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.
Contaminant source identification within a building: Toward design of immune buildings

Arash Bastani, Fariborz Haghighat, Janusz A. Kozinski

Abstract

The level of protection of a building against the intentional or accidental release of chemical agents is crucial. Both scenarios could endanger life and safety of the buildings occupants. Equipping buildings with appropriate chemical sensors can alert the building occupants about the contaminant release. The readings of these sensors can be employed to trace the location of release, and help to take the appropriate actions to minimize the casualties. However, only a limited number of them can be installed due to their initial and operating cost. Moreover, there is no information about the source strength, release time and possible source location. This paper reports the development of a methodology to identify the source location using sensors reading from limited locations. The methodology uses the artificial neural network (ANN) as a statistical analysis integrated with a multi-zone airborne contaminant transport model, CONTAM. To evaluate the applicability of this method, the contaminant dispersion within a building was modeled and the results were integrated to an ANN for the source identification. The prediction made by the trained ANN was then evaluated by predicting the source of the contaminant in 40 extra cases, which had not been seen by the network during the training session. The model was able to predict the source location in more than 90% of the cases when the building was monitored by three or more sensors. The results show that the method can be used to help building designers decide the optimum configuration of the sensors required for a space based on the accuracy level of the source detection.

1. Introduction

Recent events have highlighted both criminally introduced and accidental exposure of civilian populations to toxic chemical substances. Spreading severe acute respiratory syndrome (SARS) in a flight from Hong Kong to Beijing in 2003, and the release of nerve agent Sarin in Tokyo metro station in 1995 were such incidents caused by a sudden release of a chemical or biological agent inside an enclosed space. Superstructures, shopping centers or high-rise buildings, with a total floor space of over hundred thousand are becoming part of new cities’ landscapes. These buildings are normally accommodating several thousand occupants. Sudden intentional or accidental release of a chemical agent inside such buildings endangers the lives of its occupants. To mitigate the casualty caused by these scenarios, the protection level of the buildings needs to be increased. There are a number of options which building owners and managers have in order to increase the protection level inside the enclosed environment [14]. Equipping buildings with appropriate sensors is a logical approach to enhance the safety and security of building occupants. A sensor or a combination of them can be used to monitor the concentration level of a contaminant and warn the air quality deterioration. Moreover, the information obtained by these sensors can be applied to locate the source of the contaminant in case of an incident. Having the knowledge of the source location is required to take an appropriate action such as evacuating the building, shutting down the ventilation system, or moving the occupants to a designated shelter when the building is contaminated with a toxic gas. However, these sensors are very expensive and bulky, which limits the number of them that could be employed in a building to measure the contaminant concentration level in the entire enclosed space. Thus, the main challenge is to locate the source of a contaminant within a building using a limited number of sensors, while these readings may vary when the unknown source is located rather near or far away from the sensor. Meanwhile, the unknown source released...
time and source strength create an uncertainty in understanding the actual source location.

2. Inverse tracing of a source

2.1. Contaminant dispersion simulation

To investigate indoor environment conditions inside an enclosed space, numerical methods are a promising approach considering the great advances which have been made in computational techniques. A number of airflow models such as the computational fluid dynamics (CFD), the zonal model and the multi-zone model were developed to study the distribution of a contaminant concentration and other airflow parameters inside a space. To characterize contaminant dispersion inside a building, each airflow model requires initial and boundary conditions as input parameters. Thus, in the case where there is a source of a contaminant inside the indoor environment, the source location, as well as its release time and strength, are needed to be specified in the forward simulation to have the complete information about contaminant distribution. In source identification problems, however, the contaminant concentrations measured by the sensors are known as the consequence of an unknown causal. Hence, the distribution of contaminant concentration needs to be integrated with an inverse analysis yield to source location identification.

2.2. Inverse analysis

Finding the contaminant source location from sensors’ reading is known as an inverse analysis due to the reverse cause-effect orientation. Here, the available input data is the contaminant concentration in limited locations, and the contaminant source location is the needed information of this analysis.

Generally, when a contaminant is released inside a building, the contaminant concentration in a sensor’s location is a function of source location, source release time and source strength. Furthermore, the outdoor weather conditions, such as wind speed and direction, and the building ventilation operation influence the contaminant distribution within the building. Thus, the mentioned parameters are understood as the effective factors on the distribution profile of a contaminant and consequently, on the mounted sensors’ readings. Although changing these parameters may result in a different contaminant distribution pattern, it is possible to have the same concentration distributions which could result from different combination of these factors. For instance, a contaminant source releases \( M_1 \) (kg) of a contaminant at \( t_1 \) (s) earlier in a location could have the same contaminant distribution profile as another source releases \( M_2 \) (kg) of the same contaminant in \( t_2 \) (s) earlier in another location. Therefore, there is an uncertainty in identifying (tracking back) the source location from the sensors’ readings. This uncertainty is the main problem in the inverse analysis. To overcome this problem, a number of methods have been developed.

