

## ORIGINAL ARTICLE

Development of a modeling approach to estimate indoor-to-outdoor sulfur ratios and predict indoor PM<sub>2.5</sub> and black carbon concentrations for Eastern Massachusetts householdsChia Hsi Tang<sup>1</sup>, Eric Garshick<sup>2</sup>, Stephanie Grady<sup>2</sup>, Brent Coull<sup>3</sup>, Joel Schwartz<sup>1</sup> and Petros Koutrakis<sup>1</sup>

The effects of indoor air pollution on human health have drawn increasing attention among the scientific community as individuals spend most of their time indoors. However, indoor air sampling is labor-intensive and costly, which limits the ability to study the adverse health effects related to indoor air pollutants. To overcome this challenge, many researchers have attempted to predict indoor exposures based on outdoor pollutant concentrations, home characteristics, and weather parameters. Typically, these models require knowledge of the infiltration factor, which indicates the fraction of ambient particles that penetrates indoors. For estimating indoor fine particulate matter (PM<sub>2.5</sub>) exposure, a common approach is to use the indoor-to-outdoor sulfur ratio ( $S_{\text{indoor}}/S_{\text{outdoor}}$ ) as a proxy of the infiltration factor. The objective of this study was to develop a robust model that estimates  $S_{\text{indoor}}/S_{\text{outdoor}}$  for individual households that can be incorporated into models to predict indoor PM<sub>2.5</sub> and black carbon (BC) concentrations. Overall, our model adequately estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$  with an out-of-sample by home-season  $R^2$  of 0.89. Estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$  reflected behaviors that influence particle infiltration, including window opening, use of forced air heating, and air purifier. Sulfur ratio-adjusted models predicted indoor PM<sub>2.5</sub> and BC with high precision, with out-of-sample  $R^2$  values of 0.79 and 0.76, respectively.

*Journal of Exposure Science and Environmental Epidemiology* (2018) **28**, 125–130; doi:10.1038/jes.2017.11; published online 18 October 2017

**Keywords:** black carbon; indoor air quality; infiltration; PM<sub>2.5</sub>; sulfur tracer method

## INTRODUCTION

Air pollutant concentrations in indoor environments, where people spend most of their time,<sup>1</sup> may be the most relevant exposure metrics to include in health studies.<sup>2</sup> Because of high cost and labor demand of indoor sampling, however, studies of indoor air quality usually include data from only a few homes that may not be representative of relevant exposures. This technical limitation has led researchers to deploy models to predict indoor exposure levels using more readily available outdoor measurements.<sup>3</sup> Typically, predictive models of indoor exposure to fine particulate matter (PM<sub>2.5</sub>) are derived from mass balance equations. A key model parameter is the infiltration factor, which varies by housing characteristics (e.g., insulation, age of house, etc.) and human activities (e.g., open windows and use of fan, air purifier, or air conditioning).<sup>2,4,5</sup> However, examination of various methods and results suggest that a mass balance approach that uses reliable estimates of infiltration and includes detailed information about individual homes is not available.<sup>6</sup> As the exposure error associated with poorly characterized infiltration was found to bias health effects assessment, the statistical power of epidemiological studies to detect health effects of exposure to indoor air pollutants remains limited.<sup>7</sup>

A common quantitative approach in estimating infiltration of fine particles is to incorporate a tracer element that (a) is predominantly of outdoor origin, (b) is measured accurately in

both indoor and outdoor environments, (c) is present in relatively high levels to ensure small measurement error, (d) has a similar size distribution to PM<sub>2.5</sub>, and (e) is chemically stable.<sup>8</sup> The indoor-to-outdoor ratio of tracer elements that meet these criteria can be used as a proxy of the infiltration rate. Among major constituents of PM<sub>2.5</sub>, sulfur or sulfate is generated mostly from outdoor sources such as power plants and industrial activities.<sup>9</sup> Moreover, studies have shown that indoor sulfur sources are scarce and that indoor sulfur concentrations are highly correlated with outdoor concentrations.<sup>10,11</sup> For these reasons, sulfur has been utilized as a tracer element to assess infiltration rates,<sup>3,6,9,12,13</sup> however, infiltration rates estimated with the sulfur tracer method vary significantly and quantifying  $S_{\text{indoor}}/S_{\text{outdoor}}$  remains a challenging task.<sup>14</sup> Uncensored or underreported indoor sulfur sources (e.g., kerosene heater use and cigarette smoking) may violate the common assumption of zero indoor sulfur sources and subsequently bias the  $S_{\text{indoor}}/S_{\text{outdoor}}$  estimate.<sup>15</sup> Furthermore, previous studies often assumed the spatial variation of outdoor sulfur to be negligible and relied on sulfur measurements from central monitoring sites as a surrogate of sulfur concentration outdoors. Although a substantial amount of sulfur is generated from regional sources, local sulfur emissions have been found to contribute significantly to outdoor sulfur concentrations. For instance, in the New England area, the frequent use of oil as a home heating fuel is a notable local sulfur source and cause outdoor sulfur levels to vary significantly in this region.<sup>16</sup> Thus,

