Identifying and quantifying secondhand smoke in source and receptor rooms: logistic regression and chemical mass balance approaches

Abstract Identifying and quantifying secondhand tobacco smoke (SHS) that drifts between multiunit homes is critical to assessing exposure. Twenty-three different gaseous and particulate measurements were taken during controlled emissions from smoked cigarettes and six other common indoor source types in 60 single-room and 13 two-room experiments. We used measurements from the 60 single-room experiments for (i) the fitting of logistic regression models to predict the likelihood of SHS and (ii) the creation of source profiles for chemical mass balance (CMB) analysis to estimate source apportionment. We then applied these regression models and source profiles to the independent data set of 13 two-room experiments. Several logistic regression models correctly predicted the presence of cigarette smoke more than 80% of the time in both source and receptor rooms, with one model correct in 100% of applicable cases. CMB analysis of the source room provided significant PM$_{2.5}$ concentration estimates of all true sources in 9 of 13 experiments and was half-correct (i.e., included an erroneous source or missed a true source) in the remaining four. In the receptor room, CMB provided significant estimates of all true sources in 9 of 13 experiments and was half-correct in another two.

P. J. Dacunto$^1$, K.-C. Cheng$^{12}$, V. Acevedo-Bolton$^1$, R.-T. Jiang$^1$, N. E. Klepeis$^{13}$, J. L. Repace$^4$, W. R. Ott$^1$, L. M. Hildemann$^1$

1Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA, 2Environmental Health Laboratory, California Department of Public Health, Richmond, CA, USA, 3Graduate School of Public Health, San Diego State University Research Foundation, Center for Behavioral Epidemiology and Community Health (CBEACH), San Diego, CA, USA, 4Repace Associates, Bowie, MD, USA

Key words: Multiunit housing; Multiunit dwellings; Indoor emissions; Source apportionment; 2,5-dimethylfuran; Nicotine.

Philip J. Dacunto
Stanford University
Department of Civil and Environmental Engineering
473 Via Ortega, Rm M-4
Stanford
CA 94305
USA
Tel.: +253-363-0235
Fax: 650-725-3162
e-mail: philip.dacunto@usma.edu

Received for review 26 November 2012. Accepted for publication 25 April 2013.

Practical Implications
The utility of several popular single tracers of SHS is limited by challenges such as detection limits, indoor losses, long sampling and processing periods, and variable correlation with PM$_{2.5}$. In addition, we found measurable quantities of one proposed SHS tracer (2,5-dimethylfuran) in emissions from four non-cigarette sources, including incense, burned toast, burned microwave popcorn, and wood. This work demonstrates that tandem measures of real-time particle size, real-time particle composition, and VOCs may enable accurate identification and quantification of SHS with precise temporal resolution, even in receptor rooms adjacent to the source. This might be particularly useful inside multiunit homes where smoking occurs in a neighboring unit.

Introduction and background
The dangers of secondhand tobacco smoke (SHS) are well known. Being declared a human carcinogen by the EPA in 1992 (U.S. Environmental Protection Agency), SHS has also been named as a cause of cardiovascular disease (Taylor et al., 1992), as well as a host of acute health effects (Flouris et al., 2010). As a result, the U.S.
Surgeon General declared in a 2006 report that there is no safe level of SHS. While SHS is being eliminated from many public locations in the US and elsewhere, smoke transfer from one residence to another is becoming a concern. Currently, more than 79 million Americans (one-quarter of the US population) live in a multiunit home (MUH) (King et al., 2012). A recent study by Licht et al. (2012) showed that of U.S. MUH residents with smoke-free home rules, 44% experienced SHS intrusions from a neighbor. Exposure for affected residents is amplified by the large proportion of time that people spend at home – 69% for adults in the United States (Klepeis et al., 2001).

Quantifying the amount of SHS that ends up inside a non-smoking receptor unit (i.e., a home without a smoker) is critical to assessing the resident’s exposure. However, thus far, no proven accurate means are available for doing so, even when the SHS is at high concentrations in the receptor unit. Identifying and quantifying the pollutants derived specifically from smoking is even more difficult when the SHS in receptor units is at a lower concentration or mixed with emissions from other common indoor activities such as cooking. Qualitative reports of SHS by smell are useful only when the resident is present and awake and offer little information on the magnitude of the intrusion; furthermore, they can be perceived as biased and thus may offer insufficient proof of intrusion in the eyes of a landlord, condo association, or other entity attempting to mediate a neighbor dispute or enforce smoke-free rules. Existing methods of identifying and quantifying SHS in MUHs have potential shortcomings both for PM$_{2.5}$ measurements (non-specific for SHS) and for tracers (low concentration in receptor rooms, variable correlation with PM$_{2.5}$, and long sampling and/or processing times).