Zhang and Chen [21] applied inverse CFD modeling to locate the source of a contaminant inside an enclosed space. Here, they used quasi-reversibility (QR) method which solves the specie mass conservation equation with reversed time-step. In addition to reversing the time-step, the diffusion term in the governing equation was replaced by a fourth-order stabilization term accompanied with a stabilization coefficient. Indeed, the governing equation is not understood as a mass conservation equation anymore. Zhang and Chen [21] applied QR method to find the source of a contaminant in an aircraft cabin and an office building. They showed that the method could be used to locate the source in the aircraft with a good accuracy, while the result was more dispersive and less accurate in the case of the office building. They inferred that the replacement of the diffusion term had a negative effect in locating the source in office building with low Reynolds number.

Although, the method needs only one-time simulation, the term of time in the governing equation is a complicated issue. Herein, the reverse simulation needs to be conducted from the time that the sensor is detecting the contaminant back to the time that the contaminant was released. However, the release time of the contaminant is unknown in practice and depends on the type of the contaminant, its threshold limit to activate the alarm, the source location and its strength. Moreover, the accurate knowledge of the boundary conditions and their changes during the time is important to identify the amount of the contaminant, which left the domain after release. Therefore, the method needs a prior knowledge of the source activation time or the contaminant transport time to predict the source location. In addition, they used contaminant concentration distribution in the entire geometry as the initial condition for their model, which hardly can be achieved in a real scenario due to the limited number of sensors. To overcome this drawback, they studied the source characterization by single sensor [22]. For this purpose, they employed the QR method and added a nominal source term in the sensor’s location. The nominal source was used to find the distribution of backward location probability density function. Therefore, the location with the highest probability depicts the source location. They examined the proposed method in an aircraft with a single source of a contaminant. Moreover, they compared the QR method with another inverse model called pseudo-reversibility (PR) method. In PR method, the time is kept as forward, but the flow direction, instead, is reversed and the diffusion term is eliminated from the governing equation. As an input to their model, they used the peak-concentration measured by the sensor and the time required for the concentration to reach its peak value. Indeed, they assumed that the sensor continuously monitors the air, and it can provide the peak-concentration and the concentration profile. However, the peak-concentration is a function of the source location and its strength which are unknown. Thus, the method needs a prior knowledge of one of the source parameters to find the source location.

Liu and Zhai [7] carried out an extensive review of the available methods for identification of a contaminant source in groundwater and indoor spaces. In addition, they developed a probabilistic method to locate the source indoors. The method was adopted based on the adjoint probability method initially introduced by Neupauer and Wilson [12,13] to track a source of a contaminant in a groundwater. Liu and Zhai [7] modified and implemented the adjoint probability method for both CFD and multi-zone models to identify the contaminant source location inside an enclosed environment. The application of the CFD-based adjoint backward location probability was tested for two cases; an office building, 2-D, and an aircraft cabin, 3-D [8]. The inverse modeling was conducted with a known contaminant release time. The model showed an acceptable accuracy in predicting the contaminant source location. However, the work needs to be improved to identify the contaminant sources with both unknown location and release time. In addition, running a CFD program cannot give a fast real-time result, which is vital in an emergency situation. Hence, Liu and Zhai [9] developed the multi-zone-based adjoint probability method. The method was implemented to locate a source of a contaminant with either a single or multiple sensors and with either alarm only sensor or concentration reading sensor. Nevertheless, the method requires the knowledge of exact source activation time, which is not available in real life scenario.

Sohn et al. [16,17] proposed a probability approach to retrieve parameters of a source of a contaminant inside a building. The
proposed framework has two stages; pre-event stage and present stage. In the pre-event stage, all plausible source parameters are considered to conduct the airflow and contaminant dispersion simulation in a building. These simulations can be carried out by an airflow model, and the results are stored in a library. Each simulation corresponds to a ventilation system status, contaminant source status, contaminant species and overall, a contaminant release scenario. This stage is performed before the event. During an emergency incident, the second stage (present stage) is carried out to interpret the data obtained by the sensors and identifying the nearest matched simulation results with the measurements. For this purpose, they used Bayesian probabilistic method to compare each simulation result in the library with the measured data and assess the likelihood described by pre-simulated data. The simulation with the highest likelihood is accepted as the possible scenario for the investigated system. Thus, the inputs (source parameters) for that specific simulation are realized as the causal for the measured concentrations in sensors’ locations. This approach shows a fast real-time identification of a pollutant source without considering any prior knowledge of some source’s parameters. However, the main problem is the huge number of the required forward pre-simulations. To prepare a complete library of simulations, all plausible combinations of the causal parameters are required to be considered. The variety of these parameters depends on the degree of complexity of the system. As the character of the transport phenomena in an enclosed space is inherently complicated, this inverse technique demands lots of pre-simulations, which may fail to consider some possible combinations of causal parameters yet.

Although the mentioned methods presented a strong foundation to detect a source location inside an indoor space, some assumptions that they made need to be relaxed to be more realistic. The task imposes to characterize unknown sources of a pollutant by interpreting the observed concentration of a contaminant in limited locations inside the inspected domain. Meanwhile, the source parameters inside the area are unknown during an accident. Therefore, this study is presenting a new approach to locate a contaminant source during an accident only using the instant sensors’ readings and no prior knowledge of source parameters.

