Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Rapid identification of multiple constantly-released contaminant sources in indoor environments with unknown release time

Hao Cai a, Xianting Li b,*, Zhilong Chen a, Mingyang Wang a

a State Key Laboratory of Explosion & Impact and Disaster Prevention & Mitigation, PLA University of Science and Technology, Nanjing, 210007, PR China
b Department of Building Science, School of Architecture, Tsinghua University, Beijing, 100084, PR China

ARTICLE INFO

Article history:
Received 16 March 2014
Received in revised form 8 June 2014
Accepted 9 June 2014
Available online 18 June 2014

Keywords:
Source identification
Contaminant source
Indoor environment
Computational fluid dynamics
Sensor
Air distribution

ABSTRACT

The sudden release of airborne hazardous contaminants in an indoor environment can potentially lead to severe disasters, such as the spread of toxic gases, fire, explosion, epidemic, or radiation. To prevent and mitigate these disasters it is critical to rapidly and accurately identify the characteristics of the contaminant sources. Although remarkable achievements have been made in identifying a single indoor contaminant source in recent years, the issues related to multiple contaminant sources are still challenging. This study presents a method for identifying the exact locations, emission rates, and release time of multiple indoor contaminant sources simultaneously released at constant rates, by considering sensor thresholds and measurement errors. The method uses a two-stage procedure for rapid source identification. Before the release of contaminants, only a limited number of time-consuming computational fluid dynamics (CFD) simulations need to be conducted. After the release of contaminants, the method can be executed in real-time. Through case studies in a three-dimensional office, the method was numerically demonstrated and validated, and the results show that the method is effective and feasible. The effects of sensor threshold, measurement error and total sampling time on the source identification performance were analysed, and the limitations and applicability of the method were also discussed.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Indoor environments in modern society are persistently threatened by a huge variety of airborne hazardous contaminants, such as toxic, flammable, or explosive chemicals, harmful microbes, and radioactive substances [1–4]. The sudden release of these contaminants, accidentally or intentionally, can potentially result in exposure to toxic gases, fire, explosion, epidemic, or radiation. All these incidents would lead to significant casualties and property losses if no effective response measures were taken. In recent years, several tragic incidents, such as the Tokyo subway sarin attack, severe acute respiratory syndrome outbreak and the Fukushima nuclear leak, have aroused the concern of indoor environmental safety. To prevent and mitigate disasters due to the sudden release of hazardous contaminants, it is critical to promptly and accurately identify the characteristics of the contaminant sources, such as locations, emission rates, and release time. Source identification is an important premise for developing effective response measures, such as source elimination, dilution ventilation, air purification, evacuation, and shelter-in-place [5,6].

With known locations and emission rates of contaminant sources, the transient distribution of a contaminant can be well predicted using a dispersion model. However, in most practical applications the information on contaminant sources is incomplete or unknown, and needs to be identified using the sensor measurements of the airflow and concentration fields. This type of identification constitutes an inverse problem, which has been widely studied for several decades in the fields of heat transfer [7–9], groundwater contamination [10,11], soil pollution [12,13], and atmospheric constituent transport [14–16].

Due to increasing concern about indoor environmental safety, research on identifying indoor contaminant sources has developed rapidly in recent years. The methods available for identifying indoor contaminant sources can be roughly categorised as backward and forward methods, as shown in Table 1. To date, most backward and forward methods have shown their applicability in identifying a single indoor contaminant source [17–40]. In contrast, very few studies have discussed the scenarios involving multiple indoor contaminant sources, which are frequent and common in real-world applications [18,38].
residential or working environments, chemical contaminants are ubiquitous and continuously emitted from building materials, furniture, and equipment. In emergency situations related to earthquakes, blasts or impacts, hazardous chemicals may simultaneously escape from many leakage points. During an epidemic outbreak, many infected people may be found in hospital or other buildings. In terrorist attacks, chemical or biological agents may be deliberately released from many locations. Therefore, research on identifying multiple indoor contaminant sources is of great practical significance.

For identifying multiple indoor contaminant sources, a major challenge to the backward methods is how to ensure the uniqueness of the identification results. When multiple contaminant sources are released, different combinations of sources may result in the same or close sensor measurements at a given time point. In other words, different causes may lead to the same outcome. Because the backward methods use the same outcome (sensor measurements at a given time point) as the initial condition, it is difficult to differentiate various possible causes. Besides, the backward methods would require a significant reduction in computing time for solving inverse computational fluid dynamics (CFD) models. This potentially limits their application in real-time source identification.

As shown in Table 1, forward methods use sensor measurements during a given time period, rather than at a single moment, as inputs. In indoor environments, it is quite rare that different release scenarios will produce the same or close sensor measurements at all time points. Therefore forward methods are more promising for finding a unique solution. In addition, the two-stage procedure of the forward methods (see Table 1) also makes real-time source identification possible. A major constraint of the forward methods is that most of these methods only work well using a fast contaminant transport model, such as a multi-zone model, because a large number of simulations are usually required before a release incident to cover all possible scenarios. The multi-zone model generally assumes a uniform contaminant distribution in one zone [41,42]. Therefore the forward methods using the multi-zone model cannot identify the exact locations of contaminant sources in each zone or room.

In a previous study we presented a theoretical model for quickly identifying the exact locations and emission rates of multiple constant contaminant sources indoors, using ideal sensors [38]. By using an analytical expression of indoor contaminant dispersion, the theoretical model only requires a limited number of CFD simulations. Because CFD simulations can provide detailed information of contaminant dispersion, the exact locations of contaminant sources can be identified by the theoretical model. The major constraint of this model is that it cannot work by using real sensors, mainly because the model cannot work when there is no response of sensors due to sensor thresholds. In real-world situations, the use of real sensors will make the problem of source identification much more difficult, primarily due to three issues: (1) no response of sensors due to sensor thresholds; (2) inevitable measurement errors; and (3) unknown release time of contaminant. This research aims to fill the gap by developing a more sophisticated method that can promptly identify the locations, emission rates, and release time of multiple indoor contaminant sources simultaneously released at constant rates, by considering the sensor threshold and measurement errors.

