Finally, we specified $IG(0.01, 0.01)$ priors distributions on the variance parameters, $\{\psi_p\}$, $\{\sigma^2_{\eta_k}\}$, and $\{\sigma^2_{b_k}\}$.

All models were fit using a Bayesian Markov Chain Monte Carlo (MCMC) algorithm as implemented in WinBUGS (Spiegelhalter, Thomas, and Best 2000). For each model fit, we ran 25,000 iterations, discarding 20,000 as burn-in and thinning by five, for a total of 1,000 posterior samples. In several test runs, we examined diagnostic trace and autocorrelation plots and found satisfactory convergence.

We ran 100 simulations to evaluate the effect of ignoring covariate effects and temporal correlation in the estimation of the source profiles and the source contributions, and to examine the impact of ignoring temporal correlation on the estimated covariate effects. Table 5 displays the estimated source profiles obtained with the different source apportionment models. We find no major differences in estimated source profiles and all models appear to estimate the source profiles well.

To evaluate model performance with respect to source activity, for each simulation we computed

Table 5: Estimated Source Profiles

	Model	Si	S	Ni	OC	Al	Ti	Ca	SULF	Se	V	Br	BC	EC
RD	Truth	0.98	0.01	0	0.06	1	1.02	0.99	0.01	0	0.18	0.09	0.04	0
	(1)	0.99	0.02	0	0.09	1	1.04	0.98	0.02	0	0.17	0.13	0.07	0
	(2)	0.99	0.02	0	0.09	1	1.03	0.98	0.02	0	0.17	0.13	0.07	0
	(3)	0.99	0.02	0	0.09	1	1.03	0.98	0.02	0	0.17	0.13	0.07	0
PP	Truth	0.11	1	0	0.40	0	0.14	0.03	0.9	0.52	0.03	0.82	0.46	0
	(1)	0.13	1	0	0.45	0	0.17	0.09	0.9	0.50	0.07	0.82	0.47	0
	(2)	0.13	1	0	0.44	0	0.17	0.09	0.9	0.49	0.07	0.82	0.46	0
	(3)	0.13	1	0	0.43	0	0.17	0.09	0.9	0.49	0.07	0.82	0.46	0
OC	Truth	0.06	0	1	0	0	0.28	0.30	0.00	0.03	0.90	0.35	0.45	0.43
	(1)	0.07	0	1	0	0	0.31	0.31	0.01	0.08	0.85	0.36	0.46	0.44
	(2)	0.08	0	1	0	0	0.31	0.31	0.01	0.08	0.85	0.37	0.47	0.47
	(3)	0.08	0	1	0	0	0.31	0.32	0.01	0.08	0.85	0.37	0.47	0.47
MV	Truth	0.30	0.02	0	1	0	0.49	0.46	0	0.24	0.05	0.46	1.10	1.79
	(1)	0.31	0.04	0	1	0	0.50	0.46	0	0.26	0.10	0.50	1.21	2.05
	(2)	0.30	0.04	0	1	0	0.48	0.44	0	0.26	0.10	0.48	1.16	1.94
	(3)	0.30	0.03	0	1	0	0.48	0.44	0	0.26	0.10	0.48	1.16	1.93

the sum of squared errors in the estimated source contributions,

$$SSE_m = \sum_{i=1}^{50} \sum_{j=1}^{3} \sum_{k=1}^{4} (\eta_{ijk} - \widehat{\eta}_{ijk}^m)^2$$

where η_{ijk} is the simulated source contribution and $\widehat{\eta}_{ijk}^m$ is the estimated source contribution based on model m ($m = 1, 2, 3$) for block i, day j, source k. Table 6 summarizes the average SSE (over the 100 simulations) for each model as well as the number of times each model has the largest SSE, indicating worst performance, and the number of times each model had the smallest SSE, indicating best performance. As expected, the simple model (1), excluding the covariate and ignoring temporal correlation, had the overall worst performance with the largest average SSE as well as the largest SSE 78% of the time, whereas the complete model (3), adjusting for covariate and temporal correlation, had the best overall performance, with the smallest average SSE as well as the smallest SSE 71% of the time.

Finally, to evaluate the impact of ignoring temporal correlation on the covariate effects, we compare model (2) which ignores the temporal correlation in the data, and model (3) that appropriately adjusts for the correlation. Tables 7 and 8 summarize the results for models (2) and (3), respectively.

Table 6: Model Comparison Based on SSE in the Source Contributions

Model	Includes Covariate	Random Effect for Block (of Days)	Average SSE	Count with Largest SSE	Count with Smallest SSE
(1)	No	No	32.53	78	5
(2)	Yes	No	31.08	11	24
(3)	Yes	Yes	30.13	11	71

Table 7: Estimated Covariate Effects Ignoring Temporal Correlation

Source	α_k	Mean $\widehat{\alpha}_k$	Reject H_0: $\alpha_k = 0$	$\alpha_k \in$ 95% CI
Road Dust	0	-0.0010	20	80
Power Plants	0.04	0.0432	100	88
Oil Combustion	-0.03	-0.0301	93	83
Motor Vehicles	0.03	0.0305	100	88

The tables show that the different models provide similar effect estimates that are close to the truth, but differ with respect to inference as demonstrated by the 95% credible intervals. First, we incorrectly reject the null hypothesis, $H_0 : \alpha_{RD} = 0$, 20% of time for model (2), but only 6% of the time for model (3). This finding indicates that failing to adjust for the temporal correlation in the exposures inflates the type I error rate. Secondly, the complete model provides better coverage of the true parameters. The proportion of 95% credible intervals that contain the true α's corresponding to power plants, oil combustion, and motor vehicles are 98%, 94%, and 96%, respectively for model (3), whereas the coverages for model (2) are 88%, 83%, and 88%, respectively. Lastly, model (3) provides estimates of the source-specific correlations that are very close to the truth.