2.3. Artificial neural network (ANN) and its application in inverse analysis

As mentioned earlier, the unknown source parameters create an uncertainty in locating a contaminant source using the sensors’ readings. In a real event, neither the contaminant source parameters nor the changes in boundary conditions as a function of time are available. Hence, the system which inversely relates the contaminant concentration in sensors’ locations and the source location is not well-defined. One approach to characterize this system is the application of artificial neural network (ANN) in which the network needs to be trained based on the available data of the system. Then, the system can be replaced by the trained network to predict the system output based on the information provided by the sensor (monitored input data).

Many researchers have applied the ANN approach to predict the concentration of a pollutant such as SOx, NOx, and Ozone. Boznar et al. [2] applied ANN to predict hourly SO2 concentration in a polluted area in Slovenia. Hasham et al. [6] established an ANN to predict hourly concentration of NOx in Edmonton, Canada. They studied the effect of meteorological data, traffic, and industrial emission on the concentration of this component. The result shows a good forecasting ability of ANN. The application of ANN was not only limited to the outdoor environment. In indoor application, the airflow models were replaced by ANN which was integrated with an optimization algorithm to optimize the energy consumption inside the office buildings [23,24,10]. Vukovic and Srebric [18] applied ANN to find a source of a contaminant in a building by training a separate neural network (NN) for each separate zone of the building. Each zone was monitored by a sensor which measured transient concentration of the contaminant as the inputs of the NN assigned to that zone. To identify the source location, the distance between the NN’s prediction and the zone which that particular NN was assigned to was calculated, and then the zone which had the NN with the smallest error was recognized as the source location. Although they introduced an effective approach in source identification procedure, monitoring all building zones with sensors was not realistic. In addition, by having knowledge of the transient concentration of a contaminant in all zones the source would be clearly in the zone with the sudden change in concentration occurs and there is no need for additional analysis. Vukovic et al. [19] improved their work to detect the source location using limited number of sensors. They replaced the transient concentration of the contaminant in each zone with an exponential equation accompanied by three coefficients. Three consecutive readings of the sensor were used to calculate the coefficients of the equation, which were used as the NN inputs instead of the sensor’s readings. While the method could find the source of contaminant with unknown source parameters, the detection is only limited to the cases in which the source is not located in sensors’ zones. Their suggested equation cannot present the trend of the transient concentration in source location, and consequently the coefficient of the equation of the released zone cannot be calculated. In other words, they assumed that the source is not located in the zones which are monitored by a sensor.

While the work conducted by Vukovic et al. [19] is acknowledged, this research is attempting to improve this method to have the source detection in every possible zone in a building. Therefore, this study employs the application of ANN to find the source of contaminant release using instant sensor’s reading in limited locations.

3. Methodology

3.1. Application of artificial neural networks (ANN) to track a source of contaminant

This study developed a methodology based on the application of ANN to characterize the relation between the contaminant concentrations in sensors’ locations and the source location within an indoor environment. Generally, an ANN consists of a number of layers, which include a number of neurons. The first layer and the last one are called input and output layers, respectively. In addition, there can be one or multi-layer between the first and last layers which are known as hidden layers. In an ANN with feed forward structure, the neurons of each layer are connected to the neurons in adjacent layers. The connection between these neurons is called weight, and the magnitude of the weights represents the strength of the connection. Based on the architecture of ANN, the inputs of neurons in the current layer are transferred to the connected neurons in the next layer. For this purpose, the input of each neuron is multiplied by the weight of the connection and summed with the other neurons in the same layer. Then, the summation is passed to the transfer function to calculate the output which is recognized as the input of the neuron in the next layer. This procedure is explained mathematically as follows:

$$S_{ui} = \sum (x_iW_i)$$  \hspace{1cm} (1)

$$x_j = y_i = f(S_{ui})$$  \hspace{1cm} (2)
where \( i \) is the current layer, \( j \) is the next layer, \( x_i \) is the input of the neuron in the current layer, \( W_i \) is the weight of the connection between the neurons, \( y_j \) is the output of the current layer and \( f(x) \) is the transfer function. A linear or non-linear function can be selected as the transfer function based on the physics of the system.

Taking this network structure into account, the mapping procedure utilizes a pattern which relates the inputs to the outputs. The procedure is as follows:

- The available inputs and corresponding outputs are employed to train the network and define the relation between the inputs and the outputs.
- The trained network is tested to have an acceptable level of accuracy for predicting the outputs.

After training and testing stages, the constructed network can be applied to calculate the outputs of the inputs which the network were not exposed to during the training stage. Therefore, the complex interaction in the system can be modeled employing the trained network. This trained network is used as a tool to locate the source of release during an incident. Hence, the effort of this research is to develop a methodology based on the application of ANN to identify a contaminant source location inside an enclosed environment knowing the instant contaminant concentration in limited locations. The procedure can be classified as follows:

1. Collecting measured data presenting the contaminant dispersion inside an enclosed space to visualize its distribution inside the domain.
2. Applying a reliable airflow model to simulate the component dispersion inside the enclosed space and extracting the influential parameters in contaminant dispersion.
3. Performing adequate simulations based on the validated baseline case from the experiment to capture the impact of the influential parameters on the contaminant dispersion and preparing the required database for training the ANN.
4. Develop and train an ANN with suitable structure (specified inputs, outputs and hidden layers) to map the pattern relating the inputs to the contaminant source location as the outputs, and finally
5. Examine the accuracy of the trained ANN in contaminant source identification for situations with unseen inputs parameters.

