

A Hybrid Sensor System for Indoor Air Quality Monitoring

Yun Xiang*, Ricardo Piedrahita[†], Robert P. Dick*, Michael Hannigan[†], Qin Lv[‡], Li Shang[§]

*EECS Department, University of Michigan, Ann Arbor, U.S.A.

[†]Dept. of Mechanical Engineering, University of Colorado Boulder, CO 80309, U.S.A.

[‡]Dept. of Computer Science, University of Colorado Boulder, CO 80309, U.S.A.

[§]Dept. of ECEE, University of Colorado Boulder, CO 80309, U.S.A.

{xiangyun, dickrp}@umich.edu, {ricardo.piedrahita, hannigan, qin.lv, li.shang}@colorado.edu

Abstract—Indoor air quality is important. It influences human productivity and health. Personal pollution exposure can be measured using stationary or mobile sensor networks, but each of these approaches has drawbacks. Stationary sensor network accuracy suffers because it is difficult to place a sensor in every location people might visit. In mobile sensor networks, accuracy and drift resistance are generally sacrificed for the sake of mobility and economy. We propose a hybrid sensor network architecture, which contains both stationary sensors (for accurate readings and calibration) and mobile sensors (for coverage).

Our technique uses indoor pollutant concentration prediction models to determine the structure of the hybrid sensor network. In this work, we have (1) developed a predictive model for pollutant concentration that minimizes prediction error; (2) developed algorithms for hybrid sensor network construction; and (3) deployed a sensor network to gather data on the airflow in a building, which are later used to evaluate the prediction model and hybrid sensor network synthesis algorithm. Our modeling technique reduces sensor network error by 40.4% on average relative to a technique that does not explicitly consider the inaccuracies of individual sensors. Our hybrid sensor network synthesis technique improves personal exposure measurement accuracy by 35.8% on average compared with a stationary sensor network architecture.

I. INTRODUCTION

Indoor air quality strongly influences the well-being of individuals. People spend more than 90% of their time indoors. Moreover, pollutant concentrations for some pollutants are usually much higher indoors than outdoors. Many indoor pollutants, such as volatile organic compound (VOC), carbon monoxide, and particulate matter, are closely related to chronic diseases, cancers, and human mortality [13], [27]. Other indoor pollutants, such as carbon dioxide (CO₂), can have significant impacts on the productivity, performance, and health [28], [31] of office workers and students. Therefore, accurate estimation of personal pollutant exposure is important.

Currently, stationary and accurate air quality monitoring sensors [30] are the primary tools for personal pollution exposure measurement. However, such sensors are usually too expensive for large-scale and fine-grained deployment. Moreover, stationary sensors generally require periodic calibration, which can significantly increase maintenance cost. Thus, they are usually scarce. In many cases, researchers and building managers have to rely upon data reported from sensors far away from the occupants to estimate personal exposure.

The problem with this approach is that indoor pollutant distribution can be spatially non-uniform. Indoor pollutant concentrations may vary significantly even within the same building, e.g., indoor VOC concentrations can differ by more than 7 times for different rooms in a same building [22]. Therefore, accurately estimating the indoor personal pollution exposure would require a sensor network composed of many stationary sensors, which would be prohibitively

expensive and thus impractical.

To address this problem, recent research has focused on inexpensive and mobile personal air quality sensor networks [17], [34]. Mobile sensor networks are composed of many low-cost, power-efficient, and miniature sensors carried by individuals. As sensor technologies and markets mature, the prices of such sensors are dropping. For example, a Figaro TGS2602 VOC sensor costs only about \$11. Mobile sensor networks can measure the environments immediately surrounding their users, thus eliminating the problems caused for stationary sensor networks by non-uniform spatial pollutant distributions.

A main drawback of mobile sensors is their susceptibility to drift error [29]. Drift is the gradual deviation of a sensor's readings from the correct value. It is affected by many factors that change the sensing surface, including material degradation, exposure to sulfur compounds or acids, aging, or condensate on the sensor surface [2], [14]. Mobile sensors are generally more susceptible to drift than stationary sensors due to tradeoffs made for compactness and economy. This problem is amplified because it is difficult to frequently calibrate mobile sensors, especially when they are carried by non-specialists. Given a fixed budget, one must trade off the (spatial variation) inaccuracies of stationary sensor networks with the (drift) inaccuracies of mobile sensor networks.

We propose a hybrid sensor network architecture composed of both accurate stationary sensors and inaccurate mobile sensors. Stationary sensors can provide accurate readings and more importantly, calibration opportunities for the mobile sensors. Mobile sensors carried by individuals can measure more relevant personal exposure data. Note that although our technique focuses on the hybrid sensor network architecture, it can also be used to design mobile-only or stationary-only sensor networks.

The purpose of this paper is to provide a comprehensive solution for hybrid air quality sensor network architecture analysis and design. Network performance analysis is challenging because it is difficult to predict actual concentrations given only readings from other locations and drift-influenced readings. The challenge for network synthesis is to maximize accuracy via sensor selection and allocation given a fixed budget. Our work addresses both analysis and synthesis problems.

This work makes the following contributions:

- 1) we formulate the problem of indoor pollutant concentration estimation and propose a Bayesian analysis based optimal solution taking sensor inaccuracies into account;
- 2) we describe algorithms for automatically designing hybrid sensor networks; and
- 3) we demonstrate how to use real-world CO₂ measurement data to estimate the airflow inside a building, and use these estimates to evaluate our analysis and synthesis techniques.

To the best of our knowledge, this is the first work addressing the problem of optimal concentration prediction with inaccurate sensors

and automated design for hybrid (mobile/stationary) air quality sensor networks.

The rest of the paper is organized as follows. Section II discusses previous related work. Section III provides a motivating example. Section IV describes models to predict the indoor pollutant concentration optimally and estimate the prediction error. Section V presents algorithms to select and allocate different types of sensors to minimize average sensor network error. Section VI describes our deployment and evaluation results.

II. RELATED WORK

This section summarizes the prior works on sensor network architecture, indoor environment modeling, and sensor noise reduction.

Sensor network architecture. Postolache et al. [26] described an ad hoc sensor network for indoor and outdoor air quality monitoring. Jiang et al. [17] described MAQS, a mobile environmental sensing network utilizing portable, indoor location tracking sensors. Common Sense [34], designed by Willett et al., tried to establish an environmental sensing network based on the response from communities. The placement problem of stationary sensors has also been well studied [3], [7]. Krause et al. [20] proposed a sensor placement algorithm based on sensing quality and communication cost prediction. In their approach, the sensor nodes are all stationary, while we consider both stationary and mobile nodes. Recently, Xiang et al. [29] proposed a mixed-integer linear programming based placement algorithm for stationary sensors in a hybrid sensor network. Our technique differs from previous work in that we consider both stationary and mobile sensors in our network design and exploit the cost and accuracy trade-off between them. Moreover, in contrast with prior work, we assume no prior knowledge of the types and quantities of sensors or the carriers of the mobile sensors. Instead of relying upon an established sensor network architecture, we describe how to construct hybrid sensor networks from scratch.

