Abstract
The objective of this study was to develop and evaluate a methodology to identify individual sources of emissions based on the measurements of mixed air samples and the emission signatures of individual materials previously determined by Proton Transfer Reaction-Mass Spectrometry (PTR-MS), an online analytical device. The methodology based on signal processing principles was developed by employing the method of multiple regression least squares (MRLS) and a normalization technique. Samples of nine typical building materials were tested individually and in combination, including carpet, ceiling material, gypsum board, linoleum, two paints, polyolefine, PVC and wood. Volatile Organic Compound (VOC) emissions from each material were measured in a 50-liter small-scale chamber. Chamber air was sampled by PTR-MS to establish a database of emission signatures unique to each individual material. The same task was performed to measure combined emissions from material mixtures for the application and validation of the developed signal separation method. Results showed that the proposed method could identify the individual sources under laboratory conditions with two, three, five and seven materials present. Further experiments and investigation are needed for cases where the relative emission rates among different compounds may change over a long-term period.

Keywords
Material emission signature; PTR-MS; VOC; Source identification; Signal processing

1. Introduction
Source identification has been a challenging research topic in the area of indoor and outdoor air quality because this issue may have the key to open a new gate to develop optimal control protocols for protecting and improving human’s life and welfare for the coming ages. Over the last several decades, emissions of Volatile Organic Compounds (VOC) and Particulate Matter (PM) have been of global concern because of their significant impact on human’s health, comfort and performance. Exposure to these air pollutants has been reported to have various critical impacts on human performance and cardiovascular/respiratory related diseases in a negative way (Pope and Dockery, 2007; Wargocki and Wyon, 2006). For example, Corbett et al. (2007) examined the link between PM emissions from ships and human’s health, assessed their potential impact leading up to 60,000 premature deaths on a global scale annually and expected a further mortality increase by 40% by 2012 under current regulations and activity conditions. Moreover, this widespread concern is expected to increase in the future because of the fast growing rate of global-scale commercial trade and worldwide distributed manufacturing, the associated increase of vehicle/ship/air traffics and the corresponding massive use of diesel/gasoline fuels (Corbett et al., 2007; EPA, 2009).

Although the predicted threats of these emissions have succeeded in attracting worldwide attention for reducing these emissions, it has been generally agreed that the present countermeasures for dealing with these emissions were not sufficient to protect human’s health (EPA, 2009). Because different sources of air pollution may affect the exacerbation of human’s health effects in different ways, it has been suggested as a desirable countermeasure to identify and quantify the partial contribution each type of major air pollution source makes to the overall air quality condition (Lin et al., 2010). In addition, the accurate and complete apportionment of these pollution sources is regarded as an essential step to the development of optimal control protocols for global and built environments (Lin et al., 2010).
Source identification and apportionment has been an active research subject for outdoor air quality, but existing methods have shortcomings. Many receptor modeling studies have been performed on ambient air quality data collected for this apportionment purpose (Cass, 1998; Cincinelli et al., 2003; Didyk et al., 2000; Edwards et al., 2001; Graham et al., 2004; Hagler et al., 2006; Lewis et al., 2004; Simoneit, 2002; Wu et al., 2007). Two receptor-based apportionment methods, Chemical Mass Balance model (CMB) and Positive Matrix Factorization (PMF), are most common techniques used nowadays for the source apportionment purpose. CMB has a limitation in identifying any new or unknown sources (Fujita et al., 2003; Schauer et al., 1996). On the other hand, the limitation of PMF is that the characteristics of the profiles for several key factors identified (Begum et al., 2004; Buzcu and Fraser, 2006; Kim et al., 2003, 2004, 2008; Paatero and Tapper, 1994; Ramadan et al., 2000; Viana et al., 2009). Moreover, both methods are known to have high bias for some cases because of high variability and complexity of VOC/PM emissions, which makes the results from these methods as rough estimates on the profile of sources not knowing the true profile of emission sources (Zhao et al., 2007). Most compounds can be emitted from multiple types of sources, and thus identifying and quantifying sources on the basis of their correlation with several elemental data is limited and difficult because of similar emission characteristics of different sources.