Three tracers, nicotine, 3-ethenylpyridine (3-EP), and 2,5-dimethylfuran (2,5-DMF), have shown promise in SHS identification, but they all require lengthy sampling and laboratory processing times, making measurements with precise temporal resolution difficult. There are other disadvantages as well. Two prior studies have focused on the identification of SHS in MUHs through passive measurements of nicotine (Bohac et al., 2011; Kraev et al., 2009). However, nicotine’s semi-volatility and unpredictable sorption to surfaces in locations where smoking is infrequent make it difficult to quantify the corresponding amount of SHS actually entering a nonsmoker’s apartment at fine timescales (Van Loy et al., 1998). The tracer 3-EP has sorption challenges similar to those of nicotine (Singer et al., 2002, 2003). In contrast, 2,5-DMF has been shown to be consistently proportional to other measures of SHS (e.g., PM$_{2.5}$) and not strongly affected by sorption (Charles et al., 2008). However, despite the moderate 2,5-DMF source strength from cigarettes of 127 $\mu$g/cigarette (Bi et al., 2005), concentrations may not reach detectable levels for larger room sizes or air exchange rates, or in an adjacent room or apartment.

To date, a single published study has attempted to quantify the amount of SHS passing between units. This study used real-time laser photometers measuring PM$_{2.5}$ in a total of 30 units in 11 different buildings in Buffalo, New York (King et al., 2010). As PM$_{2.5}$ is not specific to SHS, all residents in the immediate area of the measurement site, including the smoker, needed to fill out accurate logs indicating all sources of PM$_{2.5}$. By coupling daily activity logs with PM$_{2.5}$ measurements, this study found that 2 of the 14 nonsmoking units received SHS from their neighbors. But even if the logs were assumed to be accurate, it would be challenging to determine the amount of SHS transfer if other particle sources were emitting simultaneously.

The present study seeks to overcome the shortcomings of existing methods using three distinguishing features of SHS: particle size distribution, particle composition, and gaseous composition. Our goal was to develop an SHS ‘fingerprint’ that can be used for both identification and quantification with precise temporal resolution. We present here the results of controlled experiments to evaluate, test, and calibrate this fingerprinting method, which determine which combination(s) of SHS tracers (described in more detail below) show suitable promise for future field testing in real MUHs.

### SHS particle size distribution

Previous research by Klepeis et al. (2003) has demonstrated that most fresh SHS particle mass falls between 0.02–2 $\mu$m, with a mass median diameter (MMD) of 0.2 $\mu$m. This finding is in agreement with results from other studies, which found SHS MMDs from 0.2–0.5 $\mu$m and count median diameters of 0.1–0.27 $\mu$m (summarized in the study of Nazaroff and Klepeis, 2003). Wallace and Ott (2011) show that the vast majority of SHS particles by number count are below 0.4 $\mu$m.

This distinctive size distribution contrasts with those from other common indoor sources that generate coarser particles, such as vacuuming, dusting, sweeping, and movement of people (Abt et al., 2000). Combustion of natural gas from stoves or clothes dryers produces aerosol mostly in the ultrafine particle (UFP) range of 0.001–0.1 $\mu$m (Wallace, 2006). The vast majority of cooking particles are in the ultrafine and accumulation mode (<1.0 $\mu$m) as well, but significant numbers exist above 1 $\mu$m, making the size distribution broader than that for SHS (Abt et al., 2000; Dennekamp et al., 2001; Fortmann et al., 2001; Wallace et al., 2004; Yeung and To, 2008).

Thus, excluding the ultrafines, it may be possible to use size distributions to distinguish SHS from other indoor sources. A comparison of different size bins of particles (e.g., number in >1.0 $\mu$m range vs. number in...
Identifying and quantifying SHS in source and receptor rooms

Overview

The first stage of this study involved 60 single-room experiments using a variety of instruments and samplers, to define a characteristic measurement ‘signature’ for multiple common indoor sources. We applied these ‘signatures’ to two modeling approaches: logistic regression (to identify the probability of SHS) and chemical mass balance analysis (for quantification). Finally, we conducted 13 independent two-room experiments to test the utility of the models in identifying and quantifying SHS in both source and receptor rooms.

Experimental methods

Overview

The seven two-room experiments at the Redwood City motel were carried out in adjacent ground-floor smoking (72 m³) and nonsmoking (80 m³) rooms,
both of which opened to the outside. After testing a variety of ventilation conditions, we performed four final experiments with a large (0.25 m$^2$) fan blowing out of the receptor room window in combination with the receptor room bathroom exhaust fan (0.031 m$^2$), resulting in an inter-room pressure difference of 12–13 Pa. This pressure difference (just beyond the upper end of the range of typical building differentials mentioned above) drew air into the source room wall through pre-existing pathways, such as a $\sim 10 \times 10$ cm hole under the bathroom sink and cracks around wall fittings. Air moved from the source room into the receptor room through two electrical outlets and a phone jack at a calculated flow rate of 12–14 m$^3$/h, which is in the range reported by Bohac et al. (2011).

Instruments measuring particle composition

Besides the aforementioned particle PAH and active surface monitors, we used the aethalometer to measure the unique composition of SHS particles. To account for the optical saturation effect that occurs immediately after a 10-min change of the aethalometer quartz tape filter, we used the method of Virkkula et al. (2007) to adjust UV and BC data, with instrument-specific optical compensation factors provided from 13 SHS chamber tests conducted by Bohac et al. (2010). In addition, certain high-concentration experiments required a modified Virkkula method from Bohac et al. (2010).