Indoor environment modeling. The single compartment mass balance-based model, developed by Hayes [15], [16], is widely used in modeling indoor pollutant distributions [11], [12], [23] and was validated using real-world measured data [8]. Liu et al. [21] gave a detailed description of the model and used it together with a probability-based adjoint inverse method to back-track indoor pollution sources. In this work, we build an extended model based on the mass balance-based model. In most prior work, it is assumed that the readings reported by the sensors are always accurate, and the mass balance model is mainly used to interpolate the pollutant concentrations at the locations without sensors. However, this assumption is not true in real-world applications using low-cost sensors. We extend the current model by considering and optimally compensating for the drift error.

Sensor noise reduction. One major problem for the low-cost sensors is their unreliable readings caused by long-term drift. To reduce the sensor noise, Tsujita et al. [32], [33] proposed using accurate stationary sensors to calibrate mobile sensors. Bychkovskiy et al. [6] proposed a two-phase post-deployment calibration technique. Miluzzo et al. [24] proposed an auto-calibration algorithm for mobile sensor networks. Elnahrawy et al. [10] described a sensor noise cleaning framework based on Bayes' theorem. In this work, we evaluate the impact of sensor noise to the synthesis and construction process of sensor networks. In contrast with prior work, our model incorporates indirect observations, i.e., concentration levels of adjacent locations, and thus improves accuracy of the network.

III. MOTIVATING EXAMPLE

This example describes the previously unanswered indoor pollutant concentration estimation and sensor network construction questions

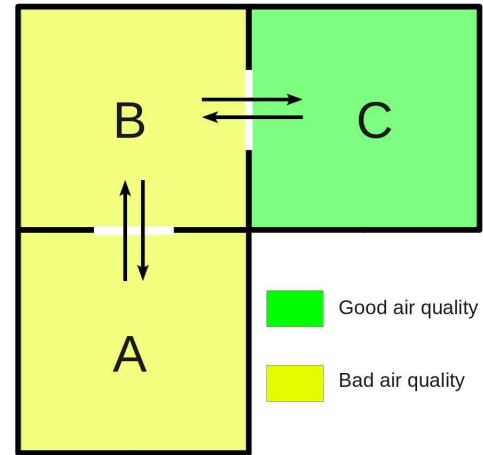


Fig. 1. Motivating example.

that motivate our work. The rest of this paper will provide answers to the questions appearing in this section.

Assume that a research team wants to deploy a small sensor network in the building shown in Figure 1. The building contains 3 rooms: A, B, and C. The rooms are connected and some have airflow between them. However, the air quality in each room differs. Assume that the budget is limited and the team can only afford one accurate sensor, which is placed in room A. The first question is, **“How should the pollutant concentrations in rooms B and C be predicted based on the reading in room A?”**

Then a somewhat inaccurate sensor is placed in room B. Suppose that one day the sensor reports a reading of 0.8 parts per million (PPM) pollutant concentration, while the estimates based on A's measurement suggests that the concentration in room B should be 0.5 PPM. The second question is, **“How can these two estimates be reconciled to minimize the error of the expected concentration?”**

Given a method of estimating pollutant concentrations, the problem of determining the numbers and types of mobile and stationary sensors remains. Subject to budget constraints, there are multiple options. In this case, one might deploy one stationary sensor and four mobile sensors, or two stationary sensors and two mobile sensors. The third question is, **“How should the numbers, types, and positions/carriers of sensors be determined to minimize the error of the measured personal pollutant exposure?”**

In this work, we aim to answer the three questions considered above. The first two questions led us to develop an optimal pollutant concentration prediction model based on analysis of indoor airflow and knowledge of pollutant source generation rate and sensor drift distributions. The third question led us to develop a hybrid sensor network synthesis algorithm that considers human mobility patterns and sensor costs.

IV. POLLUTANT CONCENTRATION PREDICTION MODELS

In this section, we describe the details of our optimal pollutant concentration prediction model. Section IV-A gives a problem definition. Section IV-B and Section IV-C introduce various concentration and error estimation models. Section IV-D describes the optimal model.

IV.A. Problem and Term Definitions

The deployment field, which is typically a building, is divided into multiple zones with inhabitants moving inside. Within the same

zone, the pollutant distribution is well-mixed and uniform. This can be achieved by subdividing zones when necessary. Depending on the pollutant type and ventilation conditions, a zone can be part of a room, an entire room, or multiple closely connected rooms.

A sensor network is deployed in a building so that a subset of the zones are covered, i.e., contain sensors. There are two potential causes of inaccurate concentration predictions. First, it is necessary to (imperfectly) estimate the pollutant concentrations of zones that are not covered. Second, sensor readings for covered zones may be inaccurate due to drift. We describe a model that takes into consideration both error sources and minimizes the expected value of prediction error.

We now define error. The error of the estimated concentration for zone i , denoted as e_i , is the difference between the predicted concentration and the ground truth. Since the estimation error is a random number, it can not be used to directly evaluate models. Therefore, we use expected error, which is the standard deviation of the distribution that e_i follows, as the evaluation criteria. The expected error is denoted as δ_i , and its relationship with estimation error e_i is

$$\delta_i = \text{std}(e_i). \quad (1)$$

Thus, an optimal pollutant concentration prediction is the concentration estimation with the minimal expected error.

The multi-zone pollutant concentration modeling problem can be defined as follows. Assume knowledge of the following deployment field information: inter- and intra-zone airflow, ventilation conditions, corresponding human motion patterns, pollutant source generation rates, and sensor drift information. A sensor network architecture, i.e., the types and quantities of the sensors, the locations of the stationary sensors, and the carriers of the mobile sensors, is deployed. Find a model to estimate the pollutant concentrations of all zones so that the average expected error is minimized.

IV.B. Concentration Prediction without Sensors

Assume that we want to evaluate the pollutant concentrations of all the zones in a building where no sensor is deployed. In general, the dynamic concentration change rate can be modeled using the following multi-zone pollutant transport equation [21], which is based on the single compartment mass balance-based model.

$$\begin{aligned} \frac{dC_i}{dt} &= \left[\sum_{j=1, \neq i}^n \left(\frac{F_{j,i}(1 - \eta_{j,i})}{Q_i} \cdot C_j \right) - \frac{\sum_{j=1, \neq i}^n F_{i,j}}{Q_i} \cdot C_j \right] \\ &+ \left[\frac{s_i}{Q_i} + \frac{F_{i,0} \cdot C_0}{Q_i} \right] \\ &= \sum_{k=1}^n a_{ik} \cdot C_k + B_i. \end{aligned} \quad (2)$$

The coefficients in Equation 2 are

$$a_{ik} = \begin{cases} -\frac{\sum_{j=1, \neq i}^n F_{i,j}}{Q_i} & (k = i) \\ \frac{F_{k,i}(1 - \eta_{k,i})}{Q_i} & (k \neq i) \end{cases} \quad \text{and} \quad (3)$$

$$B_i = \frac{s_i}{Q_i} + \frac{F_{i,0} \cdot C_0}{Q_i}, \quad (4)$$

where C_i is the concentration for zone i , C_0 is the outdoor concentration, n is the total number of zones, $F_{i,j}$ is the airflow rate from zone i to j , $F_{i,0}$ is the net airflow rate between zone i and outdoor environment, $\eta_{i,j}$ is the efficiency of the pollutant filters in the heating, ventilation, and air conditioning (HVAC) system, Q_i is the air volume in zone i , and s_i is the local pollutant source

generation rate. Note that the airflow rate $F_{i,j}$ is directional and $F_{i,j}$ is not necessarily equal to $F_{j,i}$. In our problem formulation, we neglect the kinetics among various pollutants and local removal rate. Those parameters can be easily incorporated into the model if the information is known.