Different from the area of outdoor air quality, indoor source identification has rarely been studied with a few exceptions (Arhami et al., 2010; Jia et al., 2010; Liu and Zhai, 2007, 2008; Zhang and Chen, 2007). The problem has its challenge because of both indoor and outdoor sources, affected by outdoor air via traffic emissions and long-range transport pollutants (Edwards et al., 2001; Zhao et al., 2007) as well as indoor sources. In many cases, it has been known that the strength of indoor emissions has a more significant influence on indoor VOC concentrations than the effect of infiltration from outdoor air (Kim et al., 2001). One of the recent important research challenges in the field of indoor air quality is the identification of emission sources of indoor VOCs. Several field studies were performed in residential buildings to identify several possible active compounds with great potency to human’s health and perception, and their chemical measurements were reported. However, it was hard to trace the emission sources of the detected compounds clearly. As an example, Hodgson et al. (2000) identified acetic acid as one of the important compounds, but were not certain about the sources of acetic acid in the studied houses. A first step toward a source identification would be to determine a material emission signature if it is unique for each material or each type of material. This is similar to human’s fingerprint for personal identification. A study previously conducted to deal with this issue (Han et al., 2010) showed that unique emission patterns appeared to exist for different types of building materials. These patterns could be established by Proton Transfer Reaction - Mass Spectrometry (PTR-MS). Several studies performed in other research areas such as food engineering, medical research, forensic investigation, etc. implied that pollutant mass spectra measured by PTR-MS, could be useful tools for detection and identification purposes (Granitto et al., 2007, 2008; Lirk et al., 2004; Mayr et al., 2003; Moularat et al., 2008; Van Ruth et al., 2007; Wehinger et al., 2007; Whyte et al., 2007). With the definition of material emission signatures, a signal processing technique for pinpointing the materials responsible for certain indoor VOCs may be feasible. The objectives of this study were to explore the feasibility of a signal processing methodology for emission source identification, validated by measurement-based Monte-Carlo simulations, and to apply the method to actual combined emission measurements for separating and identifying the individual sources of emissions. To our best knowledge, this study is the first of its kind on how to clearly pinpoint indoor VOC emission sources.
2. Materials and methods

2.1. Overview and basic assumptions

In a previous study (Han et al., 2010), chamber exhaust air polluted by the emissions from nine individual building materials in a 50-l small-scale chamber was sampled by PTR-MS at low, medium and high airflow rates to establish a set of material emission signatures stable over different sampling time and over different area-specific ventilation rates, and to determine their variances because of noise. The emission signature is the PTR-MS ion mass \( m/z \), which is a physical quantity denoting the mass-to-charge ratio widely used in the electrodynamics of charged species) spectrum of the air sampled from each material, followed by subtracting the background air signature from the measured and removing any components within the measurement uncertainty (< 3 sigma of the background signal). Five actual emission cases of material mixtures were studied to validate source identification methods developed. The experimental setup is shown in Figure 1. Two signal separation methods were proposed, tested using measurement-based Monte-Carlo simulations, and validated on the five multi-material mixtures. Before going into details, it would be useful to formally state the basic assumptions the separation methods were based on: 1) Emission signature exists and is unique for each material type, 2) Interaction effects between material emissions are small and can be modeled as noise or compensated, and 3) Emission signatures for material mixtures can be established by the superposition of individual emission signatures of the co-located materials.

2.2. Algorithms for separation and identification (Source identification module)

2.2.1. Algorithm 1

Under the basic assumptions stated in the overview, the source identification problem to be examined in this study can be configured as the estimation of the set of materials present in the chamber \( \{ID\} \) expressed in terms of a set of integer material identification numbers, or IDs (i.e. \( i = 1, 2, \ldots, 9 \) defined in Table 1 for each material) and the corresponding emission concentration level of each identified material expressed as a positive real number \( \alpha_i \), which is the signal intensity multiplication factor of material \( i \). The measured output from PTR-MS is a mass spectrum of the emissions from a material mixture, consisting of ion mass \( x \) (related to
VOC) and its signal intensity \( S_{sp}(x) \) (related to concentration level) for all scanned ion masses \( x=[x_1, x_2, \ldots, x_N] \). Measurements from single material emission tests (\( \sigma \sim 2 \text{neps} \) or \(< 0.3 \text{ ppb}\) for most ion masses) have shown that measured signal noise \( w(x_n) \) of PTR-MS for each ion mass \( x_n \) (where \( n=1, 2, \ldots, N \)) can be roughly modeled as independent Gaussian noises having the same variance throughout the target ion masses (Note: Several ion masses for the background air such as \( m/z=29, 30, 32 \) and \( 37 \) had larger variances than others, so those were excluded from the target ion masses). Now, the measured emission signature of a material mixture by PTR-MS can be modeled as follows:

\[
S_{sp}(x_n) = \sum_i [\alpha_i \cdot S_i(x_n) + w_i(x_n)] \quad \text{for} \quad \forall x_n, \text{with} \quad \alpha_i > 0 \quad \forall i, n=1,2,\ldots,N \tag{1}
\]

where \( sp \) represents 'the measured sample', \( i \) indicates the material ID of the correct material set for the mixture, defined in the database of material emission signatures, \( S_{sp}(x_n) \) is the PTR-MS signal intensity (not normalized) of the measured sample for a given ion mass \( x_n \), \( S_i(x_n) \) is the normalized magnitude of emission signature for a given ion mass \( x_n \) of material \( i \) (Note: An emission signature for each studied material is recorded in the database in a normalized form having a peak magnitude of 100 as its maximum. For details, refer to Han et al., 2010), \( w_i(x_n) \) is the independent measurement noise for the ion mass \( x_n \) contained in the signal for material \( i \), and \( \alpha_i \) is the signal intensity multiplication factor of the emission signature for material \( i \).