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Exp. # & Rm & Location & Source types & Vol. [m$^3$] & Air exchange rate mean (range) [per h] & $D_P$ [Pa] & Inter-room flow mean (range) [m$^3$/h] \\
\hline
1–39 & S & Stanford & 8 cigarette \\
& & & 5 cig/cook mix \\
& & & 3 cig/incense mix \\
& & & 17 cooking \\
& & & 5 incense \\
& & & 1 candle \\
& & & & 47 & 0.2 (0.1–0.5)$^d$ \\
& & & & & 2.1 (0.8–4.2)$^d$ \\
& & & & NA & NA \\
\hline
40–49 & S & MP & 7 cooking \\
& & & 2 incense \\
& & & 1 candle \\
& & & & 60 & 0.7 (0.6–0.8) \\
& & & & NA & NA \\
\hline
50–52 & S & LA (LR) & 1 synthetic log \\
& & & 2 wood smoke \\
& & & & 138 & 0.6 (0.5–0.7) \\
& & & & NA & NA \\
\hline
53–60 & S & LA (K) & 6 cooking \\
& & & 1 incense \\
& & & 1 candle \\
& & & & 44 & 1.2 (1.1–1.5) \\
& & & & NA & NA \\
\hline
61–66 & S&R & Stanford & 2 cigarette \\
& & & 1 cig/incense mix \\
& & & 2 cooking \\
& & & 1 incense \\
& & & & S: 47 \\
& & & & R: 18 \\
& & & & S: 0.6 (0.5–0.7) \\
& & & & R: 4.6 (3.7–5.3) \\
& & & & 7$^j$ \\
& & & & 21 (14–24) \\
\hline
67–73 & S&R & RC motel & 4 cigarette \\
& & & 1 cig/incense mix \\
& & & 1 cooking \\
& & & 1 incense \\
& & & & S: 72 \\
& & & & R: 80 \\
& & & & S: 0.4 (0.4–0.5)$^g$ \\
& & & & R: 2.4 (1.8–2.7)$^g$ \\
& & & & 6–13 \\
& & & & 13 (12–14)$^g$ \\
\hline
\end{tabular}
\caption{Summary of experimental study locations, characteristics, and source types}
\end{table}

\begin{tablenotes}
\item $^a$Room: S, source, R, receptor.
\item $^b$MP, Menlo Park apt; LA (LR), Los Altos living room; LA (K), Los Altos kitchen; RC, Redwood City.
\item $^c$Receptor room air exchange rate and inter-room flow calculated as per method of Ott et al. (2003), with special case of zero reverse flow.
\item $^d$Windows closed ($n = 34$).
\item $^e$Windows open ($n = 5$).
\item $^f$Window applies to all but #64 (not available for this experiment).
\item $^g$Applies to #68–71 only, others NA.
\item $^h$Shelby and R. Shanker.
\end{tablenotes}
et al. (2010) with integrated PM$_{2.5}$ data to accurately adjust the data. Additional notes on calculation of the (UV-BC)/UV ratio are provided in Supporting Information.

VOC sampling

We used 6-L canisters to sample emissions, with analysis (by Columbia Analytical Services/ALS, Simi Valley, CA, USA) via gas chromatography for 19 VOCs of interest. We focused later analyses on 14 VOCs (Table 2), most of which had moderate to high cigarette emission rates and were more likely to remain detectable in adjacent rooms or living units. To avoid partitioning loss issues, we focused our sampling on volatile compounds (rather than on semi-volatile). Additional details of the VOC sampling protocol are provided in Supporting Information.

Other instruments

We also employed TSI Model AM-510 SidePak™Personal Aerosol Monitors (TSI Inc., Shoreview, MN, USA) to measure PM$_{2.5}$, calibrating the readings to personal Aerosol Monitors (TSI Inc., Shoreview, MN, USA) via gas chromatography for 19 VOCs of interest. We focused later analyses on 14 VOCs (Table 2), most of which had moderate to high cigarette emission rates and were more likely to remain detectable in adjacent rooms or living units. To avoid partitioning loss issues, we focused our sampling on volatile compounds (rather than on semi-volatile). Additional details of the VOC sampling protocol are provided in Supporting Information.

We also employed TSI Model AM-510 SidePak™Personal Aerosol Monitors (TSI Inc., Shoreview, MN, USA) to measure PM$_{2.5}$, calibrating the readings to gravimetric measurements based on the EPA reference method (United States Code of Federal Regulations, 1997). Our gravimetric sampler consisted of an electric pump connected to a cyclone with a 2.5-μm-size cut, and polytetrafluoroethylene (PTFE) membrane filters (Pall Corp., Ann Arbor, MI, USA). The filters were weighed after at least 24 h of equilibration at controlled humidity and temperature on a Mettler-Toledo M3 microbalance (Columbus, OH, USA). Finally, we employed Bruel and Kjær Type 1302 Multi-Gas Monitors (Bruel & Kjær, Nærum, Denmark) to measure sulfur hexafluoride tracer gas released in the source room, to measure the room air exchange rate. We also used these measurements to estimate the airflow between source and receptor rooms via the method of Ott et al. (2003), assuming zero reverse airflow from receptor to source rooms.

