

# Lawrence Berkeley National Laboratory

## Lawrence Berkeley National Laboratory

### **Title**

Bayesian based design of real-time sensor systems for high-risk indoor contaminants

### **Permalink**

<https://escholarship.org/uc/item/0jp1q1jq>

### **Author**

Sreedharan, Priya

### **Publication Date**

2007-12-01

**Bayesian based design of real-time sensor systems  
for high-risk indoor contaminants**

by

Priya Sreedharan

B.S. (University of Windsor) 1997

M.S. (University of California, Berkeley) 2001

A dissertation submitted in partial satisfaction of the

Requirements for the degree of

Doctor of Philosophy  
in

Engineering - Mechanical Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor William W Nazaroff, co-Chair

Professor Van P. Carey, co-Chair

Professor Alice M. Agogino

Professor Edward Arens

Dr. Michael D. Sohn

Fall 2007

**Bayesian based design of real-time sensor systems  
for high-risk indoor contaminants**

© 2007

by Priya Sreedharan

## **Abstract**

Bayesian based design of real-time sensor systems for high-risk indoor contaminants

by

Priya Sreedharan

Doctor of Philosophy in Engineering – Mechanical Engineering

University of California, Berkeley

Professor William W Nazaroff, co-Chair

Professor Van P. Carey, co-Chair

The sudden release of toxic contaminants that reach indoor spaces can be hazardous to building occupants. To respond effectively, the contaminant release must be quickly detected and characterized to determine unobserved parameters, such as release location and strength. Characterizing the release requires solving an inverse problem. Designing a robust real-time sensor system that solves the inverse problem is challenging because the fate and transport of contaminants is complex, sensor information is limited and imperfect, and real-time estimation is computationally constrained.

This dissertation uses a system-level approach, based on a Bayes Monte Carlo framework, to develop sensor-system design concepts and methods. I describe three investigations that explore complex relationships among sensors, network architecture, interpretation algorithms, and system performance. The investigations use data obtained from tracer gas experiments conducted in a real building.

The influence of individual sensor characteristics on the sensor-system performance for binary-type contaminant sensors is analyzed. Performance tradeoffs among sensor

accuracy, threshold level and response time are identified; these attributes could not be inferred without a system-level analysis. For example, more accurate but slower sensors are found to outperform less accurate but faster sensors.

Secondly, I investigate how the sensor-system performance can be understood in terms of contaminant transport processes and the model representation that is used to solve the inverse problem. The determination of release location and mass are shown to be related to and constrained by transport and mixing time scales. These time scales explain performance differences among different sensor networks. For example, the effect of longer sensor response times is comparably less for releases with longer mixing time scales.

The third investigation explores how information fusion from heterogeneous sensors may improve the sensor-system performance and offset the need for more contaminant sensors. Physics- and algorithm-based frameworks are presented for selecting and fusing information from noncontaminant sensors. The frameworks are demonstrated with door-position sensors, which are found to be more useful in natural airflow conditions, but which cannot compensate for poor placement of contaminant sensors.

The concepts and empirical findings have the potential to help in the design of sensor systems for more complex building systems. The research has broader relevance to additional environmental monitoring problems, fault detection and diagnostics, and system design.

co-Chair \_\_\_\_\_ Date \_\_\_\_\_

co-Chair \_\_\_\_\_ Date \_\_\_\_\_

*For my parents Jayalakshmi and Kadambi Sreedharan,  
who have always supported and believed in me.*

# Contents

<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>Acknowledgements</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Preliminaries and problem statement	1
1.2 Environmental sensor systems	2
1.3 Sensor system task: source characterization	4
1.4 Design considerations	6
1.4.1 <i>Uncertainty</i>	8
1.4.2 <i>Complexity</i>	10
1.5 Objective	11
1.6 Dissertation outline	12
1.7 Contributions	15
<b>2 Background</b>	<b>18</b>
2.1 Environmental sensor systems	18
2.1.1 <i>Sensor systems in the built environment</i>	20
2.1.2 <i>Indoor contaminant sensor systems</i>	24
2.1.3 <i>Outdoor air contaminant sensor systems</i>	25
2.2 Indoor contaminant fate and transport	27
2.2.1 <i>System level processes</i>	28

2.2.2	<i>Contaminant scale processes</i>	38
2.3	Indoor contaminant fate and transport modeling	41
2.3.1	<i>Multizone models</i>	42
2.3.2	<i>Computational fluid dynamics</i>	49
2.3.3	<i>Zonal models</i>	50
2.3.4	<i>Semi-physics based network models</i>	51
2.4	Bayes Monte Carlo	53
2.4.1	<i>Bayes' rule</i>	54
2.4.2	<i>Foundation of Bayes Monte Carlo</i>	56
2.4.3	<i>Applications of Bayes Monte Carlo</i>	59
2.4.4	<i>Two-stage Bayes Monte Carlo algorithm</i>	61
2.4.5	<i>Limitations of Bayes Monte Carlo</i>	69
2.4.6	<i>Alternative algorithms</i>	73
2.5	Closing comments	82
<b>3</b>	<b>Case study description</b>	<b>83</b>
3.1	Experiments	86
3.2	Airflow patterns and tracer analyses	87
3.3	Multizone model development	95
3.4	Scope of the case study	106
3.A.	Appendix: Supplemental case study information	109
<b>4</b>	<b>Evaluating sensor characteristics in the case of threshold sensors</b>	<b>132</b>
4.1	Introduction	132
4.2	Background	135



4.2.1	<i>BMC Review</i>	135
4.2.2	<i>Case study: Experiments and model development</i>	136
4.3	Approach	138
4.3.1	<i>Library generation</i>	139
4.3.2	<i>Hypothetical threshold sensor data</i>	139
4.4	Results and Discussion	143
4.4.1	<i>Estimating release characteristics with threshold data</i>	143
4.4.2	<i>Effects of sensor threshold level and response time</i>	152
4.4.3	<i>Effects of sensor error</i>	155
4.4.4	<i>Evaluating additional release scenarios</i>	161
4.4.5	<i>Statistical considerations for interpreting threshold data</i>	165
4.5	Conclusion	168
<b>5</b>	<b>Influence of transport and mixing time scales on sensor system performance</b>	<b>170</b>
5.1	Introduction	170
5.2	Background	172
5.2.1	<i>Multizone model; likelihood function</i>	172
5.2.2	<i>Case study review and library generation</i>	174
5.3	Transport time scales in buildings	175
5.4	The likelihood function	178
5.4.1	<i>Likelihood function development</i>	179
5.4.2	<i>Implementation of likelihood function</i>	184
5.5	Investigation of time scales by blinding data	186