The measured sample signature can also be expressed in a vector form as follows:

\[
\bar{S}_{sp}(x) = \sum_i [\alpha_i \cdot \bar{S}_i(x) + \bar{w}_i(x)] \tag{2}
\]

where \( \bar{S}_{sp}(x) \) represents the sample signal intensity vector measured for all scanned ion masses \( x \) defined as \( \bar{S}_{sp}(x) = [S_{sp}(x_1), \ldots, S_{sp}(x_n), \ldots, S_{sp}(x_N)]^T \), \( \bar{S}_i(x) \) is the normalized magnitude vector of the emission signature for material \( i \) defined as \( \bar{S}_i(x) = [S_i(x_1), \ldots, S_i(x_n), \ldots, S_i(x_N)]^T \), \( \alpha_i \) is given in Equation (1), \( \bar{w}_i(x) \) is the corresponding signal noise vector for material \( i \) defined as \( \bar{w}_i(x) = [w_i(x_1), \ldots, w_i(x_n), \ldots, w_i(x_N)]^T \), and \( x_n \) is the \( n \)th ion mass in the scanned range (Note: If there is no corresponding peak for the \( n \)th ion mass recorded in the emission signature of material \( i \), then \( S_i(x_n) = 0 \) in this representation).

To perform a signature separation and identification when \( \bar{S}_{sp}(x) \) is the measurement from PTR-MS, let us define \( Ind(\bar{S}(x)) \), where \( Ind(.) \) is the set of ion mass indices corresponding to a given emission signature intensity vector \( \bar{S}(x) \), with positive (or non-zero) signal intensities. Then, in the first step, the set of possible material candidates \( \{db\} \) in the database can be searched by comparing the ion mass components (related to VOC components in a physical sense) of each material emission signature \( Ind(\bar{S}_i(x)) \) with the measured ones \( Ind(\bar{S}_{sp}(x)) \) under the epsilon condition (\( \varepsilon \)) to be defined below. While performing the comparison between the measurement and the database signatures, if all ion mass components of a signature are found in the measurement, then that material \( i \) related to the signature will be selected as a possible material candidate (i.e. \( i \in \{db\} \)), having all VOC components to be found. However, due to noise, some ion masses that should be found in the measurement might be missed or measured falsely as any near ion masses. For example, the emission
signature for Wood (Material ID=9) has ten ion mass indices including \(m/z=33, 43, 44, 45, 47, 59, 60, 61, 62\) and 75, so \(\text{Ind}(\tilde{S}_s(x)) = \{33, 43, 44, 45, 47, 59, 60, 61, 62, 75\}\). If the component indices of the measurement from PTR-MS are given as \(\text{Ind}(\tilde{S}_{sp}(x)) = \{33, 43, 44, 46, 47, 59, 60, 61, 62, 69, 76, 83\}\), in order for Wood or \(i=9\) to be selected as a possible material candidate, the mismatched two ion masses (\(m/z=45\) and 75) should be checked whether they are present in the actual emissions, but measured wrongly (as \(m/z=46\) and 76) due to noise. For this case, the epsilon condition, which considers the tolerable square error level along the ion mass axis, was considered. We assume that the noise occurring along ion masses can be modeled as an independent Gaussian noise having much lower variance than that of signal intensity for each ion mass. The missing of any correct materials due to this noise along ion masses can now be considered in the Chi-square distribution by setting a threshold value enabling the detection of the material at a 95% confidence level. This threshold value is a one-sided detection limit and can be determined based on actual PTR-MS measurements. These considerations can be summarized as Step 1 such that:

**Step 1.** If \(\text{Ind}(\tilde{S}_{sp}(x)) \supset \text{Ind}(\tilde{S}_{db}(x))\) where \(\varepsilon = \sum_{n-no\text{ matched}} \left( x_{n-sp}^{(near)} - x_{n-db}^{(near)} \right)^2 \) and \(\varepsilon \leq \text{Thresh} \) (95% confidence level), then find \(\{db\}\).

Here, \(x_{n-sp}^{(near)}\), which is an element of \(\{x_n\}\) of the measured sample, denotes the nearest ion mass component in the sample emission signature with a positive signal intensity to an ion mass \(x_{n-db}^{(near)}\) for which there was no match in the attempted emission signature in the database because of noise. For example, \(\varepsilon = (46-45)^2 + (76-75)^2 = 2\). If the calculated threshold value for detection (in this case, for two degrees of freedom) is larger than 4, Wood or \(i=9\) can be selected as a possible material candidate. But, if the detection threshold is less than 2, Wood or \(i=9\) will not be included in the candidate set.