Course of experimental events

After preparing instruments and samplers for use (synchronizing clocks, setting logging interval, measuring flow rates, calibrating PM$_{2.5}$ monitor with high-efficiency filter, etc.), we collected a period of background data with the room under the steady-state ventilation conditions appropriate for that experiment (e.g., doors and windows closed, or fans drawing air from source room to receptor room). Subsequently, we first released sulfur hexafluoride tracer in the source room and then initiated the source emissions (details of specific sources are discussed below). After the source emissions ended, we waited ~5–15 min for the room to become well mixed and then started the gravimetric sampling apparatus and collected VOC samples. At the end of the experiment, we opened all doors and windows to flush any remaining emissions out of the room and then collected an additional 5–10 min of background data. If we were conducting multiple experiments on the same day, we left instruments running throughout the day. Specific details of every experiment are provided in Supporting Information (Table S5).

Sources

We focused on sources considered most likely to be found in real homes (Table S5). Cigarette experiments typically consisted of a single Marlboro Gold cigarette, lit by butane lighter, and smoked by a volunteer smoker. In six experiments, we burned three to six cigarettes to achieve higher concentrations; in those cases, the volunteer smoked only one or two cigarettes in sequence, while the rest smoldered in an ashtray. Cooking experiments included a broad range of methods (frying, baking, toasting, microwaving, etc.) using typical cookware with both gas and electric stoves, and ovens. Sources included a variety of fried meat or fish (hamburger, chicken, bacon, pork, sausage, salmon), as well as toast, popcorn, french fries, pizza, and oven drippings. Incense experiments incorporated four

---

**Table 2** Summary of measurements with applications to CMB analysis and logistic regression

<table>
<thead>
<tr>
<th>Instrument/sampler</th>
<th>Measurement (also CMB ‘ambient’ species)</th>
<th>Abbrev.</th>
<th>Units</th>
<th>CMB source profile$^a$</th>
<th>Logistic regression variable [–]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dylos</td>
<td># particles $&gt; 0.5$ μm</td>
<td>0.5</td>
<td>#/m$^3$</td>
<td># $&gt; 0.5$ μm/μg PM$_{2.5}$</td>
<td>2.5/0.5</td>
</tr>
<tr>
<td>Dylos</td>
<td># particles $&gt; 2.5$ μm</td>
<td>2.5</td>
<td>#/m$^3$</td>
<td># $&gt; 2.5$ μm/μg PM$_{2.5}$</td>
<td>2.5/0.5</td>
</tr>
<tr>
<td>Grimm</td>
<td># particles $&gt; 0.3$ μm</td>
<td>0.3</td>
<td>#/m$^3$</td>
<td># $&gt; 0.3$ μm/μg PM$_{2.5}$</td>
<td>1.0/0.3</td>
</tr>
<tr>
<td>Grimm</td>
<td># particles $&gt; 1.0$ μm</td>
<td>1.0</td>
<td>#/m$^3$</td>
<td># $&gt; 1.0$ μm/μg PM$_{2.5}$</td>
<td>1.0/0.3</td>
</tr>
<tr>
<td>EcoChem PC</td>
<td>particle PAH</td>
<td>PC</td>
<td>ng/g/m$^3$</td>
<td>ng PAH/μg PM$_{2.5}$</td>
<td>PC/OC</td>
</tr>
<tr>
<td>EcoChem DC</td>
<td>Active surface area</td>
<td>DC</td>
<td>mm$^2$/m$^3$</td>
<td>mm$^2$/μg PM$_{2.5}$</td>
<td>PC/OC</td>
</tr>
<tr>
<td>Aethalometer (BCC)</td>
<td>Black carbon (BC)</td>
<td>BC</td>
<td>μg/m$^3$</td>
<td>μg BC/μg PM$_{2.5}$</td>
<td>(UV-BC)/UV</td>
</tr>
<tr>
<td>Aethalometer (BCB)</td>
<td>UV-absorbing matter (BC-equiv.)</td>
<td>UV</td>
<td>μg/m$^3$</td>
<td>μg UV/μg PM$_{2.5}$</td>
<td>(UV-BC)/UV</td>
</tr>
<tr>
<td>Canister</td>
<td>VOC</td>
<td>VOC</td>
<td>μg/m$^3$</td>
<td>μg VOC/μg PM$_{2.5}$</td>
<td>VOC/PM$_{2.5}$$^b$</td>
</tr>
</tbody>
</table>

$^a$ Created by dividing measurement by PM$_{2.5}$ concentration (μg/m$^3$) from corresponding period.

$^b$ VOCs include chloromethane, 1,3-butadiene, acetonitrile, acrolein, acetone, vinyl acetate, benzene, toluene, ethylbenzene, m,p-xylene, styrene, d-limonene, isoprene, 2,5-dimethylfuran.
different types of incense, including both stick and cone varieties. Mixed-source experiments (cigarette/cooking or cigarette/incense) involved source emissions in sequence in all but one experiment (64), with one source ending before the next began. Scented candle experiments incorporated a sustained burn period of 10–20 min followed by several sequences of extinguishing and relighting the candles to generate additional PM$_{2.5}$. Finally, fireplace experiments used cherry wood or a commercial synthetic log (Duraflame, Stockton, CA, USA) with newspaper to assist with lighting the fire and/or priming the flue. Fires were extinguished using water from a bucket or spray bottle, after which the fireplace front was covered with plastic to prevent any further emissions from entering the room.