5.6	Evaluating different sensor systems	192
5.7	Statistical considerations	206
5.8	Conclusions	211
<b>6</b>	<b>Fusing information from heterogeneous sensors</b>	<b>215</b>
6.1	Introduction	215
6.2	Background	218
6.2.1	<i>Algorithm review</i>	218
6.2.2	<i>HVAC operation and control</i>	219
6.2.3	<i>Case study</i>	226
6.3	Heterogeneous sensor fusion concepts	227
6.3.1	<i>Physics perspective</i>	228
6.3.2	<i>Algorithm perspective</i>	231
6.3.3	<i>Information fusion from specific sensors</i>	239
6.4	Illustrative example	244
6.4.1	<i>Door position fusion algorithm</i>	245
6.4.2	<i>Description of investigations</i>	248
6.4.3	<i>Mechanically driven airflow conditions</i>	249
6.4.4	<i>Naturally driven airflow conditions</i>	281
6.5	Advanced fusion methods and statistical issues	299
6.5.1	<i>Sensor fusion using benign tracer gas releases</i>	300
6.5.2	<i>Dynamic library generation and advanced filtering</i>	302
6.5.3	<i>Methods that combine BMC, MCMC and SMC</i>	304
6.5.4	<i>Using the sensor system as a general system diagnostic tool</i>	305

6.6	Conclusions	307
6.A.	Appendix: Simulated data	312
<b>7</b>	<b>Conclusions and future work</b>	<b>321</b>
7.1	Summary	321
7.2	Specific research suggestions	326
7.2.1	<i>Chapter specific</i>	326
7.2.2	<i>Common to all chapters</i>	331
7.3	General future research suggestions	332
7.3.1	<i>The next generation of high-risk indoor sensor systems</i>	332
7.3.2	<i>Beyond high-risk indoor sensor systems</i>	334
	<b>References</b>	<b>336</b>

# List of Figures

Figure 1.1. Physical system description. ....	4
Figure 1.2. Forward problem solution. ....	5
Figure 1.3. Inverse problem description. ....	5
Figure 2.1. Two stage Bayes Monte Carlo framework.....	62
Figure 3.1. Plan view of case study unit. ....	84
Figure 3.2. Time-dependent normalized tracer-gas concentrations for Experiment 1. ...	90
Figure 3.3. Time-dependent normalized tracer-gas concentrations for Experiment 4. ...	92
Figure 3.4. Time-dependent normalized tracer-gas concentrations for Experiment 13. .	93
Figure 3.5. Measurements compared with multizone model predictions for Experiment 1. ....	96
Figure 3.6. Measurements compared with multizone model predictions for Experiment 4. ....	97
Figure 3.7. Measurements compared with multizone model predictions for Experiment 13. ....	98
Figure 3.8. Linear scale plot comparing measurements and model predictions for Experiment 1. ....	100
Figure 3.9. Linear scale plot comparing measurements and model predictions for Experiment 4. ....	103
Figure 3.10. Linear scale plot comparing measurements and model predictions for Experiment 13. ....	105
Figure 4.1. Time-dependent normalized tracer gas concentrations for Experiment 1...	137

Figure 4.2. Sample illustrating conversion of tracer gas concentration to threshold data: .....	142
Figure 4.3. Probability of source being in location indicated using threshold data. ....	144
Figure 4.4. Release mass and release duration estimated using threshold data. ....	145
Figure 4.5. Prior confidence intervals of the concentration profiles for three zones with concentration data. ....	147
Figure 4.6. Confidence intervals of the concentration profiles for three zones after performing Bayesian updating using 5 min of alarm-type sensor data. ....	149
Figure 4.7. Confidence intervals of the concentration profiles for three zones after performing Bayesian updating using 10 min of data. ....	151
Figure 4.8. Time required to locate the release location with varying threshold level and sensor response time. ....	153
Figure 4.9. Estimated mass released as a function of threshold level and response time. ....	154
Figure 4.10. Probability of correctly identifying the release location versus time .....	156
Figure 4.11. Probability of correctly identifying the release location using threshold data generated with 30% error.....	158
Figure 4.12. Probability of correctly identifying release location versus time since release using sensors with varying sensor response time and error and at a threshold level of 0.023. ....	160
Figure 4.13. Determination of source location and released mass using threshold signals generated from Experiment 2 data, plotted versus time since release. ....	162

Figure 4.14. Determination of source and mass using threshold signals generated from Experiment 3 data. ....	163
Figure 4.15. Determination of source and mass using threshold signals generated from Experiment 4 data. ....	164
Figure 5.1. Normalized model-to-measurement error for seven tracer gas experiments.....	180
Figure 5.2. Normalized model-to-measurement error as a function of release duration. ....	183
Figure 5.3. Probability of correctly identifying release location with blinded data sets.....	188
Figure 5.4. Estimated mass released with blinded data sets. ....	190
Figure 5.5. Probability of correctly identifying release location (Experiment 1).....	193
Figure 5.6. Probability of correctly identifying release location (Experiment 4).....	195
Figure 5.7. Estimates of the median release mass and uncertainty reduction of the 80% confidence interval (Experiment 1). ....	197
Figure 5.8. Estimates of the median release mass and uncertainty reduction of the 80% confidence interval (Experiment 4). ....	198
Figure 5.9. Probability of correctly identifying release location for 3 different sensor response times (Experiment 1). ....	202
Figure 5.10. Probability of correctly identifying release location for 3 different sensor response times (Experiment 4). ....	203
Figure 5.11. Estimates of median release mass and uncertainty reduction of the 80% confidence interval for 3 sensor response times (Experiment 1). ....	204

Figure 5.12. Estimates of median release mass and uncertainty reduction of the 80% confidence interval for 3 sensor response times (Experiment 4). .....	205
Figure 5.13. Probability of source being in location indicated for Experiment 4.....	207
Figure 5.14. Release mass estimated for Experiment 4 data. ....	208
Figure 6.1. Sample economizer control sequence for economizer outside air damper.	221
Figure 6.2. Supply fan diagram. Variable frequency drive controls the supply fan speed, .....	223
Figure 6.3. Variable air volume (VAV) units supply air to the individual zones.....	225
Figure 6.4. Inverse problem description .....	228
Figure 6.5. Prior probability algorithm for use with door position sensors.....	246
Figure 6.6. Layout indicating location of interior doors for the case study.....	250
Figure 6.7. Normalized model-to-measurement error for seven tracer gas .....	251
Figure 6.8. Experiment 1: probability of correctly locating release with and without door sensors.....	255
Figure 6.9. Experiment 1: probability of correctly locating either return intake or Room 1.2a as the release location.....	256
Figure 6.10. Experiment 1: median mass for networks with and without door-position sensors.....	259
Figure 6.11. Exp 1: Effect of door-position sensors on uncertainty reduction of mass estimate. ....	260
Figure 6.12. Experiment 4: probability of locating release with and without door sensors.....	262

