Sensor Networks for Detecting Toxic Releases in Buildings

Michael D. Sohn[1,*], Ashok J. Gadgil[1],
Priya Sreedharan[1], William W. Nazaroff[1,2]

[1]Indoor Environment Department
Lawrence Berkeley National Laboratory
Berkeley, CA

[2]Civil and Environmental Engineering
University of California, Berkeley
Berkeley, CA

March 2009

This work was partly supported by the Office of Chemical Biological Countermeasures, of the Science and Technology Directorate of the Department of Homeland Security and by the Defense Threat Reduction Agency, and performed under U.S. Department of Energy Contract No. DE-AC02-705CH11231.

Sensor Networks for Detecting Toxic Releases in Buildings

Michael D. Sohn[1,*], Ashok J. Gadgil[1], Priya Sreedharan[1], and William W. Nazaroff[1,2]

[1]Indoor Environment Department, Lawrence Berkeley National Laboratory, USA
[2]Civil and Environmental Engineering, University of California, Berkeley, USA

**Corresponding author email: mdsohn@lbl.gov*

ABSTRACT

Sudden releases of a toxic agent indoors can cause immediate and long-term harm to occupants. In order to protect building occupants from such threats, it is necessary to have a robust air monitoring system that can detect, locate, and characterize accidental or deliberate toxic gas releases. However, developing such a system is complicated by several requirements, in particular the need to operate in real-time. This task is further complicated when monitoring sensors are prone to false positive and false negative readings. We report on work towards developing an indoor monitoring system that is robust even in the presence of poor quality sensor data. The algorithm, named BASSET, combines deterministic modeling and Bayesian statistics to join prior knowledge of the contaminant transport in the building with real-time sensor information. We evaluate BASSET across several data sets, which vary in sensor characteristics such as accuracy, response time, and trigger level. Our results suggest that optimal designs are not always intuitive. For example, a network comprised of slower but more accurate sensors may locate the contaminant source more quickly than a network with faster but less accurate sensors.

INTRODUCTION

Contaminant releases in or near a building can lead to significant human exposures unless prompt response is taken. However, selecting the best response depends in part on knowing the source locations, the amounts released, and the dispersion characteristics of the pollutant. We present an approach that estimates this information in real time. The approach, called BASSET, uses Bayesian statistics to interpret sensor measurements. The algorithm determines best estimates and uncertainties for the release conditions, including the operating state of the building. Because the method is fast, it can continuously update the estimates as measurements stream in from sensors.

We developed this Bayesian approach because traditional data interpretation and parameter estimation algorithms, such as optimization, Gibbs sampling, and Kalman filtering, often rely on assumptions that are not met in buildings, or depend on inverse modeling, which must repeatedly run computationally-intensive fate and transport models to incorporate new data being collected immediately after an event has occurred. Furthermore, these techniques are difficult to apply in the face of data errors, including the presence many false positive and false negatives readings, or when data are sparse (spatially and/or temporally).

The overall objective of this research was therefore to develop algorithms and software for indoor sensor data fusion. Specific goals included (1) real-time source characterization, (2) real-time hazard assessment, and (3) pre-event optimal sensor placement. In this paper, we elucidate the Bayesian algorithm for interpreting sensor data in real time, and demonstrate the approach with two examples. In the first example, we characterize a pollutant release in a hypothetical five-room building, comparing concurrent and sequential sampling of the sensor data. In the second example, we characterize a tracer gas release in a three-story building using trigger-type sensors, i.e., sensors that only report a “yes” or a “no” depending on a pre-set trigger level.

METHODS

We developed a Bayesian approach for comparing predictions from models to sensor data. To do so, we recognized a feature of Bayesian updating not previously reported in the literature: the ability to decouple modeling from data analysis. This feature allows us to perform real-time interpretation as data stream in during a pollutant release event (Figure 1).