Assuming \(w_i(x_n)\) is an i.i.d. Gaussian noise (Note: i.i.d. stands for ‘independent and identically distributed’), the signal intensity factor of each material candidate \(\alpha_i\), where \(i \in \{db\}\), can be determined via the following linear regression approach such that:

**Step 2.** Find an optimal value of each \(\alpha_i\) for \(i \in \{db\}\), which minimizes the following performance index \(J\) (i.e. in the sense of MRLS, multiple regression least squares. For detailed descriptions, see Cohen et al., 2003).

\[
J = \min_{\{\alpha_i\}} \sum_{n-all\ scanned} \left( S_{sp}(x_n) - \sum_{i\in\{db\}} \alpha_i \cdot S_i(x_n) \right)^2
\]

Here, it should be noted that \(\alpha_i\)s are unknown control variables to be estimated in the sense of MRLS while \(S_{sp}(x_n)\) is measured by PTR-MS and \(S_i(x_n)\) is given in the database for all scanned ion masses, where \(n=1, 2, \ldots, N\).

The first step of the algorithm is the scanning of the ion mass components in the measured sample emission signature by the comparison with those in the database. Because of noise, several components can have small deviations from the exact values of ion masses. So, the algorithm tries to find any matching emission signatures in the database (a set of materials denoted as \(\{db\}\)), having the same ion mass components under the \(\varepsilon\) condition described above. Next, by using only the selected set of material emission signatures \(\{db\}\), the algorithm attempts to find the optimal set of signal intensity factors \(\alpha = [\alpha_{ID_1}, \ldots, \alpha_{ID_j}, \ldots, \alpha_{db}]\).
where ID\(_j\) is the \(j^{th}\) material ID in the set of \(\{db\}\) selected as possible material candidates. If the final value of \(\alpha_i\) is less than a specified small threshold value (e.g. 10% of the smallest signal intensity factor in the database, 0.007), then material \(i\) will be excluded from the final set of the material candidates \(\{ID\}\).

2.2.2. Algorithm 2

Let us set up the problem as \(w_i(x_n)\) being a function of the spectrum of \(x_n\), where \(n=1, 2, \ldots, N\). In Algorithm 1, the error value for a relatively small signal peak contributes less (although it might be an important peak for detection and identification) to the overall performance index than larger peaks. So, a normalization technique for adjusting the weight for each term might be useful to reflect in the overall error terms the contributions of the errors for any small peaks in the proportion comparable with those of larger peaks. The other definitions and procedure are the same as those of Algorithm 1 except for the definition of the performance index \(J\) as follows:

\[
J = \min_{\{\alpha_i\}} \sum_{n=\text{all scanned}} \left( \frac{S_{xp}(x_n) - \sum_{i \in \{db\}} \alpha_i \cdot S_i(x_n)}{S_{xp}(x_n)} \right)^2
\]

(4)

2.3. Environmental chamber setup and conditions

For both single material and mixture tests, a 50-l small-scale environmental chamber (0.5 m × 0.4 m × 0.25 m high) made of electro-polished stainless steel (Figure 1) was operated with a precise airflow controller (Alicat Scientific, accuracy ±0.1% of measured values) and a humidity controller (Vaisala INTERCAP HMP50, accuracy ±1% of measured values), using external supply air passing through a dedicated filtering assembly (Wilkerson 3-stage carbon filters with micro filtration). The chamber was maintained at a constant stable temperature in the range of 23.5-25.4°C (with a small variation of < ±0.02°C during each sampling period) and at a stable relative humidity of 31±0.1% RH during the tests. The background concentrations of individual VOCs in the empty chamber were maintained clean to be less than 1 µg/m³.

2.4. Test specimens

Nine typical building materials were used including ceiling, wood, carpet, linoleum, PVC, polyolefine, gypsum, paint 1 (water-based acrylic) and paint 2 (with linseed oil) applied on gypsum board. The detailed descriptions of the materials tested can be found elsewhere (Han et al., 2010). Specimens were cut and prepared according to the sizes specified in Table 1. The specimens except for Gypsum, Paint 1 and 2 were stapled together back to back in order that only the material upper surface was exposed to the air in the test chamber. A VOC-free aluminum tape (3M 2113) was applied to seal all edges. The prepared specimens were placed vertically, parallel with the airflow in the chamber (the same direction of the long side of the chamber). The range of area-specific ventilation rates defined as the ratio of ventilation rate to emitting surface area was set by adjusting the airflow to the chamber while keeping the size of specimen unchanged.

Table 1
2.5. PTR-MS setting
A PTR-MS device (Ionicon Analytik high-sensitivity model with a detection limit as low as 1 pptv, Austria) was operated at the standard conditions (Drift tube pressure: 2.3 ~ 2.4 mbar, PC: 455 mbar, FC: 6.5 STP cc/min, U SO: 75 V, U S: 100 V, Drift tube voltage: 600 V and Source: 6.0 mA). Detailed descriptions of the device, its principle and applicability can be found elsewhere (Blake et al., 2009; de Gouw et al., 2003; Han et al., 2010; Hewitt et al., 2003; Lindinger et al., 1998, 2001; Lirk et al., 2004; Steeghs et al., 2004). The signal intensity of VOC emissions used in the present study was measured by the instrument in the unit of ion count rates (counts per second, cps), which were then normalized by per million hydronium ion (H$_3$O$^+$) count rates to compensate the variations in the hydronium ions as other related works in this area (e.g. de Gouw and Warneke, 2007; Jobson et al., 2005; Whyte et al., 2007). This normalized product ion count rate (ncps) becomes directly proportional to the concentration level of a target VOC.