Figure 6.13. Experiment 4: median mass for networks with and without door sensors.....	264
Figure 6.14. Experiment 4: effect of door sensors on uncertainty reduction of mass estimate. ....	265
Figure 6.15. Simulated data for Experiment 4 conditions with modeled concentrations. ....	268
Figure 6.16. Simulated data and modeled concentrations for a 20 g release in Stairwell 2. ....	269
Figure 6.17. Effect of door-position sensors for identifying release location for Exp 4 simulation.....	271
Figure 6.18. Reevaluation of simulated data from Exp 4 using a newly generated library.....	272
Figure 6.19. Effect of door-position sensors for quantifying the mass released for the Experiment 4 simulation. ....	273
Figure 6.20. Experiment 4 simulation: effect of door-position sensors on uncertainty reduction of mass released.....	274
Figure 6.21. Effect of door sensors for correctly identifying Stairwell 2 as the site of the contaminant release, using simulated event data. ....	276
Figure 6.22. Time to locate release to 0.9 probability with and without door-position sensors for a simulated release event in Stairwell 2. ....	277
Figure 6.23. Effect of door-position sensors for estimating the mass released for the simulated case of a release into Stairwell 2. ....	279



Figure 6.24. Effect of door-position sensors on reducing uncertainty in the estimate of mass released for the simulated case of a release into Stairwell 2. ....	280
Figure 6.25. Effect of door-position sensors for correctly identifying the locations of releases that occurred in Room 1.3 .....	284
Figure 6.26. Time to locate release to a confidence of 0.9 or better for releases that occur in Room 1.3.....	285
Figure 6.27. Effect of door-position sensors for estimating mass released for experiments AHUoff-1 and AHUoff-3). ....	286
Figure 6.28. Effect of door-position sensors on reducing the uncertainty of the estimated mass release for experiments AHUoff-1 and AHUoff-3. ....	287
Figure 6.29. Effect of door-position sensors for identifying the location of releases that occurred in Stairwell 2 (Experiments AHUoff-2 and AHUoff-4). ....	288
Figure 6.30. Time to locate release to 0.9 probability for releases occurring in Stairwell 2 .....	289
Figure 6.31. Effect of door-position sensors for estimating mass released for Experiments AHUoff-2 and AHUoff-4). ....	290
Figure 6.32. Effect of door-position sensors on reducing uncertainty of the estimated mass release for Experiments AHUoff-2 and AHUoff-4 .....	291
Figure 6.33. Time-dependent probability of locating a release with and without door-position sensors for the Tier 2 simulated release experiments.....	293
Figure 6.34. Time-dependent estimates of mass released for Tier 2 simulated experiments, with and without door-position sensors.....	294

Figure 6.35. Prior confidence intervals of time-dependent concentrations with and without door sensors. ....	295
Figure 6.36. Updated confidence intervals of concentration profiles with and without door sensors after performing 5 minutes of Bayesian updating. ....	296
Figure 6.37. Updated confidence intervals of concentration profiles with and without door sensors after performing 10 minutes of Bayesian updating. ....	298

# List of Tables

Table 3.1 Description of twelve tracer-gas experiments .....	87
Table 4.1 Parameter values of the library of 5000 simulated contaminant releases.....	139
Table 5.1 Parameter values used to generate the library of 5000 simulated contaminant releases.....	174
Table 6.1 Sensors for consideration in heterogeneous sensor networks.....	240
Table 6.2 Parameter values used to generate the library of 5000 simulated contaminant releases.....	250
Table 6.3 Description of simulated release conditions with AHU off.....	282
Table 6.4 Tier 2 network configurations.....	292

# Acknowledgements

What an amazing experience these years at Berkeley have been! This dissertation was possible because of the excellent mentorship, encouragement and enthusiasm from my advisors and support from family and friends. A colleague once said to me that his dissertation writing year was one of the happiest times of his life. That thought seemed absurd until I experienced it for myself. Prior to this year I could not have imagined I would have enjoyed a lifestyle limited mainly to writing, sleeping, eating, and exercise. I am grateful that my family and friends were very understanding about my needs and priorities, especially during this last year.

I first wish to thank my dissertation advisors and committee. This research was not possible without the unwavering enthusiasm, encouragement and intellectual brilliance of Professor Bill Nazaroff. After each of my meetings with Bill, I left more stimulated, encouraged and enthusiastic about research. I am especially indebted to Bill for the generosity in which he shared his energy and time in writing journal articles and in working through drafts of this dissertation. I thank Professor Carey for serving as co-Chair of my dissertation committee. Professor Carey has been a constant source of encouragement since I came to Berkeley. His knack for understanding the essence of scientific problems resulted in fruitful and interesting research meetings.

This dissertation could not have been completed without the support of Dr. Michael Sohn, a scientist at Lawrence Berkeley National Laboratory (LBNL). Mike's generosity in sharing his knowledge about Bayesian methods, statistics, scientific computing and modeling were essential to the success of this research. I am grateful for the time that

Mike took to meet with me on an almost weekly basis, especially in the first three years of my research. I also thank Dr. Ashok Gadgil, group leader of the Air Pollutant Transport (APT) group at LBNL. Ashok has been a mentor to me since I first arrived at Berkeley. I've enjoyed working with him and Mike on my dissertation these last four years. I especially appreciate Ashok's willingness to discuss the research at a detailed level. I am grateful, also, for the financial support from LBNL.

I also thank Professors Alice Agogino and Ed Arens for serving on my dissertation committee. Alice has been a constant source of encouragement to me and I appreciate the interactions I've had with her and with the Berkeley Expert Systems Technology (BEST) group. I am especially appreciative for her constructive comments and help in improving upon a draft version of this dissertation. I've also enjoyed my interactions with Professor Arens and appreciate the enthusiasm he has shown for my research.

I am grateful for my interactions with the APT group at LBNL, which has a rare and unique mix of experts, all of whom are generous with their knowledge and very friendly. I especially thank David Lorenzetti, a scientist in the APT group. His suggestions on technical writing and presentation were invaluable. I am grateful to him for reviewing early drafts of this dissertation. I thank, also, the air quality group on campus (Airheads). I appreciate the relaxed and friendly atmosphere that characterizes the Airheads group meetings and will miss my regular interactions with Bev, Nasim, Arman, Ling, Seema, Dev, George and Sharon.

There are numerous additional current and former Berkeley graduate students that I've interacted with over the last many years. We studied together for class exams, preliminary exams, qualifying exams, discussed research and offered general support to

one another. I especially thank Sasha Lanning, Katherine Dunphy-Guzman, Michael Wetter, Jeffrey Siegel, Amnon Barilan, Marko Mikofski, Jessica Granderson, Rodney Martin, Amip Shah, and Julian Marshal.