Now consider a building with n zones. The estimated concentrations for all the zones in the building can be represented by a vector $C = [C_1, C_2, \dots, C_n]^T$. Thus, the pollutant transport function can be re-written as

$$\begin{aligned} \frac{dC}{dt} &= A \cdot C + B \quad \text{and} \quad (5) \\ A &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, B = [B_1, B_2, \dots, B_n]^T. \end{aligned} \quad (6)$$

In the rest of the paper, matrix A is referred to as the airflow matrix. This model is widely used and has been found to be accurate in real-world experiments [8].

For most of the pollutants, the health and/or performance impact is evaluated on a time scale varying from days to years. Moreover, if some pollutant is released and causes a sudden change in local source generation rates, the indoor environment can return to a well-mixed state quickly. For example, it takes about 80 minutes for a 238 m³ smoking lounge to become well-mixed [19]. Therefore, in personal exposure measurement applications, the dynamic variation in Equation 2 can be neglected [5], [11], leaving $\frac{dC_i}{dt} = 0$.

The equilibrium state equation for zone i can be described using the following equation.

$$\sum_{j=1}^n a_{ij} C_j + b_i s_i + k_i C_0 = 0, \quad (7)$$

where b_i equals $\frac{1}{Q_i}$ and u_i equals $\frac{F_{i,0}}{Q_i}$. In matrix form, we have

$$A \cdot C + B = 0. \quad (8)$$

If all the zones are in the well-mixed state, the pollutant concentration in any zone is a linear combination of the concentrations of other zones (including outdoor environment) and its own local source generation rate.

The airflow matrix A can be estimated using multiple methods. For example, Liu et al. [21] suggest that we can derive the airflow matrix by solving the corresponding computational fluid dynamics equations using tools such as CONTAM [25]. Another approach is to use the existing sensors, with the help of regression analysis, to estimate the airflow matrix. We will show in Section VI-A2 how to use a CO₂ sensor network to derive the average airflow matrix. The first approach does not require any existing sensor infrastructure. However, it is less accurate since it relies on the empirical estimation for parameters such as building leakages.

The inter-zone airflow may vary in time as the human behavior and ventilation conditions change, e.g., doors and windows opening and closing or changes in the state of the heating system. However, it is not necessary to derive multiple airflow matrices for all the scenarios. Since the concentration relationship between zones is linear, we can use a single averaged matrix as long as the equilibrium state assumption in Equation 7 holds.

To estimate the pollutant concentrations of uncovered zones, we need to estimate the source generation rates s_i . We assume that the source generation rates follow certain distributions with known mean values and standard deviations. The knowledge of the distributions can be obtained by analyzing the historical data or existing literature

for buildings with similar characteristics [4], [9], [22]. The error of the estimation can be captured and compensated for by sensors located in or near the zone.

Assume that the source generation rate distribution for zone i is

$$S_i = N(\mu_i, \sigma_i), \quad (9)$$

where N is the type of source generation rate distribution, which is assumed to be Gaussian, μ_i is its expected mean value, and σ_i is its standard deviation. For each zone, its actual generation rate is a random number s_i that follows distribution S_i .

The optimal generation rate prediction, for any uncovered zone i , is the mean value μ_i of its distribution. Thus, when there is no sensor deployed in the building, by solving Equation 7, the concentration of zone i can be estimated as

$$C_i = - \sum_{j=1}^n a'_{ij} (b_j \mu_j + k_j C_0), \quad (10)$$

where a'_{ij} is the element of the inverse matrix A^{-1} of the airflow matrix A .

Given that there are no sensors deployed, Equation 10 predicts pollutant concentration with minimal expected error. The ground truth concentration, denoted as ξ_i , can be calculated as

$$\xi_i = - \sum_{j=1}^n a'_{ij} (b_j s_j + k_j C_0), \quad (11)$$

where s_j is the ground truth source generation rate of zone j .

By its definition, the estimation error of zone i is the difference between the predicted concentration and ground truth and can be expressed as

$$e_i = C_i - \xi_i. \quad (12)$$

Note that e_i is a random number and its standard deviation is the expected error, δ_i .

By replacing C_i and t_i in Equation 12 with Equation 10 and Equation 11, the estimation error becomes

$$e_i = - \sum_{j=1}^n a'_{ij} \cdot b_j (\mu_j - s_j). \quad (13)$$

Note that the outdoor concentration C_0 can be measured by accurate stationary monitoring stations. Thus, it is accurate and does not cause any errors in Equation 13.

Since the term $\mu_j - s_j$ in Equation 13 is a random number that follows distribution $N(0, \sigma_i)$, we define the local generation rate vector H as

$$H = [b_1 h_1, b_2 h_2, \dots, b_n h_n]^T, \quad (14)$$

where h_i equals $\mu_i - s_i$. Assume that the estimation errors of all the zones are $e = [e_1, e_2, \dots, e_n]^T$. In the matrix form, the estimation errors can be calculated as

$$e = A^{-1} \cdot (-H). \quad (15)$$

Equation 10 gives the optimal pollutant concentration prediction with no sensors deployed. Equation 15 calculates the estimation errors for the prediction for all zones. As indicated in Equation 13, the estimation error is a random number which is the linear combination of the generation rates of all the zones.

Instead of predicting the pollutant concentration using the empirical concentration distribution of each zone directly, we estimate the distributions of source generation rates and use them to calculate the concentrations. We do this because unlike the source generation rates, the concentrations are highly correlated. For example, assume

we have two zones i and j , with estimation errors e_i and e_j respectively. The airflow between i and j is high. If there is an accurate sensor located in zone i , the prediction error in zone j , based on the observation on zone i , should decrease significantly. However, if we model their empirical pollutant concentration distributions independently, the estimation error in zone j remains the same, which greatly overestimates the error. By modeling the distributions of the independent source generation rates, one can avoid error overestimation resulting from ignoring correlations.

IV.C. Concentration Estimation with Sensors

In the previous discussion, we have derived the optimal concentration prediction model for a non-monitored building in Equation 10. Now we consider a scenario in which sensors are deployed. Specifically, we will evaluate how the deployment of sensors, both accurate and inaccurate, affects the concentration estimation accuracies for uncovered zones.

Assume that in zone i , a sensor is deployed. Thus, the predicted concentrations are

$$\begin{cases} C_i = r_i & i \in R \\ \sum_{j=1, j \notin R}^n a_{ij} C_j + \sum_{j \in R} r_j + b_i s_i + k_i C_0 = 0 & i \notin R, \end{cases} \quad (16)$$

where r_i is the reading of the sensor in zone i and R is a subset of the set of all zones Z and contains the zones that are covered by sensors. The relationship between the ground truth and the sensor reading is

$$d_i = r_i - \xi_i, \quad (17)$$

where d_i is the sensor reading error and is a random number following Gaussian distribution $N(0, \sigma_s)$, in which σ_s is the standard deviation. Note that this error, caused by sensor drift, is independent of source generation rates, and hence independent of the concentration prediction errors.