### 2. Major assumptions

The source identification problem is specified by introducing the following assumptions.

**Assumption 1.** The contaminant is dispersed as a passive gas in steady indoor airflow. The passive gas refers to the gaseous contaminant that is present in sufficiently low concentration that the effects of its dispersion on the indoor airflow field and air density can be neglected.

**Assumption 2.** The number of potential sources is limited and their locations are known. The current work does not attempt to
handle the situations where the number of the potential sources is unlimited and their locations are completely unknown.

**Assumption 3.** The emission rates of contaminant sources are constant. In practice, a contaminant may be released continuously or instantaneously. The current study only aims to identify indoor contaminant sources that are continuously released at constant rates.

**Assumption 4.** All contaminant sources are released simultaneously. That is, the current work only deals with multiple contaminant sources in synchronous release.

3. Source identification method

3.1. Overall framework

The overall framework of the identification method is depicted in Fig. 1. The method is composed of three essential components. Component I is an analytical expression of transient contaminant dispersion, which can predict the temporal-spatial distribution of a contaminant in real-time with known source information, such as source locations, emission rates, and release time. Component II is a source identification model that can identify the locations and emission rates of multiple contaminant sources with a known release time. Component III is an algorithm for identifying the release time of contaminant sources.

3.2. Major components

3.2.1. Component I: analytical expression of transient contaminant dispersion

In our previous work, an analytical expression of transient contaminant dispersion was presented to predict the transient distribution of contaminants in real-time [43]. For the dispersion of a passive gas in a steady-state airflow field, the analytical expression can be expressed as:

\[
C_p(t) = C_0 + \sum_{k=1}^{K} \left\{ (C_{S,k} - C_0) a_{S,k,p}(t) \right\} + \sum_{i=1}^{N} \left\{ \frac{S_i}{Q} a_{C,i,p}(t) \right\}
\]  

(1)

where \(C_p(t)\) is the contaminant concentration (mg/m³) at an arbitrary indoor point \(p\) and moment \(t\), \(C_{S,k}\) is the contaminant concentration (mg/m³) at the \(k\)th inlet, \(C_0\) is the initial contaminant concentration (mg/m³) indoors, \(S_i\) is the emission rate (mg/s) of the \(i\)th contaminant source, \(Q\) is the airflow rate (m³/s), \(K\) is the number of indoor supply air inlets, \(N\) is the number of potential indoor contaminant sources, \(a_{S,k,p}(t)\) is the transient accessibility of supply air (TASA) from the \(k\)th inlet to \(p\) at moment \(t\), and \(a_{C,i,p}\) is the transient accessibility of contaminant source (TACS) from the \(i\)th source to \(p\) at \(t\).

The TASA and TACS indices are dimensionless numbers that quantify the effects of the supply air and contaminant source on an indoor position at different moments, respectively. TASA is a function of flow characteristics, regardless of contaminant type and source. In a ventilated space, assume that only air from the \(k\)th inlet contains a hypothetical contaminant, and there is no source of the contaminant in the space. The TASA from the \(k\)th inlet to point \(p\) at moment \(t\) was defined as [43]:

\[
a_{S,k,p}(t) = \frac{C_p(t)}{C_{S,k}}.
\]  

(2)

TACS is a function of both the flow characteristics and source location, regardless of the emission rate and contaminant type. Assume that the initial concentration was zero, all inlet concentrations were zero, and only the \(i\)th contaminant source existed. The TACS from the \(i\)th source to \(p\) at \(t\) was defined as [43]:

\[
\]
\[ a_{Cl,p}(t) = C_{p}(t)/C_{e,i} \]

where \( C_{e,i} \) is the average exhausted concentration (mg/m\(^3\)) under steady-state conditions.

Using CFD simulations, the temporal-spatial distributions of TASA and TACS from each supply air inlet and contaminant source can be obtained in advance. With the TASA and TACS data, transient contaminant distributions under different combinations of supply air concentrations and source emission rates can be calculated in real-time through the simple algebraic calculations in Eq. (1). This feature of the analytical expression enables the real-time simulation of contaminant dispersion.

### 3.2.2. Component II: identification model with known release time

When an initial concentration and all supply inlet concentrations are zero, Eq. (1) can be reduced to:

\[ C_p(t) = \frac{1}{Q} \sum_{i=1}^{N} S_i a_{Cl,i}(t) \]

Assuming the identified emission rate of the \( i \)th contaminant source is \( S_i^* \), then

\[ C_p^c(t) = C_p(t) + e = \sum_{i=1}^{N} S_i^* \left( \frac{a_{Cl,i}(t)}{Q} \right) + e, \]

where \( C_p^c(t) \) is the measurement of sensor (mg/m\(^3\)) at \( p \) and \( t \); \( C_p(t) \) is the calculated concentration (mg/m\(^3\)) at \( p \) and \( t \) by substituting \( S_i^* \) into Eq. (4); and \( e \) is the difference between \( C_p^c(t) \) and \( C_p^c(t) \) (mg/m\(^3\)). Ideally, \( e \) is equal or close to zero.

Using Eq. (5), the emission rate of each potential source can be identified by solving a linear programming model, which can be expressed in a canonical form as follows:

\[
\begin{align*}
\min f(x) &= c^T x \\
\text{s.t.} \quad Ax &= b \\
\quad x &\geq 0
\end{align*}
\]

where \( f(x) \) is the objective function, \( Ax = b \) and \( x \geq 0 \) are the constraint functions, \( x \) represents the vector of variables to be determined, \( c \) and \( b \) are vectors of known coefficients, \( A \) is a known matrix of coefficients, and \( (\cdot)^T \) is the matrix transpose.