2.6. Test procedure
The nine building materials were previously studied at three different area-specific ventilation rates (Measurements #1, #2 and #3) as shown in Table 1 to establish a database of emission signatures by PTR-MS specific to each individual material. Five multi-material mixture tests were conducted (Measurements #4-#8) in the present study to obtain combined emission signatures for the studied mixtures and to validate the feasibility of the proposed source identification methods. The mixture tests were performed within two weeks after the single material measurements were finished. For each measurement, the mass spectra for the background emission signal from the empty chamber and for the sample emission signal with each prepared specimen inside the chamber were measured all after three volumetric air changes from the start of ventilation to allow concentrations in the chamber to reach over 95% of the quasi-steady state level. PTR-MS was set to scan from m/z = 21 to m/z = 250 once every 12 s with an ion mass resolution interval of 50 ms. The total sampling period was 10 min (600 s) with 50 ion mass spectra collected for each dataset. During each measurement, another set of duplicate mass spectra was scanned to verify the collected data.

3. Results
3.1. Relative signal intensity (α) and variance (ε) of material emission signatures
The previous study (Han et al., 2010) determined the identification of each ion mass with a related VOC by using GC/MS and PTR-MS together. Now, each ion mass (m/z) of PTR-MS
represents a VOC, and its signal intensity (ncps) corresponds to the concentration level of VOC.

To perform a signal separation simulation using the Monte-Carlo method, some reasonable ranges of signal intensities (mean) and noise levels (variance) for material emission signatures are needed. To conduct a measurement-based simulation, the experimental data collected using PTR-MS through Measurements #1, #2 and #3 were used to get the range information. The pattern of material emission signatures determined by PTR-MS with a special filtering method employed were found to be consistent and stable in a normalized mass spectrum domain even under different area-specific ventilation rates tested and also over different sampling time (Han et al., 2010). There were also identifiable differences in the VOC emission signatures measured by PTR-MS among different types of building materials. The differences include the type of ion mass present, its relative amount (signal intensity) or both. This study used the emission signatures collected at Measurement #2 (medium airflow rate) as a reference, and the variance of signal intensity for each material was determined based on the variations of all collected data from the corresponding reference signature. Figure 2 summarizes the results.

3.2. Separation simulations of the algorithms for various material mixtures

To compare the performances of the suggested algorithms, Monte-Carlo simulations were conducted by setting the design parameters based on PTR-MS measurements (for relative signal intensities and noise levels). The Monte-Carlo method is a class of computational algorithms that rely on repeated random sampling to compute the performance results for a given method when simulating physical and mathematical systems, which tend to be unfeasible to compute an exact result with a deterministic algorithm. The noise component, \( \sigma \), was modeled as white Gaussian random noise, and its variance was determined by the corresponding percentage of noise level (\( \varepsilon \)) which represents the standard deviation of the
noise with respect to the maximum peak value of the corresponding reference material emission signature. The number of random samples, $N$, was determined to make the variance of the final simulation results less than 1% (e.g. $\sigma$ of Success Rate < 1% out of 100%) with a computational time as small as possible. With these considerations, $N=1000$ was selected in this study. For the performance comparison of the algorithms, three indices - error norm expectation for the estimation performance of emission strengths, success rate for the detection performance of source material IDs, and success score expectation as an auxiliary index for success rate, were introduced as follows:

$$ Err \equiv E \left\| \overline{\alpha}_{est} - \overline{\alpha}_{true} \right\|_2, \quad Score \equiv E \{scr(N)\}, \quad SR \equiv \left( N - N_{fail} \right) \times 100 / N $$

where $Err$ is the expected value of the 2-norm of the difference between the estimated value of signal intensity factor vector and the ground truth value, $N_{fail}$ indicates the number of material identification failures that occurred during the simulation (If every material in the sample is correctly identified with several wrong candidates, this case is classified as success. However, if there is any missing material identified in the final set of the material candidates, then it is declared as failure. For example, if [5 6 8], the ID result is [1 3 5 6 8], this is considered as success. But, if the result is [3 5 6 7], this is failure), $Score$ is the expected value of success score ($scr$) which is defined as follows: for each simulation, if the ID result is success with exact identification, then $scr = 100$; if success with $n$ wrong candidates, then $scr = 100 - 10 \times n$; and if failure, $scr = 0$, and SR represents the success rate in material identification defined in percentage.