Thus, the airflow matrix A is

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & a_{ii} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nn} \end{bmatrix}. \quad (18)$$

In general, if a sensor is placed in zone i , all the elements $a_{ij, j \neq i}$ should be 0. The prediction error of zone i , instead of Equation 13, is calculated as

$$e_i = \begin{cases} d_i & i \in R \\ - \left(\sum_{j=1, j \notin R}^n a'_{ij} \cdot b_j h_j + \sum_{k \in R} a'_{ik} d_k \right) & i \notin R, \end{cases} \quad (19)$$

where each a'_{ij} is an element of the inverse, A^{-1} , of the modified airflow matrix. The source generation rate vector H is

$$H = [b_1 h_1, \dots, b_i h_i = d_i, \dots, b_n h_n]^T. \quad (20)$$

With the modified coefficients shown in Equations 18 and 20, Equation 15 is still valid. The equations described above can be applied to both stationary sensors and mobile sensors. The stationary sensors are typically more accurate. Thus, the standard deviation σ_s of sensor error should be much smaller for stationary sensors than mobile ones.

IV.D. Optimal Concentration Prediction Model

So far, we have derived concentration prediction and error estimation models with and without sensors. However, when the deployed sensor in a zone is inaccurate, the current solution is sub-optimal. In this section, we will discuss how to optimally predict concentrations with inaccurate sensors.

IV.D.1) Bayesian Analysis: For many types of low-cost mobile sensors, drift error eventually dominates the error of the predicted concentration based on empirical data. For example, the 4-month uncompensated drift error of Figaro TGS2602 VOC sensor is about 0.8 PPM on average [29], while the expected error of VOC in many environments is only around 0.3 PPM [9], [22]. Therefore, there is significant noise in the sensor readings, which greatly limits the overall accuracy of the sensor network.

To improve the concentration prediction accuracy, the sensor noise needs to be minimized. It is achieved by combining the known estimates, i.e., sensor readings and empirical source generation rate estimates, of the measured zone and its connected zones. Specifically, assume that for a certain zone, r is the observed reading from the noisy sensor and ξ is the ground truth concentration. According to Bayes' theorem, the following equation holds [10].

$$f(\xi|r) = \frac{f(r|\xi)g(\xi)}{\int_{v \in \Theta} f(r|v)g(v)dv}, \quad (21)$$

where f is the sampling distribution of r , g is the prior distribution, and Θ is the domain of g . To minimize the sensor error, we should pick the truth estimator ξ so that the probability that ξ is observed given sensor reading r is maximized. In other words, $f(\xi|r)$ should be maximized.

Therefore, the optimal concentration prediction $C_{optimal}$ can be derived using the following equation.

$$C_{optimal} = \arg \max_{\xi} f(r|\xi)g(\xi). \quad (22)$$

According to the problem definition, the sampling distribution f is the distribution of the sensor drift, while the prior distribution g is equal to the distribution of concentration prediction as described in Equation 16. It is shown in Section VI-A1 that both sensor drift and the prior distribution, which is the spatial pollutant distributions among zones, can be modeled using Gaussian distribution. In that case, Equation 22 have a closed form solution.

$$C_{optimal} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2} \mu_e + \frac{\sigma_e^2}{\sigma_s^2 + \sigma_e^2} r, \quad (23)$$

where σ_e and μ_e is standard deviation and mean of the estimated concentration assuming no sensor is present, σ_s is the standard deviation of sensor drift, and r is the sensor reading. Note that if either the sensor drift or the prior pollutant distribution does not follow Gaussian, Equation 23 may not be valid. We should refer to Equation 22 to find the optimal combination of the readings, and the results differ according to the detailed distributions.

IV.D.2) Optimal Prediction Model: We derive our predictive model based on Equation 23, which optimally balances the weightings of the inaccurate sensor readings and the similarly inaccurate source generation rate estimates to improve the overall prediction accuracy. Assuming that $w = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$, the estimation of the concentration for zone i can be described as

$$C_i = w_i \cdot C_{estimate} + (1 - w_i) \cdot C_{sensor} \\ = -\frac{w_i}{a_{ii}} \left(\sum_{j=1, j \neq i}^n a_{ij} C_j + b_i \mu_i + k_i C_0 \right) + (1 - w_i) r_i, \quad (24)$$

where $C_{estimate}$ is the estimated concentration for zone i assuming no sensor is located in that zone, C_{sensor} is the sensor reading for zone i , and w_i can be considered as assigned weight, which ranges from 0 to 1. If w_i equals 0, the sensor reading is considered accurate and hence determines the concentration for the zone. If w_i equals 1, it means there is no sensor located in the zone.

The ground truth concentration for zone i can be re-written as

$$\xi_i = -\frac{w_i}{a_{ii}} \left(\sum_{j=1, j \neq i}^n a_{ij} \cdot t_j + b_i s_i + k_i C_0 \right) + (1 - w_i) t_i. \quad (25)$$

Thus, the estimation error, defined as $C_i - \xi_i$, is

$$e_i = -\frac{w_i}{a_{ii}} \left(\sum_{j=1, j \neq i}^n a_{ij} \cdot e_j + b_i h_i \right) + (1 - w_i) d_i, \quad (26)$$

Therefore, the airflow matrix is

$$A = \begin{bmatrix} a_{11} & a_{12}w_1 & \cdots & a_{1n}w_1 \\ a_{21}w_2 & a_{22} & \cdots & a_{2n}w_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}w_n & a_{22}w_n & \cdots & a_{nn} \end{bmatrix}, \quad (27)$$

in which except for a_{ii} , each element in the i th row is multiplied by w_i .

By solving Equation 26, we have

$$e_i = -\sum_{j=1}^n a'_{ij} \cdot (w_j \cdot b_j h_j - (1 - w_j) \cdot a_{jj} d_j). \quad (28)$$

Thus, the local generation rate vector H is

$$H = \begin{bmatrix} w_1 \cdot b_1 h_1 - (1 - w_1) \cdot a_{11} d_1 \\ w_2 \cdot b_2 h_2 - (1 - w_2) \cdot a_{22} d_2 \\ \vdots \\ w_n \cdot b_n h_n - (1 - w_n) \cdot a_{nn} d_n \end{bmatrix}. \quad (29)$$

Equation 15 can be used to calculate the estimation errors of all the zones.

In general, Equation 24 gives the optimal concentration predictions. Equation 28 allows us to calculate the estimation error of the optimal predictor. Note that although we have presented equations for zones containing a single sensor, it is easy to extend the current solutions to cases where multiple sensors are co-located in a same zone.

V. HYBRID SENSOR NETWORK SYNTHESIS

In this section, we describe algorithms to solve the hybrid sensor network synthesis problem based on our optimal prediction model. Section V-A generalizes the problem and provides definitions. Section V-B discusses the reasoning and underlying observations for the synthesis algorithm. Section V-C describes the algorithm in detail.