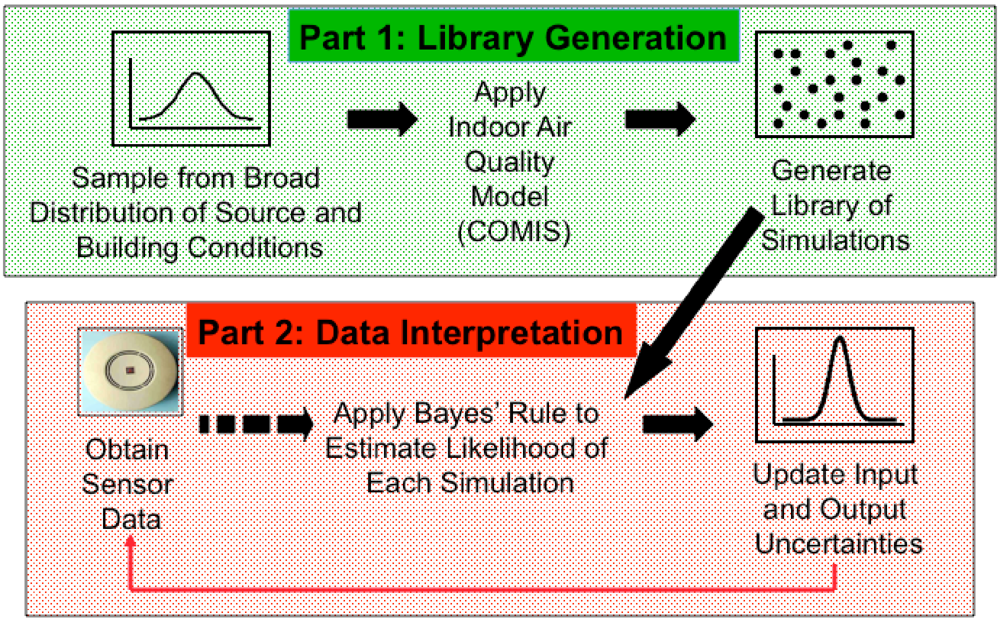

Figure 1: Illustration of Bayesian updating procedure.

Our variation of Bayesian updating, which is referred to in the literature as “Bayes Monte Carlo Updating,” comprises two stages. In the *Library Generation* stage, the practitioner develops a fate and transport model of the building, characterizes uncertainties in the model inputs, and simulates many hypothetical airflow and pollutant transport scenarios. These time-consuming tasks are completed before a pollutant release occurs. In the *Data Interpretation* stage, the algorithm evaluates the relative agreement between each model simulation and the incoming sensor data, using traditional Bayesian updating procedures. This second stage may be conducted as data stream in from the sensors, in real time.

We provide a brief introduction to the procedure here; we refer the reader to [1-3] for a more detailed discussion of the theory and application of this work.

Stage 1: Pre-event computations

Before a release event, the practitioner develops a model of the building's indoor airflow and pollutant transport. Best estimates for model inputs are generated from, for example, previous building characterization exercises, tracer gas flow experiments and modeling, published literature, and professional judgment. Any uncertain model parameter (e.g., effective leakage areas) or variable input (e.g., outdoor temperature and door positions) is assigned a probabilistic distribution of possible values. Release characteristics (e.g., the location, duration, and amount of pollutant released in an incident) also are assigned uncertainty distributions.

The practitioner next generates a library of model simulations, by repeatedly sampling the distributions of the model inputs, and predicting the airflow and pollutant transport for each resulting model. The final library may comprise many thousands of such simulations, or *realizations*. Each realization represents a single possible combination of building configuration, weather condition, and pollutant release scenario, and comprises time-sequences of predicted pollutant concentrations at multiple locations in the building.

Because this library will be used to assess sensor data during an actual event, it is important to (i) characterize the uncertainty and variability in the model inputs properly; and (ii) draw sufficient samples from the distributions. Artificially narrow uncertainty distributions may not cover an actual release. Similarly, insufficient sampling may miss a combination inherent in the original distributions. One method for ensuring sufficient sampling is to increase the sample size until summary statistics (e.g., means, variances, coefficients of variation) of the model predictions no longer change.

Stage 2: During-event data interpretation

During an actual release, the algorithm compares data streaming in from sensors to each realization in the library of model simulations. Each realization in the library is compared against the data to assess, quantitatively, the likelihood that the realization describes the event in progress. A realization agreeing well with the data is assigned a high likelihood. This in turn suggests that the model inputs used to generate that realization have high probability of describing the event in progress. By evaluating the relative fits for each realization to all available data, the Bayesian method estimates the model inputs and outputs, including uncertainties. Mathematically, we employ an empirical version of Bayesian updating to calculate the statistics. We refer the reader to [1] for detailed explanation of the calculations, but will discuss the important features, and potential pitfalls, of applying the technique here.

First, care must taken in designing a proper likelihood function. The function should define the error structure of the data, i.e., the difference between the data and the model predictions resulting from measurement error, from spatial and temporal averaging or correlations, and from imperfect model representation. For unbiased measurements with a (log-) normally distributed error structure, the function is Gaussian. Other types of sensor data, such as data from trigger sensors, will require an alternate likelihood function [1, 3].