Note: To differentiate the ID performance result with less wrong candidates from that with more wrong candidates (e.g. For [5 6 8], [1 5 6 8] ($scr = 90$) vs. [1 3 4 5 6 8] ($scr = 70$)), $Score$ index was introduced.

Table 2

<table>
<thead>
<tr>
<th>Algorithm 1</th>
<th>Algorithm 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>0.02</td>
</tr>
<tr>
<td>Success Rate</td>
<td>99.9%</td>
</tr>
<tr>
<td>Success Score</td>
<td>95</td>
</tr>
</tbody>
</table>

The results in Table 2 indicate five noticeable aspects concerning the performance of the two algorithms. The algorithm performance for signature separation and identification may rely mainly on key factors such as the number of materials in a mixture, the signal intensity ratio and the variance of each composing signature. Algorithm 1 always showed by far the better performance than Algorithm 2 in the present simulations in terms of error expectation ($Err$), which implies that Algorithm 2 may get affected more susceptibly to the variation of emission signature because of noise than Algorithm 1. For some conditions, Algorithm 2 could show a better performance than Algorithm 1 in the sense of success rate and score, which suggests
that the performance of the two algorithm can vary depending on given environmental and material conditions affecting the performance difference. A mixture case having no possible false ID can show a better performance in terms of success rate and score than that with several possible false IDs potentially under similar conditions (e.g. [2 8 9] vs. [3 6 8] or [1 7 9]). The material number in a mixture seemed to strongly affect the performance difference because of complexity increase (e.g. the low success rate of the 9-material mixture case), but when focusing only on several major materials having higher signal intensities, the two algorithms suggested in this study still showed reasonably high success rates (>95%).

3.3. Monte-Carlo simulations for the nine-material mixture

The results in Table 2 suggest that the two proposed algorithms in this study exhibited a different quality of performance depending on given conditions of the two key factors – the maximum ratio of signal intensity factors (\(\gamma\)) and the variance (\(\varepsilon\)). To examine this aspect of the two algorithms as the two key parameters change, the most complicated case in this study, 9-material mixture, was selected, and the success rates of the two for this case were explored at various combinations of \(\gamma\) and \(\varepsilon\), varying within their corresponding realistic ranges (i.e. \(1.0 \leq \gamma \leq 100, \ 0 \leq \varepsilon \leq 5.0\%\)). Figure 3 shows the results.

As expected, the trends in Figure 3a and 3b represent the decaying performance of both algorithms as the maximum ratio and the variance increase, although the slope of the decreasing performance of each algorithm is slightly different (i.e. The performance of Algorithm 1 in terms of success rate decays much faster in an exponential manner than that of Algorithm 2). Both algorithms exhibited high success rates within narrow decent ranges of the maximum ratio (\(1.0 \leq \gamma \leq 20\)) and the variance (\(0 \leq \varepsilon \leq 1.0\%\)) even for this complicated mixture. Within these decent ranges, Algorithm 1 showed a better performance than Algorithm 2 in terms of success rate. However, out of these ranges (more general ranges), the latter seemed better than the former in the sense of success rate (Figure 3c).
3.4. Experimental results of multi-material mixtures

- Case 1: [7 8] = Carpet + Linoleum (Measurement #4)

The algorithms validated by measurement-based Monte-Carlo simulations were applied to actual emission measurements obtained from five multi-material mixtures. For comparison, the ground truths for the correct material IDs and the corresponding emission levels (represented by signal intensities) were known and obtained by the optimal separation of each measured signature in terms of MRLS (The exemplary separation profile of the signature for Case 1 is shown in Figure 4). Case 1 was comprised of Carpet (Material ID=7) and Linoleum (ID=8). Both algorithms identified the correct sources of material emissions with some false materials: for Algorithm 1, Gypsum (ID=4), so success score (scr)=90; for Algorithm 2, Gypsum (ID=4) and Wood (ID=9), so success score (scr)=80. In addition, the corresponding emission levels from the two materials could be estimated by both algorithms with reasonably small errors (Error expectations, or Err, were <1.0 for both cases). The performance results of the two algorithms for the actual emission measurement from Carpet and Linoleum could be summarized as below in terms of the performance indices defined in the previous section:

Algorithm 1: \( \mathbf{ID}_{est1} = [4 \ 7 \ 8] \), \( \tilde{\mathbf{\alpha}}_{est1} = [0.68 \ 0.50 \ 13.44] \), \( Err1 = 0.81 \), \( scr1 = 90 \)

Algorithm 2: \( \mathbf{ID}_{est2} = [4 \ 7 \ 8 \ 9] \), \( \tilde{\mathbf{\alpha}}_{est2} = [0.28 \ 0.53 \ 13.63 \ 0.02] \), \( Err2 = 0.39 \), \( scr2 = 80 \).