This achievement could not have been possible without the support of my family. First and foremost I thank my parents. I can't imagine a more supportive and encouraging set of parents. They were positive when I found it hard to be. I am quite certain they thought about my dissertation as much as I did these last four years. My brother has always been a source of reason and calm. I also thank my cousin Dilip who was adamant that I return to Berkeley and complete my PhD.

This research was supported, in part, by fellowships from the National Science Foundation and Soroptimist International Founder Region, and 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 (U.S. Department of Energy Contract No. DE-AC02-05CH11231).

# Chapter 1

## 1 Introduction

### 1.1 Preliminaries and problem statement

Society has become concerned about the consequences of high-risk contaminant releases into densely populated areas. Such releases may be accidental or intentional and precedents exist for both. The Bhopal release of methyl isocyanate killed 6000 people and left more than 50,000 with long-term health impairment (Dhara et al., 2002). Two releases of sarin in Japan, the first in the residential neighborhoods of Matsumoto, and the second on the Tokyo subway, together caused 19 deaths and injuries to over 6000 people (Yanagisawa et al., 2006).

Exposure to contaminants can occur in diverse locations. Buildings are of particular concern. People spend 90% of their time indoors and buildings are occupied at high density, thus providing a relatively efficient exposure pathway. An additional reason for concern is that certain high-risk contaminants, such as sarin, are very toxic and can cause harm over short time periods (Singer et al., 2005). In the occurrence of an unexpected contaminant release, an accurate and rapid sensor system would be useful towards preventing prolonged contaminant exposure and minimizing harmful effects.

Developing a sensor system involves making decisions about several important aspects of the system, including the type and number of individual sensors, their

placement, the communication infrastructure, and sensor interpretation algorithms. The research reported in this dissertation is based on the premise that to understand the benefits of any one design decision, a system-level analysis is required to evaluate its benefit to the overall sensor system. For example, a network composed of more rapid sensors may not yield a significant improvement in the overall sensor system performance due to the inherent dynamics of the contaminant transport. This dissertation explores and develops statistically based system-design methods for sensor systems that aim to detect and diagnose, in real-time, high-risk contaminant releases indoors.

## **1.2 Environmental sensor systems**

Researchers are interested in monitoring human exposure to environmental contaminants such as pesticides, industrial chemicals, chemical constituents in consumer products, and radiological releases. Accompanying the increased interest in characterizing environmental exposure have been advances in micro-scale sensors, wireless networking, and computing devices. These technological advances are creating new opportunities to monitor, collect and process environmental information that were previously unavailable (Zhao and Guibas, 2004).

While environmental sensing technologies have improved markedly, there have been significantly fewer advances in methods for integrating sensors into environmental systems that monitor pollutants, determine exposures, and guide responses to release events. Whicker et al. (2003) demonstrate the need for system design methods for detecting radiological releases in laboratories. Their work suggests that exposure can be reduced if sensor system designs consider the effects of sensor placement in relation to



contaminant transport. While the risk analysis literature presents techniques to estimate exposure and health effects from environmental contaminants, there is less discussion on design methods for monitoring such contaminants. The atmospheric release literature has been largely confined to the treatment of theoretical data (Smith and French, 1993; Sohn et al., 2002a). The use of real experimental data is recent and limited to the outdoor atmosphere (Politis and Robertson, 2004).

Current research trends in chemical and biological species sensor design focus on optimizing sensor performance without considering system-level effects. However, the value of specific performance improvements, such as lower detection limits and faster response times, depend on the system-level monitoring objectives and constraints. Without a system-centered design approach, it is unclear how to assess the design trade-offs among different performance metrics and thus to optimize a sensor system.

This dissertation focuses on indoor air sensor systems. However, the methods may have application beyond this domain, in other sensor interpretation applications, such as for outdoor sensor systems, energy monitoring, groundwater contamination systems, and intelligent transportation systems. In the future, physical phenomena of all types, such as wildlife habitat conditions, forest fires, or pedestrian traffic conditions may be monitored by inexpensive and easily deployed sensors (Zhao and Guibas, 2004). As sensors become nearly ubiquitous, the performance of sensing systems of all types may be limited by our ability to creatively and effectively use the information that sensors provide, an aspect that necessitates a systems perspective.

### 1.3 Sensor system task: source characterization

A sensor system is more useful for response guidance and control strategy planning if it can provide information beyond simply detecting the occurrence of a release event. Characteristics of the release, such as the source location, strength, duration, and future concentrations, are necessary to guide an effective response and, hence, to minimize exposure and resulting health consequences.

The task of identifying unknown release parameters and system variables is referred to as source characterization and is accomplished using an inverse problem formulation. The inverse problem formulation is contrasted with the behavior of the real system and a forward problem solution.

When a contaminant is released, a set of physical processes ( $H$ ) acts upon the release characteristics ( $X$ ) to generate time and spatially varying concentrations ( $H(X)$ ) (Figure 1.1). If the release conditions ( $X$ ) are known, a model of the physical system ( $H'$ ) can be exercised over  $X$  to yield an approximation of the true concentrations from a contaminant release ( $H'(X)$ ) (Figure 1.2). The extent to which the forward solution can reproduce the true concentrations ( $H(X)$ ) is dependent on the model accuracy.

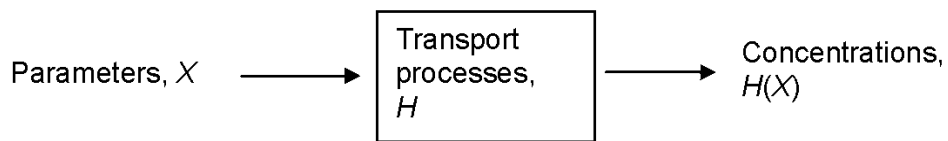


Figure 1.1. Physical system description.  $X$  consists of a set of release conditions, such as the release location, source, mass and building operating conditions.

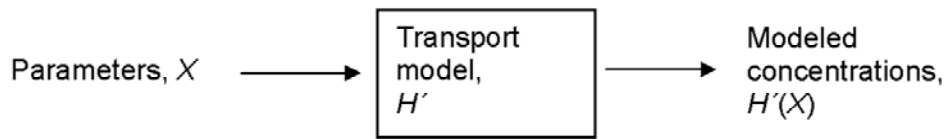


Figure 1.2. Forward problem solution. A transport model is exercised over a set of input parameters,  $X$ , to approximate the concentrations from a contaminant release.

In contrast to the forward solution, an inverse problem uses the observed concentrations  $H(X)$  to estimate the unobserved parameters,  $X$  (Figure 1.3) using an operator  $G$ . An inverse modeling solution inverts the forward model and applies this inverted model directly to a set of outputs to generate a set of inputs. A contaminant fate and transport model is not easily inverted. If the model cannot be inverted, then  $G$  must somehow use the model  $H'$  to solve for  $X$ . For example, the forward model may be exercised repeatedly until  $H'(X')$  is sufficiently close to  $H''(X)$  (i.e., optimization).