V.A. Problem Definition

In a hybrid sensor network, there might be multiple types of sensors with varying accuracies, long-term drift rates, lifespans, and prices. Our work mainly focuses on the trade-off between accuracy and price. In other words, given the same budget, we want to minimize the personal exposure estimation error of the sensor network.

Note that the exposure error, denoted as E_i , is different from the estimation error e_i and expected error δ_i as defined in Section IV-A. In real world applications, we are interested in personal exposure rather than indoor concentrations. Thus, the value of a sensor should be determined both by its measurement accuracy and the number of people it serves. For example, if a sensor is placed in an isolated

zone with no people in it, even if its reading is accurate, it does not improve the quality of personal exposure measurement.

We define the exposure error for zone i as

$$E_i = \sum_{m=0}^k \delta_i(t_0 + m\Delta t) \cdot P_i(t_0 + m\Delta t) \cdot \Delta t, \quad (30)$$

where E_i is the exposure error, $\delta_i(t_0 + m\Delta t)$ is the expected error of zone i during time interval from $t_0 + m\Delta t$ to $t_0 + (m+1)\Delta t$, $P_i(t_0 + m\Delta t)$ is the number of people in zone i during the same time interval, Δt is a time interval during which the number of people and expected error of each zone are considered to be constant, and k is the total number of such time intervals in a day. Note that the expected error is a function of time because of the motion of sensor carriers. We also assume that the information of sensor readings is shared by everyone in the same zone, with or without sensors. Thus, their personal pollutant exposure measurement have the same expected error δ_i .

The problem of hybrid sensor network synthesis can be described as follows: given a certain budget b , the set of zones Z , the set of mobile sensor carrier candidates J , the set of sensor prices T , and assuming that the total cost of the deployed sensors is c , find a sensor network architecture, i.e., $z_{1,2,\dots} \in Z$ and $j_{1,2,\dots} \in J$ where sensors are deployed, so that the total exposure error, $\sum_{i=1}^n E_i$, is minimized under the constraint $c < b$. One could modify this definition if the accuracy were more important for some people than others, e.g., those with respiratory health problems.

V.B. Synthesis Overview

To construct a hybrid sensor network, we need to determine the types and quantities of sensors first. This problem is similar to the knapsack problem, in which we have a budget and a list of items. Each item has a weight and value, and we need to find the set of items that maximizes value while meeting a weight budget. If each type of sensor has a fixed value, i.e., amount of exposure error reduction, the problem is equivalent to the knapsack problem and hence NP-hard.

In our problem formulation, the exposure error improvement of each type of sensor is not fixed. It is dependent on the inter-zone airflow, sensor location, sensor drift distribution, source generation rate distribution, and the sensor architecture. For example, different placement locations for a sensor can lead to significantly different exposure error improvement results. Therefore, to determine the correct value of each sensor, we must perform sensor placement and allocation algorithms during the process of sensor selection. However, the sensor placement problem, even for the stationary sensors, is also NP-hard [29].

To address this problem, we rely on the observation that the price of the accurate stationary sensors is much higher than that of the inaccurate mobile sensors. For example, an accurate photo-ionization detector (PID) based VOC sensor may cost \$600, while a metal oxide VOC sensor costs only \$11. Even after considering the cost of all the peripheral components, the stationary sensors are still several times more expensive than the mobile sensors. Moreover, the stationary sensors need to be manually calibrated frequently, which increases maintenance cost.

Therefore, we decompose the synthesis problem into two sub-problems. The first sub-problem is the selection and placement of the stationary sensors, which we solve by exhaustively searching all the possible selection and placement schemes. There are two reasons for this design decision: (1) the high cost of the stationary sensors constraints the quantities that can be deployed in the sensor network and (2) stationary sensors can provide calibration opportunities for

the mobile sensors, thus help to improve the accuracy of the entire network.

The second sub-problem is the allocation and assignment of the mobile sensors. Because of the relatively large quantity of the mobile sensors, exhaustive search would be computationally intractable. We use a heuristic in which we choose one sensor per iteration based on its unit value. Unit value is defined as the exposure error reduction per unit cost. This is repeated until the budget is met. Note that the inaccurate and low-cost sensors can be used as stationary sensors. In that case, they are treated as mobile sensors each of whose trajectories are confined in a single zone.

V.C. Algorithm

Algorithm 1 Hybrid sensor network synthesis algorithm

```

Require:  $Z$  // set of rooms
Require:  $S_M$  // set of mobile sensors
Require:  $S_{ST}$  // set of stationary sensors
Require:  $J$  // set of mobile sensor carriers candidates
Require:  $U$  // set of source generation rate distributions
Require:  $D$  // set of sensor drift error distributions
Require:  $T$  // set of sensor prices
Require:  $M$  // set of mobility patterns of all the individuals
Require:  $b$  // budget
 $e_{min} = \infty$  // minimal personal exposure error
 $Y_{min} \leftarrow \{\}$  // sensor network architecture of  $e_{min}$ 
 $Y_{ST} \leftarrow \text{placement\_search}(S_M, b)$  //  $Y_{ST}$  is the set of all the possible stationary sensor placement schemes under current budget
 $\forall Y \in Y_{ST}, W(Y) \leftarrow \text{weight\_calculation}(Y, D, U)$  //  $W$  is the weight table
for  $Y \in Y_{ST}$  do
   $e_{pre} \leftarrow \text{error\_calculation}(Y, U, D, M, W(Y))$ 
   $Y_{pre} \leftarrow Y$ 
   $c \leftarrow \text{total\_cost}(Y, T)$ 
  while  $c < b$  do
     $\Delta e_{int} \leftarrow 0$ 
    for  $s \in S_M$  do
      for  $j \in J$  do
         $X \leftarrow Y_{pre} \cup (s, j)$ 
         $W(X) \leftarrow \text{weight\_calculation}(X, D, U)$ 
         $e_{cur} \leftarrow \text{error\_calculation}(X, U, D, M, W(X))$ 
         $\Delta e_{cur} = \frac{e_{pre} - e_{cur}}{T(s)}$ 
        if  $\Delta e_{cur} \geq \Delta e_{int}$  then
           $\Delta e_{int} \leftarrow \Delta e_{cur}$ 
           $e_{int} \leftarrow e_{cur}$ 
           $Y_{int} \leftarrow X$ 
        end if
      end for
    end for
     $e_{pre} \leftarrow e_{int}$ 
     $Y_{pre} \leftarrow Y_{int}$ 
     $c \leftarrow \text{total\_cost}(Y_{pre}, T)$ 
  end while
  if  $e_{pre} < e_{min}$  then
     $e_{min} \leftarrow e_{pre}$ 
     $Y_{min} \leftarrow Y_{pre}$ 
  end if
end for

```

The detailed algorithm is shown in Algorithm 1. The algorithm first searches all the possible assignments for stationary sensors within the budget limit. For each stationary sensor assignment, a greedy algorithm is used to assign mobile sensors. As long as the budget is not exceeded, the greedy algorithm tries to find the mobile sensor and the corresponding carrier so that the exposure error reduction per unit price, ΔE , is maximized. ΔE is defined as

$$\Delta E = \frac{E_{pre} - E_{cur}}{T(s)}, \quad (31)$$

where E_{pre} is the previous average exposure error before assigning the new sensor, E_{cur} is the current average exposure error after the assignment, and $T(s)$ is the price of the sensor to be assigned. When

the algorithm ends, it returns a sensor network architecture, i.e., a sensor selection and the location/carrier of each sensor, with the minimal exposure error that the algorithm can find within the budget limit.