In summary, incorporating covariate effects on the source contributions and adjusting for temporal correlation did not appear to impact estimation of the source profiles. This more elaborate source characterization did, however, improve estimation of the source activity. Furthermore, while failure to adjust for temporal correlation in exposures measured on consecutive days may not influence the effect estimates, it does affect inference. Methods that ignore the temporal correlation are too liberal, whereas methods that adjust appropriately are of the correct size. Additionally, the correctly

Table 8: Estimated Covariate Effects Adjusting for Temporal Correlation

Source	α_k	Mean $\widehat{\alpha}_k$	Reject H_0: $\alpha_k = 0$	$\alpha_k \in$ 95% CI	ρ_k	Mean $\widehat{\rho}_k$
Road Dust	0	-0.0010	6	94	0.54	0.5278
Power Plants	0.04	0.0431	100	98	0.32	0.3285
Oil Combustion	-0.03	-0.0301	87	94	0.46	0.4449
Motor Vehicles	0.03	0.0305	100	96	0.60	0.6046

specified model provides better coverage of the true effects.

5 Data Analysis

In this section, we apply our mixed multiplicative factor analysis model defined in (1), (3), and (8) to analyze the complete standardized HSPH PM exposure data (N=139). We conducted these analyses with two objectives in mind. First, to evaluate the multiplicative formulation, we fit the simple additive and multiplicative factor analysis models, excluding covariates and ignoring temporal correlation. Second, to explore the influence of meteorology on source activity, we fit a suite of models with various combinations of the meteorological covariates (described in Section 2.2), in all cases adjusting for temporal correlation with the random effect for block of consecutive days.

All models were fit via MCMC methods in WinBUGS (Spiegelhalter, Thomas, and Best 2000), as motivated in Section 2.3.4. We defined $logN(-0.5, 0.588)$ priors on the unconstrained $\{\lambda_{pk}\}$, $N(0, 100)$ priors on $\{\mu_{\eta_k}\}$ and $\{\boldsymbol{\alpha}_k\}$, and $IG(0.01, 0.01)$ priors on $\{\psi_p\}$, $\{\sigma^2_{\eta_k}\}$, and $\{\sigma^2_{b_k}\}$. We ran 80,000 iterations, discarding 70,000 as burn-in and thinning by ten, for a total of 1,000 posterior samples. We examined diagostic trace and autocorrelaion plots and found satisfactory convergence.

Table 9 provides a summary of the models that were fit as well as the DIC for each model. We use the DICs to compare the model fits, where a smaller DIC indicates better model fit to the data. First, the simple multiplicative model (0b) has a DIC of 943.945 compared to the simple additive model (0a) which has a DIC of 1145.790. This finding suggests that the factor analysis with multiplicative error describes the exposure data better than the standard additive factor analysis model. Secondly,

Table 9: Summary of Model Fits to PM Exposure Data^{+}

Model	Error Structure	DIC	Temp Block	Wind Speed	Pressue	Relative Humidity	Temp (^{o}C)*
(0a)	Add	1145.790	No				
(0b)	Mult	943.945	No				
(0c)	Mult	948.105	Yes				
(1a)	Mult	955.404	Yes	X			
(1b)	Mult	947.381	Yes		X		
(1c)	Mult	960.620	Yes			X	
(1d)	Mult	950.551	Yes				X
(2a)	Mult	943.358	Yes	X	X		
(2b)	Mult	945.271	Yes	X		X	
(2c)	Mult	930.028	Yes	X			X
(2d)	Mult	954.744	Yes		X	X	
(2e)	Mult	947.806	Yes		X		X
(2f)	Mult	940.065	Yes			X	X
(3a)	Mult	935.230	Yes	X	X	X	
(3b)	Mult	950.162	Yes	X	X		X
(3c)	Mult	944.670	Yes	X		X	X
(3d)	Mult	942.925	Yes		X	X	X
(4)	Mult	937.850	Yes	X	X	X	X

$+$ separate effects for each pollution source

$*$ quadratic effect of temperature in degrees Celcius

the multiplicative model that adjusts for the temporal correlation but does not include covariates (0c) has a larger DIC (948.105) than the simple multiplicative model (0b). While we would expect this adjustment to improve model fit, perhaps four separate correlations, one for each source, are unnecessary and fewer parameters would be sufficient. Finally, with respect to the models that include covariates, we find that model (2c), adjusting for wind speed, temperature in degrees Celcius, and temporal correlation, has the smallest DIC (930.028) and thus provides the best fit the exposure data.

The estimated covariate effects and corresponding credible intervals for each model may be found in Appendix C. In general, these models suggest the following relationships between meteorology and source activity. Oil combustion and motor vehicle contributions are negatively associated with wind speed. PM contributions from power plants, oil combustion, and motor vehicles are positively associated with relative humidity, whereas road dust contributions are negatively associated with relative humidity. Oil combustion contributions have negative linear relationship with temperature and motor vehicle contributions have a positive linear relationship with temperature; power plant contributions have a significant non-linear relationship with temperature in degrees Celcius.

For the purposes of model selection, we refit the model with the smallest DIC, model (2c), repeatedly, constraining non-significant effects to zero. Table 10 summarizes the posterior medians and 95% credible intervals for the final model on the source contributions.

The estimates of μ_{η_k} represent the mean log standardized contribution for each source when both wind speed and temperature are equal to zero. The estimated contributions for each source are on the scale of a standard deviation of the element constrained to 1 on the corresponding source profile. For instance, in (8), the loading for aluminum is constrained to 1 on the road dust profile. Therefore, given null wind speed and 0^oC, we estimate a road dust contribution of $exp(-0.3521) = 0.70$ standard deviations of aluminum. Likewise, under these same conditions, we would expect a power plant contribution of $exp(-1.3930) = 0.25$ standard deviations of sulfur, an oil combustion contribution of $exp(0.2267) = 1.25$ standard deviations of nickel, and a motor vehicle contribution of $exp(-0.6722) = 0.51$ standard deviations of organic carbon.