The linear programming model identifies the emission rate of each potential source by minimizing the objective function with constraint functions. The objective function is the sum of absolute errors between concentrations calculated by Eq. (4) and those detected by sensors. The constraint functions are derived from Eq. (5). The detailed derivation procedure of the linear programming model is given in Appendix section. After solving the linear programming model, the emission rate of each potential source can be determined, and then the number, locations, and emission rates of indoor sources can be identified, because it is assumed that the locations of all potential sources are known. A major constraint of this model is that it can only work with a known release time of contaminant sources. In practice, the release time is usually unknown and must be identified.

The linear programming model excludes negative emission rates by using constraints \( x \geq 0 \). Ideally, any potential source with an identified emission rate greater than zero can be thought as a source. However, the identification results will inevitably contain some errors in practice because of the errors of numerical simulations and sensor measurements. If the identified emission rate of a potential source is a tiny number, a quantitative criterion is required to help determine whether the potential source actually release any contaminant.

In a ventilated space, if only the \( i \)th contaminant source existed and all the inlets concentrations would be zero, then the average exhausted contaminant concentration under steady-state conditions is:

\[ C_{e,i} = S_i/Q \]

where \( S_i \) is the emission rate of the \( i \)th potential source (mg/m\(^3\)).

According to Eq. (7), a quantitative criterion is proposed as:

\[ S_{thd} = \zeta \times C_{thd} \times Q \]

where \( S_{thd} \) is the emission rate threshold (mg/s) under which the release of a source is suggested to be neglected; \( \zeta \) is a factor far less than 1 (e.g., \( \zeta = 0.01 \)); \( C_{thd} \) is the concentration threshold of the contaminant (mg/m\(^3\)) above which a threat to people or property may exist during a specific time period.

The emission rate threshold \( S_{thd} \) is a function of \( C_{thd} \) and \( Q \), and is significantly scaled down by a small number \( \zeta \). If the identified emission rate of a potential source is less than \( S_{thd} \), it suggests that the release of the potential source can be neglected because it alone will not cause a severe threat to people or property. For most contaminants, the value of \( C_{thd} \) can be determined by referring to public exposure guidelines, such as AEGs (Acute Exposure Guideline Levels), ERPGs (Emergency Response Planning Guidelines), and TEELs (Temporary Emergency Exposure Limits) [44,45]. In some special cases, if the value of \( C_{thd} \) cannot be easily determined, it somewhat makes sense to use a number small enough as the emission rate threshold. The reason is that if the identified emission rate is small enough, whether or not considering this source will not significantly affect the concentration field and corresponding response measures. In addition, the emission rate threshold can also be set as zero for conservative reasons.

### 3.2.3. Component III: algorithm for identifying release time

The release time of a contaminant is denoted by \( t_r \), which refers to the time when the contaminant release event starts. After the contaminant release, the sensor that first detects the contaminant is called the first responding sensor (FRS), and the time when the contaminant is detected by the FRS is called the alarming time of FRS and denoted by \( t_{td} \). The time interval between \( t_r \) and \( t_{td} \) is called the lag time of the FRS, and is expressed as \( t_{FRS} = t_{td} - t_r \), as shown in the following figure. Fig. 2

In practice, \( t_{td} \) is known, while \( t_r \) and \( t_{FRS} \) are to be determined. The presented algorithm begins by assuming \( t_{FRS} = 0 \) s, and continues by increasing \( t_{FRS} \) at a time step \( \Delta t_{FRS} \). As shown in Fig. 1, with an assumed release time and the measurements of all sensors over a period \( \tau \) after \( t_r \), the contaminant concentrations at the FRS can be predicted by sequentially executing Components II and I. An index called SLT is proposed to quantify the match degree between the predicted and measured concentrations at the FRS over period \( \tau \). The SLT index is defined as:

**FRS: First responding sensor**

\[
\text{Lag time of FRS} \quad t_{FRS} = t_{td} - t_r
\]

\[ t_r: \text{Release time of contaminant} \]

\[ t_{td}: \text{The moment when the contaminant was detected by FRS} \]

**Fig. 2. Timeline of the contaminant release event.**
where $SLT_j$ is the SLT index (m$^3$/mg) when $t_{FRS} = j \times \Delta t_{FRS}$; $C_{1,k}$ is the $k$th measured concentration at the FRS (mg/m$^3$); $C_{2,k}$ is the $k$th predicted concentration at the FRS when $t_{FRS} = j \times \Delta t_{FRS}$ (mg/m$^3$); and $N_{FRS}$ is the number of measured concentrations at the FRS.

The denominator of Eq. (9) is essentially a widely used concept, namely the distance between two data sets. A smaller distance between the predicted and measured data will lead to a larger value of $SLT$, which further indicates a higher match degree between the two types of data. In practice, the errors in numerical simulations and sensor measurements are inevitable, and thus the distance between the predicted and measured data is always greater than zero, which ensures that the definition of SLT is always meaningful.

By increasing $t_{FRS}$ continuously, and running Components II and I iteratively, a curve of SLT changing with $t_{FRS}$ can be plotted. The peak point of the curve indicates a best match between the two data sets. With the SLT curve, the release time $t_r$ can be identified by finding the peak point of the curve. For example, if the peak point corresponds to a lag time $l \times \Delta t$, then the identified release time was $t_r = l \times \Delta t$. Furthermore, the source locations and emission rates can be determined, which were the outputs of Component II using $t_{FRS} = l \times \Delta t_1$ as the assumed lag time of the FRS.

3.3. Identification procedure

In practice, the presented identification method operates in a two-stage procedure, as illustrated in Fig. 3. In the first stage, before contaminant release, CFD is used to simulate a limited number of benchmark scenarios of contaminant dispersion. In each scenario only one of the potential sources is released at a constant emission rate. Therefore the number of CFD simulations is equal to the number of potential sources. After the simulations, the TACS data from each potential source to each sensor are calculated and stored in a database.

In the second stage, after contaminant release, the source locations, emission rates and release time are identified using the measurements of all sensors over a specific period after the release time as the inputs. In this stage, Components I and II are operated iteratively to identify the release time of the contaminant. Nevertheless, this stage will be completed in a few seconds, because Components I and II are very fast.
In the identification procedure, the analytical expression (Component II) plays an important role in reducing the computational load. For an indoor space with a limited number of potential contaminant sources, there are theoretically an infinite number of contaminant dispersion scenarios, by considering all the possible combinations of source parameters, such as source numbers, source locations, and emission rates. Because of the analytical expression, only a limited number of time-consuming CFD simulations are required for covering all the possible scenarios, which makes it possible to use CFD in the forward methods of source identification.