<sup>1</sup>Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA; <sup>2</sup>Pulmonary, Allergy, Sleep, and Critical Care Medicine Section, VA Boston Healthcare System, Boston, Massachusetts, USA and <sup>3</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. Correspondence: Dr Chia Hsi Tang, Department of Environmental Health, Harvard School of Public Health, Landmark Center West, Room 412J, 401 Park Drive, Boston, MA 02115, USA. Tel.: +6176455332. Fax: +6173848823.

E-mail: cht710@mail.harvard.edu

Received 25 October 2016; accepted 22 May 2017; published online 18 October 2017

ignoring the spatial variability of outdoor sulfur concentrations likely introduces exposure errors and subsequently lowers the predictive power of  $S_{\text{indoor}}/S_{\text{outdoor}}$  models.

Numerous investigations have linked increased health risks with exposure to PM<sub>2.5</sub>,<sup>17–20</sup> and a reliable estimation of indoor PM<sub>2.5</sub> exposure levels may inform the development of effective regulation and control strategies to reduce health risks. The aim of this work is to develop a paradigm to estimate indoor air pollution levels in specific homes at times when direct measurements are not available, using sulfur as a measure of infiltration, and accounting for the spatial variability of outdoor sulfur concentrations. We constructed a robust model that accounts for potential sources of exposure errors to estimate  $S_{\text{indoor}}/S_{\text{outdoor}}$ . We then applied the modeled  $S_{\text{indoor}}/S_{\text{outdoor}}$  to predict indoor PM<sub>2.5</sub> and black carbon (BC) concentrations. Finally, we conducted cross-validation on all models to examine their predictive power specifically in situations where indoor measurements are not available.

## MATERIALS AND METHODS

### Study Design

Between November 2012 and December 2014, we collected 405 indoor samples in 130 homes of subjects with COPD who were participating in a study assessing the health effects of indoor particulate matter. Subjects were veterans living in Eastern Massachusetts who received care through the VA Boston Healthcare System and non-Veterans who were recruited via advertisement. Subjects who reported current smoking, indoor sources of smoke, including regular candle burning or use of a fireplace, wood stove, or other indoor combustion source, or known sources of indoor sulfur (such as from a humidifier) were excluded from the study. After excluding homes with reported indoor sulfur sources and those with fewer than two samples, our analysis included 328 indoor samples from weeklong sampling sessions collected at 102 residences. Samples were collected at least twice per home in different seasons to measure indoor sulfur, PM<sub>2.5</sub>, and BC concentrations.

### Data Collection

**Outdoor air pollution.** Daily outdoor PM<sub>2.5</sub> samples were collected at the central monitoring site located on the roof of the Countway Library of the Harvard Medical School in downtown Boston throughout the study period. Details on outdoor PM<sub>2.5</sub> measurements at the supersite were reported in a previous study.<sup>21</sup>

**Indoor air pollution.** A Harvard School of Public Health Micro-environmental Automated Particle Sampler (MAPS) was placed in each subject's home for indoor sampling. The MAPS includes an inertial impactor that collects PM<sub>2.5</sub> on Teflon filters at a low flow rate of 1.8 l/min. Teflon filters were weighed on an electronic microbalance (MT-5 Mettler Toledo, Columbus, OH, USA) before and after field measurements. Indoor BC concentrations were analyzed by measuring filter blackness of the Teflon filter using a smoke stain reflectometer (model EEL M43D, Diffusion Systems, UK). Sulfur concentration in the indoor PM<sub>2.5</sub> samples was determined using X-ray fluorescence spectroscopy (model Epsilon 5, PANalytical, The Netherlands).<sup>22</sup> We defined the season that each sample was obtained as winter (December–February), spring (March–May), summer (June–August), and fall (September–November).

**Questionnaires.** Subjects completed a questionnaire at study entry to collect information about building age, type of heating fuel, number of air conditioning units, and type of heating system (i.e., forced air heating). Following each sampling period, participants were asked to report specific activities and home characteristics during the sampling period that could generate indoor air pollution or alter the penetration of outdoor air pollution into the home. The following questions were included: (1) how many hours were the windows open during the sampling session? (2) How many hours did you use an electric space heater during the sampling session? (3) Did you use an air purifier during sampling session and for how long? Responses to these questions and the baseline questionnaire were used to examine the impact of activities on the sulfur ratio for each residence.