Table 10: Estimates of the Mixed Model Parameters in the Final Model

Source	$\widehat{\mu}_{\eta_k}$	Covariate	$\widehat{\alpha}_k$	95% Credible Interval	$\widehat{\rho}_k$
Road Dust	-0.3521	—	—	—	0.5474
Power Pl	-1.3930	Temp	-0.0149	(-0.0616 , 0.0277)	0.3030
		Temp^2	0.0023*	(0.0007 , 0.0039)	
Oil Comb	0.2267	Wind Speed	-0.0501*	(-0.0819 , -0.0164)	0.4691
		Temp	-0.0306*	(-0.0469 , -0.0149)	
Vehicles	-0.6722	Wind Speed	-0.0453*	(-0.0672 , -0.0250)	0.5427
		Temp	0.0346*	(0.0234 , 0.0460)	

With respect to the covariate effects, we found a significant decrease in both oil combustion and motor vehicle contributions associated with increasing wind speed. For instance, for a ten unit increase in wind speed, we estimate a drop in oil combustion PM by a factor of $exp(-0.501) = 0.61$ (standard deviations of nickel), and a drop in motor vehicle PM by a factor of $exp(-0.453) = 0.64$ (standard deviations of organic carbon). This finding may suggest that the wind effectively disperses the particles coming from these sources. We also estimate that oil combustion contributions decrease by a factor of $exp(-0.306) = 0.74$, while motor vehicle contributions increase by a factor of $exp(0.346) = 1.41$ for a 10^oC increase in temperature. These findings support less home heating and perhaps more traveling during warmer weather. Power plant activity has a significant parabolic relationship with temperature as demonstrated in Figure 3 with elevated exposures at low and high temperatures.

6 Discussion

In this paper, we considered methods to assess factors associated with pollution source activity. Our primary objectives were (1) to allow source contributions to depend on covariates and (2) to adjust for temporal correlation in the exposures. To meet these aims, we proposed a multiplicative factor analysis with a mixed model on the latent source contributions. The multiplicative error structure facilitated the modeling framework and also served to (3) respect the non-negativity of the measured

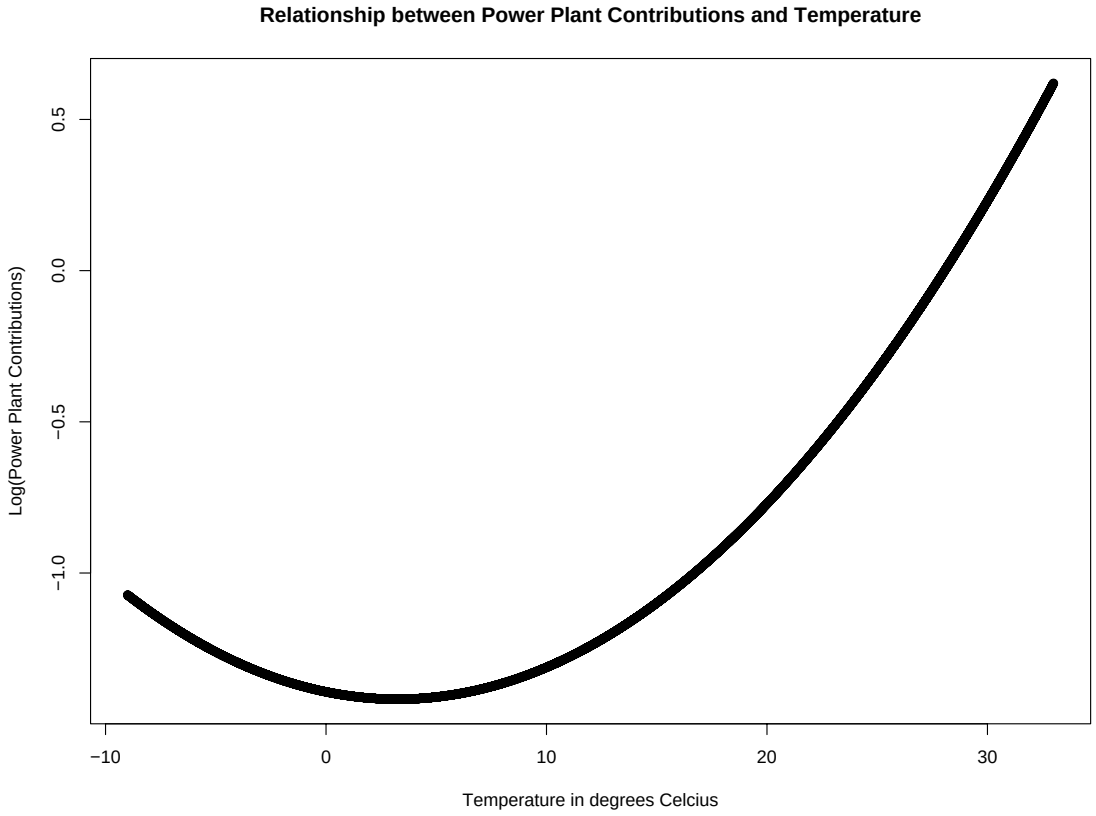

Figure 3: Quadratic Effect of Temperature on Power Plant Contributions

chemical concentrations. The mixed multiplicative factor analysis model extends the model proposed by Wolbers and Stahel (2005; and Billheimer 2001) by imposing mixed models on the unobserved source contributions and taking a Bayesian approach to model fitting.

We conducted a simulation study with two goals in mind; (1) to examine the impact of incorporating weather and design information on the estimation of the source profiles and the source activities, and (2) to assess the impact of ignoring temporal correlation on the estimated covariate effects. We demonstrated that including covariate effects and adjusting for temporal correlation does not have a large influence on the estimation of the source profiles, but does in general improve estimation of the source activities. This is an important advantage in the context of health effect studies since the quality of the estimated health effects depends on the estimation of the source-specific contributions from complex mixtures of exposure. We also showed that ignoring temporal correlation in the expo-

sures affects the tests of significance for the covariate effects.

Using our proposed methods, we analyzed the HSPH exposure data. We found that the multiplicative factor analysis fits the data better than the standard additive factor analysis model. We also demonstrated how this new modeling framework enables us to explore the role of important factors, such as meteorological conditions and temporal correlation, that influence source dynamics.

One limitation of our modeling framework is that the assumption of log-normality does not directly allow for null chemical concentrations. Wolbers and Stahel (2005) propose adding a small non-negative vector into source apportionment model to handle zero concentrations. Alternatively, one could also consider extensions to incorporate left truncated concentrations, such as in limit of detection problems.

7 Acknowledgements

This research was supported by NIEHS grant ES07142 (MCN), American Chemistry Council grant 2843 and NIEHS grant ES012044 (BAC), and NIH grant ES012972 (JJG).