4. A validation case

4.1. Case setup

A three-dimensional office was used to validate the presented identification method (Fig. 4). The office (X \times Y \times Z = 9.6 \text{ m} \times 3.2 \text{ m} \times 5 \text{ m}) was ventilated by a mixing ventilation (MV) system and a personalized ventilation (PV) system. The MV system was used to control the temperature, humidity and contaminant concentration of the whole room. The air supplied by the MV system entered the room from two supply air inlets (0.4 \text{ m} \times 0.4 \text{ m}) at 16 \text{ °C} and 0.4 \text{ m/s} and exited from an exhaust air outlet (0.8 \text{ m} \times 0.4 \text{ m}). The air exchange rate was 3 ACH (air changes per hour) or 0.128 \text{ m}^3/\text{s}. The PV system was used to supply fresh air directly to the six people in the office through six table-mounted supply air terminal devices (ATDs). For each person sitting behind a table, there was an ATD that is close to his/her head. In the office, each computer, person, lamp, and the single window contributed a table, there was an ATD that is close to his/her head. In the office, each computer, person, lamp, and the single window contributed 108, 75, 34, and 220 W heat to the room, respectively. The four computers, person, lamp, and the single window contributed 108, 75, 34, and 220 W heat to the room, respectively. The four walls, ceiling, and floor were set as adiabatic boundaries.

It is assumed that a toxic gas was intentionally released into the PV system in a terrorist attack. The toxic gas could potentially be delivered to the breath zone of the indoor people from one or more of the six ATDs. In this case, all the ATDs were potential contaminant sources, and denoted as CS1–CS6. Five sensors (SR1–SR5) were installed to identify the contaminated ATDs that were releasing the toxic gas (Fig. 4). The positions of the potential sources and sensors are listed in Table 2. A toxic gas spreading scenario in which the toxic gas was released from CS3 and CS5 at 50 mg/s was studied to test the presented identification method.

4.2. Numerical methods

A commercial CFD program, Airpak, was used to simulate the indoor airflow field and contaminant dispersion. Airpak is customised from a general-purpose program, Fluent, for indoor environment simulations [46]. Airpak has been widely used and validated in indoor airflow and contaminant dispersion studies, as reported by Xu [47].

An indoor zero-equation turbulence model was used to simulate the indoor turbulent flow [48]. An ideal gas law model was used to account for temperature-dependent density effects, and a surface-to-surface radiation model was used to calculate radiative heat transfers [46]. For simplicity and illustration purpose, each contaminated ATD was modelled as a contaminant source, rather than a supply air inlet, in CFD simulations. After the test of mesh independence, the office room was divided into 121,038 hexahedral control volumes. The momentum equations were solved on non-uniform staggered grids using a SIMPLE algorithm [49]. A second-order upwind scheme and linear under-relaxation iteration was used in the solution process.

In the presented identification method, Component II identifies the emission rate of each potential contaminant source by solving a linear programming model (Eq. (6)). Linear programming was first presented by Dantzig in 1947 to refer to the optimization problems in which both the objective function and the constraints are linear [50,51]. After decades of development, linear programming has become mature and widely used. The most popular algorithm in linear programming is called the simplex algorithm [51]. In this study, the simplex algorithm provided by Matlab optimization toolbox 2.3 was used to solve the linear programming model.

4.3. Validation procedure

The validation procedure includes four major steps.

Step I: prepare TACS data

First the steady-state airflow field indoors was simulated using CFD. Next, with the same airflow field, six benchmark scenarios of contaminant dispersion were simulated using CFD. In each scenario, only one of the six potential sources (Table 2) was released at a constant rate. Finally, using the concentration data simulated, the TACS data from each potential source to each sensor were calculated by Eq. (3).

Step II: simulate a toxic gas spreading scenario for testing the identification method

A scenario in which CS3 and CS5 were simultaneously released was simulated using CFD. The simulated concentrations at the positions of five sensors (SR1–SR5) were sampled directly. These concentration data are called “exact measurements”.

Step III: identify the contaminant sources

To mimic the sensor measurements in a real-world experiment, the “exact measurements” were modified into “perturbed measurements” using a data pre-processing procedure. First, the exact measurements below a specific value, the sensor threshold, are eliminated. Next, the “exact measurements” were added with normally distributed random errors using a perturbation scheme, which is described in Section 4.4. The “perturbed measurements” were further taken as the inputs of the presented identification method to identify the source locations, emission rates and release time.

Step IV: evaluate the source identification performance

The source identification performance can be evaluated by comparing the identified results with those predefined in Step II. A

| No. | Position (m) | X   | Y   | Z   |
|-----|--------------|-----|-----|-----|
| CS1 | 3.00         | 0.95| 0.85|     |
| CS2 | 5.00         | 0.95| 0.85|     |
| CS3 | 7.00         | 0.95| 0.85|     |
| CS4 | 3.00         | 0.95| 4.05|     |
| CS5 | 5.00         | 0.95| 4.05|     |
| CS6 | 7.00         | 0.95| 4.05|     |
| SR1 | 3.30         | 2.20| 2.00|     |
| SR2 | 5.30         | 2.20| 2.00|     |
| SR3 | 7.30         | 2.20| 2.00|     |
| SR4 | 5.30         | 2.20| 1.00|     |
| SR5 | 5.30         | 2.20| 3.00|     |

* CS1–CS6: potential contaminant sources; SR1–SR5: sensors.
series of indices, the scale of relative errors (SRE), were introduced to quantify the identification performance. The definitions of SRE indices are provided in Section 4.4.