3.2. Selection of the airflow model and the input parameters to train ANN

To prepare the required database for training the network, the application of a reliable airflow model is needed to simulate the distribution of a contaminant inside the investigated area. Once an airflow model is validated with experimental data, it can be reliably employed to predict contaminant distribution inside a domain with different initial and boundary conditions. The multi-zone model is a common airflow model employed in building simulations, which can simulate the airflow profile and contaminant distribution much faster than CFD. In this model, the whole building is divided into a number of zones with a uniform temperature and contaminant concentration. These zones are interconnected by airflow paths with user defined leakage characteristics [4]. CONTAM [20] is one of the most widely used and validated multi-zone software utilized to study the airflow and contaminant distribution inside various buildings (residential and commercial) with different ventilation strategies [4,5].

To determine a specie concentration in a zone, CONTAM solves the mass balance equation for that specific component. Thus, the equation consists of inflow of the compound, outflow of the compound, generation and removal of that component in a zone which is considered as a control volume.

Neglecting the possible adsorption or reaction of a component, the main driving forces in the distribution of a compound are its source parameters and the inter-zonal air movement. The airflow rate (inter-zonal air movement) is a function of pressure drop across the airflow path and mostly has been determined using the power law equation [20] as follows:

\[
F_{ij} = C \times (\Delta P)^n
\]

where \( C \) and \( n \) are constant and \( \Delta P \) is the pressure drop across the airflow path. In addition to the zone pressure, stack effect and wind pressure contribute to air movement between two zones. Therefore, the conditions of the following parameters are directly/indirectly affecting the distribution of a contaminant inside an enclosed space: wind direction and velocity, outdoor and indoor air conditions, type of ventilation and its operating schedule, removal efficiency of the filter procedure in the mechanical ventilation system and the type of opening and the cracks between the zones and in the building envelope. For a given building some of these data such as the filter efficiency and building envelope characteristics can be measured experimentally and other information about the indoor and outdoor air can be monitored. Hence, the mentioned parameters, additional to the sensor’s reading, are considered as the inputs of the network to locate the source while the output of the network is the location of the source.

4. Results and discussion

4.1. Case study

The application of the proposed method is studied inside an apartment as a possible contaminated space. The selected target building is OPTIBAT which is a real-scale experimental chamber replica of an apartment within a building located in the laboratory hall of the National Institute of Applied Science in Lyon [1]. Fig. 1 shows the OPTIBAT experimental facility as it was presented to CONTAM for airflow and contaminant distribution analysis [4]. Two facades of this flat are guarded by two climatic chambers (Fig. 1) to have a controlled environment. The temperature in these two chambers can be varied between –10 °C and 30 °C while the relative humidity may be set from 30% to 80%. Moreover, the pressure drop between these two façades can reach up to 200 Pa, which is equivalent to a wind velocity of 70 km/h [1]. The other facades are kept in the same conditions as indoors. Haghighat and Li [5] simulated the inter-zonal airflow inside OPTIBAT and validated their simulation with measured data. The measurement was done with fan pressurization test, and for both summer and winter conditions, which were imposed by the climatic chambers. This study adopts their simulation for summer conditions as the validated baseline to create the required database to train the ANN for source identification purpose. The indoor and outdoor conditions, the boundary conditions and the airflow coefficient of the inter-zonal airflow paths were obtained from Haghighat and Li [5]. Fig. 2 shows the airflow profile inside the apartment and the magnitude of the net airflow rate between the zones.

As mentioned earlier, this building was selected as the plausible area, which is suddenly contaminated by a toxic component. CONTAM simulation of this building was employed to calculate the contaminant concentration in sensors’ locations, which set as the inputs of the ANN. It needs to be mentioned that simulation is based on constant environmental conditions, and it is assumed that the airflow profile is in a steady-state condition which, also, can be seen in an interior zone of a mechanically ventilated building.
Moreover, this study assumed that background concentration of the contaminant is zero inside the zones and there is only one source of a contaminant which suddenly releases a determined amount of the contaminant inside the building. In addition, the source is modeled as a burst of the contaminant and there is no continuous generation of the component during the simulation.

4.2. Establishing and training ANN

4.2.1. Inputs and outputs of the network

According to the objective of this study, the ANN needs to predict the location of a source of a contaminant inside a building. Obviously, the output of the network is the zone which contains the source of the contaminant. For the inputs, they have to be the parameters which are available during the incident, and they directly/indirectly affect the distribution of the contaminant or visualize the dispersion profile of the released component. In this study, the only variables are the sensor’s readings concluded from variable source parameters. Indeed, the influential parameters on the airflow profile such as wind condition are not presented to the network as the inputs due to the steady-state condition that the airflow profile has in this building with the constant boundary conditions.

4.2.2. Training data preparation

To predict the source location, the ANN needs to be trained by available data to learn the relation between the sensors’ reading and the source location. These data can be generated using the validated baseline simulation as a tool to determine the contaminant concentration in sensors’ locations. Indeed, instead of measurements, the modeling was applied to calculate the concentration of the contaminant in sensors’ locations for different
source parameters; location, release time and release strength. In training session, the sensors' readings are determined for known source parameters, and they are set as the inputs of the network, while the corresponding source locations are set as the outputs. These inputs and outputs train the network to predict the unknown source location in a real incident utilizing the instant sensors' readings. The prediction ability of the network highly depends on the quality of the training data.