VI. EXPERIMENTAL RESULTS

This section describes the evaluation of our model and synthesis algorithms. Section VI-A describes the experimental setup and the CO₂ experimental measurement for an office building. Section VI-B shows the evaluation results of our pollutant concentration predictor. Section VI-C presents the simulation results of our hybrid sensor network synthesis algorithm.

VI.A. Simulation Setup

The sensing error of each zone is determined by the distributions of sensor drifts and the distributions of the source generation rates of all the connected zones, as shown in Equation 26. Thus, real-world deployment, with each deployment only providing one sampling point of the error distribution, is insufficient to estimate the accuracy of the sensor network. Therefore, we use simulation with parameters derived from real-world measurements to evaluate our techniques.

In this section, we talk about the selection of the simulation parameters, including pollutant generation rate and sensor drift distributions, sensor prices, and human motion traces. We also describe the deployment of a CO₂ sensor network in a building, which is used to determine the inter-room airflow.

VI.A.1) Parameter Selections: The standard deviation of pollutant generation rate estimation is assumed to be around 0.3 PPM based on the indoor VOC concentration measurement of an industrial area building [22]. The data have passed the Lillie normality test. Therefore, we assume that the distributions of indoor source generation rates are Gaussian. We estimate the sensor drift error based on existing work [29]. The drift error of Figaro TGS2602 VOC sensors, after compensation, is about 0.24 PPM. The drift error data have also passed the Lillie normality test and hence its distribution is assumed to be Gaussian.

In our synthesis, the mobile sensor is modeled on Figaro TGS2602 VOC sensors, which cost about \$11 each. The stationary sensor is modeled on Baseline-MOCON VOC sensors. Its measurement error is limited by the resolution of the analog-to-digital converter interface, and is assumed to be 0.03 PPM. The cost of the accurate Baseline-MOCON sensor is about \$600. Both of these sensors require peripheral circuitry to gather and transmit data and perform proximity detection. The cost of such supporting circuit is about \$150 [17]. Thus, in this work, we assume that the total costs of the mobile and stationary sensor nodes are \$150 and \$750, respectively.

Mobile sensors are automatically calibrated when in the same zone with a stationary sensor. We assume that once calibrated, the mobile sensors' drift can be removed. Note that our technique can be used even if the calibration is imperfect, i.e., some residual drift error remains after calibration.

To evaluate the sensor network performance and select appropriate mobile sensor carriers, human motion traces are needed. In this work, we generate motion traces using the human mobility model described by Kim et al. [18]. Their mobility model is based on a statistical survey of the existing literature and U.S. Bureau of Labor Statistics data. The details of the distributions and parameters can be found in the existing literature [18].

VI.A.2) A CO₂ Sensor Network Deployment and Analysis: To estimate the airflow in a building, we performed a field experiment in which eight air quality sensing platforms were distributed throughout

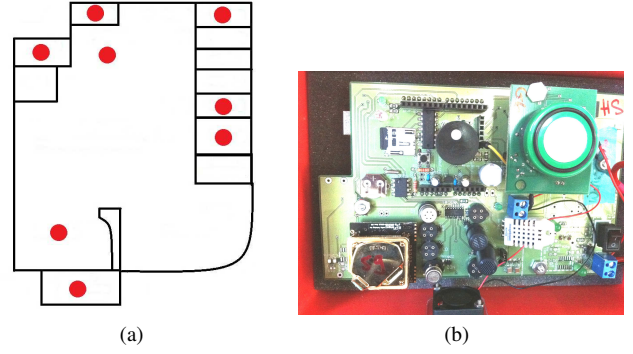


Fig. 2. Deployment environment and equipment: (a) building for deployment and (b) custom-built CO₂ measurement equipment.

an office building. The sensor nodes, as shown in Figure 2(b), are custom-built with a processor-communication architecture based on the Arduino platform [1]. The sensor nodes are equipped with multiple sensors, including the non-dispersive infrared S100 CO₂ sensor from ELT. This sensor has high accuracy, low drift, and low sensitivity to temperature and humidity. The CO₂ concentration is sampled at 0.2 Hz and is stored with a time stamp on a micro-SD card. A fan is used to pull air through the sensors at a constant rate of around 1 liter per minute.

Sensor calibrations were performed in a gas chamber before deployment. Gas mixtures in the chamber were precisely set using mass flow controllers operated through a LabVIEW control system. We performed the calibration at 3 different CO₂ levels: 0 PPM, 730 PPM, and 2,268 PPM. The exposure at each concentration level lasted 60 minutes.

Figure 2(a) shows the floorplan of the deployment building and sensor locations. The building is divided into eight zones, and contains room types such as single-occupancy office, large office with multiple occupants, and conference room. A sensor node was placed in each zone and collected data continuously from 8 June 2012 through 21 June 2012. The platforms were generally positioned near the locations frequented by room occupants, while trying to ensure they were far enough away to not be a nuisance, or be disturbed.

The measurement data from the deployment are used to derive the indoor airflow matrix A . The daily average concentration for zone i can be estimated as

$$(-a_{ii}) \begin{bmatrix} C_i(1) \\ \vdots \\ C_i(l) \end{bmatrix} = \sum_{j=1, j \neq i}^n a_{ij} \begin{bmatrix} C_j(1) \\ \vdots \\ C_j(l) \end{bmatrix} + \begin{bmatrix} G_i(1) \\ \vdots \\ G_i(l) \end{bmatrix}, \quad (32)$$

where $C_i(t)$ is the average concentration for day t , l is the total duration of the experiment, and $G_i(t)$ is a constant determined by the daily outdoor concentration and indoor generation rates. Linear regression analysis is applied to Equation 32 to estimate the airflow matrix A . The airflow matrix is later used in simulations to evaluate our concentration prediction and synthesis techniques in Section VI-B and Section VI-C.