Periods of analysis

For each experiment, we calculated background-subtracted averages for each species of interest from the initial portion of the well-mixed decay period, typically 60 min. In the first 60 single-room experiments, we had two of every instrument except the active surface (DC) monitor, yielding numerous collocation experiments. When two identical instruments were collocated, we used the mean result. In cases where instruments were not collocated (as in the two-room experiments), we adjusted the results to the mean using the collocation-derived relationship. Low concentrations encountered in the receptor room for two-room experiments required us to develop a ‘low-end’ mean adjustment derived from only lower-concentration collocations; otherwise, the adjustment to the overall mean would introduce too much error for measurements near zero. The selection of the ‘low-end’ collocations was based on the range of concentrations of that species encountered in the receptor room and ranged from 36 to 54% of the total number of collocations for each instrument.

We took VOC samples (and corresponding PM$_{2.5}$ measurements for VOC/PM$_{2.5}$ ratios) in the first 15 min of the well-mixed decay period, aiming to capture detectable concentrations of as many of the VOCs as possible. The exceptions to this were the VOC samples from the first three single-chamber experiments, which extended over the first 60 min of the well-mixed period.

Detailed explanations of the calculation of measurement uncertainty, as well as a discussion of missing data, are provided in Supporting Information.

Modeling approaches

Logistic regression modeling (SHS identification)

To identify the presence or absence of SHS, we created logistic regression models, which offer the advantage of being relatively easy to employ. These are statistical models that return as their response variable a probability ($\pi$) between 0 (is not SHS) and 1 (is SHS) based on the observed data (Chatterjee and Hadi, 2006). Once created, these can be used to identify SHS simply by substituting measurements into an equation:

$$
\pi = \frac{e^{x}}{1 + e^{x}} = \frac{e^{\beta_0 + \beta_1 V_1 + \beta_2 V_2 + \ldots + \beta_p V_p}}{1 + e^{\beta_0 + \beta_1 V_1 + \beta_2 V_2 + \ldots + \beta_p V_p}}.
$$

In these models, the variables ($V_1,\ldots,V_p$) are formed from our measures of interest; we created 18 in total from which to choose in the model formulation (Table 2; note that the 14 VOC/PM$_{2.5}$ ratios are listed in footnote b). We used the programming language R (R Development Core Team, 2010) with associated packages to fit parameters ($\beta_0,\ldots,\beta_p$) to data from the first 60 single-room experiments (Table 3); we used the final 13 two-room experiments as an independent data set on which to test the models. As many methods of statistical analysis work best with a full data set, we used ‘k-nearest neighbor’ imputation to fill in missing data from these experiments, with $k=6$ (R Package ‘imputation’, Wong, 2011). For each model, a returned probability of greater than 50% indicated that SHS presence was likely. Detailed information on model selection and refinement is provided in Supporting Information.

Chemical mass balance analysis (SHS quantification)

To quantify the amount of SHS and to determine what fraction of the aerosol came from other sources, we applied chemical mass balance (CMB) analysis, using our source emission ‘signatures’. CMB compares several measured ‘ambient’ species values in an unknown aerosol with previously determined ‘profile’ values of those species in possible sources and numerically determines the portion of the aerosol generated by each source (summarized in Coulter, 2004). The tool we used, the EPA’s CMB 8.2 program (2004), utilizes effective variance weighted least squares to compute source contributions and related uncertainties (Coulter, 2004). CMB 8.2 does not require a full set of ambient measurements to perform source apportionment and thus continues to be useful in cases where data are incomplete or missing.

For our CMB species, we used all the measurements and uncertainties from the instruments and samplers discussed above (Table 2). We created source profiles, with uncertainties, using data from the 60 single-room experiments; again, we reserved the final 13 two-room experiments as independent test cases. To create profiles, we calculated the mean and standard deviation of species measurements/PM$_{2.5}$ for each of seven source categories (cigarettes, cooking, cone incense, stick incense, wood, synthetic log, and candles), which were
determined via hierarchical cluster analysis of the pure source single-room experimental data. For the synthetic log source profile uncertainty (m = 1), we used propagation of error of each species and PM$_{2.5}$ uncertainties. As CMB analysis did not require a full data set, we did not use imputation to fill in the missing data as we did for logistic regression. Additional details of the hierarchical cluster analysis and the employment of CMB 8.2 are provided in Supporting Information.

### Results and discussion

#### Single-room experiments

We present a selection of some of the best logistic regression models and their cross-validation rates (cv, %) in Table 3. A model's ‘cross-validation rate’ is an assessment of its predictive accuracy; additional details on this topic are provided in Supporting Information. Generally, those variables with the most predictive ability by themselves were the particle size ratios, the aethalometer (UV-BC)/UV ratio, and several VOC ratios including acetonitrile/PM$_{2.5}$, toluene/PM$_{2.5}$, and m,p-xylene/PM$_{2.5}$. From the logistic regression models consisting of these single variables, we determined a ‘cutoff’ of the variable value for which the probability of SHS is 50% (Table 3, ‘notes’ column).