In this study, the contaminant concentration for the given building in all sensors' zones is a function of the source intensity, location and the release time. Therefore, by changing these three parameters, various combinations of contaminant concentration in sensor locations can be obtained. Indeed, the concentration in a zone can be presented as the following function:

\[
Conc_k = f(M, Z, \Delta t)
\]

where \(Conc_k\) is the simulated concentration at the sensor's location, \(M\) is the released mass of the contaminant in the source-zone, \(Z\) is the source-zone and \(\Delta t\) is the time duration between the release and sensor detection. The combinations of these parameters are given to the CONTAM model as inputs to predict contaminant concentrations in the sensors' locations for training the network. As a matter of fact, to improve the performance of the network the prepared data for the training section must include the effect of all these three parameters. For this purpose, all six internal zones in the dwelling are considered as the possible location of the source (\(Z\)) and utilized as source-zone during the training session. For released mass and release time, two series of magnitudes are considered; \([0, 10, 20 ... 100]\) units and \([5, 10, 15, 20, 25]\) minute for \(M\) and \(\Delta t\), respectively. The injected mass of the contaminant is selected as a percentage of the maximum potential mass of the studied contaminant which can be released in real scenario, and it is introduced to the network as \(0-100\) units. Therefore, the network is trained for all potential amounts of the released contaminant from the source.

For simulation purposes, the source was assumed to be a gaseous contaminant. The combinations of the concentrations in sensors' locations and the corresponding source-zone built up the inputs and the outputs required for training section. Indeed, the forward simulation results are employed to train the network inversely to map the relation between the contaminant distribution and the source location. Fig. 3 presents the whole procedure of the ANN training stage.

4.2.3. The ANN structure

Regarding the inputs and the outputs of the network, the number of neurons on the input layer is equal to the number of the sensors. Also, the output layer has only one neuron which gives the contaminant source location. In between, there are two hidden layers with 20 and 30 neurons to bridge the inputs to the output layer. As discussed before, the input of each neuron is multiplied by the weight which provides the connection between two neurons and then summed with the output of the other neurons on the same layer. This summation has to be transferred to the neuron on the next layer by utilizing a transfer function. Here, in this network, a hyperbolic tangent sigmoid function is assigned as the transfer function to the intermediate layers and linear function for the output layer.

During the training session, the ANN learns to map the relations between inputs and outputs. This procedure effectively trains the network to predict the results when it is faced with new input data of interest. "Back-propagation" is a well known supervised training algorithm consisting of two steps [15]. In the first step, the inputs are propagated forward to calculate the outputs and error of the network. Weights, biases and transfer functions perform the major roll in this forward section. In the second step, the error is propagated backward to change the weights and biases yield to appropriate magnitudes of the weights to achieve the least error.

---

1 This interval can be modified according to a sensor’s reading interval.
Gradient descent is the approach adopted to get to the least error in back-propagation learning technique [3]. Essentially, the trained network must have generalization capability to foresee the results of unseen inputs and avoid the over-fitting problem. “Early stopping” is a common generalization method adopted in training procedure [11]. Herein, the early stopping training is implemented in conjunction with a Levenberg–Marquardt back-propagation training function in MATLAB. It is a method to check the generalization of the network. In this technique, the training database is categorized in three subsets; training subset, validation subset and testing subset [11]. While the primary subset is used to determine the best weights and biases, the validation subset is monitored by training algorithm to avoid over-fitting problem. When the error associated with this subset is increasing continuously the training is stopped and the weights and biases of the minimum validation error are re-adopted [11]. The error associated with the testing subset is monitored for further check on the generalization capability of the network. However, this error does not have any impact on the training procedure. In this study 70% of the available data is randomly assigned to the training subset. Validation and testing subsets consist of 20% and 10% of the rest of the data, respectively.

4.3. ANN prediction

The OPTIBAT building is considered as the possible target building and its air flow simulation is utilized as the baseline to create the training database. Although the locations of the sensors are crucial in fast detection of a contaminant, this study does not plan to find the optimum location of the sensors. Its aim is to develop a method to find the source of release without knowing its parameters (location, time and strength) and validate the applicability of this procedure. Therefore, it was assumed that the building is monitored by three sensors located in zone-1, zone-4 and zone-6, respectively. Indeed, the selection of the sensors’ locations is reasonable regarding the air flow movement in this building. The air is infiltrated through the first three zones, moves to zone-4 and finally ex-infiltrated from the last two zones. Thus, the selected configuration of the sensors’ network has the ability to provide information about the contaminant concentration. Following the training procedure presented in Fig. 3, the contaminant concentrations in sensors’ locations were determined and set as the inputs to the ANN while the corresponding source location was set as the output of the network for training purpose. The convergence of the training algorithm was checked based on two criteria. The first one is the training performance evaluated by the mean square error (MSE) associated with the prediction of the network for the data categorized as the training subset. When the MSE stabilized over