- Case 2: [8 9] = Linoleum + Wood (Measurement #5)

The optimal signal intensity factors for this case were \( \tilde{\mathbf{\alpha}}_{true} \) for [8 9] = [18.42 8.23] in the MRLS sense. The following are the performance results:

Algorithm 1: \( \mathbf{ID}_{est1} = [8 \ 9] \), \( \tilde{\mathbf{\alpha}}_{est1} = [19.20 \ 9.63] \), \( Err1 = 1.60 \), \( scr1 = 100 \)

Algorithm 2: \( \mathbf{ID}_{est2} = [8 \ 9] \), \( \tilde{\mathbf{\alpha}}_{est2} = [20.47 \ 4.82] \), \( Err2 = 3.98 \), \( scr2 = 100 \).
• Case 3: [6 8 9] = Paint 2 + Linoleum + Wood (Measurement #6)

The optimal signal intensity factors for this case were \( \tilde{\alpha}_{\text{true}} \) for [6 8 9] = [6.09 5.96 4.41] by MRLS. The following are the performance results:

Algorithm 1: \( \text{ID}_{\text{est}_1} = [6 8 9] \), \( \tilde{\alpha}_{\text{est}_1} = [1.71 6.51 5.78] \), \( \text{Err}_1 = 4.62 \), \( \text{scr}_1 = 100 \)

Algorithm 2: \( \text{ID}_{\text{est}_2} = [6 8 9] \), \( \tilde{\alpha}_{\text{est}_2} = [3.63 6.76 3.75] \), \( \text{Err}_2 = 2.67 \), \( \text{scr}_2 = 100 \).

• Case 4: [3 6 7 8 9] = Ceiling + Paint 2 + Carpet + Linoleum + Wood (Measurement #7)

The optimal signal intensity factors were \( \tilde{\alpha}_{\text{true}} \) for [3 6 7 8 9] = [3.11 1.28 0.00 4.44 3.42] by MRLS (The detailed signature profile is given in Figure 5). The following are the performance results:

Algorithm 1: \( \text{ID}_{\text{est}_1} = [1 8 9] \), \( \tilde{\alpha}_{\text{est}_1} = [0.16 4.28 3.58] \), \( \text{Err}_1 = 0.28 \)

Algorithm 2: \( \text{ID}_{\text{est}_2} = [1 3 6 8 9] \), \( \tilde{\alpha}_{\text{est}_2} = [0.38 2.33 1.84 4.87 1.51] \), \( \text{Err}_2 = 2.21 \).

• Case 5: [2 3 5 6 7 8 9] = PVC + CEI + PT1 + PT2 + CAR + LIN + WOD (Measurement #8)

The optimal signal intensity factors for this case were \( \tilde{\alpha}_{\text{true}} \) for [2 3 5 6 7 8 9] = [0.00 1.05 0.00 0.19 1.38 2.30] in terms of MRLS. The following are the performance results:

Algorithm 1: \( \text{ID}_{\text{est}_1} = [6 7 8 9] \), \( \tilde{\alpha}_{\text{est}_1} = [0.40 0.32 1.26 2.32] \), \( \text{Err}_1 = 0.44 \)

Algorithm 2: \( \text{ID}_{\text{est}_2} = [1 3 6 7 8 9] \), \( \tilde{\alpha}_{\text{est}_2} = [0.02 1.04 2.27 0.04 1.50 1.63] \), \( \text{Err}_2 = 2.37 \).

4. Discussion

The experimental results imply three important aspects concerning the proposed technique, which may provide a new insight on adsorption effects occurring under material mixture conditions and open a new gate to quantitatively analyze and assess this adsorptive phenomenon inherent in material emissions.

First, although there were some interactions among VOC emissions from a material mixture (which was observed in the measured signatures of the tested material mixtures having several different ion masses other than the ones found in the emissions from the original individual materials, with very small signal intensities), the effect appeared to be small enough to allow
the algorithms developed for separation and identification on the basis of the superposition assumption of the co-located materials’ emission signatures. In addition, the emission signatures of the mixtures maintained the individual emission signature pattern of each material almost unchanged even in a mixture. For example, note the similar pattern of the measured signature (Refer to the enlarged figure shown in Figure 6b) of a material mixture (Case 1 - Carpet and Linoleum) to the superposed signature of the two individual materials (Figure 6a), considering the given area-specific ventilation rates in Table 1. Generally, it was observed that the similar patterns of the measured signatures to the superposed ones were maintained for all mixture cases tested, but the measured signal intensities were reduced significantly (meaning a lowered level of VOC concentrations emitted from material mixtures) compared with the superposed ones, mainly because of sorption effect among materials. Another possible reason for this reduction of signal intensity may be the effect of VOC emission decay over time, but this might be small or negligible considering that the mixture tests were performed right after the measurements of the single materials. It is hypothesized that the compounds from a main emitting material were adsorbed on the surfaces of adsorptive materials, and in a quasi-equilibrium emission state, the adsorbed compounds may re-emit from the adsorptive surfaces with reduced emission intensities. In this process, the individual shapes of emission signatures were maintained and could be generated by the superposition of the initial library of material emission signatures established. If the major compounds having high concentration levels are heavy, this superposition may not be valid. However, the major compounds for the materials studied were mostly light (< m/z=150). Because of these special phenomena, the algorithms could be applied to the actual combined emissions of the studied material mixtures and yield satisfactory performances.