**Land use parameters.** Land use parameters often serve as surrogates of anthropogenic PM<sub>2.5</sub> sources; in this study, land use parameters were used to quantify the spatial variability of PM and BC concentrations outside participating homes. The percentage of urban spaces in a grid of 1 km × 1 km cells covering the study area was obtained from the 2011 collection of the National Land Cover Database. Major road (A1–A3) density was gathered from the StreetMap USA database using the Feature Class Code (A1–A4) classification from the U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing system. Annual averaged traffic count for major roads was obtained from the Highway Performance Monitoring System database. The built-in Kernel density algorithm<sup>23</sup> from ArcMap was used to calculate traffic count weighted for major road density within 1 km<sup>2</sup>. Population density was calculated within 1 km<sup>2</sup> from the census track database of year 2000.

### Statistical Analysis

We considered the dynamics of PM<sub>2.5</sub> involving infiltration (or inflow), exfiltration (or outflow), indoor emission, and deposition removal (Figure 1).

On the basis of the relationships illustrated in Figure 1, we can obtain a simplified mass balance equation as follows:

$$\frac{dC_{\text{indoor}}(t)}{dt} = (F_{\text{in}} + E) - (F_{\text{out}} + R) \quad (1)$$

where  $dC_{\text{indoor}}(t)$  is the change in indoor concentration of particle mass ( $\mu\text{g}/\text{m}^3$ ) during the time interval  $dt$ ,  $F_{\text{in}}$  represents particles infiltrated from outdoor,  $E$  is the indoor emission,  $F_{\text{out}}$  is particle exfiltration to the outdoors, and  $R$  is the indoor removal by deposition, with  $d$  as the deposition rate. Dockery and Spengler<sup>24</sup> first developed the mass balance equation for indoor particles based on the assumption that air exchange rate ( $a$ ), penetration rate ( $p$ ), and deposition rate ( $k$ ) remain constant over time period  $dt$  to obtain the following equation:

$$\frac{dC_{\text{indoor}}(t)}{dt} = \frac{a \times p \times dC_{\text{outdoor}}(t)}{a + d} + \frac{E}{V(a + d)} - \frac{dC_{\text{indoor}}(t)}{dt} \times (a + d) \quad (2)$$

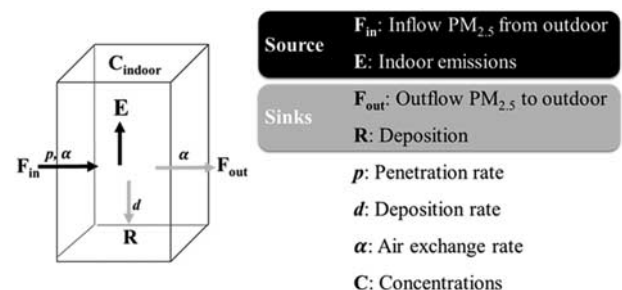
where  $dC_{\text{outdoor}}(t)$  is the change in particle concentration outside the household ( $\mu\text{g}/\text{m}^3$ ),  $V$  is the volume of the house ( $\text{m}^3$ ),  $p$  is the penetration factor (dimensionless), and  $d$  is the deposition rate (dimensionless). The product of the penetration factor and air exchange rate is equivalent to the infiltration factor,  $F_{\text{infr}}$ , illustrated in Figure 1. Under the assumption that indoor air is well mixed and outdoor concentration is constant,<sup>14</sup> the steady-state indoor concentration can be expressed as follows:

$$C_{\text{indoor}} = \frac{E/V}{(a + d)} + \frac{a \times p}{a + d} C_{\text{outdoor}} \quad (3)$$

As air exchange rate is usually much faster than the indoor deposition rate ( $d$ ),<sup>25</sup> we can further rearrange Eq. 3 as follows:

$$C_{\text{indoor}} = \frac{E/V}{a} + p C_{\text{outdoor}} \quad (4)$$

From the above derivation, it is apparent that the key parameter allowing us to use outdoor particle concentrations to predict indoor concentrations is the infiltration factor ( $F_{\text{infr}}$ ). The traditional method to estimate infiltration is to measure the building tightness of each house, which is unrealistic due to the large number of complex physical factors involved. For this reason, the indoor-to-outdoor ratio of tracer elements is often used to quantify



**Figure 1.** Dynamics of inflow, outflow, emission, removal of PM<sub>2.5</sub> in the indoor environment.

infiltration. In this study, we used sulfur as the tracer element and constructed a sulfur model to estimate the indoor-to-outdoor sulfur ratio ( $S_{\text{indoor}}/S_{\text{outdoor}}$ ) for individual households. Subsequently, we incorporated the estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$  to predict indoor PM<sub>2.5</sub> and BC concentrations based on the mass balance concept (Eq. 4). Finally, we examined the predictive ability of all three models using cross-validation by home and season.