With the likelihoods calculated, the algorithm applies Bayes' Rule to calculate posterior probabilities for the library (one for each realization). These posterior probabilities apply to all random variables in the library, and so can be used to revise/update uncertainties in all model parameters, variable inputs, release conditions (location, amount, strength), and so on. Furthermore, the posterior probabilities also apply to model projections, which allow us to predict, and revise, the future migration of the plume at the event unfolds.

This second stage of the approach is mathematically simple and can be executed very quickly-- typically much quicker than the rate at which new data arrive from sensors.

ILLUSTRATIVE EXAMPLES

We illustrate the approach with two demonstrations. In the first, we locate and characterize a hypothetical pollutant release in a five-room building [1]. We also compare concurrent and sequential sampling, and examine how noise in the sensor data degrades a sensor network's performance. In the second, we focus on the performance of a network consisting of trigger- or alarm-type sensors, rather than continuous-output devices [2, 3]. While both demonstrations are taken from recently published journal publications, our purpose here is to introduce these concepts as they may relate to building-ventilation research.

Demonstration one: Study of a five-room building

The study building is a single story structure comprising three rooms, a common area (CA), and a bathroom (Figure 2). The building does not have a ventilation system, and the status of one of the CA windows and the door between the CA and Room 3 is "unknown" to the data fusion algorithm (e.g., owing to failed position sensors at these locations). We used a multizone model to predict whole-building airflows and pollutant transport, treating each of the five rooms as a distinct well-mixed zone.

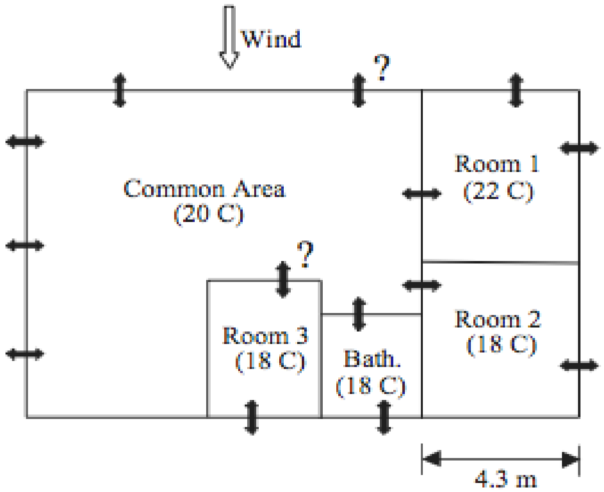

Figure 1: Plan of the five-room building. The arrows represent windows or doors; question marks indicate unknown open or closed status.

We also generated synthetic data, with errors, to represent measurements that might stream in from air monitoring sensors placed in the building. The synthetic data were based on an airflow and pollutant transport simulation that represents a single possible pollutant release event; this simulation was excluded from the Bayes Monte Carlo library of 5000 simulations. From this excluded simulation, we generated high-, medium-, and low-quality synthetic data, using progressively larger magnitudes of errors. We also evaluated two data collection plans. The first, *concurrent sampling*, provides synthetic sensor data to the BASSET algorithm from all five zones simultaneously, at five-minute intervals. The second, *sequential sampling*, provides measurements from one zone at a time, at five-minute intervals. Sequential sampling might represent a situation where a single (expensive) sensor is multiplexed to several sampling tubes.

As part of the pre-event calculation, we generated 5000 airflow and pollutant simulations, each of them equally likely, using Latin Hypercube sampling.

Figure 2 shows the estimation of the source location for the three qualities of data. With concurrent sampling of medium- or high-quality data, BASSET correctly identified the source location at t=5 min., when five measurements were obtained. With low-quality data, the identification of the source location was slower, requiring more measurements, and thus more time, to overcome the error in the data. Again, the medium- and high-quality data permit dramatic uncertainty reductions at t=5 min., in all cases converging to the correct answers. Sequential sampling collects data five times slower than concurrent sampling. In consequence, the medium- and high-quality data did not locate the source until all of the rooms were sampled once (t=25 min.), though reasonably good estimates were generated as early on as t=10 min. The low-quality data, however, did not locate the source even after 30 min.