4.4. Data pre-processing and performance evaluation

The perturbation scheme used to generate random errors was defined as:

\[ C_{\text{pert}}(t) = C_{\text{exct}}(t) + \varepsilon_r = C_{\text{exct}}(t) + \xi \times \text{STD} \times C_{\text{exct}}(t), \]

where \( C_{\text{pert}}(t) \) and \( C_{\text{exct}}(t) \) are perturbed and exact measurements (mg/m³) at time \( t \), respectively (mg/m³); \( \varepsilon_r \) is a random variable that follows a normal distribution with mean = zero and standard deviation = \( \xi \times C_{\text{exct}}(t) \) (mg/m³); \( \xi \) is a fraction to adjust the level of measurement error (0 ≤ \( \xi \) ≤ 1); STD is the standard normal deviation.

Eq. (10) indicates that all concentrations were measured by sensors with the same relative error; i.e. a larger concentration may have a greater measurement error. To reflect the level of measurement errors more clearly, two parameters, \( \delta_1 \) and \( \delta_2 \), were defined as:

\[ \delta_1 = \max_{1 \leq i \leq M} \left( \frac{C_{\text{pert},i} - C_{\text{exct},i}}{C_{\text{exct},i}} \right) \times 100\%, \] (11)

\[ \delta_2 = \left( \frac{\sum_{i=1}^{M} \frac{C_{\text{pert},i} - C_{\text{exct},i}}{C_{\text{exct},i}}}{M} \right) \times 100\%, \] (12)

where \( \delta_1 \) and \( \delta_2 \) are the maximum and average absolute relative error between perturbed and exact measurements, respectively; \( C_{\text{pert}} \) and \( C_{\text{exct}} \) are the \( i \)th perturbed and exact measurements, respectively; and \( M \) is the number of exact measurements.

The SRE indices for performance evaluation were defined as follows [38]:

\[ \text{SRE}_1 = \left| \frac{S_i}{S_i^*} \right| \max_{1 \leq i \leq N} (S_i), \] (13)

where \( S_i \) and \( S_i^* \) are the identified and actual emission rates (mg/s) of the \( i \)th source, respectively; and \( N \) is the number of potential sources. \( \text{SRE}_1 \) indicates the absolute relative error between the identified and actual emission rate of the \( i \)th source. A lower value of \( \text{SRE}_1 \) indicates the more accurate identification of the \( i \)th source.

\[ \text{SRE}_2 = \frac{\sum_{i=1}^{N} \text{SRE}_1}{N} \] (14)

\( \text{SRE}_2 \) indicates the average level of \( \text{SRE}_1 \) (1 ≤ \( i \) ≤ \( N \)). A lower value of \( \text{SRE}_2 \) indicates the more accurate identification of all potential sources with respect to the overall average.

\[ \text{SRE}_3 = \max_{1 \leq i \leq N} \text{SRE}_1 \] (15)

\( \text{SRE}_3 \) indicates the maximum level of \( \text{SRE}_1 \) (1 ≤ \( i \) ≤ \( N \)). A lower value of \( \text{SRE}_3 \) indicates the more accurate identification of all potential sources with respect to the maximum relative error.

5. Results and discussion

5.1. CFD simulation results

The steady-state airflow field indoors was first simulated, which shows a typical distribution pattern of mixing ventilation, as shown in Fig. 5. With this airflow field, six benchmark scenarios of contaminant dispersion were further simulated over 180 s with a time step 0.5 s. In each scenario, only one of the six potential sources (Table 2) was released at a constant rate. By using the simulated concentrations and Eq. (3), the TACS from each potential source to each sensor was calculated. Fig. 6 presents an example of TACS variation, from potential source CS3 to each sensor. The TACS curves from CS3 to SR3 and SR4 varied in magnitude and shape, while the curves from CS3 to SR1, SR2, and SR5 overlapped one another and had much lower values. This indicates that SR3 and SR4 can more easily detect CS3 than the rest sensors.

The toxic gas spreading scenario for testing the presented identification method was simulated over 180 s with a time step of 0.5 s. In this scenario, CS3 and CS5 were simultaneously released at a constant rate of 50 mg/s. Fig. 7 shows the variation in concentration distribution in this scenario. Using the simulated concentrations, the lag times of each sensor with different sensor thresholds were obtained directly, and are summarised in Table 3. For this scenario the FRS was sensor SR3 at sensor thresholds of 1.0, 5.0, and 10.0 mg/m³, and its lag times were 11, 14, and 15 s respectively.

5.2. Source identification with different sensor thresholds

The exact measurements at the three sensor threshold levels, 1.0, 5.0, and 10.0 mg/m³, were used to test the presented identification method. In each test, the total sampling time of the FRS was 150 s, the assumed lag time of the FRS was increased from 0 s at a step of 1 s, and the sampling interval of each sensor was 5 s. The measurements of all sensors were taken as the inputs to the identification method. Table 4 summarises the number of measurements from each sensor at different sensor thresholds. It is
shown that the total number of measurements from all sensors decreased with the increase in sensor threshold.

Fig. 8 presents the SLT curves calculated using exact measurements at different sensor thresholds. It is shown that each curve has a sharp single peak. By finding the horizontal coordinate value of each peak point, the lag time of the FRS at each sensor threshold was identified and listed in Table 5. It is shown that the identified lag time of the FRS was identical to the actual ones for each sensor threshold.

With the identified lag times, the emission rate of each potential source was further determined and listed in Table 5. Two indices, SRE2 and SRE3, were provided in Table 5 to help evaluate the identification performance. The results indicate that the identified emission rates were very accurate, even for the worst identification results corresponding to the sensor threshold 10.0 mg/m³. A possible explanation for the high accuracy is the use of exact measurements. The identification accuracy of emission rates also presented a declining trend with the increase in sensor threshold. This may be due to the fact that fewer measurements were used in source identification at a higher sensor threshold (see Table 4).