| Case number | Source-zone | Released mass [unit] | Dt [min] | Prediction of ANN | Predicted zone | Prediction relative error [%] |
|-------------|-------------|----------------------|---------|-------------------|---------------|-----------------------------|
| 1 | 3 | 97 | 20 | 3.05 | 3 | 1.73 |
| 2 | 6 | 32 | 10 | 5.97 | 6 | 0.47 |
| 3 | 6 | 26 | 25 | 6.05 | 6 | 0.86 |
| 4 | 3 | 72 | 20 | 2.93 | 3 | 2.37 |
| 5 | 4 | 12 | 15 | 3.02 | 3 | 24.52 |
| 6 | 6 | 59 | 10 | 6.00 | 6 | 0.06 |
| 7 | 2 | 68 | 20 | 1.99 | 2 | 0.57 |
| 8 | 5 | 63 | 25 | 5.04 | 5 | 0.71 |
| 9 | 2 | 86 | 10 | 1.95 | 2 | 2.31 |
| 10 | 1 | 8 | 5 | 0.92 | 1 | 8.28 |
| 11 | 1 | 27 | 25 | 0.96 | 1 | 3.94 |
| 12 | 4 | 62 | 15 | 3.96 | 4 | 0.99 |
| 13 | 1 | 64 | 15 | 1.03 | 1 | 3.26 |
| 14 | 2 | 83 | 15 | 1.98 | 2 | 2.81 |
| 15 | 2 | 1 | 25 | 3.29 | 3 | 64.48 |
| 16 | 4 | 92 | 20 | 3.99 | 4 | 0.26 |
| 17 | 2 | 48 | 5 | 2.12 | 2 | 5.92 |
| 18 | 5 | 78 | 5 | 5.00 | 5 | 0.05 |
| 19 | 3 | 14 | 15 | 2.39 | 3 | 20.26 |
| 20 | 4 | 53 | 5 | 3.96 | 4 | 1.13 |
| 21 | 1 | 93 | 5 | 0.98 | 1 | 1.51 |
| 22 | 2 | 96 | 5 | 2.39 | 2 | 19.55 |
| 23 | 2 | 13 | 10 | 2.35 | 2 | 17.40 |
| 24 | 4 | 26 | 5 | 4.85 | 5 | 21.28 |
| 25 | 3 | 19 | 20 | 2.52 | 3 | 16.14 |
| 26 | 6 | 41 | 25 | 5.97 | 6 | 0.54 |
| 27 | 2 | 74 | 25 | 1.91 | 2 | 4.30 |
| 28 | 5 | 2 | 25 | 5.05 | 5 | 1.07 |
| 29 | 4 | 17 | 25 | 3.84 | 4 | 4.13 |
| 30 | 4 | 38 | 10 | 3.92 | 4 | 1.97 |
| 31 | 5 | 84 | 15 | 5.04 | 5 | 0.84 |
| 32 | 1 | 34 | 20 | 0.98 | 1 | 1.55 |
| 33 | 3 | 97 | 10 | 3.31 | 3 | 10.24 |
| 34 | 5 | 8 | 15 | 4.82 | 5 | 3.56 |
| 35 | 4 | 46 | 15 | 3.90 | 4 | 2.46 |
| 36 | 5 | 61 | 20 | 5.02 | 5 | 0.35 |
| 37 | 2 | 93 | 25 | 2.17 | 2 | 8.73 |
| 38 | 1 | 44 | 25 | 1.01 | 1 | 0.76 |
| 39 | 5 | 48 | 20 | 5.00 | 5 | 0.08 |
| 40 | 5 | 57 | 10 | 4.99 | 5 | 0.29 |

A. Bastani et al. / Building and Environment 51 (2012) 320–329

Fig. 3. Best Linear Fit: $Y = 0.9723X + 0.109$.

Fig. 4. Training performance graph for three sensors.

Table 1

The source parameters of the extra cases and their prediction by the trained ANN.

Fig. 5. The ANN prediction for the extra cases (with sensors in zones 1,4,6).
certain iterations (epochs), the algorithm was considered to reach its convergence. The second criterion was imposed by the training function to satisfy the generalization of the trained network. As mentioned earlier, the "early stopping" was adopted in this study and it was obliged by the algorithm when the MSE of the validation subset did not reduce over 100 epochs. Fig. 4 shows the MSE graph of three subsets utilized for training the ANN with three sensors inside the building. The MSE of the training subset dropped less than 0.04 and the MSE of the validation subset did not reduce than 0.08 over 100 epochs. Also, the validation and testing curves were very similar which one can conclude that there was no possibility of overfitting. The training process converged in 248 epochs and it took one minute on a dual-processor Pentium IV workstation (Dell Precision 380) with CPU 3.4 GHz.

To evaluate the capability of the trained ANN to locate an unseen contaminant source, 40 extra different scenarios are generated with different release mass than the ones employed in training procedure. Different source strength creates different concentration in sensors' locations and consequently different input parameters of the ANN. Therefore, the accuracy of the prediction of those extra cases implies the functionality of the method in real applications. These 40 extra cases are considered as the possible incidents occurred inside the building. In those incidents, the source parameters are assumed to be unknown and the only available information is the sensors' readings when one of them is alarming the situation with over threshold limit concentration. However, these source parameters are given to the CONTAM model to determine the sensors' readings in those incidents. The released masses and the corresponding release time and source-zone are selected randomly for those extra scenarios, and the contaminant concentrations in the sensor zones are calculated at the corresponding release time. Afterward, these concentrations are given to the trained ANN and the prediction of the network is compared with the actual source location. Fig. 5 presents the ANN prediction of the extra cases. Also, the source parameters and the relative error of the prediction for those cases are tabulated in Table 1.