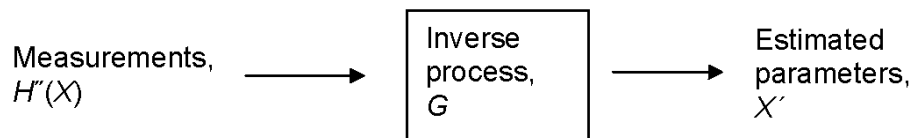


Figure 1.3. Inverse problem description.  $G$  uses the concentration measurements to estimate the unknown release conditions.

The solution to the inverse problem is complicated by the limitations of the model (i.e.,  $H' \neq H$ ) and by errors in the observed concentrations (i.e.,  $H(X) \neq H''(X)$ ). The choice of both  $H'$  and the sensors that generate  $H''(X)$  will affect the inverse problem solution. In essence, this dissertation seeks to elucidate these relationships.

Inverse problems can be solved using a variety of mathematical techniques. Optimization is one approach, but it is not ideal for the present application because it

typically involves an iterative solution and is thus suboptimal for real-time applications. An alternative to optimization solutions are state estimation or observer methods, which are used extensively in controls engineering. Such methods are also generally unsuitable for the present purposes, because they often impose linearity and other model restrictions on the measurements and parameter distributions.

In this dissertation, I use a Bayesian statistical approach to solve the inverse problem. The Bayes Monte Carlo (BMC) algorithm provides a framework to explicitly incorporate the variability and uncertainty that characterize pollutant transport and sensor information. The algorithm is noniterative and therefore amenable for real-time monitoring; it can describe nonlinear and non-Gaussian systems; and it imposes no restrictions on the way errors are treated. Like all Bayesian tools, it provides a natural framework to incorporate expert knowledge and combine this knowledge with data. Finally, it is appropriate for decision analysis because it provides confidence intervals on all estimated parameters. The algorithm is described in Chapter 2.

## **1.4 Design considerations**

An overall sensor system is composed of several subsystems that include the sensors, sensor network architecture, communication systems, and sensor interpretation algorithms. In this dissertation, I am primarily concerned with design decisions that relate to sensor selection and network architecture. These include the type, number and placement of sensors, and the sensor interpretation algorithms that convert sensor signals into useful information.

Another type of design problem would seek to maximize the information extracted from a given set of sensors. While the research presented here applies equally well to this more restricted problem, the dissertation focuses more on the broader question of sensor selection.

Understanding the effects of the number of sensors, sensor type, and sensor interpretation algorithms on system performance is a challenging goal. For any particular sensor network architecture, two interpretive algorithms may yield different results. Similarly, for a given algorithm, networks that differ in the total number of sensors or in their placement may characterize the release differently. In addition to the hardware and statistical interpretation algorithms, the model  $H'$  is an important design variable. Any sensor system requires a model to map the system input parameters to resulting concentrations. The selection of sensors should be influenced by the ability of the model to effectively use that information. Ultimately, the benefit of any one design decision should be evaluated by the value it adds to achieving the overall system objectives. Each chapter in this dissertation explores different facets of the design problem and expands on these concepts. An understanding of these issues is important for identifying an optimal or near-optimal design.

The design problem in this application is challenging specifically because of two phenomena: uncertainty and complexity. In the absence of uncertainty and complexity, it would be relatively simple to develop an optimized sensor system. Thus, these phenomena deserve some elucidation.

### 1.4.1 Uncertainty

*“Uncertainty: Lack of knowledge about the “true” value of a quantity, lack of knowledge about which of several alternative model representations best describes a biological/chemical/physical/other mechanism of interest, or lack of knowledge about which of several alternative probability density functions should represent a quantity of interest. Uncertainty is a property of the analyst.”* (Cullen and Frey, 1999) Uncertainty exists when an analyst has incomplete knowledge about the state of the system under investigation. For example, if an analyst had a perfect particle measuring device, there would be no uncertainty associated with the true particle concentration for the particular time and location monitored. However the extrapolation of this measurement to indicate particle concentrations for additional times and locations is characterized by uncertainty. Per the quoted definition, uncertainty also exists if the analyst is unsure about what model to use to represent reality. Such models can refer to a physical (or deterministic) model, or to a statistical model that represents measurement error.

The uncertainty owing to an imperfect measuring device is referred to as measurement uncertainty. Model uncertainty refers to condition of having an imperfect physical or statistical model. In this dissertation, model uncertainty mainly refers to physical model uncertainty. While measurement uncertainty prevents the analyst from knowing the true value of some attribute of a system, model uncertainty prevents the analyst from applying that value perfectly in interpretation or extrapolation. A concentration generated from a fate-and-transport model may differ from the true concentration owing to a combination of measurement uncertainty and model uncertainty.

Variability, a property of the physical system, contributes to model uncertainty. *“Variability results from heterogeneity of values over time, space, or different members of a population. Variability is a property of nature.”* (Cullen and Frey, 1999). For example, temporal variability in processes that govern contaminant transport and fate will cause repeated measurements, from a perfect measuring device at a fixed location, to differ. There are few realistic instances in which variability does not exist. In the case of contaminant transport, temporal and spatial variability of contaminant concentrations would generally cease to exist only in the case of a well-sealed room that is at thermal equilibrium and at uniform concentration. A model may be unable to represent the variability that describes certain physical phenomena, such as the spatial variations of species concentration within a room. Thus, variability can be associated with model uncertainty. Variability is a phenomenon to be considered in analyzing physical systems.

The performance of an environmental sensor system is affected by both measurement uncertainty and model uncertainty. Commonly, research efforts are devoted to improving sensor performance, thus reducing measurement uncertainty. In some cases model uncertainty may be more important than measurement uncertainty, which would constrain the extent to which benefits could be realized from the use of additional information from more precise or accurate instrumentation. Similarly, the computational effort needed to execute a higher resolution indoor fate-and-transport model may be unstrategic if there is an insufficient number of sensors to capture the variability across the region in space. These concerns form part of the motivation for the research reported here. Within the dissertation, I consider the influences of both model and measurement uncertainty.

## 1.4.2 Complexity

Uncertainty usually increases with increasing system complexity. One source describes a complex system as one with many strongly-coupled degrees of freedom in which local and global phenomena interact in nonlinear ways ([http://en.wikipedia.org/wiki/Complex\\_system](http://en.wikipedia.org/wiki/Complex_system)).