8 References

1. Aitchison, J., and Ho, C. (1989). The multivariate Poisson-log normal distribution. *Biometrika* **76**, 643–653.
2. Bandeen-Roche, K. (1994). Resolution of additive mixtures into source components and contributions: A compositional approach. *Journal of the American Statistical Association* **89**, 1450–1458.
3. Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association* **91**, 1450-1460.
4. Bedrick, E. J., Christensen, R., and Johnson, W. (1997). Bayesian binomial regression: predicting survival at a trauma center. *American Statistician* **51**, 211-218.
5. Billheimer, D. (2001). Compositional receptor modeling. *Environmetrics* **12**, 451–467.
6. Casella, G., and Berger, R. (1990). *Statistical Inference.* Belmont, CA: Duxbury Press.

7. Christensen, W. F. and Sain, S. R. (2002). Accounting for dependence in a flexible multivariate receptor model. *Technometrics* **44**, 328–337.

8. Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). *The Analysis of Longitudinal Data: 2nd Edition*. Oxford, England: Oxford University Press.

9. Kavouras, I. G., Koutrakis, P., Cereceda-Balic, F., and Oyola, P. (2001). Source apportionment of PM10 and PM2.5 in five Chilean cities using factor analysis. *Journal of the Air & Waste Management Association* **51**, 451–464.

10. Kim, E., Hopke, P. K., Larson, T. V., and Covert, D. S. (2004). Analysis of ambient particle size distributions using unmix and positive matrix factorization. *Environmental Science & Technology* **38**, 202–209.

11. Koutrakis, P. and Spengler, J. D. (1987). Source apportionment of ambient particles in Steubenville, Ohio using specific rotation factor analysis. *Atmospheric Environment* **21**, 1511–1519.

12. Nikolov, M. C., Coull, B. A., Catalano, P., and Godleski, J. J. (2006). An informative Bayesian structural equation model to assess source-specific health effects of air pollution.

13. Oh, J. A., Suh, H. H., Lawrence, J. E., Allen, G. A., and Koutrakis, P. (1997). Characterization of particulate mass concentrations in South Boston, MA. Proceedings of AWMA/EPA Symposium on "Measurement of Toxic and Related Air Pollutants", April 29-May 1, 1997, Research Triangle Park, NC. AWMA publication number VIP-74 (Pittsburgh, PA), 397–407.

14. Park, E. S., Guttorp, P., and Henry, R. C. (2001). Multivariate receptor modeling for temporally correlated data by using MCMC. *Journal of the American Statistical Assocation* **96**, 1171–1183.

15. Park, E. S., Spiegelman, C. H., and Henry, R. C. (2002). Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models. *Environmetrics* **13**, 775–798.

16. Seigneur, C., Pai, P., Hopke, P. K., and Grosjean, D. (1999). Modeling atmospheric: Particulate matter. *Environmental Science & Technology* **33**, 80A–86A.

17. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, **64**, 583–639.

18. Spiegelhalter, D., Thomas, A., and Best, N. (2000). *WinBUGS Version 1.3. User's Manual*. MRC Biostatistics Unit. Institute of Public Health, Cambridge. `http://www.mrc-bsu.cam.ac.uk/bugs`.

19. Wolbers, M. and Stahel, W. (2005). Linear unmixing of multivariate observations: A structural model. *Journal of the American Statistical Assocation* **100**, 1328–1342.

APPENDIX A

Derivation of Covariance Expressions

Multiplicative Factor Analysis Model:

$$\mathbf{Y}_{ij} = (\mathbf{\Lambda}\boldsymbol{\eta}_{ij}) \circ \boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}$$

Distributional Assumptions on the Errors:

$$log(\boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}}) \overset{iid}{\sim} MVN_P(\mathbf{0}, \mathbf{\Psi})$$
$$\mathbf{\Psi} = diag(\boldsymbol{\psi})$$

Properties of the Log-Normal Distribution (Casella and Berger 1990; Aitchison and Ho 1989):

$$log(x_1) = y_1 \overset{iid}{\sim} N(\mu_{y_1}, \sigma_{y_1}^2) \quad \Rightarrow \quad x_1 \overset{iid}{\sim} logN(\mu_{y_1}, \sigma_{y_1}^2)$$
$$log(x_2) = y_2 \overset{iid}{\sim} N(\mu_{y_2}, \sigma_{y_2}^2) \quad \Rightarrow \quad x_2 \overset{iid}{\sim} logN(\mu_{y_2}, \sigma_{y_2}^2)$$

$$E(x_i) = e^{\mu_{y_i} + \frac{\sigma_{y_i}^2}{2}}$$
$$Var(x_i) = e^{2(\mu_{y_i} + \sigma_{y_i}^2)} - e^{2\mu_{y_i} + \sigma_{y_i}^2}$$
$$Cov(x_1, x_2) = (e^{\sigma_{y_1, y_2}} - 1)(e^{\mu_{y_1} + \mu_{y_2} + \frac{\sigma_{y_1}^2 + \sigma_{y_2}^2}{2}})$$

Moments of the Log-Normal Errors:

$$E(\boldsymbol{\varepsilon}_i) = \begin{pmatrix} e^{\frac{\psi_1}{2}} \\ e^{\frac{\psi_2}{2}} \\ \vdots \\ e^{\frac{\psi_P}{2}} \end{pmatrix}, \quad V(\boldsymbol{\varepsilon}_i) = \begin{pmatrix} e^{2\psi_1} - e^{\psi_1} & & & \\ & e^{2\psi_2} - e^{\psi_2} & & \\ & & \ddots & \\ & & & e^{2\psi_P} - e^{\psi_P} \end{pmatrix}$$

Covariance Terms in the Multiplicative FA Model:

$$V(\mathbf{Y}_{ij}) = V((\mathbf{\Lambda}\boldsymbol{\eta}_{ij}) \circ \boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}})$$

$$V(\mathbf{Y}_{ij}) = V((\begin{pmatrix} \boldsymbol{\lambda}_1^{(r)T} \\ \boldsymbol{\lambda}_2^{(r)T} \\ \vdots \\ \boldsymbol{\lambda}_P^{(r)T} \end{pmatrix} \boldsymbol{\eta}_{ij}) \circ \boldsymbol{\varepsilon}_{ij}^{\mathbf{Y}})$$