**Sulfur model.** As previously described, in situations where there are no indoor sulfur emissions, we can calculate the indoor-to-outdoor sulfur ratio (Eq. 5) directly and use it as a surrogate of the infiltration factor ( $F_{\text{inf}}$ ):

$$F_{\text{inf}} \propto \frac{S_{\text{indoor}}}{S_{\text{outdoor}}} \quad (5)$$

However, given that indoor sulfur emissions may still have occurred despite efforts to exclude these homes by questionnaire data, we estimated the infiltration factor with the following model (Eq. 6):

$$S_{\text{indoor}} = \alpha_{0i} + \beta_0 + (\alpha_{1i} + \beta_1)S_{\text{outdoor}} + \sum_{j=1}^4 \beta_{j+1}(S_{\text{outdoor}} \times I_j) \quad (6)$$

where  $S_{\text{indoor}}$  and  $S_{\text{outdoor}}$  are the sulfur concentration measured indoors and outdoors, respectively,  $i$  is the identifier of the homes included in the study, and  $j$  is the season identifier.  $I_j$  is an indicator variable for season  $j$ . As sulfur was not measured outside each individual household, we used the sulfur concentration measured at the central site as a surrogate of the sulfur concentration outside homes. The exposure error introduced by using sulfur measured at a central site was accounted for in the subsequent indoor PM<sub>2.5</sub> and BC model. Here we included a random slope by home ( $\alpha_{1i}$ ) to generate  $S_{\text{indoor}}/S_{\text{outdoor}}$  for individual homes and to take into account the spatial variability of the sulfur concentrations outside homes that are not captured when using central site measurements. We also included fixed intercept ( $\beta_0$ ) and random intercept by home ( $\alpha_{0i}$ ) to capture potential indoor sulfur emissions such as underreported indoor smoking and humidifier use. Finally, we included an interaction term between outdoor sulfur concentration and season to capture the seasonal variation of the infiltration. Thus, the estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$  equals the sum of the estimated fixed ( $\beta_1$ ) and random slope ( $\alpha_{1i}$ ) and the slopes of the season interaction terms ( $\beta_j$ ).

**Indoor PM<sub>2.5</sub> and BC model.** Combining the estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$  (Eq. 6) and mass balance (Eq. 3) we can predict indoor concentrations of PM<sub>2.5</sub> and BC with the following equation:

$$C_{\text{indoor}} = \alpha_{0i} + \beta_0 + (\alpha_{1i} + \beta_1) \times \frac{S_{\text{indoor}}}{S_{\text{outdoor}}} \times C_{\text{outdoor}} \quad (7)$$

where  $C_{\text{indoor}}$  ( $\mu\text{g}/\text{m}^3$ ) is the indoor concentration of particle mass and  $C_{\text{outdoor}}$  ( $\mu\text{g}/\text{m}^3$ ) is the particle concentration outside the individual household. In this model, we included a random intercept ( $\alpha_{0i}$ ) by home ( $i$ ) to account for indoor PM<sub>2.5</sub> and BC emissions. The random slope ( $\alpha_{1i}$ ) by home corrects any bias in the sulfur ratio due to underestimating the spatial variation of sulfur by using central site sulfur measurement as a surrogate of outdoor sulfur concentrations.

To a greater extent than sulfur, the spatial distribution of PM<sub>2.5</sub> and BC can vary by location significantly, even within a small metropolitan area. The PM<sub>2.5</sub> and BC concentrations measured at the central site may not reflect the local generation of particles outside participating homes, as BC, and to a lesser extent PM<sub>2.5</sub>, are highly associated with traffic volume and metropolitan activities. For this reason, we used land use variables, including major road density, percent urban space, and the distance between home and the central measurement site to predict outdoor PM<sub>2.5</sub> and BC concentrations in addition to the measured concentrations. The final indoor model (Eq. 8) is therefore formulated as

$$C_{\text{indoor}} = (\alpha_{0i} + \beta_0) + (\alpha_{1i} + \beta_1) \times \left( \frac{S_{\text{indoor}}}{S_{\text{outdoor}}} \right) \times C_{\text{countway}} \times (\beta_2 \text{roaddensity} + \beta_3 \% \text{urban} + \beta_4 \text{distance btw home } i \text{ and central}) \quad (8)$$

where  $C_{\text{countway}}$  ( $\mu\text{g}/\text{m}^3$ ) is the measured concentration at the central monitoring site, road density (number of vehicles-distance traveled in km/1 km<sup>2</sup>) is the traffic density of A1–A3 roads within a 1 km<sup>2</sup> grid in which the

participating home falls, and %urban is the percentage of urban spaces within a 1 km<sup>2</sup> grid surrounding the homes.