VI.B. Concentration Prediction Model Evaluation

In this section we evaluate our pollutant concentration prediction models. Since the stationary sensors are accurate and hence always have fixed weights of 0, we do not include stationary sensors in this evaluation. We have randomly selected 5 carriers from the motion traces and varied the number of mobile sensors. Based on the resulting sensor network architectures, we apply different methods to

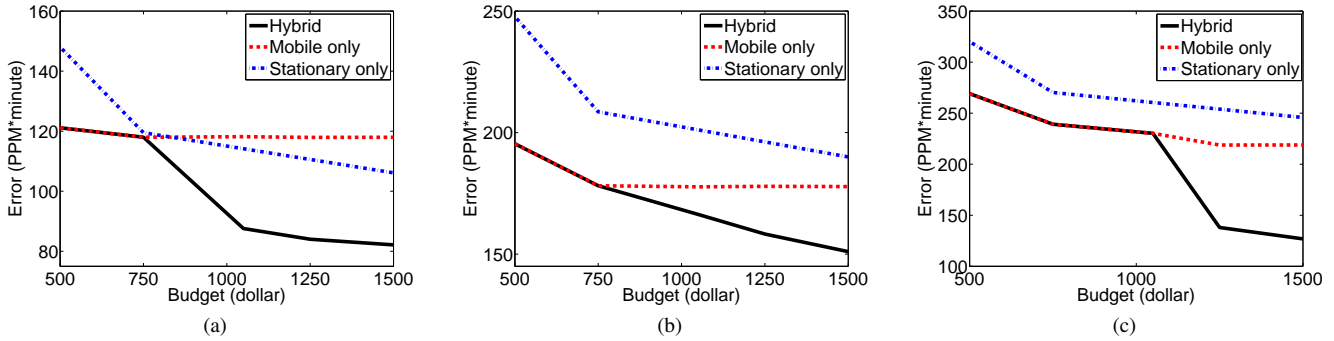


Fig. 5. The synthesis results for (a) small, (b) medium, and (c) large human motion traces.

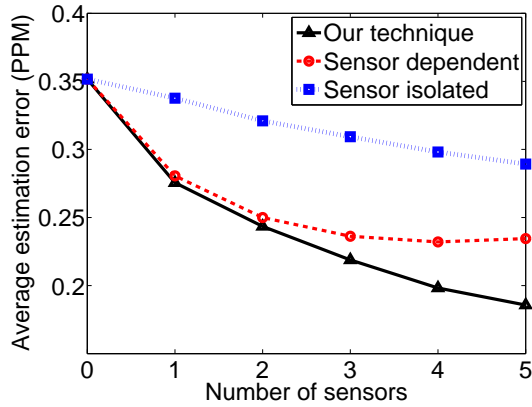


Fig. 3. The average error for different error estimation schemes.

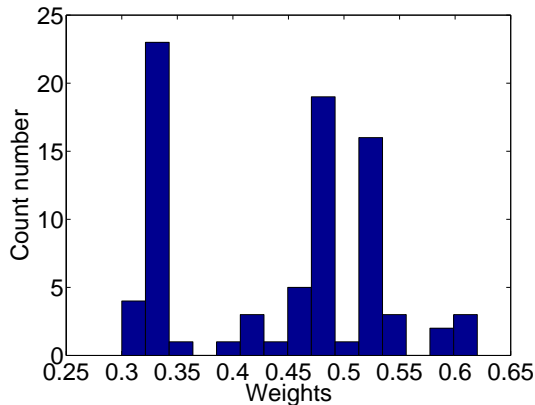


Fig. 4. The optimal sensor drift compensation weight distribution.

predict the pollutant concentrations of all the zones. During sensor network construction, the weights and average expected errors are recorded.

Figure 3 shows the expected errors of various concentration prediction schemes. The “sensor isolated” scheme assumes that a sensor’s readings are not used to aid in estimating concentrations in other zones. The “sensor dependent” scheme uses sensor readings to aid in estimates for distant zones; the prediction error is calculated based on Equation 19. In contrast to our technique, neither of the two schemes use weights to trade off position error and drift error. As a result, our technique improves the prediction accuracy by 40.4% on average compared with the “sensor isolated” method, and by 11.2% on average compared with the “sensor dependent” method. When there are 5 sensors deployed, the “sensor dependent” method incurs 26.3% more error compared with our optimal technique. The results

TABLE I
COMPARISON BETWEEN THE HEURISTIC AND OPTIMAL SOLUTION

Budget (\$)	Average error (PPM×minute)		Differences (%)
	Heuristic	Optimal	
750	108.93	108.93	0
900	88.72	88.72	0
1050	81.26	66.97	17.58
1200	73.32	62.24	15.11
1350	68.62	60.13	12.38
1500	59.00	59.00	0
Average			7.51

show that both indoor airflow modeling and weight adjustment are important.

Figure 4 shows the distribution of all the weights. The X axis gives the weight values and the Y axis gives the frequency of appearance. The wide spread of weights explains why our technique, which adjusts weights dynamically and optimally, outperforms other models with fixed weights.

VI.C. Hybrid Sensor Network Evaluation

We compare hybrid sensor network architecture accuracy against that of two other architectures. The first contains only mobile, inaccurate, low-cost sensors. The second contains only stationary, accurate, expensive sensors. All of the three approaches use the algorithm described in Algorithm 1 to construct the network.

Figure 5 presents the simulation results. The simulation is performed on small, medium, and large human motion traces. There are 20 individuals and 4 sensor carrier candidates in the small trace, 30 individuals and 6 sensor carrier candidates in the medium trace, and 40 individuals and 10 sensor carrier candidates in the large trace. When the budget is less than \$750, stationary sensors are unaffordable, thus the solution is the same for both the mobile-only and hybrid schemes. As the budget increases, the hybrid solution starts to outperform the other two solutions. Note that the stationary-only solution is optimal (but for a constrained problem definition), while the mobile only and hybrid solutions are heuristic due to the problem decomposition described in Section V-B.

When the budget is very limited, the mobile-only solution outperforms the stationary-only solution since no stationary sensor can be afforded. When we have a large enough budget, the stationary-only solution gives the most accurate measurement by placing an accurate sensor in every zone. The hybrid sensor network architecture, however, provides the best solution when the budget is between these extremes. In our simulation, when the budget is no less than \$750 (thus can afford at least one stationary sensor), the hybrid architecture improves the sensor network accuracy by 23.9% on average compared with the mobile-only architecture, and by 35.8% on average compared with the stationary-only architecture.

Even though our proposed algorithm can significantly improve the personal exposure measurement accuracy, it is not optimal. We compared the algorithm with the optimal solution for a small trace with 20 individuals and 5 carrier candidates (computational cost prevented us from finding optimal solutions for larger problem instances). The optimal solution was found using exhaustive search for both the stationary sensors and mobile sensors. The results are shown in Table I. In 3 of the 6 test cases, our heuristic returns the optimal solution. In the worst case, it has 17.58% more error. On average, our heuristic has about 7.5% error compared with optimal.

VII. CONCLUSION

We have described a synthesis and evaluation framework for hybrid sensor networks. This framework is composed of an optimal indoor concentration prediction and its error estimation model, and a hybrid sensor network synthesis algorithm. A field experiment was used to measure the inter-zone airflow. Compared with approaches that do not consider inter-zone airflow, our model improves accuracy by 40.4% on average by considering both the location-dependent and drift-dependent measurement errors. Simulations indicate that our hybrid sensor network architecture on average is 23.9% more accurate than the mobile-only architecture and 35.8% more accurate than the stationary-only architecture.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under awards CNS-0910816, CNS-0910995, and CNS-1059372.