Indoor contaminant fate-and-transport processes are complex. Several processes — including thermal, fluid and mass transfer, and chemical transformation processes — can interact in nonlinear but coupled ways and influence the evolution of concentrations during a release event. Each process is characterized by some level of uncertainty, which propagates through interconnected processes and therefore increases the overall uncertainty.

Complexity increases with building size because, for example, the number of transport pathways increases. The number of sensor system design options also increases with building scale. Consider a 3-room building and consider that only 2 sensors are available. The decision of where to place the 2 sensors is relatively simple. There are only 3 possible combinations if each room is limited to one sensor. Compare this example to a 20 floor, 200 room office building, with 30 available sensors. The uncertainty that describes contaminant transport is larger in this case, and the optimal placement of the 30 sensors from  $\sim 10^{35}$  possible combinations is a daunting challenge.

With increasing complexity, the designer may find the task of allocating resources more difficult. Understanding the performance trade-offs among different design alternatives will help in this task. However, studying such tradeoffs becomes more difficult with increased uncertainty. Thus, methods are needed to help analyze more



complex systems and identify system-level effects of different design decisions. In addition, knowledge of generalizable and scalable phenomena or performance patterns may help reduce the total parametric design space to a manageable subset.

## 1.5 Objective

This research is motivated by the thesis that a system-level analysis and design approach is essential to identify appropriate and effective design choices for sensor systems that aim to monitor high-risk indoor contaminant events. Broadly construed, this dissertation seeks to advance the knowledge of sensor system design through empirical, conceptual, and methodological developments by elucidating the relationships among different components of the system. These components include the sensor network design, model selection, and statistical algorithms.

To review the inverse problem formulation, using Figures 1.1-1.3, the goal of a sensor system is to determine the nature of  $X$  using all available measurements  $H(X)$  and predict  $H(X)$  for future times. The set of input parameters,  $X$ , may consist of an observable component,  $X_1$ , and unobservable component,  $X_2$ . The first objective is to understand how measurements,  $H(X)$ , and measurable release conditions,  $X_1$ , should be selected to facilitate the solution of the unobservable parameters,  $X_2$ . The second objective is to identify generalizable truths that relate the overall sensor system performance to the physical system, which can be applied towards selecting sensors or models for other buildings. If general phenomena can be identified, the lessons and methods may be applicable towards modeling larger and more complex contaminant transport systems.

Each chapter in this dissertation explores one aspect of the complex space that describes Figure 1.3. The combined characteristics of contaminant sensors, sensors for other building parameters, and model limitations are explored in the context of their effects on system-level performance. Chapters 4 and 5 explore how limitations of both  $H'$  and  $H''(X)$  are related to the inverse problem solution. Chapter 4 investigates the use of binary measurements and specifically how characteristics of this class of sensors may influence the inverse problem solution. Chapter 5 seeks to understand how the estimation of  $X$  may be tied to the time-based value of  $H''(X)$  and simultaneously how a multizone model for  $H'$  poses challenges to realizing the information content from  $H''(X)$ . Chapter 6 complements  $H''(X)$  with a measurable subset of  $X$  ( $X_1$ ) and explores how this heterogeneous set of information can be used to characterize the unmeasurable parameters of  $X$  ( $X_2$ ).

While the statistical algorithms are a key design component, this research does not strive to develop optimal interpretive algorithms. Rather, it uses an existing algorithm to explore the relationships among sensor selection, network architecture, and model selection. Instead of maximizing the information content from a fixed set of sensors through algorithmic enhancements, this work addresses broader design challenges.

## **1.6 Dissertation outline**

Chapter 2 is a background chapter that discusses the application and methodology. Chapter 3 describes the experimental data that comprise the case study. The dissertation “core” consists of three distinct investigations (Chapters 4-6), followed by a concluding

chapter (Chapter 7). In the following paragraphs, I broadly outline Chapters 4-7, summarizing the objective, methods, and achievements of each.

The investigations in Chapters 4-6 are executed using experimental data obtained from one building. These data are used to generate several sets of sensor signals from a hypothetical release event. The Bayes Monte Carlo algorithm is the main analytical tool. Details of the algorithm and the case study are presented in Chapters 2 and 3.

Chapter 4 investigates how the individual sensor characteristics can be selected in such a way to improve the overall system performance. A class of chemical sensors that produce binary or threshold-type signals is investigated. The performance trade-offs among different sensor characteristics — including threshold levels, response time, and sensor error — are evaluated.

Several sets of threshold sensor signals were generated by exercising the tracer data from one release over different values of threshold levels, response time and error. The success of each set of sensor characteristics is evaluated by the ability of the system to characterize the release through reducing the uncertainty of each model parameter. The effects of each sensor characteristic are explained through physical insight.

This chapter demonstrates that a system-level approach may be beneficial in selecting an effective mix of sensor characteristics for sensor system application. The Bayes Monte Carlo algorithm is evaluated using signals that are generated from real experimental data, rather than from simulated data.

While Chapter 4 considers the limitations imposed by the sensors, Chapter 5 considers the limitations imposed by transport physics and the associated model. Chapter 5 explores the concept that the information a sensor system may learn about a

release is ultimately constrained by the physics that govern contaminant transport and therefore also by the physical model used by the algorithm. Understanding these relationships may lead to better design decisions.

The connections among the governing physics, the transport model, and system performance are investigated from the perspective of the contaminant transport and mixing time scales. A discussion describes how the time scales may influence the ability of the sensor system to characterize the release location and strength. The time scales are then investigated in two ways: first, by blinding portions of the data to the algorithm in order to assess the value of the blinded information; and second, by evaluating the performance of several sensor networks.

The relationship between the time scales and the transport model suggests that the statistical algorithms must carefully treat model uncertainty. The resulting likelihood function, though developed for the particular case study and BMC algorithm, may be useful for representing model uncertainty for other algorithms using a similar modeling approach.

Chapter 5 provides further insight into how the transport physics may help explain the relationship between sensor characteristics and sensor-system performance. While the investigation is necessarily limited to the case study, the time-scale concepts and illustrative likelihood function may help guide design of sensor systems for more complicated systems, and may also help identify and frame the relevant questions.

The concepts introduced in Chapter 5 emphasize the importance of the transport physics and model in selecting sensors. While Chapters 4 and 5 considered contaminant sensors, Chapter 6 considers the fusion of information from other kinds of sensors.

Networks employing non-chemical sensors can reduce the overall information costs, or may improve performance relative to a network that exclusively uses chemical sensors.

Chapter 6 describes a Bayesian framework for fusing information from heterogeneous types of sensors. While the framework is general, it is evaluated using the case study to investigate the benefit of door position switches.

The results from using these sensors are interpreted physically. I discuss how fusing information from heterogeneous sensors may help address the challenge of system complexity. I identify where the BMC and modeling framework may impose limitations to extracting otherwise potentially useful information from other sensor types, and offer suggestions for future research directions.