5.3. Source identification with different measurement errors

The identification method was further tested with two levels of measurement error, which were generated using the perturbation scheme described in Section 4.4. As an example, Fig. 9 shows the exact, highly perturbed, and lowly perturbed measurements of the FRS (SR3). According to Eqs. (10)–(12), the level of measurement error was determined by , characterised by and . For the highly perturbed measurements, and were 0.1, 27.1% and 7.4%, respectively; whereas for lowly perturbed measurements these parameters were 0.01, 3.3% and 1.0%, respectively.

Fig. 10 presents the SLT curves using highly perturbed measurements at different sensor thresholds. The SLT curves in Fig. 10 were much flatter than those in Fig. 8. Moreover, multiple peak points appeared in the curve corresponding to the sensor threshold of 10.0 mg/m³. A comparison between Figs. 10 and 8 indicates that

Table 3

| Threshold (mg/m³) | SR1 | SR2 | SR3 | SR4 | SR5 |
|------------------|-----|-----|-----|-----|-----|
| 1.0              | 110 | 91  | 11  | 55  | 42  |
| 5.0              | 136 | 116 | 14  | 63  | 53  |
| 10.0             | 153 | 130 | 15  | 67  | 60  |

* First responding sensor (FRS) highlighted in grey.

Table 4

| Threshold (mg/m³) | Number of measurements from each sensor |
|------------------|----------------------------------------|
|                  | SR1 | SR2 | SR3 | SR4 | SR5 |
| 1.0              | 10  | 14  | 30  | 21  | 23  |
| 5.0              | 5   | 9   | 30  | 20  | 22  |
| 10.0             | 2   | 7   | 30  | 19  | 21  |

* First responding sensor (FRS) highlighted in grey.
the measurement errors would decrease the accuracy of identifying the release time of the contaminant.

Table 6 summarises the identification results using highly perturbed measurements. Compared with the identification results using exact measurements (Table 5), when highly perturbed measurements were used the identified lag times of the FRS were generally less accurate, and the accuracy of identified emission rates significantly decreased. These results indicate that the measurement errors can significantly degrade the identification performance.

The identification method was further tested with lowly perturbed measurements. The SLT curves are plotted in Fig. 11 and the identification results summarised in Table 7. By decreasing the level of measurement error, the SLT curves became steeper and more symmetrical, and each curve only had a single peak. The identified lag times of the FRS were generally more accurate, and the accuracy of the identified emission rates increased significantly. These results indicate that the identification performance can be improved by reducing the level of measurement errors.

In contrast to the results using exact measurements (Table 5), the results using perturbed measurements (Tables 6 and 7) did not present the trend that the identification accuracy will decline with an increase in sensor threshold. In fact, the results corresponding to a sensor threshold of 10.0 mg/m$^3$ in Tables 6 and 7 were unexpectedly good. The difference between the results may be explained from two aspects. Firstly, the measurement errors were generated in a random manner, which resulted in the randomness of the identification results. Secondly, in the early stage of contaminant dispersion, the concentrations are normally low and changing rapidly, which can lead to large errors in the simulation results. The high sensor threshold may be helpful for eliminating these simulation errors.

### 5.4. Source identification with different total sampling times

The identification method was also tested by changing the total sampling times of the FRS. Table 8 summarises the total sampling time of each sensor. It is shown that one or many sensors were unable to provide concentration data when the total sampling time of the FRS was less than 150 s.

Table 9 summarises the identification results using lowly perturbed measurements. The threshold and sampling interval of each sensor was 5 s. The toxic gas was actually released from CS3 and CS5 at 50 mg/s. First responding sensor (FRS) used a total sampling time of 150 s. Sampling interval of each sensor was 5 s.

### Table 5

| Threshold (mg/m$^3$) | Lag time of FRS (s) | Identified emission rate (mg/s) | Performance index (%) | Computation time (s) |
|----------------------|--------------------|---------------------------------|-----------------------|---------------------|
| Actual               | Identified         | CS1    | CS2    | CS3    | CS4    | CS5    | CS6    | SRE2   | SRE3   |
| 1.0                  | 11                 | 11     | <0.001 | 0.001  | 49.999 | <0.001 | 49.999 | 0.001  | 0.002  | 2.63   |
| 5.0                  | 14                 | 14     | <0.001 | <0.001 | 49.999 | <0.001 | 49.997 | 0.021  | 0.008  | 2.09   |
| 10.0                 | 15                 | 15     | <0.001 | <0.001 | 49.999 | 0.398  | 49.998 | <0.001 | 0.134  | 1.45   |

* The toxic gas was actually released from CS3 and CS5 at 50 mg/s. First responding sensor (FRS) used a total sampling time of 150 s. Sampling interval of each sensor was 5 s.

### Table 6

| Threshold (mg/m$^3$) | Lag time of FRS (s) | Identified emission rate (mg/s) | Performance index (%) | Computation time (s) |
|----------------------|--------------------|---------------------------------|-----------------------|---------------------|
| 1.0                  | 11                 | 11     | 0.728  | 0.055  | 48.734 | <0.001 | 48.637 | 33.394 | 12.269 | 66.788 | 1.86   |
| 5.0                  | 14                 | 16     | 0.912  | <0.001 | 48.635 | <0.001 | 46.214 | 46.260 | 17.441 | 92.520 | 1.63   |
| 10.0                 | 15                 | 12     | 0.498  | 0.564  | 49.018 | <0.001 | 49.948 | <0.001 | 0.698  | 1.964  | 1.42   |

* The toxic gas was actually released from CS3 and CS5 at 50 mg/s. First responding sensor (FRS) used a total sampling time of 150 s. Sampling interval of each sensor was 5 s.

SRE2, and SRE3 are performance indexes of identification results, referring to Eqs. (14) and (15).
sensor were 5 mg/m³ and 5 s, respectively. The FRS was SR3, and its actual lag time was 14 s (see Table 3). When the total sampling time of the FRS was less than 50 s, the identification method failed to obtain accurate results. The failure of the model may be explained by two reasons. Firstly, except the measurements from the FRS, there are no measurements from the other sensors, or the measurements from the other sensors are not enough. Secondly, the measurements from the other sensors are probably important to the source identification.