Letting $\boldsymbol{\lambda}_p^{(r)T} = \begin{pmatrix} \lambda_{p1} & \lambda_{p2} & ... & \lambda_{pK} \end{pmatrix}$ represent the pth row in $\mathbf{\Lambda}$,

$$Var(Y_{ijp}) = Var((\boldsymbol{\lambda}_p^{(r)T}\boldsymbol{\eta}_{ij}) * \varepsilon_{ijp}^{Y})$$
$$Cov(Y_{ijp}, Y_{ijp'}) = Cov((\boldsymbol{\lambda}_p^{(r)T}\boldsymbol{\eta}_{ij}) * \varepsilon_{ijp}^{Y}, (\boldsymbol{\lambda}_{p'}^{(r)T}\boldsymbol{\eta}_{ij}) * \varepsilon_{ijp'}^{Y})$$

Variances:

$$Var(Y_{ijp}) = Var(E(Y_{ijp}|\varepsilon_{ijp}^{Y})) + E(Var(Y_{ijp}|\varepsilon_{ijp}^{Y}))$$

$$= Var(E((\boldsymbol{\lambda}_p^{(r)T}\boldsymbol{\eta}_{ij}) * \varepsilon_{ijp}^{Y}|\varepsilon_{ijp}^{Y})) + E(Var((\boldsymbol{\lambda}_p^{(r)T}\boldsymbol{\eta}_{ij}) * \varepsilon_{ijp}^{Y}|\varepsilon_{ijp}^{Y}))$$

$$= Var((\boldsymbol{\lambda}_p^{(r)T}E(\boldsymbol{\eta}_{ij})) * \varepsilon_{ijp}^{Y}) + E((\boldsymbol{\lambda}_p^{(r)T}V(\boldsymbol{\eta}_{ij})\boldsymbol{\lambda}_p^{(r)}) * (\varepsilon_{ijp}^{Y})^2)$$

$$= (\boldsymbol{\lambda}_p^{(r)T}E(\boldsymbol{\eta}_{ij}))^2 Var(\varepsilon_{ijp}^{Y}) + (\boldsymbol{\lambda}_p^{(r)T}V(\boldsymbol{\eta}_{ij})\boldsymbol{\lambda}_p^{(r)})E((\varepsilon_{ijp}^{Y})^2)$$

$$= (\boldsymbol{\lambda}_p^{(r)T}E(\boldsymbol{\eta}_{ij}))^2 Var(\varepsilon_{ijp}^{Y}) + (\boldsymbol{\lambda}_p^{(r)T}V(\boldsymbol{\eta}_{ij})\boldsymbol{\lambda}_p^{(r)})(Var(\varepsilon_{ijp}^{Y}) + (E(\varepsilon_{ijp}^{Y}))^2)$$

$$= (\boldsymbol{\lambda}_p^{(r)T}E(\boldsymbol{\eta}_{ij}))^2 (e^{2\psi_p} - e^{\psi_p}) + (\boldsymbol{\lambda}_p^{(r)T}V(\boldsymbol{\eta}_{ij})\boldsymbol{\lambda}_p^{(r)})(e^{2\psi_p} - e^{\psi_p} + (e^{\frac{\psi_p}{2}})^2)$$

Therefore,

$$Var(Y_{ijp}) = (\boldsymbol{\lambda}_p^{(r)T}E(\boldsymbol{\eta}_{ij}))^2 (e^{2\psi_p} - e^{\psi_p}) + (\boldsymbol{\lambda}_p^{(r)T}V(\boldsymbol{\eta}_{ij})\boldsymbol{\lambda}_p^{(r)})(e^{2\psi_p})$$

Covariances:

Let $\boldsymbol{\varepsilon}_{ij}^{Y^*} = \begin{pmatrix} \varepsilon_{ijp}^{Y} \\ \varepsilon_{ijp'}^{Y} \end{pmatrix}$, $W_p = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and $W_{p'} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Then, $W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*} = \varepsilon_{ijp}^{Y}$ and $W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*} = \varepsilon_{ijp'}^{Y}$.

$$Cov(Y_{ijp}, Y_{ijp'}) = Cov((\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij}) * \varepsilon_{ijp}^{Y}, (\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij}) * \varepsilon_{ijp'}^{Y})$$

$$= Cov((\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij})(W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*}), (\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij})(W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*}))$$

$$= E((\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij})(W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij})(W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*})) -$$
$$E((\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij})(W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*}))E((\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij})(W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*}))^T$$

$$= E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\varepsilon}_{ij}^{(Y^*)T} W_{p'})(\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij})(\boldsymbol{\eta}_{ij}^T \boldsymbol{\lambda}_{p'}^{(r)})) -$$
$$E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij}))E((W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij}))^T$$

$$= E(E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*} \boldsymbol{\varepsilon}_{ij}^{(Y^*)T} W_{p'})(\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij} \boldsymbol{\eta}_{ij}^T \boldsymbol{\lambda}_{p'}^{(r)}) | \boldsymbol{\varepsilon}_{ij}^{Y^*})) -$$
$$E(E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij}) | \boldsymbol{\varepsilon}_{ij}^{Y^*}))E(E((W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij}) | \boldsymbol{\varepsilon}_{ij}^{Y^*}))^T$$

$$= E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*} \boldsymbol{\varepsilon}_{ij}^{(Y^*)T} W_{p'})E((\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij} \boldsymbol{\eta}_{ij}^T \boldsymbol{\lambda}_{p'}^{(r)}) | \boldsymbol{\varepsilon}_{ij}^{Y^*})) -$$
$$E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})E((\boldsymbol{\lambda}_p^{(r)T} \boldsymbol{\eta}_{ij}) | \boldsymbol{\varepsilon}_{ij}^{Y^*}))E((W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*})E((\boldsymbol{\lambda}_{p'}^{(r)T} \boldsymbol{\eta}_{ij}) | \boldsymbol{\varepsilon}_{ij}^{Y^*}))^T$$

$$= E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*} \boldsymbol{\varepsilon}_{ij}^{(Y^*)T} W_{p'})(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij} \boldsymbol{\eta}_{ij}^T) \boldsymbol{\lambda}_{p'}^{(r)})) -$$
$$E((W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij})))E((W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_{p'}^{(r)T} E(\boldsymbol{\eta}_{ij})))^T$$