**Out-of-sample cross-validation.** We conducted cross-validation by home and season to evaluate the predictive power of the sulfur model (Eq. 6) and indoor PM<sub>2.5</sub> and BC model (Eq. 8) described above. For each home, we first held out one measurement out of the two to four total samples measured in the same home during the study period. We then trained the sulfur, indoor PM<sub>2.5</sub>, and BC models using the rest of the samples from that specific home and all samples from other homes to predict the held-out data. We iterated the same procedure until all data were predicted once and examined the  $R^2$  of the cross-validated prediction vs the observed measurements. We designed the cross-validation to utilize measurements from the same home to provide home-specific information and samples from other homes to provide data that were temporally unavailable. In essence, this cross-validation examined whether our models could accurately predict  $S_{\text{indoor}}/S_{\text{outdoor}}$ , indoor PM<sub>2.5</sub>, and indoor BC during the periods with no indoor measurements.

## RESULTS AND DISCUSSION

### General Characteristics

Samples with an observed sulfur ratio higher than 1.12 (95th percentile of the total 405 samples) and homes with fewer than 2 samples were excluded for model predictability and cross-validation purposes. This resulted in a data set collected during 328 weeklong sampling sessions at 102 residences that were matched by date to a central monitoring site (countway supersite). The distribution of measured indoor and outdoor concentrations is summarized in Table 1. Indoor PM<sub>2.5</sub> concentrations were generally higher than those measured outdoors suggesting significant indoor sources of particles. On the contrary, indoor BC concentrations were lower than those measured outdoors, which is likely due to the lack of indoor sources of BC and potentially larger amount of traffic emissions and biomass burning activities at the central site.<sup>26,27</sup> More importantly, the mean difference of the indoor and outdoor concentration of PM<sub>2.5</sub> and BC varied considerably. Previous studies have attributed this variation to multiple factors, including indoor particle emission, air exchange rate, and natural ventilation.<sup>9</sup> Together, these findings indicate the need to consider variability of the infiltration factors in the modeling process of indoor PM<sub>2.5</sub> and BC exposure levels.

Table 2 displays the land use environment and household characteristics of the 102 homes in the Greater Boston included in the study data sample. The land use setting of the study homes varied significantly, suggesting that the spatial variation of outside PM<sub>2.5</sub> and BC is likely to vary from home to home. Furthermore, the prevalence of air conditioning and forced air heating varied considerably. Moderate variability was also found in natural ventilation, specifically window-opening hours. These factors are expected to influence the infiltration of particles and were found to lower the model predictive power in previous studies.<sup>28</sup> More importantly, the large variation among land use and household characteristics reinforces the notion that assuming no indoor sulfur and homogeneous outdoor sulfur concentration may contribute to reduced model fit.

**Table 1.** Distribution of measured indoor and outdoor (central site) PM<sub>2.5</sub> concentrations and its BC and sulfur constituents.

	Indoor		Central site	
	Mean	SD	Mean	SD
PM <sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ )	8.8	6.5	6.5	2.2
BC ( $\mu\text{g}/\text{m}^3$ )	0.24	0.26	0.58	0.24
Sulfur ( $\mu\text{g}/\text{m}^3$ )	0.30	0.15	0.45	0.19

**Table 2.** Distribution of home characteristics, surrounding land use, and questionnaire variables related to indoor air pollution.

	Mean	SD	Min	Median	Max
<i>Variables</i>					
Percent urban spaces within 1 km <sup>2</sup> grid (%)	67	27	0	74	100
Major road density within 1 km <sup>2</sup> grid (no. of vehicle-km/km <sup>2</sup> )	2.2	1.6	0.0	2.0	11.5
Distance to central site (km)	27.7	20.2	1.2	24.5	88.2
Building age (years)	62.8	31.9	9.0	51.0	171.0
No. of air conditioning	2.2	0.7	1.0	2.0	4.0
Use of forced air heating (# subject-week)		Yes		No	
		66		262	
<i>Questionnaire</i>					
(1) How many hours was the window open during sampling session?	31.4	57.1	0.0	0.0	168.0
(2) How many hours did you use an electric space heater during sampling session?	5.5	6.8	0.2	3.0	24.0
(3) How many hours did you use an air purifier during sampling session?	10.1	33.6	0.0	0.0	168.0
(4) What fuel do you use for heating? (# homes)	Gas	Electric		Other/oil	
	44	21		30	

**Table 3.** Cross-validated  $R^2$  and corresponding RMSE values for predicted indoor  $S_{\text{indoor}}/S_{\text{outdoor}}$ , PM<sub>2.5</sub>, and BC.