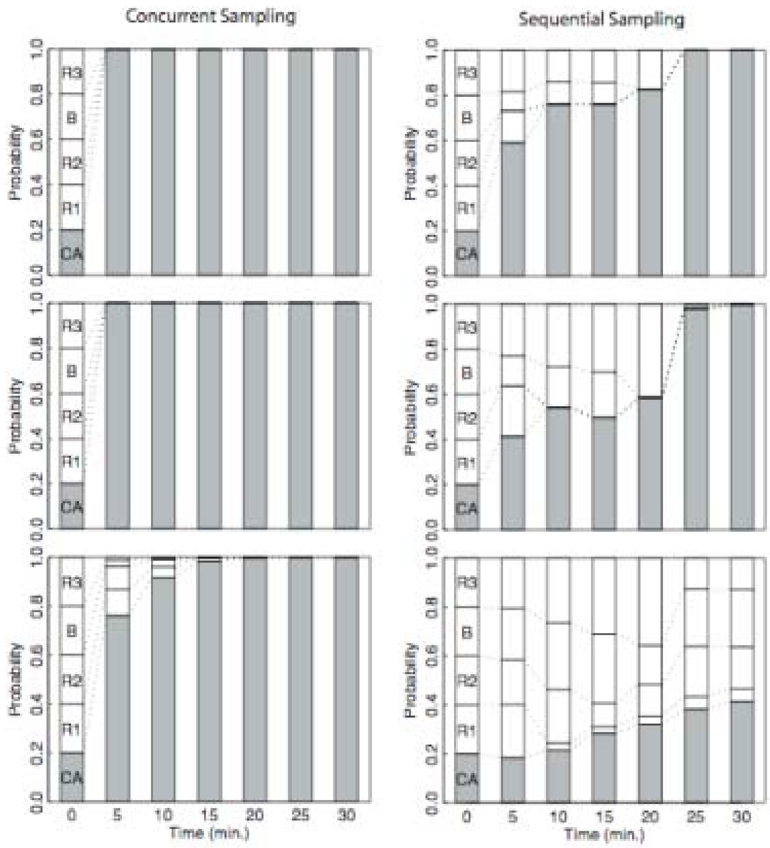

Figure 2: Locating the source using high-, medium-, and low-quality measurements (upper, middle, and lower plots, respectively). The gray bars represent the probability of release in the Common Area as data are interpreted by BASSET. The data are for a simulated release in this area.

Though it is tempting to interpret Figure 2 purely in terms of the success or failure of the interpretation approach, it is important to emphasize that the results merely illustrate the types of data interpretation and "what-if" analyses that may be conducted using Bayesian updating. For example, cases where the algorithm cannot identify the release scenario with high probability may be interpreted as due to insufficient information in the measurements.

Demonstration two: Study of a three-floor building

We now analyze a real building, consisting of three floors, which we characterized using tracer gas experiments [4]. We consider the following problem. A contaminant is released somewhere in a building, or near its indoor air intakes (i.e., in or near the HVAC return). A network of trigger or alarm-type sensors operates to identify the release. We seek to understand how sensor characteristics such as threshold level and response time affect the ability of the BASSET algorithm to quickly detect and characterize the contaminant release.

The study building consists of 660 m^3 of interior volume and approximately 280 m^2 of floor area on three levels. A mechanical air-handling unit (AHU) supplies air to the first and second floors. The AHU is a 100% recirculating unit (i.e., there is no deliberate outside air intake), and it returns air from the first floor.

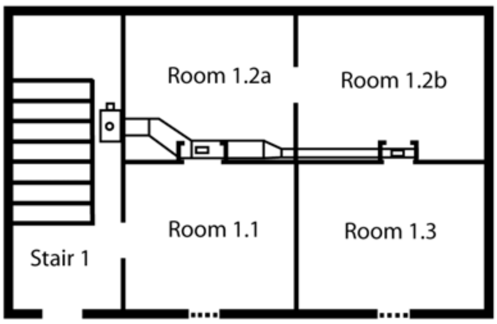

Figure 3: First floor plan view. The intake for the recirculating fan unit is in the stairwell. Outlets are in Rooms 1.2a, 1.2b, and two rooms on the second floor.