When the total sampling time of the FRS was equal to or greater than 50 s, all identification results, except that corresponding to 90 s, were very accurate and varied in a narrow range. As shown in Table 8, with the increase in total sampling time of the FRS, more sensors provided concentration data for source identification. These results indicate that extending the total sampling time of the FRS or using measurements from more sensors can contribute to improving the identification performance. In practice, the source identification method can be executed repeatedly by increasing the total sampling time of the FRS. A minimum or appropriate total sampling time of the FRS can be determined when the identification results tend to be stable. The results corresponding to 90 s were much worse than expected. This may be because the measurement errors were randomly generated.

5.5. Computational requirements

All calculations reported were conducted on a personal computer (CPU: Intel(R) Core(TM)2 T7200 @ 2.00 GHz). As shown in Tables 5–7 and 9, the computation times for each identification procedure in this study were only a few seconds. Therefore, the presented identification method is fast enough to be integrated into a real-time identification system.

5.6. Limitations and applicability

The identification method presented was developed based on four assumptions. Recognition of the underlying assumptions is thus important to understand the limitations and applicability of the method. In the following, the underlying assumptions are analysed, and the limitations and applicability of the method are discussed.

The first assumption is that the contaminant can be treated as a passive gas and the indoor airflow field is in steady-state. For most indoor environments equipped with mechanical ventilation, the airflow field can reach a steady-state much faster than the contaminant dispersion, and thus the airflow field can be regarded as steady during the dispersion of contaminant. Normally, indoor airflow is turbulent and the contaminant concentration is low enough that the contaminant dispersion only has a trivial effect on the airflow field and air density [43]. This assumption is, therefore, reasonable for most indoor environments. For quiescent environments, the presented method may also be applicable. The reason is that a static environment can also be thought with fixed airflow mode, in which pure molecular diffusion transports contaminant. It is worth noting that, the presented method is not applicable to the situations where the dispersion of contaminant is significantly different from that of passive gas, such as the releases of particles, aerosols, and a massive leakage of dense or light gas. In addition, the method cannot be applied to the indoor environments where the airflow is always changing, such as the naturally ventilated environments.

This work also assumes that there are a limited number of potential sources and their locations are known. This assumption excludes the application of the method in situations where the number of the potential sources is unlimited and their locations are completely unknown. Nevertheless, a variety of contaminant dispersion scenarios that satisfy this assumption can still be found in practice, such as the emission of volatile organic compounds from furniture, leakage of toxic or flammable gas from suspected leak points, viruses spreading from infected people at specific locations, toxic agents released from supply air inlets by terrorists.

The third assumption is that the contaminant sources are continuously released at constant rates. According to this

---

**Table 7**

| Threshold (mg/m³) | Lag time of FRS (s) | Identified emission rate (mg/s) | Performance index (%) | Computation time (s) |
|------------------|---------------------|--------------------------------|----------------------|---------------------|
|                  | Actual              | Identified                      | CS1                  | CS2                  | CS3 | CS4 | CS5 | CS6 | SRE2 | SRE3 |                  |
| 1.0              | 11                  | 10                              | 0.332                | 0.103                | 50.024 | <0.001 | 50.675 | <0.001 | 0.378 | 1.350 | 1.89 |
| 5.0              | 14                  | 14                              | <0.001               | <0.001               | 49.974 | <0.001 | 50.242 | <0.001 | 0.089 | 0.484 | 1.55 |
| 10.0             | 15                  | 15                              | <0.001               | <0.001               | 50.029 | <0.001 | 49.855 | <0.001 | 0.058 | 0.290 | 1.34 |

* The toxic gas was actually released from CS3 and CS5 at 50 mg/s. First responding sensor (FRS) used a total sampling time of 150 s. Sampling interval of each sensor was 5 s.

b SRE2, and SRE3 are performance indexes of identification results, referring to Eqs. (14) and (15).
The presented method is applicable to constantly-released contaminant sources. In addition, the method can only be used to identify multiple contaminant sources simultaneously released at constant rates, by assuming that all contaminant sources are released at the same cation. This assumption also indicates that the presented method is not applicable to the situations where the dispersion of contaminant is significantly different from that of passive gas, such as the releases of particles, aerosols, and a massive leakage of dense or light gas. Further validations of the method in a laboratory chamber and real indoor environments are ongoing and will be reported in the near future. Some major limitations of the method are also worth attention. Firstly, the method cannot be applied to situations where the dispersion of contaminant is significantly different from that of passive gas, such as the releases of particles, aerosols, and a massive leakage of dense or light gas. Second, the method is not applicable for identifying instantaneous releases or continuous releases with fast changing rates. In addition, the method can only be used to identify multiple contaminant sources that are simultaneously released. Therefore, further research is required to develop more sophisticated techniques that can be applied to more complex situations of contaminant dispersion.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grant no. 50908128), the Natural Science Foundation of Jiangsu Province (Grant no. BK20131067), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (Grant no. 51321064), and the National Science Foundation for Distinguished Young Scholars of China (Grant no. 51125030).

Appendix. Derivation of linear programming model

In Eq. (5), $C_p^M(t)$ and $Q$ can be measured, and $a_{Ci,j}(t)$ can be obtained by CFD simulations or tracer gas experiments. In contrast, $S_j^T$ and $e$ are unknown and interrelated. With Eq. (5), when there are $M$ sensor measurements the following linear equations can be constructed:

$$
\begin{bmatrix}
  a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,M}x_M + e_1 = b_1 \\
  a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,M}x_M + e_2 = b_2 \\
  \vdots \\
  a_{M,1}x_1 + a_{M,2}x_2 + \cdots + a_{M,M}x_M + e_M = b_M
\end{bmatrix},
$$

where $a_{ij} = d_{ij}/Q$, $b_j = C_p^M$, $x_i = S_j^T$, $i \in [1, N]$, and $j \in [1, M]$, and $C_p^M$ is the $j$th measurement from a specific sensor at a specific moment. $d_{ij}$ is the TACS from the $i$th source to a specific sensor at the specific moment when $C_p^M$ was measured.