The second aspect examined in this study is that the sorption effect among materials can be quantitatively assessed in an accurate manner for each mixture by using this technique (Refer to another previous study on the assessment of sorption impact among material mixtures on
perceived indoor air quality by human assessments, which is done by Sakr et al., 2006). The approach derived from the present study is based on the reduced percentage of the estimated signal intensity of emission signature from a material mixture when compared with that of the stand-alone emission signature for each individual material obtained under the same emission conditions (RH, temperature and area-specific ventilation rate). For example, the mixture of Carpet and Linoleum (Case 1) should have the relative signal intensity factors as $\alpha_7 = 0.53$ for Carpet and $\alpha_8 = 29.74$ for Linoleum under the given mixture emission conditions in Table 1 (Measurement #4) if there are no interaction and no sorption effect between the two. However, the experimental results showed that the intensity factors were different as $\alpha_7 = 0.72$ for Carpet and $\alpha_8 = 13.82$ for Linoleum (Figure 4). Because of large porous areas in the surface of Carpet, the major portion of the VOC emissions from Linoleum might be trapped (i.e. adsorbed) in those porous areas of Carpet because of adsorption, and the reduction effect of VOC emissions from Linoleum could be estimated at the percentage of 53.5% (from 29.74 to 13.82). This sorption effect can be enhanced as sorptive materials such as Carpet (material ID = [7]) and Gypsum ([4]) or gypsum-based materials such as Ceiling ([3]), Paint 1 ([5]) and Paint 2 ([6]) are added more in the studied material mixture. This enhanced effect could be seen in the consecutive study of material mixtures as shown in Figure 7. The initial signature for Linoleum (the main emitting material in this case) could be measured again with a similar level of signal intensities after other added materials were taken out of the chamber as shown in Figure 7h.

Lastly, by using this technique, emission sources and their corresponding signal intensities (i.e. concentration levels of VOC emissions) can be identified and estimated in a laboratory condition. PTR-MS emission signatures were specific to each material tested, and the interactions among different material emissions may not significantly affect the pattern change of the individual emission signatures at least for the studied material mixtures. Therefore, it is possible to identify the related material based on the emission pattern of the
same material measured by PTR-MS at least in a conditioned laboratory environment. The differences in emission patterns may potentially be used to identify emission sources (e.g. in buildings) based on air samples measured by PTR-MS. However, emission signatures may change over the course of long-term emissions.

If there is a VOC related problem in an indoor air environment, different relevant sources can be identified and screened individually by PTR-MS. Finding the source(s) would help eliminating the problem efficiently and effectively. An advantage of this technique is that it may find the sources invisible or hidden when a building with problems of indoor air quality is suspected. Further studies are needed for extending this technique into a practical tool for emission source identification in real indoor environments because of the higher number of possible emission sources, the more complicated adsorption and desorption effects and the change of emission pattern over time for a given source. The pattern change of material emission signatures over a long-term period may be accounted for by using appropriate emission source models.

5. Conclusions
The results of this study demonstrate the feasibility of identifying emission sources with high success rates when multiple materials are present indoors by utilizing the PTR-MS and an effective signal processing method under laboratory indoor conditions. The following conclusions can be derived:

• In a controlled environment, the identification of indoor emission sources was found feasible with the estimations of their individual relative source strengths when the developed technique was utilized with a limited number of materials for composing a mixture.

• The effect of VOC mixture emissions might be superposed in the mass spectrum domain of PTR-MS because of small interactions among material emissions and the conservation of individual material emission pattern even in the presence of emission interactions such as sorption.

• The sorption effect among material emissions could be quantitatively assessed using the new technique proposed.

The source identification methodology presented in this study should consider the possible change of emission signatures over a long period of time as a main possible contributor of the change in a field condition. However, there are other possible causes affecting the change of emission signatures such as temperature, relative humidity, surface velocity over materials, large portion of chemical reactions among material emissions, etc. The source identification technique may break down in a mixture condition having a certain large number of materials (e.g. 30 materials) or in a highly reactive environment like an ozone-initiated chemical-reaction dominant space. In addition, for some materials, any stable emission signature may not be established somehow. If some materials have heavy VOCs as major compounds for their emission signatures, the measured signature from their mixtures may be distorted from the superposed emission signature mainly due to different adsorption phenomena, leading to a breakdown of the method. For this reason, a broader range and different types of materials should be investigated before this new approach can be widely applied.

Acknowledgement
We thank the Syracuse Center of Excellence in Environmental and Energy Systems (SyracuseCoE) for funding the present study.
References