	$R^2$ (unitless)	RMSE ( $\mu\text{g}/\text{m}^3$ )	Intercept bias ( $\mu\text{g}/\text{m}^3$ )	Slope bias (unitless)
Indoor sulfur model	0.89	0.038	-0.02	1.05
Indoor PM <sub>2.5</sub> model	0.79	1.695	0.38	0.90
Indoor BC model	0.76	0.122	0.01	1.01

Estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$

The performance of the mixed-effects model (Eq. 6) to predict indoor sulfur using central site measurements was excellent, with a high cross-validated  $R^2$  of 0.89, a non-biased slope, and a negligible intercept bias when comparing the predicted to the measured indoor sulfur concentration (Table 3).

Table 4 displays the relationship between the predicted sulfur ratio and home characteristics and human activities. The  $S_{\text{indoor}}/S_{\text{outdoor}}$  increased with higher natural ventilation (window opening) and urban density. Use of an air purifier, air conditioning, and forced air heating had an opposite effect on the predicted  $S_{\text{indoor}}/S_{\text{outdoor}}$  (-0.06 to -0.28). These findings are consistent with those of previous investigations.<sup>29,30</sup> As expected, homes using oil as primary heating fuel had a higher sulfur ratio (0.07) compared to those that using gas or electric fuel. Outdoor furnaces emitting sulfur-rich particles may have added to the spatial variability of the sulfur concentration outdoors. Consequently, the  $S_{\text{indoor}}/S_{\text{outdoor}}$  for these oil-using homes was slightly overestimated (0.07), due to the relatively lower concentrations of sulfur at the central site compared to outside the home (Table 4). However, inclusion of a random slope by home in the indoor PM<sub>2.5</sub> and BC model to account for indoor sulfur heterogeneity induced bias (Eq. 8). Finally, the predicted sulfur ratio did not vary by electric space heater use and there was a borderline association with traffic density.

Predicted Indoor PM<sub>2.5</sub> and BC

Using the estimated  $S_{\text{indoor}}/S_{\text{outdoor}}$  as the infiltration proxy, we predicted indoor PM<sub>2.5</sub> and BC concentrations. The cross-validated  $R^2$  for indoor PM<sub>2.5</sub> and BC was 0.79 and 0.76, respectively, showing strong predictive power (Table 3). Model performance was superior to previous indoor models, specifically those that did not use a sulfur tracer as an infiltration surrogate (Table 5). We relaxed the assumption regarding absence of indoor sulfur sources and used fixed and random intercepts (Eq. 6) to filter

**Table 4.** Relationship between the predicted  $S_{\text{indoor}}/S_{\text{outdoor}}$  ratio and household characteristics and behaviors.

	Estimate	Std. error	t-value	P-value
<i>Heat fuel</i>				
Other/oil vs gas/electric	0.07	0.020	3.165	0.001
<i>Window</i>				
Open vs closed	0.10	0.020	4.340	< 0.001
Road density	0.03	0.017	1.790	0.07
Electric space heater use	0.003	0.030	0.096	0.92
<i>Forced-air heating use</i>				
Yes vs no	-0.06	0.030	-1.930	0.05
<i>Land cover</i>				
Urban vs rural	0.20	0.043	4.477	< 0.001
<i>Air purifier use</i>				
Yes vs no	-0.28	0.095	-2.884	0.004
<i>Air conditioning use</i>				
Yes vs no	-0.06	0.026	-2.396	0.017

noise attributable to potential indoor sulfur emissions. This approach is likely improved our model's performance. Furthermore, the number of participating homes in this study was higher than that of previous studies and, therefore, enhanced our model's predictive ability with a relatively larger sample size.

While indoor PM<sub>2.5</sub> concentrations differed across households, this variability was not explained by housing characteristics except for the significantly lower PM<sub>2.5</sub> concentration (-6.3  $\mu\text{g}/\text{m}^3$ ) when air purifiers were used. Unexplained variations in indoor particle concentrations could have been generated by cooking and unreported incense burning, which was found to contribute to indoor particle concentrations in previous studies.<sup>28,31,32</sup> On the other hand, we found a significant relationship between indoor BC of outdoor origin and traffic density surrounding homes. This is consistent with previous reports that traffic pollutants, including BC, can easily penetrate buildings.<sup>21,33-35</sup>

Our model's predictive ability could be further improved by extending the sample size, either by measuring daily instead of weeklong indoor concentrations or by expanding the study area. We also did not take into account the influence of chemical transformations, such as changes in gas-to-particle partitioning during the infiltration of volatile organic compounds, nitrate, or ammonium,<sup>36,37</sup> which may also affect the performance of indoor

**Table 5.** Comparison of the current study and previous approaches to modeling infiltration and indoor PM<sub>2.5</sub> concentrations.