During the training and testing procedure, the source-zone was introduced to the network as the nominal source number \( C_0 \). Indeed, the multi-zone model assumes that the air is fully mixed and the whole zone has a uniform concentration. Therefore, the area of each zone was normalized as one unit and the range of \( C_0 \) was considered for each zone to cover total area of the zone containing the contaminant source. Regarding this definition, the trained ANN could detect the source of contaminant in 90% of the unseen cases (36 out of 40). Moreover, the mean average error of the prediction was 8.49%, the median of the errors was 2.82% and the first and third quartiles were 0.90% and 8.83%, respectively. Obviously, the trained network had a significant accuracy in predicting the source location without knowing the source parameters and the history of contaminant concentration from the release time. The main deviation from the accurate source detection was seen for the cases with low released mass magnitude. Cases 5, 15, 19 had source strength less than 15 unit and the network could not correctly predict the source-zone. This failing attributes to the inaccuracy associated with ANN to predict the target which its inputs have magnitude near the constraints of their range. ANN can only be used for interpolation and its accuracy is negatively affected when the magnitude of the inputs is near the lower bound of its range. Generating contaminant with low strength results in a low concentration level in sensors' locations and may create a significant discrepancy in interpolating to the target value. Nevertheless, the sources located in zone-5 were all identified correctly even with low released mass (case 28 and case 34). The contaminant generated by the sources located in this zone can only be seen by the sensor located in zone-6 and the first two sensors show zero

![Fig. 6. The prediction error of the ANN for different number of sensors.](image)

### Table 2

| Situation | Number of sensors | Sensors' zones      | Prediction mean average error [%] | Number of cases with detected source |
|-----------|-------------------|---------------------|----------------------------------|--------------------------------------|
| 1         | 6                 | 1,4,6,3,2,5         | 2.72                             | 39 out of 40                          |
| 2         | 5                 | 1,4,6,3,2           | 7.52                             | 38 out of 40                          |
| 3         | 4                 | 1,4,6,3             | 7.86                             | 38 out of 40                          |
| 4         | 3                 | 1,4,6               | 8.49                             | 36 out of 40                          |
| 5         | 2                 | 1,4                 | 14.17                            | 30 out of 40                          |
concentration. Therefore, the network learns when the magnitude of the first two inputs is zero, the source may be in zone-5 or zone-6 and the system has less complexity to be mapped. Overall, the method could find the source location with a good accuracy and regardless of its location whether it was located in sensors’ zone or in adjacent zones.

Furthermore, the effect of the number of sensors was evaluated on the accuracy of the ANN prediction. Indeed, the number of sensors is equal to the number of inputs to the ANN and the higher number of the sensors results in more input information to the ANN. However, these sensors are very bulky and expensive and utilizing the minimum number of them would highly be desired, while the occupants' safety should not be sacrificed. Therefore, to quantify the effect of sensor’s number on the prediction, the ANN was trained and evaluated for five situations which consider the building monitored with two to six sensors. Moreover, the trained ANN was evaluated for source identification for the same 40 extra cases presented in Table 1. Table 2 shows those five situations with their mean average error of source prediction and the number of cases in which the source was located correctly. It was observed that the mean average error was reduced as more sensors were installed. Moreover, similar to the case with three sensors, the source in case 15 could not be identified in other situations which attributes to the inaccuracy associated with ANN to predict the target which its inputs have magnitude near the lower bound of its range.

To analyze the distribution of the prediction error, Fig. 6 presents the median, the first quartile and third quartile of the error of 40 extra cases for all of these five situations. Except the situation with the two sensors, the median of the other situations was less than 4%. Also, by increasing the number of the sensors, the density of the prediction error near the median was increased. The method has the least accuracy in detecting the source with two sensors. Although, the network could locate the source in 30 out of 40 cases with two sensors, in 10 cases it had a prediction error of more than 17%. Overall, the method presented a good performance in locating a source of contaminant without having prior knowledge of the source parameters. Moreover, its application can be utilized to evaluating the effect of the number of the sensors which can help building designers decide the optimum configuration of sensors required for a space based on the accuracy level of the source detection without sacrificing occupants' safety.

5. Conclusions

Identification of a contaminant source inside an enclosed space is critical to protect the occupants’ health and safety against an accidental or intentional chemical agent release. Finding the source location using the sensors’ reading is an inverse problem. However, the inverse relation between contaminant concentration in limited locations and the source location is not mathematically defined. Thus, the application of ANN was proposed to map this relation and its capability to find the contaminant source was evaluated for a simple scenario which deals with an incidental release of a contaminant in a one-story building. CONTAM multi-zone model was employed to analyze the airflow profile and contaminant dispersion inside that building. The validated model was set as the baseline to create the required data in this study. As the main effort in this research, an ANN was created and trained to detect a source of contaminant by using the contaminant concentration provided by the sensors. In training session, a database was created consisting of contaminant concentrations in sensors’ locations and the corresponding source locations. The attempt was to train a generalized ANN which can span the whole range of source strength, release time and release location of the source. Afterward, the prediction ability of the trained ANN for unseen scenarios was evaluated with 40 extra cases as the hypothetical incidents inside the building. This investigation brought up the following conclusions:

- The method was able to locate the source of contaminant in more than 90% of the tested cases. Either the source was located in the sensors’ zones or the adjacent zones, the method could detect it while the building was monitored by three or more sensors.
- The input parameters to the detection system were only the instant sensors’ readings which are always available in real scenario.
- The source location identification was made with unknown source parameters; source strength, and release time and location.
- The training of the ANN and its evaluation for 40 cases were carried out in around one minute, which is a reasonable time for fast detection in an emergency situation.