REFERENCES

- [1] Arduino open-source electronics prototyping platform. <http://www.arduino.cc/>.
- [2] K. Arshak, E. Moore, G. M. Lyons, J. Harris, and S. Clifford. A review of gas sensors employed in electronic nose applications. *Sensor Review*, 24(2):181–198, 2004.
- [3] J. Berry, L. Fleischer, W. Hart, C. Phillips, and J. Watson. Sensor placement in municipal water networks. *J. Water Resources Planning and Management*, 131(3):237–243, 2005.
- [4] S. K. Brown, M. R. Sim, M. J. Abramson, and C. N. Gray. Concentrations of volatile organic compounds in indoor air – a review. *Indoor Air*, 4(2):123–134, 1994.
- [5] J. M. Burke, M. J. Zufall, and H. Ozkaynak. A population exposure model for particulate matter: case study results for PM_{2.5} in Philadelphia, PA. *J. Exposure Analysis and Environmental Epidemiology*, 11(6):470–489, 2001.
- [6] V. Bychkovskiy, S. Megerian, D. Estrin, and M. Potkonjak. A collaborative approach to in-place sensor calibration. In *Proc. Int. Conf. Information Processing in Sensor Networks*, pages 301–316, Apr. 2003.
- [7] K. Chakrabarty, S. Iyengar, H. Qi, and E. Cho. Grid coverage for surveillance and target location in distributed sensor networks. *IEEE Trans. Computers*, 51(22):1448–1453, Dec. 2002.
- [8] A. Chaloulakou and I. Mavroidis. Comparison of indoor and outdoor concentrations of CO at a public school. Evaluation of an indoor air quality model. *Atmospheric Environment*, 36(11):1769–1781, 2002.
- [9] L. E. Ekberg. Volatile organic compounds in office buildings. *Atmospheric Environment*, 28(22):3571–3575, 1994.
- [10] E. Elnahrawy and B. Nath. Cleaning and querying noisy sensors. In *Proc. Int. Conf. Wireless Sensor Networks and Applications*, pages 78–87, 2003.
- [11] P. G. Georgopoulos, S. S. Isukapalli, and K. Krishnan. *Modeling exposures to chemicals from multiple sources and routes*, pages 315–351. John Wiley & Sons, Ltd, 2010.
- [12] R. Goyal and M. Khare. Indoor air quality modeling for PM₁₀, PM_{2.5}, and PM_{1.0} in naturally ventilated classrooms of an urban indian school building. *Environmental Monitoring and Assessment*, 176:501–516, 2011.
- [13] H. Guo, S. C. Lee, L. Y. Chan, and W. M. Li. Risk assessment of exposure to volatile organic compounds in different indoor environments. *Environmental Research*, 94(1):57–66, 2004.
- [14] J.-E. Haugen, O. Tomic, and K. Kvaal. A calibration method for handling the temporal drift of solid state gas-sensors. *Analytica Chimica Acta*, 407(1–2):23–39, Feb. 2000.
- [15] S. R. Hayes. Estimating the effect of being indoors on total personal exposure to outdoor air pollution. *J. Air Pollution Control Association*, 39(11):1453–1461, 1989.
- [16] S. R. Hayes. Use of an indoor air quality model (IAQM) to estimate indoor ozone levels. *J. Air & Waste Management Association*, 41(2):161–170, 1991.
- [17] Y. Jiang, K. Li, L. Tian, R. Piedrahita, Y. Xiang, O. Mansata, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang. MAQS: A personalized mobile sensing system for indoor air quality monitoring. In *Proc. Int. Conf. Ubiquitous Computing*, pages 271–280, Sept. 2011.
- [18] J. Kim, V. Sridhara, and S. Bohacek. Realistic mobility simulation of urban mesh networks. *Ad Hoc Networks*, 7(2):411–430, 2009.
- [19] N. E. Klepeis. Validity of the uniform mixing assumption: determining human exposure to environmental tobacco smoke. *Environmental Health Perspectives*, 107(Suppl. 2):357–363, 1999.
- [20] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg. Near-optimal sensor placements: maximizing information while minimizing communication cost. In *Proc. Int. Conf. Information Processing in Sensor Networks*, pages 2–10, 2006.
- [21] X. Liu and Z. J. Zhai. Prompt tracking of indoor airborne contaminant source location with probability-based inverse multi-zone modeling. *Building and Environment*, 44(6):1135–1143, 2009.
- [22] M. R. Mannino and S. Orecchio. Polycyclic aromatic hydrocarbons (PAHs) in indoor dust matter of Palermo (Italy) area: extraction, GC–MS analysis, distribution and sources. *Atmospheric Environment*, 42(8):1801–1817, 2008.
- [23] S. Miller-Leiden, C. Lohascio, W. W. Nazaroff, and J. Macher. Effectiveness of in-room air filtration and dilution ventilation for tuberculosis infection control. *J. Air & Waste Management Association*, 46(9):869–882, 1996.
- [24] E. Miluzzo, N. D. Lane, A. T. Campbell, and R. Olfati-Saber. CaliBree: a self-calibration system for mobile sensor networks. In *Proc. Int. Conf. Distributed Computing in Sensor Networks*, pages 11–14, June 2008.
- [25] NIST. CONTAM: A multizone airflow and contaminant transport analysis software. <http://www.bfrl.nist.gov/IAQanalysis/CONTAM/index.htm>.
- [26] O. Postolache, J. Pereira, and P. Girao. Smart sensors network for air quality monitoring applications. *IEEE Trans. Instrumentation and Measurement*, 58(9):3253–3262, Sept. 2009.
- [27] A. Rabl and J. Spadaro. Public health impact of air pollution and implications for the energy system. *Annual Review of Energy and the Environment*, 25:601–628, 2000.
- [28] D. G. Shendell, R. Prill, W. J. Fisk, M. G. Apte, D. Blake, and D. Faulkner. Associations between classroom CO₂ concentrations and student attendance in Washington and Idaho. *Indoor Air*, 14(5):333–341, 2004.
- [29] Y. Xiang, L. S. Bai, R. Piedrahita, R. P. Dick, Q. Lv, M. P. Hannigan, and L. Shang. Collaborative calibration and sensor placement for mobile sensor networks. In *Proc. Int. Conf. Information Processing in Sensor Networks*, pages 73–84, Apr. 2012.
- [30] P. Tans and K. Thoning. How we measured background CO₂ levels on Mauna Loa. http://www.esrl.noaa.gov/gmd/ccgg/about/co2_measurements.html.
- [31] D. Tsai, J. Lin, and C. Chan. Office workers’ sick building syndrome and indoor carbon dioxide concentrations. *J. Occupational and Environmental Hygiene*, 9(5):345–351, 2012.
- [32] W. Tsujita, H. Ishida, and T. Moriizumi. Dynamic gas sensor network for air pollution monitoring and its auto-calibration. In *IEEE Proc. Sensors*, pages 56–59, Oct. 2004.
- [33] W. Tsujita, A. Yoshino, H. Ishida, and T. Moriizumi. Gas sensor network for air-pollution monitoring. *Sensors and Actuators, B, Chemical*, 110(2):304–311, Oct. 2005.
- [34] W. Willett, P. Aoki, N. Kumar, S. Subramanian, and A. Woodruff. Common sense community: scaffolding mobile sensing and analysis for novice users. In *Pervasive Computing*, volume 6030, pages 301–318.