	<i>Current study</i>	<i>Allen et al.<sup>37</sup></i>	<i>Baster et al.<sup>27</sup></i>	<i>Wallace and Williams<sup>6</sup></i>
Study area	102 homes Boston	353 homes 7 cities across US	39 homes Boston	37 homes North Carolina
Infiltration proxy	Sulfur	Sulfur	GIS and housing characters	
Model technique	Mixed-effects model	Multivariable regression	Bayesian model	
Predictive power infiltration proxy	CV $R^2 = 89.3\%$	CV $R^2 = 30\text{--}60\%$	—	
Predictive power on PM <sub>2.5</sub>	CV $R^2 = 79.1\%$	—	CV $R^2 = 0.37$	$R^2 = 0.74$

Abbreviation: CV, cross-validation.

PM<sub>2.5</sub> and BC exposure models. Nevertheless, the models developed in this study had superior predictive power compared to models described the literature and, therefore, may provide more reliable exposure data to study the health effects of particles of indoor vs outdoor origin.

## CONCLUSION

In this paper, we presented a new paradigm that utilizes a relatively small number of indoor measurements to predict exposures for individual households when indoor measurements are not available. Our methods strengthen the estimation of infiltration factor, indoor PM<sub>2.5</sub>, and BC concentrations for individual residences in the Greater Boston Area. Our models may improve indoor PM<sub>2.5</sub> exposure assessments for future health effects studies and could serve as an exposure model framework more generally for other locations.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This research was supported by U.S. EPA grant RD83587201 and NIH 1 R01 ES109853-A1. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the U.S. EPA. Further, the U.S. EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

## REFERENCES

- Adgate JL, Ramachandran G, Pratt GC, Waller LA, Sexton K. Spatial and temporal variability in outdoor, indoor, and personal PM<sub>2.5</sub> exposure. *Atmos Environ* 2002; **36**: 3255–3265.
- Morawska L, He C, Hitchins J, Gilbert D, Parappukkaran S. The relationship between indoor and outdoor airborne particles in the residential environment. *Atmos Environ* 2001; **35**: 3463–3473.
- Sarnat A, Long C, Koutrakis P, Coull B, Schwartz J, Suh H. Using sulfur as a tracer of outdoor fine particulate matter. *Environ Sci Technol* 2002; **36**: 5305–5314.
- Van Der Zee SC, Strak M, Dijkema MB, Brunekreef B, Janssen NA. The impact of particle filtration on indoor air quality in a classroom near a highway. *Indoor Air* 2016; **27**: 291–302.
- Meng QY, Spector D, Colome S, Turpin B. Determinants of indoor and personal exposure to PM(2.5) of indoor and outdoor origin during the RIOPA Study. *Atmos Environ* 2009; **43**: 5750–5758.
- Wallace L, Williams R. Use of personal indoor outdoor sulfur concentrations to estimate the infiltration factor and outdoor exposure factor for individual homes and persons. *Environ Sci Technol* 2005; **39**: 1707–1714.
- Zeger S, Thomas D, Dominici F, Samet J, Schwartz J, Dockery D *et al*. Exposure measurement error in time series studies of air pollution: concepts and consequences. *Environ Health Perspect* 2000; **108**: 417–426.
- Wilson WE, Mage DT, Grant LD. Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: Why and how. *J Air Waste Manag Assoc* 2000; **50**: 1167–1183.
- Long CM, Sarnat JA. Indoor-outdoor relationships and infiltration behavior of elemental components of outdoor PM<sub>2.5</sub> for Boston-area homes. *Aerosol Sci Technol* 2004; **38**: 91–104.
- Ozkaynak H, Xue J, Spengler J, Wallace L, Pellizzari E, Jenkins P. Personal exposure to airborne particles and metals: Results from the particle team study in Riverside, California. *J Expo Anal Environ Epidemiol* 1996; **6**: 57–78.
- Koutrakis P, Briggs SL, Leaderer BP. Source apportionment of indoor aerosols in Suffolk and Onondaga Counties, New York. *Environ Sci Technol* 1992; **26**: 521–527.
- Gaffin J, Petty C, Hauptman M, Kang C, Coull B, Schwartz J *et al*. Modeling indoor particulate exposure in inner city school classrooms. *J Expo Sci Environ Epidemiol*. 2016; **27**: 451–457.
- Geller M, Chang M, Sioutas C, Ostro B, Lipsett M. Indoor outdoor relationship and chemical composition of fine and coarse particles in the Southern California deserts. *Atmos Environ* 2002; **36**: 1099–1110.
- Diapouli E, Chaloulakou A, Koutrakis P. Estimating the concentration of indoor particles of outdoor origin: a review. *J Air Waste Manag Assoc* 2013; **63**: 1113–1129.
- Wallace L, Williams R, Suggs J, Jones P. *Estimating Contributions of Outdoor Fine Particles to Indoor Concentrations and Personal Exposures: Effects of Household Characteristics and Personal Activities*. US Environmental Protection Agency, 2006.
- Energy Information Administration: Petroleum & Other Liquids: Consumption and Sales Data. [Internet]. Available at <http://www.eia.gov/petroleum/data.cfm#consumption>. Last accessed date January 2016.
- Dominici F. Airborne particulate matter and mortality: timescale effects in four US cities. *Am J Epidemiol* 2003; **157**: 1055–1065.
- Simkhovich BZ, Kleinman MT, Kloner RA. Air pollution and cardiovascular injury. *J Am Coll Cardiol* 2008; **52**: 719–726.
- Bell M, Ebisu K, Peng RD, Samet JM, Dominici F. Hospital admissions and chemical composition of fine particle air pollution. *Am J Respir Crit Care Med* 2009; **179**: 1115–1120.
- Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M *et al*. Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environ Sci Technol* 2012; **46**: 652–660.
- Kang CM, Koutrakis P, Suh HH. Hourly measurements of fine particulate sulfate and carbon aerosols at the Harvard-U.S. Environmental Protection Agency Supersite in Boston. *J Air Waste Manag Assoc* 2010; **60**: 1327–1334.
- Agency USEP. EPA positive matrix factorization (PMF) 3.0 fundamentals and user guide. In: *Development*. USEOoRa (ed). Available at <http://www.epa.gov/ttnamti1/files/ambient/inorganic/mthd-3-3.pdf>. 2008.
- Silverman B *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: New York, NY, USA, 1986.
- Dockery DW, Spengler JD. Indoor-outdoor relationships of respirable sulfates and particles. *Atmos Environ* 1981; **15**: 335–343.
- Thatcher TL, Lunden MM, Revzan KL, Sextro RG, Brown NJ. A concentration rebound method for measuring particle penetration and deposition in the indoor environment. *Aerosol Sci Technol* 2003; **37**: 847–864.
- Patel MM, Chillrud SN, Correa JC, Feinberg M, Hazi Y, Kc D *et al*. Spatial and temporal variations in traffic-related particulate matter at New York City High Schools. *Atmos Environ* 2009; **43**: 4975–4981.
- Saleh R, Robinson ES, Tkacik DS, Ahern AT, Liu S, Aiken AC *et al*. Brownness of organics in aerosols from biomass burning linked to their black carbon content. *Nat Geosci* 2014; **7**: 647–650.
- Baxter LK, Clougherty JE, Paciorek CJ, Wright RJ, Levy JI. Predicting residential indoor concentrations of nitrogen dioxide, fine particulate matter, and elemental carbon using questionnaire and geographic information system based data. *Atmos Environ* 2007; **41**: 6561–6571.
- Cyrys J, Pitz M, Bischof W, Wichmann HE, Heinrich J. Relationship between indoor and outdoor levels of fine particle mass, particle number concentrations and black smoke under different ventilation conditions. *J Expo Anal Environ Epidemiol* 2004; **14**: 275–283.
- Bell M, Ebisu K, Peng R, Dominici F. Adverse health effects of particulate air pollution: modification by air conditioning. *Epidemiology* 2009; **20**: 682–686.