Moreover, the impact of sensors’ number on the performance of the network was studied. It was concluded that the higher number of sensors results in higher detection accuracy. Although there was a plan to find the optimum number of the sensors in this study, the method can evaluate the detection accuracy based on the number of employed sensors. Therefore, the method can help building managers to decide the optimum configuration of the sensors required for the investigated area.

Acknowledgments

The authors would like to express their gratitude to the Natural Sciences and Engineering Research Council of Canada for the financial support through the strategic grant programme.

References

[1] Amara F, Depecker P, Allard F. OPTIBAT: a real scale cell in simulated climatic environment for multizone air flow patterns in buildings. Coventry, Great Britain, Air Infiltration and Ventilation Center, Ventilation for energy efficiency and optimum indoor air quality: Proceedings 13th AIVC conference, Nice, France, 15–18 September, 1992.
[2] Bozorg M, Lesjak M, Makar P. A neural network-based method for the short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain. Atmos Environ B 1993;27:221–30.
[3] Braspenning PJ, Thijsman F, Weijsters AJMN. Artificial neural networks: an introduction to ANN theory and practice. Berlin, Heidelberg, New York: Springer-Verlag; 1995.
[4] Haghhighat F, Megri AC. A comprehensive validation of two airflow models – COMIS and CONTAM. Indoor Air 1996;6:278–88.
[5] Haghhighat F, Li H. Building airflow movement validation of three airflow models. J Archit Plan Res 2004;21:331–49.
[6] Hasham FA, Kindzierski WB, Stanley SJ. Modeling of hourly NOx concentrations using artificial neural networks. J Environ Eng Sci 2004;3:111–9.
[7] Liu X, Zhai Z. Inverse modeling methods for indoor airborne pollutant tracking: literature review and fundamentals. Indoor Air 2007;17:419–38.
[8] Liu X, Zhai Z. Location identification for indoor instantaneous point contaminant source by probability-based inverse Computational Fluid Dynamics modeling. Indoor Air 2008;18:2–11.
[9] Liu X, Zhai Z. Prompt tracking of indoor airborne contaminant source location with probability-based inverse multi-zone modeling. Build Environ 2009;44:1125–43.
[10] Magnier L, Haghhighat F. Multiobjective optimization of building design using TRNSYS simulation, genetic algorithm, and artificial neural network. Build Environ 2010;45:739–46.
[11] Matlab. R2010b math work documentation. Math Works Inc.; 2010.
[12] Neupauer RM, Wilson JL. Backward probabilistic model of groundwater contamination in non-uniform and transient flow. Adv Water Resour 2002;25:733–46.
[13] Neupauer RM, Wilson JL. Backward location and travel time probabilities for a decaying contaminant in an aquifer. J Contam Hydrol 2003;66:39–58.
[14] NIOSH. Guidance for protecting building environments from airborne chemical, biological, or radiological attacks. National Institute of Occupational Safety and Health, DHHS (NIOSH); 2002. publication No. 2002-2139.
[15] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by backpropagating errors. Nature 1986;323:533–6.

[16] Sohn MD, Reynolds P, Gadgil AJ, Sextro RG. Rapidly locating sources and predicting contaminant dispersion in buildings. In: Proceedings of the indoor air 2002 conference, Monterey, CA, vol. 4; 2002. p. 211–6.

[17] Sohn MD, Sextro RG, Gadgil AJ, Daisey JM. Responding to sudden pollutant releases in office buildings: 1. Framework and analysis tools. Indoor Air 2003; 13:267–76.

[18] Vukovic V, Srebric J. Application of neural networks trained with multizone models for fast detection of contaminant source position in buildings. ASHRAE Trans 2007;113:154–62.

[19] Vukovic V, Tabares-Velasco PC, Srebric J. Real-time identification of indoor pollutant source positions based on neural network locator of contaminant sources and optimized sensor networks. J Air Waste Manage Assoc 2010;60:1037–48. doi:10.3155/1047-3289.60.9.1034.

[20] Walton GR, Stuart W. NISTIR 7251 CONTAM 2.4c user guide and program documentation. National Institute of Standards and Technology, Building Environment Division; 2005.

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

[22] Zhang T, Chen Q. Identification of contaminant sources in enclosed spaces by a single sensor. Indoor Air 2007;17:439–49.

[23] Zhou L, Haghighat F. Optimization of ventilation system design and operation in office environment, Part I: Methodology. Build Environ 2009;44:651–6.

[24] Zhou L, Haghighat F. Optimization of ventilation systems in office environment, Part II: Results and discussions. Build Environ 2009;44:657–65.