Chapter 7 concludes the dissertation with a summary and a discussion of ideas for future research on the themes explored here.

## **1.7 Contributions**

The achievements in this research are empirical, conceptual, and methodological. On an empirical level, this research uses experimental data rather than synthetic data. It also explores the use of binary sensors, concentration sensors, and non-chemical sensors.

In terms of conceptual development, a framework was developed for understanding how sensor system performance may be related to the underlying transport physics and its model representation. I discuss concepts for fusing heterogeneous sensor types. While both of these frameworks are explored within the context of the BMC algorithm, they are likely to have relevance to other sensor interpretation algorithms, as they are connected to the underlying transport physics.

Methodologically, a likelihood function was developed to characterize model uncertainty for the illustrative example. The functional form may be applicable for other buildings. While the likelihood function is not optimal, opportunities are identified to further develop a optimal likelihood function, which is more robust to model and sensor uncertainties.

Many of the investigations reported here were conducted using data from a case study. A benefit from using a case study approach is that the research is grounded in real truth, rather than highly idealized conditions. However, a drawback from using a case study approach is that the generalizability from the case study approach is not always assured. Generalizability exists in degrees, and the specific extent of generalizability of truths revealed from a case-study approach must be assessed carefully.

Thus, the generalizability of this dissertation's findings to different release conditions and different building types is not certain. It is likely the findings will have high relevance to single, instantaneous releases for buildings that are served by central, overhead air distribution. The extent to which the findings may be applicable to other releases and ventilation systems requires further investigation. Because the insight and frameworks presented in each chapter are supported by physical explanations, it is possible to hypothesize how the findings may apply to different systems. At a minimum, this dissertation is likely to provide insight into the questions that may be relevant when considering more complex buildings and airflow conditions. Even in the event that the direct methodologies fail when applied to different systems, such failure would provide valuable insight into improvement opportunities, and is therefore relevant.

This dissertation considers the development of sensor systems for high-risk contaminants that are released in or near a building. The methods explored also may be applicable to other types of environmental monitoring, such as energy monitoring, or fugitive greenhouse gas emission characterization. Additional design problems may also benefit from a systems-centered, Bayesian approach. While the specific findings may have limited direct relevance, the concepts, challenges, and questions that inspired the investigations are likely to be highly pertinent in many and diverse applications.

# Chapter 2

## 2 Background

This dissertation develops real-time sensor systems for characterizing indoor high-risk contaminant releases. The task of the sensor system is to transform concentration measurements into information that describes an unknown release conditions. A Bayesian statistical approach is used to achieve this task. Several subjects are relevant to this topic that include the physics of airborne environmental contaminant fate and transport, building systems and operation, contaminant transport modeling, and statistical methods. It is important, also, to understand the research accomplishments in indoor sensing systems and airborne contaminant sensor systems as the application considered in this dissertation sits within a broader domain. This chapter provides background material for these relevant topics. Sections 2.3.1 and 2.4.4 describe the specific modeling and statistical tools used in the dissertation (i.e., multizone model and two-stage Bayes Monte Carlo algorithm).

### 2.1 Environmental sensor systems

There is an increasing need for methods to select and place sensors into systems and for interpreting sensor information (Zhao and Guibas, 2004). This need has arisen in part because of advances in three broad areas: miniaturized sensor technologies, wireless communication systems, and computing power.



While sensor technologies have advanced significantly, there are comparably fewer methods for incorporating sensors into monitoring systems. To illustrate this point, a recent literature search on scholar.google.com with keywords “chemical sensor” revealed 12,200 references. A more narrow search combining the keywords “chemical sensor” with “sensor system” resulted in 1580 references, which is a small fraction of the total research articles on chemical sensors. Few of these 1580 references refer to the integration of sensors into monitoring systems. Rather, the term “system” in these references typically refers to hardware design, such as the use of arrays of sensors on a single device for improving the ability of a sensor to distinguish among different chemical species (Barrettino et al., 2003; Bedoya et al., 2004; Hunter et al., 2002). In addition to referring to species selectivity, the term “smart sensor system” may refer to the immediate processing of signals using pattern recognition or other statistical and artificial intelligence techniques for improving the accuracy with which chemical species are identified; the term “smart” does not refer to intelligent system-level processing (Grate et al., 1993).

Among this group of 1580 references, only two references discuss systems-level issues. One is a paper that reviews fire detection systems (Liu and Kim, 2003), which is discussed later in this chapter, and the second uses fuzzy logic to convert sensor measurements into qualitative air quality categories (Wide et al., 1997).

Furthermore, a search combining the keywords “chemical sensor” with “indoor” resulted in only 290 references. These references discuss topics such as electronic noses (Bourgeois et al., 2003) and robotic systems for plume-tracking (Ishida et al., 2001; Russell, 2001). Among these 290 references, none (excepting a published article by this

author) address the problem of incorporating sensors into systems for protecting occupants of indoor environments.

In the ensuing paragraphs, I review sensor systems for the built environment and airborne contaminant sensor systems for indoor and outdoor applications. Each application is considered within the following contexts. Does the sensor system operate in real-time? Is the primary function detection, or does the process include system characterization or state estimation? Do the methods consider the uncertainties associated with the sensor measurements and physical system? It is important, also, to note that in this dissertation, the term “sensor system” refers to the monitoring system, which includes the sensor network design, sensor selection and interpretation algorithms, rather than only the sensor network, or sensor hardware systems.

### **2.1.1 Sensor systems in the built environment**

Sensor systems can play useful roles towards achieving several different objectives in the built environment. These objectives include energy conservation, improved indoor air quality, and, more recently, protecting buildings from intentional and harmful releases of chemical or biological contaminants. Along the lines of building “security”, fire and earthquakes also pose physical threats and systems are developed and deployed in response to these concerns. To promote energy efficient, healthy and secure building operation, real-time monitoring systems can be extremely helpful and in some cases may be essential. For example, it would be difficult to optimize thermal-mechanical system performance according to building thermal loads without sensing and control systems. Similarly, sensor systems can be invaluable by identifying localized security threats such

as fire and chemical releases. This section reviews sensor systems that have been employed for achieving these objectives and provides the broader context in which sensor systems play a role in the built environment.

Increasing energy efficiency has been a strong motivation in the development of indoor sensor and monitoring systems. Heating, ventilating and cooling (HVAC) systems, along with lighting systems, account for a large percentage of commercial building energy consumption. The advent and widespread deployment of digital technology initiated a transformation in the 1980s in how HVAC systems are controlled: pneumatic control systems were replaced with digital control systems.

With digital controls, an energy management and control system (EMCS) acts as a supervisory controller. In addition to controlling thermal equipment, the EMCS also can collect and monitor the performance of building energy systems. However, research has shown that “typical” practice does not fully utilize the features of the EMCS (Piette et al., 2001), and, thus, the hopes of reduced energy consumption with digital technology have yet to be substantially realized.