- 31 Habre R, Coull B, Moshier E, Godbold J, Grunin A, Nath A *et al*. Sources of indoor air pollution in New York City residences of asthmatic children. *J Expo Sci Environ Epidemiol* 2014; **24**: 269–278.
- 32 Pokhrel AK, Bates MN, Acharya J, Valentiner-Branth P, Chandyo RK, Shrestha PS *et al*. PM2.5 in household kitchens of Bhaktapur, Nepal, using four different cooking fuels. *Atmos Environ* 2015; **113**: 159–168.
- 33 Dorizas PV, Assimakopoulos MN, Helmis C, Santamouris M. An integrated evaluation study of the ventilation rate, the exposure and the indoor air quality in naturally ventilated classrooms in the Mediterranean region during spring. *Sci Total Environ* 2015; **502**: 557–570.
- 34 Park SK, O'Neill MS, Stunder BJ, Vokonas PS, Sparrow D, Koutrakis P *et al*. Source location of air pollution and cardiac autonomic function: trajectory cluster analysis for exposure assessment. *J Expo Sci Environ Epidemiol* 2007; **17**: 488–497.
- 35 Maynard D, Coull BA, Gryparis A, Schwartz J. Mortality risk associated with short-term exposure to traffic particles and sulfates. *Environ Health Perspect* 2007; **115**: 751–755.
- 36 Lunden MM, Kirchstetter TW, Thatcher TL, Hering SV, Brown NJ. Factors affecting the indoor concentrations of carbonaceous aerosols of outdoor origin. *Atmos Environ* 2008; **42**: 5660–5671.
- 37 Hering SV, Lunden MM, Thatcher TL, Kirchstetter TW, Brown NJ. Using regional data and building leakage to assess indoor concentrations of particles of outdoor origin. *Aerosol Sci Technol* 2007; **41**: 639–654.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2018