$$= E(W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*} \boldsymbol{\varepsilon}_{ij}^{(Y^*)T} W_{p'})(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij} \boldsymbol{\eta}_{ij}^T) \boldsymbol{\lambda}_{p'}^{(r)}) -$$
$$E(W_p^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}))(E(W_{p'}^T \boldsymbol{\varepsilon}_{ij}^{Y^*})(\boldsymbol{\lambda}_{p'}^{(r)T} E(\boldsymbol{\eta}_{ij})))^T$$

$$= (W_p^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*} \boldsymbol{\varepsilon}_{ij}^{(Y^*)T}) W_{p'})(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij} \boldsymbol{\eta}_{ij}^T) \boldsymbol{\lambda}_{p'}^{(r)}) -$$
$$(W_p^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*}))(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}))((W_{p'}^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*}))(\boldsymbol{\lambda}_{p'}^{(r)T} E(\boldsymbol{\eta}_{ij})))^T$$

$$= \{W_p^T (V(\boldsymbol{\varepsilon}_{ij}^{Y^*}) + E(\boldsymbol{\varepsilon}_{ij}^{Y^*}) E(\boldsymbol{\varepsilon}_{ij}^{Y^*})^T) W_{p'}\} \{\boldsymbol{\lambda}_p^{(r)T} (V(\boldsymbol{\eta}_{ij}) + E(\boldsymbol{\eta}_{ij}) E(\boldsymbol{\eta}_{ij})^T) \boldsymbol{\lambda}_{p'}^{(r)}\} -$$
$$(W_p^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*}))(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}))\{(W_{p'}^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*}))(\boldsymbol{\lambda}_{p'}^{(r)T} E(\boldsymbol{\eta}_{ij}))\}^T$$

$$= \{W_p^T V(\boldsymbol{\varepsilon}_{ij}^{Y^*}) W_{p'} + W_p^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*}) E(\boldsymbol{\varepsilon}_{ij}^{Y^*})^T W_{p'}\} \{\boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_{p'}^{(r)} + \boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}) E(\boldsymbol{\eta}_{ij})^T \boldsymbol{\lambda}_{p'}^{(r)}\} -$$

$$(W_p^T E(\boldsymbol{\varepsilon}_{ij}^{Y^*}))(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}))(E(\boldsymbol{\eta}_{ij})^T \boldsymbol{\lambda}_{p'}^{(r)})(E(\boldsymbol{\varepsilon}_{ij}^{Y^*})^T W_{p'})$$

$$= (Cov(\varepsilon_{ijp}^Y, \varepsilon_{ijp'}^Y) + E(\varepsilon_{ijp}^Y) E(\varepsilon_{ijp'}^Y))(\boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_{p'}^{(r)} + \boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}) E(\boldsymbol{\eta}_{ij})^T \boldsymbol{\lambda}_{p'}^{(r)}) -$$

$$E(\varepsilon_{ijp}^Y) E(\varepsilon_{ijp'}^Y)(\boldsymbol{\lambda}_p^{(r)T} E(\boldsymbol{\eta}_{ij}) E(\boldsymbol{\eta}_{ij})^T \boldsymbol{\lambda}_{p'}^{(r)})$$

Therefore,

$$Cov(Y_{ijp}, Y_{ijp'}) = (e^{\frac{\psi_p + \psi_{p'}}{2}})(\boldsymbol{\lambda}_p^{(r)T} V(\boldsymbol{\eta}_{ij}) \boldsymbol{\lambda}_{p'}^{(r)})$$

APPENDIX B

Parameter Settings for Simulation Study

$$\mathbf{\Lambda} = \begin{pmatrix} & RoadDust & PowerPlants & OilCombustion & MotorVehicles \\ Si & 0.98 & 0.11 & 0.06 & 0.30 \\ S & 0.01 & 1.00 & 0.00 & 0.02 \\ Ni & 0.00 & 0.00 & 1.00 & 0.00 \\ OC & 0.06 & 0.40 & 0.00 & 1.00 \\ Al & 1.00 & 0.00 & 0.00 & 0.00 \\ Ti & 1.02 & 0.14 & 0.28 & 0.49 \\ Ca & 0.99 & 0.03 & 0.30 & 0.46 \\ SULF & 0.01 & 0.90 & 0.00 & 0.00 \\ Se & 0.00 & 0.52 & 0.03 & 0.24 \\ V & 0.18 & 0.03 & 0.90 & 0.05 \\ Br & 0.09 & 0.82 & 0.35 & 0.46 \\ BC & 0.04 & 0.46 & 0.45 & 1.10 \\ EC & 0.00 & 0.00 & 0.43 & 1.79 \end{pmatrix}$$

$$\boldsymbol{\psi} = \begin{pmatrix} Si & S & Ni & OC & Al & Ti & Ca & SULF & Se & V & Br & BC & EC \\ 0.05 & 0.01 & 0.14 & 0.10 & 0.05 & 0.06 & 0.11 & 0.01 & 0.73 & 0.22 & 0.21 & 0.08 & 0.04 \end{pmatrix}$$

Settings for parameters in mixed model

Source	k	μ_{η_k}	α_k	$\sigma^2_{\eta_k}$	$\sigma^2_{b_k}$	$\rho_k = \frac{\sigma^2_{b_k}}{\sigma^2_{b_k}+\sigma^2_{\eta_k}}$
Road Dust	1	-0.32	0.00	0.51	0.60	0.54
Power Plants	2	-1.36	0.04	0.56	0.26	0.32
Oil Combustion	3	-0.18	-0.03	0.45	0.38	0.46
Motor Vehicles	4	-1.07	0.03	0.12	0.18	0.60

APPENDIX C

Summary of Model Fits in Data Analysis

Model (1a) - Wind Speed Only

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5763	Wind Speed	-0.0380*	(-0.0735 , -0.0021)
Power Plants	0.3688	Wind Speed	-0.0116	(-0.0573 , 0.0286)
Oil Combution	0.5209	Wind Speed	-0.0408*	(-0.0740 , -0.0099)
Motor Vehicles	0.5910	Wind Speed	-0.0519*	(-0.0748 , -0.0300)

Model (1b) - Pressure Only

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5674	Pressure	0.4985*	(0.0370 , 1.0181)
Power Plants	0.3972	Pressure	-0.5103	(-1.0310 , 0.0315)
Oil Combution	0.5217	Pressure	0.1346	(-0.3227 , 0.6130)
Motor Vehicles	0.6245	Pressure	-0.2309	(-0.6390 , 0.1782)