The PC/DC ratio was not helpful in distinguishing SHS; while it was useful in distinguishing SHS from cooking, too many other non-SHS sources (incense, candles, wood smoke) had PC/DC ratios in the same range as SHS for it to be a useful indicator. In addition, the ratio of 2,5-dimethylfuran/PM$_{2.5}$ was not a useful indicator of SHS for two reasons: (i) its emission rate from cigarettes was too low for the VOC to be reliably detectable, even in the source room, and (ii) we also found measurable quantities of 2,5-dimethylfuran in emissions from other sources. Thus, not only were we unable to detect the VOC in one experiment involving bacon and cigarette emissions mixed (LOD of 1.4 µg/m$^3$), but we found measurable quantities in seven experiments without a cigarette source: two cone incense, two burned toast, one burned microwave popcorn, and two wood smoke experiments.

![Table 3 Selected logistic regression models with parameters ($\beta$) and variables ($V$)](image)

- **#** Model name | cv [%]$^a$ | $\beta_x$ | $V_y$ | $\beta$ | $V$ | $\beta$ | $V$ | Notes
- 1 | Aeth. (A) | 97 | -15.1 | A | 16.1 | | | Cutoff: $\geq 0.94 = \text{SHS}^c$
- 2 | Dylos (D) | 75 | 0.3 | D | -31.7 | | | Cutoff: $\leq 0.01 = \text{SHS}^c$
- 3 | Grimm (G) | 78 | 1.0 | G | -499.8 | | | Cutoff: $\leq 0.002 = \text{SHS}^c$
- 4 | Acetonitrile | 92 | -4.2 | ACE | 142.5 | | | Cutoff: $\leq 0.03 = \text{SHS}^c$
- 5 | A & D | 82 | -10.0 | A | 6.8 | G | -376.8 | Use if VOC sample NA
- 6 | A & G | 83 | -5.3 | A | 6.8 | G | -376.8 | Use if VOC sample NA
- 7 | A, D, VOC 1 | 100 | -2.9 | A | 1.9 | D | -7.2 | ACE | 71.5
- 8 | A, G, VOC 1 | 100 | -3.0 | A | 2.0 | G | -38.4 | ACE | 75.5
- 9 | A, D, VOC 2 | 100 | -3.3 | A | 2.0 | D | -7.3 | ACE | 71.2
- 10 | A, G, VOC 2 | 100 | -3.5 | A | 2.1 | G | -40.5 | ACE | 77.9
- 11 | A, D, VOC 3 | 88 | -25.7 | A | 20.3 | D | 16.9 | ACE | 316.7
- 12 | A, G, VOC 3 | 90 | -4.8 | A | 2.1 | G | -40.1 | ACE | 43.4
- 13 | A, D, VOC 4 | 97 | -3.3 | A | 1.3 | D | -4.3 | ACE | 32.1
- 14 | A, G, VOC 4 | 97 | -3.4 | A | 1.3 | D | -23.2 | ACE | 32.3

*aCorrect prediction rate by leave-one-out cross-validation (does not account for revised parameters from ridge regression).

*bA-aethalometer (UV-BC)/UV; D-Dylos 2.5/0.5; G-Grimm 1.0/0.3; ACE-acetonitrile/PM$_{2.5}$; BEN-Benzene/PM$_{2.5}$; ACTN-acetone/PM$_{2.5}$; STY-styrene/PM$_{2.5}$; MPX-m,p-xylene/PM$_{2.5}$; ISO-isoprene/PM$_{2.5}$; 13B-1,3-butadiene/PM$_{2.5}$; TOL-toluene/PM$_{2.5}$; ETH-ethylbenzene/PM$_{2.5}$; LIM-d-limonene/PM$_{2.5}$.

**cCutoff**: variable value at which SHS probability is 50%.

**dRidge**: ridge regression used to refine parameter estimates (see Supporting Information).
emissions and prevents an anomaly in one type of measurement from skewing the results. In light of this, some of the most useful logistic regression models are those that include results from the Grimm or Dylos, the aethalometer, and several VOCs. The VOC sample would probably be the most challenging to obtain in a field setting, although it would not be impossible (as residents could be instructed to open the sampler valve when they smell smoke, and background samples could be taken by researchers at another time). However, even without VOCs, we achieved quite useful results with models consisting of only the aethalometer and the Dylos or Grimm (cross-validation rate of 82–83%).

For the first 60 experiments, CMB returned reasonable estimates of all true sources in 54 of 60 (90% of) experiments (Supporting Information). In two experiments (#18 and #30), it was missing a source, in another two (#49 and #52), it included very small amounts of an extra source, and in one (#39), it misidentified the stick incense as cone incense. In only one experiment (#60, consisting of baking frozen pizza in an electric oven) did the results not include any correct source classification. Instead, the program did not return a significant result, likely because the emissions were so low (average of 4 μg/m³ PM₂.₅ above background for the period of analysis). Overall, we obtained good fits, with 75% of the $R^2$ values greater than 0.80 and 97% of the $χ^2$ values in the optimum range of <2.00.

To assess the source apportionment for the eight mixed-source tests, we compared the CMB results with those obtained by independent calculation of source contributions for each experiment, which was possible because the sources occurred in succession (Ferro et al., 2004). Using the measured PM₂.₅ decay rate for both sources combined, we projected the PM₂.₅ concentration due to only the first source from the beginning of the second source emission period to the end of the experiment. Subtracting this amount from the total measured concentration gave the estimated contribution of the second source alone. Finally, calculating the average concentration due to each source during the period of interest enabled us to estimate the fractional contribution of each. This method assumes the same decay rate for the PM₂.₅ from both sources (given that both involve biomass combustion). In three of the five experiments where CMB correctly predicted the presence of both sources (#15, #25, #37), results of both methods indicated a similar prevalence of one source over the other (Table 4).