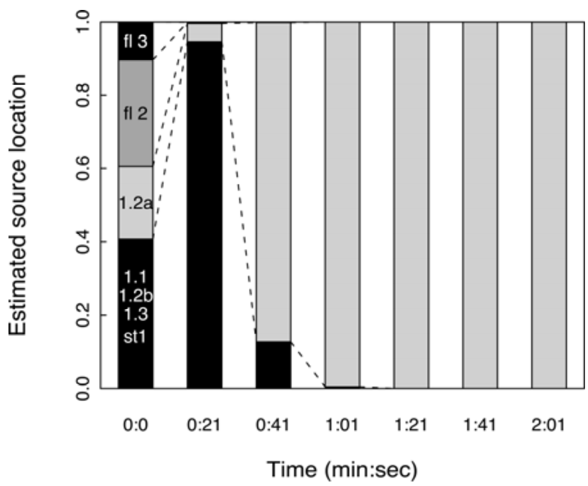

Fig. 4 Probability of source being in location indicated, as estimated with the BASSET algorithm using threshold data with response time of 20 s, threshold level of 2.3%, and without added error. The actual release location is Room 1.2a. Time is referenced to the instantaneous release event.

We generated a library of 5000 multizone model realizations by sampling from statistical distributions for a set of key input parameters. Variables included release location, source strength, and duration.

We generated synthetic threshold sensor data by interpreting the tracer data from the experiments [4] as if they were concentrations to which surface acoustic wave (SAW) sensors were exposed. Three sensor attributes were varied: threshold level, response time, and error.

In this implementation of Bayes' rule, the likelihood function is based on the assigned probability that we used to generate the false positives and negatives. In practice, the designer of the sensor system should have reliable information on the sensor's actual rate of false positives and false negatives.

For brevity, this paper only shows the ability of the sensor system to estimate the release location. Interested readers are referred to [2, 3] for more results and discussion.

Figure 4 depicts the time required to identify the release location (Room 1.2a). At time zero, every zone is assumed to be equally likely as the release location. As trigger data arrive, BASSET adjusts these probabilities, locating the release location with greater than 90% confidence within one minute. While the information content in threshold sensor data is significantly less than that in direct concentration measurements, the sensor system can successfully reconstruct the source, at least in some circumstances. For example, if rapid response hinges on locating a source very quickly, this example suggests that threshold sensors may be acceptable for real-time monitoring.

DISCUSSION AND CONCLUDING REMARKS

Real-time environmental monitoring systems have the potential to help protect high-value building occupants in the event of a toxic pollutant release. Here, we have demonstrated — albeit for a limited set of circumstances — that a network of continuous or single-level threshold sensors can be used to determine the location and magnitude of the release within a Bayes Monte Carlo framework. More importantly, the Bayesian approach naturally produces a systems-level view of the sensor network, which may lead to better tradeoffs studies between sensor characteristics, such as response time and error, than might be possible when considering sensors individually. The approach also allows us to compare such network options as deploying numerous low-accuracy and/or slow sensors versus fewer but higher accuracy and/or faster sensors.

However, with more complex buildings, system characterization will be more challenging, and also more expensive. Further work is needed to test the feasibility of the BASSET algorithm in such buildings. Research also is needed on hybrid methods, which could augment the sensor system with prior knowledge and continuous tuning of the airflow model.

Such advances would not only be beneficial for designing indoor monitoring systems, but may potentially be extended to the real-time optimization of building energy use, to error diagnostics in building mechanical systems, and to the detection, and characterization, of outdoor pollutants entering a building. Such approaches also hold the promise of improving building performance with respect to thermal comfort and indoor air quality.

ACKNOWLEDGEMENT

This work was partly supported by the Office of Chemical Biological Countermeasures, of the Science and Technology Directorate of the Department of Homeland Security and by the

Defense Threat Reduction Agency, and performed under U.S. Department of Energy Contract No. DE-AC02-705CH11231.

REFERENCES

1. Sohn, M D, Reynolds, P, Singh, N and Gadgil, A J. 2002. Rapidly Locating and Characterizing Pollutant Releases in Buildings. J. Air & Waste Manage. Assoc. Vol. 52, pp. 1422-1432.
2. Sreedharan, P, Sohn, M D, Gadgil, A J and Nazaroff, W W. 2006. Systems approach to evaluating sensor characteristics for real-time monitoring of high-risk indoor contaminant releases. Atmospheric Environment. Vol. 40, pp. 3490-3502.
3. Sreedharan, P. Sohn, M D, Gadgil, A J and Nazaroff, W W. 2007. Influence of indoor transport and mixing time scales on the performance of sensor systems for characterizing contaminant releases. Atmospheric Environment. Vol. 41, pp. 9530-9542.
4. Sextro, R G, Daisey, J M, Feustel, H E, Dickerhoff, D J and Jump C. 1999. Comparison of modeled and measured tracer gas concentrations in a multizone building. Proceedings of the 8th International Conference on Indoor Air Quality and Climate – Indoor Air 99, Vol 1, pp 696-701. Indoor Air 99, Edinburgh.