Model (1c) - Relative Humidity Only

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4630	RelHum	-0.0256*	(-0.0361 , -0.0153)
Power Plants	0.3536	RelHum	0.0190*	(0.0072 , 0.0308)
Oil Combution	0.4084	RelHum	0.0264*	(0.0169 , 0.0360)
Motor Vehicles	0.6103	RelHum	0.0128*	(0.0055 , 0.0200)

Model (1d) - Temperature (^{o}C) Only

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5026	Temp	-0.0259	(-0.0703 , 0.0231)
		$Temp^2$	0.0017	(-0.0001 , 0.0033)
Power Plants	0.2996	Temp	-0.0160	(-0.0634 , 0.0288)
		$Temp^2$	0.0023*	(0.0007 , 0.0039)
Oil Combution	0.4985	Temp	-0.0068	(-0.0517 , 0.0358)
		$Temp^2$	-0.0008	(-0.0025 , 0.0009)
Motor Vehicles	0.5794	Temp	0.0320*	(0.0030 , 0.0643)
		$Temp^2$	0.0002	(-0.0009 , 0.0013)

Model (2a) - Wind Speed and Pressure

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5873	Wind Speed	-0.0343	(-0.0675 , 0.0018)
		Pressure	0.4062	(-0.0760 , 0.9097)
Power Plants	0.3996	Wind Speed	-0.0179	(-0.0555 , 0.0223)
		Pressure	-0.5462*	(-1.0821 , -0.0210)
Oil Combution	0.4839	Wind Speed	-0.0437*	(-0.0763 , -0.0078)
		Pressure	-0.0186	(-0.4801 , 0.4538)
Motor Vehicles	0.5939	Wind Speed	-0.0574*	(-0.0836 , -0.0332)
		Pressure	-0.5143*	(-0.8793 , -0.1246)

Model (2b) - Wind Speed and Relative Humidity

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4780	Wind Speed	-0.0340*	(-0.0668 , -0.0035)
		Rel Hum	-0.0250*	(-0.0346 , -0.0150)
Power Plants	0.3376	Wind Speed	-0.0117	(-0.0500 , 0.0267)
		Rel Hum	0.0194*	(0.0074 , 0.0306)
Oil Combution	0.5099	Wind Speed	-0.0438*	(-0.0749 , -0.0136)
		Rel Hum	0.0269*	(0.0179 , 0.0349)
Motor Vehicles	0.6003	Wind Speed	-0.0522*	(-0.0735 ,-0.0307)
		Rel Hum	0.0134*	(0.0065 , 0.0196)

Model (2c) - Wind Speed and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5397	Wind Speed	-0.0389*	(-0.0749 , -0.0042)
		Temp	-0.0282	(-0.0751 , 0.0181)
		$Temp^2$	0.0018*	(0.0001 , 0.0035)
Power Plants	0.2845	Wind Speed	-0.0051	(-0.0415 , 0.0337)
		Temp	-0.0155	(-0.0576 , 0.0269)
		$Temp^2$	0.0023*	(0.0008 , 0.0039)
Oil Combution	0.4863	Wind Speed	-0.0453*	(-0.0794 , -0.0127)
		Temp	-0.0087	(-0.0513 , 0.0337)
		$Temp^2$	-0.0008	(-0.0024 , 0.0007)
Motor Vehicles	0.5386	Wind Speed	-0.0473*	(-0.0688 , -0.0251)
		Temp	0.0248	(-0.0025 , 0.0523)
		$Temp^2$	0.0004	(-0.0006 , 0.0014)

Model (2d) - Pressure and Relative Humidity

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4832	Pressure	0.3624	(-0.0975 , 0.8454)
		Rel Hum	-0.0247*	(-0.0352 , -0.0143)
Power Plants	0.3624	Pressure	-0.4508	(-0.9167 , 0.0606)
		Rel Hum	0.0180*	(0.0062 , 0.0291)
Oil Combution	0.4278	Pressure	0.2797	(-0.1763 , 0.7645)
		Rel Hum	0.0272*	(0.0173 , 0.0361)
Motor Vehicles	0.6111	Pressure	-0.1218	(-0.5058 , 0.2730)
		Rel Hum	0.0123*	(0.0050 , 0.0194)

Model (2e) - Pressure and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5127	Pressure	0.6003*	(0.0855 , 1.0901)
		Temp	-0.0203	(-0.0689 , 0.0230)
		$Temp^2$	0.0017*	(0.0002 , 0.0033)
Power Plants	0.3351	Pressure	-0.3369	(-0.8231 , 0.1163)
		Temp	-0.0179	(-0.0603 , 0.0252)
		$Temp^2$	0.0023*	(0.0007 , 0.0039)
Oil Combution	0.5120	Pressure	-0.0017	(-0.4685 , 0.4606)
		Temp	-0.0076	(-0.0503 , 0.0339)
		$Temp^2$	-0.0008	(-0.0023 , 0.0008)
Motor Vehicles	0.5746	Pressure	-0.0449	(-0.4136 , 0.3404)
		Temp	0.0333*	(0.0027 , 0.0626)
		$Temp^2$	0.0002	(-0.0009 , 0.0012)

Model (2f) - Relative Humidity and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4366	Rel Hum	-0.0240*	(-0.0342 , -0.0134)
		Temp	0.0031	(-0.0411 , 0.0492)
		$Temp^2$	0.0005	(-0.0012 , 0.0020)
Power Plants	0.1971	Rel Hum	0.0289*	(0.0189 , 0.0389)
		Temp	-0.0525*	(-0.0936 , -0.0121)
		$Temp^2$	0.0038*	(0.0023 , 0.0053)
Oil Combution	0.3603	Rel Hum	0.0270*	(0.0172 , 0.0368)
		Temp	-0.0444*	(-0.0846 , -0.0018)
		$Temp^2$	0.0007	(-0.0007 , 0.0024)
Motor Vehicles	0.4755	Rel Hum	0.0160*	(0.0099 , 0.0225)
		Temp	0.0120	(-0.0178 , 0.0391)
		$Temp^2$	0.0010*	(0.0000 , 0.0020)