Examining the modified pseudo-inverse normalized (MPIN) matrix from CMB 8.2 for species absolute values >0.5 showed which species were most influential in source apportionment. In general, the most influential species for cooking source apportionment were active surface (DC) and particle size, with the DC and both Dylos size species influential 100% of the time when cooking sources were correctly apportioned. Selected VOCs (acetonitrile, benzene, m,p-xylene, and isoprene) were influential for correct cigarette apportionment in 67–80% of cases. Finally, aethalometer UV and active surface (DC) measurements were influential in correct

Table 4 Evaluation of CMB apportionment, experiments 1–60 (mixed sources)

<table>
<thead>
<tr>
<th>Experiment &amp; sources</th>
<th>PM₂.₅ source contribution estimate</th>
<th>Independent estimate via projected decay (see text)</th>
<th>Correctb</th>
<th>Half-correctb</th>
</tr>
</thead>
<tbody>
<tr>
<td>15: Bacon &amp; cigarette</td>
<td>42% cooking 56% cigarette</td>
<td>23% cooking 77% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>18: Burger &amp; cigarette</td>
<td>61% cooking 39% stick incense</td>
<td>39% cooking 61% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>21: Bacon &amp; cigarette</td>
<td>54% cooking 45% cigarette</td>
<td>29% cooking 71% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>25: Chicken &amp; cigarette</td>
<td>54% cooking 39% cigarette</td>
<td>36% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>30: Burned toast &amp; cigarette</td>
<td>85% cigarette</td>
<td>38% cooking 62% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>35: Stick incense &amp; cigarette</td>
<td>62% stick incense 38% cigarette</td>
<td>45% stick incense 55% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>37: Cone incense &amp; cigarette</td>
<td>39% cone incense 61% cigarette</td>
<td>30% cone incense 70% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>39: Stick incense &amp; cigarette</td>
<td>7% cone incense 72% cigarette</td>
<td>37% stick incense 63% cigarette</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total 5 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6If percentages do not add up to 100%, remainder could not be apportioned in a statistically significant manner.
6Correct (all true sources identified) and 'half-correct' (one of two true sources identified, or includes incorrect extra source) refers only to source ID, not whether source contributions matched independent estimates.
apportionment of wood or synthetic log emissions in all fireplace experiments.

Two-room experiments

We subsequently applied the logistic regression models and CMB source profiles created in the single-room experiments to both the source and receptor rooms of the two-room experiments (Table 5). For those experiments with a full data set (#62-66 and #68-73), several logistic regression models were >80% correct in both source and receptor rooms. Of particular note, the model including the aethalometer, Dylos, and measurements of acetonitrile, m,p-xylene, and isoprene was correct in all 11 of the full data set experiments.

CMB results were generally accurate as well, when evaluating its performance on whether or not it provided a significant estimate of the true sources. Of

<table>
<thead>
<tr>
<th>Source</th>
<th>#</th>
<th>Rm</th>
<th>Logistic regression: probability SHS by model (%)</th>
<th>CMB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Cigarette</td>
<td>61</td>
<td>S</td>
<td>53</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>51</td>
<td>–</td>
</tr>
<tr>
<td>Cigarette</td>
<td>62</td>
<td>S</td>
<td>54</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td>Cigarette</td>
<td>67</td>
<td>S</td>
<td>16*</td>
<td>43*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>56</td>
<td>73</td>
</tr>
<tr>
<td>Cigarette</td>
<td>68</td>
<td>S</td>
<td>57</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>41*</td>
<td>62</td>
</tr>
<tr>
<td>Cigarette</td>
<td>69</td>
<td>S</td>
<td>51</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>58</td>
<td>59</td>
</tr>
<tr>
<td>Cigarette</td>
<td>70</td>
<td>S</td>
<td>36*</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>58</td>
<td>77</td>
</tr>
<tr>
<td>Stick incense</td>
<td>63</td>
<td>S</td>
<td>37</td>
<td>62*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>Cone incense</td>
<td>71</td>
<td>S</td>
<td>48</td>
<td>65*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>48</td>
<td>67*</td>
</tr>
<tr>
<td>Chicken</td>
<td>64</td>
<td>S</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Chicken</td>
<td>65</td>
<td>S</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Bacon</td>
<td>72</td>
<td>S</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>54*</td>
<td>42</td>
</tr>
<tr>
<td>Cig/st. incense</td>
<td>66</td>
<td>S</td>
<td>42*</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>50*</td>
<td>49*</td>
</tr>
<tr>
<td>Cig/cone incense</td>
<td>73</td>
<td>S</td>
<td>43*</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>50*</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S&amp;R</td>
<td>20/26</td>
<td>19/24</td>
</tr>
</tbody>
</table>