In Eq. (16), if the identified emission rate of each potential source $x_i$ approaches or is equal to its actual value, $\sum_{j=1}^{M}|e_j|$ will be close to zero or a minimum. Therefore, $x_i$ or $S_j^T$ can be identified using a nonlinear programming model, as follows:

$$
\min f(x) = \sum_{i=1}^{M}|x_{N+1}|
\begin{cases}
  \text{s.t.} & Ax = b \\
  & x_i \geq 0, \quad i \in [1, N]
\end{cases}
$$

where $x = (x_1, x_2, \ldots, x_N, x_{N+1}, \ldots, x_{N+M})^T$, $A = (a_{ij})_{MN+M}$, $b = (b_1, b_2, \ldots, b_M)^T$, $x_{N+1} = e_j$, and $a_{N+i} = 1$, $0 \leq i \leq N$, $0 \leq j \leq M$.
To reduce the solution complexity, the nonlinear programming model was further transformed into a linear programming model, as follows:

\[
\min f(x) = \frac{M}{2} \sum \left( \frac{X_{n+i}}{x} + \frac{x}{X_{n+i}} \right) + C_i, \\
\text{s.t.} \quad Ax = b, \quad \begin{cases} \frac{x}{X_{n+i}} \geq 0, \quad X_{n+i} \geq 0, \quad \sum x_{n+i} \\
= 0, \quad i \in [1, N + 2M] \end{cases}
\]

where \( x = (x_1, x_2, \cdots, x_m, x_{m+1}, \cdots, x_{m+N}, \cdots, x_{m+M}) \),

\[X_{n+i}^{\min} = \left( \begin{array}{cccc} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{array} \right), \quad j = 1, 2, \cdots, M,
\]

\[(a_{i,N}) = \left( \begin{array}{cccc} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{array} \right), \quad j = 1, 2, \cdots, M.
\]

The objective function of the linear programming model is the sum of absolute errors between concentrations calculated by Eq. (4) and those measured by sensors. By minimizing the objective function, the emission rate of each potential source can be solved.

References

[1] Siddiqui M, Jayanti S, Swaminathan T. CFD analysis of dense gas dispersion in indoor environment for risk assessment and risk mitigation. J Hazard Mater 2012;209-210(1):177–85.

[2] Xu X, Zhang Z, Tian G, Wang G. Numerical simulation of combustible gas diffusion in the confined room with obstacles. Nat Gas Ind 2010;30(5):119–22.

[3] Jones AP. Indoor air quality and health. Atmos Environ 1999;33(28):4535–64.

[4] Målström MF, Campos MP, Mazzilli BP, Villaverde FL. Assessment of external pollution sources in urban areas using a multivariate statistical approach: role of indoor transport and mixing time scales on the performance of sensor systems for characterizing contaminant releases. Atmos Environ 2007;41(40):9340–503.

[5] Musser A. An analysis of combined CFD and multizone IAQ model assembly for greenhouse gas emission modeling. Indoor Built Environ 2009;18(1):275–82.

[6] Zhang TF, Yin S, Wang S. An inverse method based on CFD to quantify the temporal release rate of a continuously released pollutant source. Atmos Environ 2013;77(0):62–77.

[7] Zhang T, Chen Q. Identification of contaminant sources in enclosed environments by inverse CFD modeling. Indoor Air 2007;17(3):167–77.

[8] Huang C, Wang S. A three-dimensional inverse heat conduction problem in a three-layered medium. J Heat Mass Transf 1999;42(18):3387–92.

[9] Alifanov OM. Inverse heat transfer problems. New York: Springer-Verlag; 2005.

[10] Sorensen JH, Shumpert BL, Vogt BM. Planning for protective action decision making: evacuate or shelter-in-place. J Hazard Mater 2004;109(1–3):101–14.

[11] Jones AP. Indoor air quality and health. Atmos Environ 1999;33(28):4535–64.

[12] Målström MF, Campos MP, Mazzilli BP, Villaverde FL. Assessment of external pollution sources in urban areas using a multivariate statistical approach: role of indoor transport and mixing time scales on the performance of sensor systems for characterizing contaminant releases. Atmos Environ 2007;41(40):9340–503.

[13] Musser A. An analysis of combined CFD and multizone IAQ model assembly for greenhouse gas emission modeling. Indoor Built Environ 2009;18(1):275–82.

[14] Bady M, Kato S, Huang H. Identification of pollution sources in urban areas using reverse simulation with reversed time marching method. J Asian Arch Build Eng 2009;8(1):177–92.

[15] Zheng X, Chen Z. Back-calculation of the strength and location of hazardous materials releases using the pattern search method. J Hazard Mater 2010;183(1–3):474–81.

[16] Amannuzzo AJ, Young GS, Haupt SE. Utilizing state estimation to determine the source location for a contaminant. Atmos Environ 2012;46(0):580–9.

[17] Huang C, Wang S. A three-dimensional inverse heat conduction problem in a three-layered medium. J Heat Mass Transf 1999;42(18):3387–92.

[18] Alifanov OM. Inverse heat transfer problems. New York: Springer-Verlag; 2005.
[45] Woodall GM, Smith RL. The air toxics health effects database (ATHED). Toxicol Appl Pharm 2008;233(1):20–4.

[46] Fluent Inc. Airpak 2.1 user's guide. Lebanon, NH: Fluent Inc; 2002.

[47] Xu L. Effectiveness of hybrid air conditioning system in a residential house [Ph.D. thesis]. Tokyo, Japan: Waseda University; 2003.

[48] Chen Q, Xu W. A zero-equation turbulence model for indoor airflow simulation. Energy Build 1998;29(2):137–44.

[49] Patankar SV. Numerical heat transfer and fluid flow. Washington, DC: Hemisphere Publishing Corp; 1980.

[50] Dantzig GB. Reminiscences about the origins of linear programming. Oper Res Lett 1982;1(2):43–8.

[51] Ravindran A, Reklaitis GV, Ragsdell KM. Engineering optimization: methods and applications. 2nd ed. New York: Wiley; 2006.