Model (3a) - Wind Speed, Pressure, and Relative Humidity

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4833	Wind Speed	-0.0321	(-0.0646 , 0.0016)
		Pressure	0.2808	(-0.2172 , 0.7684)
		Rel Humid	-0.0247*	(-0.0345 , -0.0143)
Power Plants	0.3569	Wind Speed	-0.0177	(-0.0593 , 0.0202)
		Pressure	-0.4802	(-0.9796 , 0.0186)
		Rel Humid	0.0185*	(0.0073 , 0.0301)
Oil Combution	0.5220	Wind Speed	-0.0423*	(-0.0739 , -0.0103)
		Pressure	0.1816	(-0.2884 , 0.6088)
		Rel Humid	0.0269*	(0.0181 , 0.0361)
Motor Vehicles	0.5838	Wind Speed	-0.0584*	(-0.0828 , -0.0377)
		Pressure	-0.3882*	(-0.7693 , -0.0137)
		Rel Humid	0.0125*	(0.0057 , 0.0192)

Model (3b) - Wind Speed, Pressure, and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.5473	Wind Speed	-0.0287	(-0.0639 , 0.0068)
		Pressure	0.4935	(-0.0168 , 0.9976)
		Temp	-0.0213	(-0.0657 , 0.0211)
		$Temp^2$	0.0016*	(0.00004 , 0.0032)
Power Plants	0.3270	Wind Speed	-0.0113	(-0.0485 , 0.0274)
		Pressure	-0.3698	(-0.8575 , 0.1442)
		Temp	-0.0194	(-0.0613 , 0.0266)
		$Temp^2$	0.0024*	(0.0008 , 0.0039)
Oil Combution	0.4573	Wind Speed	-0.0513*	(-0.0851 , -0.0186)
		Pressure	-0.2071	(-0.6919 , 0.2547)
		Temp	-0.0133	(-0.0570 , 0.0251)
		$Temp^2$	-0.0007	(-0.0021 , 0.0009)
Motor Vehicles	0.5495	Wind Speed	-0.0509*	(-0.0745 , -0.0284)
		Pressure	-0.2822	(-0.6605 , 0.0740)
		Temp	0.0239	(-0.0022 , 0.0529)
		$Temp^2$	0.0003	(-0.0006 , 0.0013)

Model (3c) - Wind Speed, Relative Humidity, and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4680	Wind Speed	-0.0335	(-0.0656 , 0.0014)
		Rel Humid	-0.0239*	(-0.0349 , -0.0135)
		Temp	0.0021	(-0.0419 , 0.0457)
		$Temp^2$	0.0005	(-0.0011 , 0.0021)
Power Plants	0.1442	Wind Speed	-0.0047	(-0.0371 , 0.0284)
		Rel Humid	0.0292*	(0.0194 , 0.0384)
		Temp	-0.0541*	(-0.0948 , -0.0125)
		$Temp^2$	0.0039*	(0.0024 , 0.0054)
Oil Combution	0.3983	Wind Speed	-0.0509*	(-0.0818 , -0.0216)
		Rel Humid	0.0268*	(0.0186 , 0.0360)
		Temp	-0.0478*	(-0.0844 , -0.0082)
		$Temp^2$	0.0008	(-0.0007 , 0.0022)
Motor Vehicles	0.5045	Wind Speed	-0.0485*	(-0.0665 , -0.0298
		Rel Humid	0.0166*	(0.0105 , 0.0226)
		Temp	0.0033	(-0.0238 , 0.0288)
		$Temp^2$	0.0012*	(0.0003 , 0.0022)

Model (3d) - Pressure, Relative Humidity, and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4572	Pressure	0.4584	(-0.0152 , 0.9041)
		Rel Humid	-0.0224*	(-0.0339 , -0.0115)
		Temp	0.0060	(-0.0369 , 0.0476)
		$Temp^2$	0.0005	(-0.0012 , 0.0021)
Power Plants	0.1920	Pressure	-0.1393	(-0.6026 , 0.3512)
		Rel Humid	0.0286*	(0.0179 , 0.0384)
		Temp	-0.0527*	(-0.0934 , -0.0092)
		$Temp^2$	0.0038*	(0.0022 , 0.0053)
Oil Combution	0.3473	Pressure	0.1948	(-0.2892 , 0.6208)
		Rel Humid	0.0272*	(0.0173 , 0.0369)
		Temp	-0.0443*	(-0.0858 , -0.0018)
		$Temp^2$	0.0008	(-0.0008 , 0.0022)
Motor Vehicles	0.4594	Pressure	0.1478	(-0.2136 , 0.5214)
		Rel Humid	0.0167*	(0.0097 , 0.0233)
		Temp	0.0127	(-0.0184 , 0.0413)
		$Temp^2$	0.0010	(-0.0000 , 0.0021)

Model (4) - Wind Speed, Pressure, Relative Humidity, and Temperature

Source	Temp Corr	Covariate	Post Median	95% Credible Interval
Road Dust	0.4759	Wind Speed	-0.0260	(-0.0616 , 0.0068)
		Pressure	0.3744	(-0.0925 , 0.8528)
		Rel Humid	-0.0224*	(-0.0329 , -0.0119)
		Temp	0.0025	(-0.0471 , 0.0487)
		$Temp^2$	0.0006	(-0.0011 , 0.0023)
Power Plants	0.1850	Wind Speed	-0.0072	(-0.0457 , 0.0262)
		Pressure	-0.1624	(-0.6708 , 0.2975)
		Rel Humid	0.0286*	(0.0183 , 0.0386)
		Temp	-0.0539*	(-0.0934 , -0.0077)
		$Temp^2$	0.0038*	(0.0020 , 0.0053)
Oil Combution	0.4151	Wind Speed	-0.0505*	(-0.0828 , -0.0201)
		Pressure	-0.0238	(-0.4976 , 0.4085)
		Rel Humid	0.0264*	(0.0174 , 0.0351)
		Temp	-0.0460*	(-0.0855 , -0.0066)
		$Temp^2$	0.0007	(-0.0007 , 0.0021)
Motor Vehicles	0.4967	Wind Speed	-0.0512*	(-0.0723 , -0.0311)
		Pressure	-0.0854	(-0.4617 , 0.2755)
		Rel Humid	0.0165*	(0.0106 , 0.0225)
		Temp	0.0039	(-0.0233 , 0.0293)
		$Temp^2$	0.0012*	(0.0003 , 0.0022)