Table 5 Application of logistic regression models and CMB analysis to two-room experiments

aRoom: S, source; R, receptor.
bProbability > 50% indicates SHS; asterisk (*) and bold font indicate incorrect result; ‘=’ indicates model cannot be used due to missing measurement(s).
cPilot study: Dylos, PM2.5 measurements only.
dHalf-correct: missing 1 true source or includes erroneous extra source.
ePilot study: no VOC sample.
fCorrect identification of 100% incense, although incense type apportionment not fully correct.
gProbabilities for model 2 receptor room #66 and #73 were both >50% (50.2% and 50.1%, respectively).
hIndependent estimation of apportionment by source strength for #66 source room was 29% incense, 71% cigarette. Estimate was not possible for #73 because sources were ignited simultaneously.
Dacunto et al.

the 26 cases (13 experiments, two rooms each), 18 were fully correct and six (four source room and two receptor room experiments) were half-correct, with either an additional source incorrectly included or a true source missing. In two cases, CMB provided an incorrect source estimation, both in the receptor room (#67 and 72).

While the overall magnitude of each species was always significantly lower in the receptor room as expected, the change in relative species concentrations as they passed from room to room was difficult to predict. For example, ratios of VOC/PM$_{2.5}$ sometimes increased in the receptor room due to low PM$_{2.5}$ concentrations, despite the significant drop in both species. Dylos (>2.5 µm)/(>0.5 µm) ratios decreased in 8 of 13 cases indicating a preferential loss of larger particles, but actually increased in three cases. Finally, the aethalometer (UV-BC)/UV decreased in every case in the Stanford building, as UV measurements decreased more significantly than did BC. However, the ratio actually increased in 5 of 7 motel experiments, decreasing or remaining equal in the other two; this result indicates that in that setting, the BC measurements decreased significantly as well. Overall, while species did undergo changes as they passed from the source to receptor rooms, they remained conserved enough that logistic regression models and CMB source profiles created from single-room experiments were useful in the receptor room as well.

CMB vs. logistic regression

The CMB approach provides a very sophisticated analysis of the data, but it requires more effort and expertise to employ. In contrast, logistic regression models (once formulated) are simpler to employ and offer a means of quickly assessing the presence or absence of SHS; also, if the probability of SHS is high, one can even estimate its quantity via the associated PM$_{2.5}$ measurements. While neither the logistic regression nor the CMB approach was perfect in all cases, they were most useful when used together. For example, if several regression models indicated the presence of SHS, they supported a CMB result that included a cigarette source.

Field applications

While the number of instruments and samplers required for obtaining a full data set is significant, useful results can still be obtained with only a few instruments and samplers. For example, the logistic regression model using data from only the Dylos and aethalometer has a cross-validation rate of 82%, and CMB continues to be useful with such reduced data sets (provided PM$_{2.5}$ data are included as well). Larger epidemiological studies that seek to examine the association between SHS transfer in MUHs and adverse health outcomes could use an even more limited (and inexpensive) approach, employing only a nicotine sampler, Dylos, and PM$_{2.5}$ monitor. The nicotine sampler provides proof of SHS intrusion at some point; subsequently, the Dylos data with the associated logistic regression model can be used to identify which PM$_{2.5}$ peaks are most likely SHS. Finally, the magnitude and duration of the intrusion can be measured from the PM$_{2.5}$ time series data with a real-time particle monitor.

Limitations

The source profiles and regression models presented here are based upon a reasonably small sample size ($n = 60$) with only a single type of cigarette and undoubtedly do not adequately represent or model every possible condition in the field. Further data from experiments in additional source and receptor rooms, with even more types of sources, would contribute to improved models and source profiles. In addition, these methods need to be evaluated in a field study to see whether they are equally useful in real MUHs with SHS intrusion problems. However, the initial testing and calibration of the methods under controlled settings indicate that they show some promise for such applications in the field.

Conclusion

We have developed several logistic regression models that can be used to accurately predict the presence or absence of cigarette smoke in both source and receptor rooms. In addition, the profiles of common indoor sources (presented as Supporting Information to this article) can be used in CMB analysis to determine estimates of SHS concentrations in both source and receptor rooms.

This work demonstrates that accurate identification and quantification of SHS with precise temporal resolution may be feasible in multunit housing applications. It suggests that an approach using multiple measures of particle size, particle composition, and VOCs in only the receptor apartment might augment the use of single tracers to better estimate exposure. For example, a week-long nicotine sample would indicate that SHS did drift into an apartment at some point in time, but this offers no insights into the timing or duration of the SHS exposure, nor on the relative concentration levels of other SHS pollutants also present, such as PM$_{2.5}$. However, with data from the appropriate instruments, investigators could determine which PM$_{2.5}$ peaks were most likely SHS using logistic regression models, and subsequently quantify the SHS using CMB. Such an approach would enable a more accurate assessment of the duration and intensity of exposures for nonsmoking residents of multunit homes.
Acknowledgements

This research was supported by funds from the California Tobacco-Related Disease Research Program of the University of California, grant number 19CA0123. We thank Prof. Jonathan Taylor of the Stanford Department of Statistics for assistance with logistic regression modeling. We also thank ClearWay, Minnesota, for their loan of the aethalometers and Martha Hewett, Dave Bohac, and Josh Novacheck of the Center for Energy and the Environment for their extensive assistance with the aethalometers.

References


Dacunto et al.