The energy efficiency research community has been addressing the problem of poor control and operation of building systems and underutilization of available technology. Examples of research efforts include the deployment of field-scale monitoring systems (Piette et al., 2001) and development of automated diagnostics of HVAC systems (Katipamula and Brambley, 2005a and 2005b).

The literature in the automated diagnostics area, sometimes referred to as fault detection and diagnostics, is extensive. The concept is to use sensor information — sometimes available through the HVAC control systems themselves — and to determine

if there is faulty operation. Faulty operation can lead to increased energy consumption, additional maintenance costs, and decreased occupant comfort. Katipamula and Brambley (2005a; 2005b) provide a comprehensive review of the advances in this field.

Basic monitoring efforts have illustrated energy savings opportunities through simple improved operation. Piette et al. (2001) describe the use of a web-based tool that operates in parallel with the EMCS and diagnoses problems in building HVAC systems.

More recently, efforts have been made to use wireless sensor networks for improving building control, both in lighting and in HVAC systems. Kintner-Meyer and Conant (2004) describe the integration of wireless sensors into HVAC controls. Granderson et al. (2004) describe the integration of wireless sensor networks for advanced lighting control based on incorporating user preferences through a Bayesian decision tool.

Among the research efforts described in this subsection, methods employed in fault detection and diagnostics most closely resemble the research presented in this dissertation. One approach to automated diagnostics uses sensor information to estimate unobserved system parameters and then uses these parameters to indicate the presence of faulty operation. The problem is often formulated as an inverse problem, since the parameters of interest are not typically directly measured.

Compared to the time scales of interest in this dissertation, the time scales encountered in fault detection and diagnostics are usually longer. For example, the fouling of a heat exchanger is likely to occur over time scales of months to years. As a result, real-time computational methods may not be required. Methods range from physical and analytical modeling, to those using artificial intelligence or statistical techniques. Qualitative methods, such as rule-based systems, have also been used

(Katipamula and Brambley, 2005a and 2005b). So far, it does not appear that Bayesian methods have been applied in the HVAC fault detection and diagnostics efforts.

In addition to the energy-related research, there are extensive literatures on intelligent buildings and on fire protection sensor systems. Jablonski et al. (2003) provide an overview of issues in designing an intelligent building. They discuss how, in an intelligent building, a unifying centralized information system controls all functions of the building, including fire protection, thermal condition, IT systems, and office automation. The realization of a complete intelligent building requires cooperation among a diverse and disparate range of disciplines and designers, which is rare within the building industry, particularly in the US.

Sensor systems for fire protection have many parallels with sensor systems for detecting high-risk contaminants. Liu and Kim (2003) provide an excellent summary of recent advances in automated systems used for fire protection, many of which are enabled by advances in sensor technology and by the use of artificial intelligence techniques for sensor information processing.

In summary, there are many objectives that have motivated or should motivate the development of sensor systems and the adoption of information technology in the built environment. The harnessing of information technology for increasing energy efficiency has not reached its full potential. In part, progress has been hampered by the lack of sufficient economic incentives, and because inefficient thermal and lighting systems are not life-threatening. In contrast to energy systems, the failure of security systems, such as fire protection or chemical detection systems, poses acute risks. It is possible that overall improvement in achieving building security objectives may encourage an

improvement in performance of other systems, thereby resulting in better performing buildings overall.

### **2.1.2 Indoor contaminant sensor systems**

Exposure to indoor airborne contaminants generally receives less attention compared to outdoor exposure in discussions on the health effects of air pollution. However, as Nazaroff and Weschler (2001) observe, the indoor environment is especially important towards understanding health effects of airborne pollutants for several reasons: people spend most of their time indoors; there is a lack of regulatory intervention on indoor pollutants; there is a higher surface area to volume ratio indoors, which facilitates surface reactions; and the intake fraction indoors is much larger than the outdoor intake fraction (Bennett et al., 2002).

The release of high-risk contaminants poses a different set of problems compared to the release of other contaminants. Here, high-risk contaminants are defined as those that may result in serious health consequences from acute exposures, including death, or severe illness or injury. Such releases may include accidental chemical releases, laboratory radiological releases, or chemical or biological releases into or near a commercial building that are intended to cause harm.

Despite the wealth of literature in energy monitoring and automation, intelligent buildings, and wireless sensors, real-time monitoring of indoor air pollutants in commercial buildings has been restricted mainly to carbon dioxide monitoring for ventilation control and energy management (known as demand-controlled ventilation). Demand-controlled ventilation has become more common in recent years because of

indoor air quality building codes (such as ASHRAE Standard 62, or California's energy code, Title 24) and because of efforts to make buildings more sustainable. The LEED green building rating system, which is well documented on the U.S. Green Building Council's website ([www.usgbc.org](http://www.usgbc.org)), has been partly responsible for the latter effort. Fisk and De Almeida (1998) review sensor-based demand control ventilation.

Among the papers surveyed, only three address methods for developing real-time, high-risk indoor contaminant sensor systems. Whicker et al. (2003) highlight the need for methodologies to determine the optimal number and location of air monitors for facilities housing radiological substances. They present a method that identifies optimal sensor locations by detailed experiment and illustrate how even slight variations in placement can result in increased exposure. A second paper (Arvelo et al., 2002) investigates the use of multizone and computational fluid dynamics models, coupled with genetic algorithms, to optimally place sensors. These authors are concerned mainly with detection, rather than real-time characterization. In a third paper, Sohn et al. (2002a) propose a two-stage Bayesian interpretation scheme that can be used to characterize pollutant releases in buildings in real-time. They illustrated this method using simulated data from a five-room building. A subsequent effort demonstrated the two-stage approach using tracer data from a building (Sohn et al., 2002b). These papers present a promising framework. This dissertation advances the ideas introduced in these works.

### **2.1.3 Outdoor air contaminant sensor systems**

An early motivation for outdoor real-time monitoring was for detecting radiological releases from nuclear fallout. Bayesian methods have been proposed for interpreting data

from accidental radioactivity releases into outdoor air (Politis and Robertson, 2004; Smith and French, 1993). Smith and French conducted their study using simulated data, while Politis and Robertson used data generated from tracer releases. Both studies used sequential analysis techniques: calculations were made at each time interval as data were acquired. These studies focused on algorithm development.

Some researchers have used Bayesian and other statistical methods to evaluate outdoor dispersion models. Bergin and Milford (2000) used Bayesian methods to reduce uncertainty estimates in Lagrangian photochemical air quality models. Koracin et al. (2007) used statistical methods, specifically maximum likelihood, to estimate error characteristics from the MM5 Lagrangian dispersion model. The work of Koracin et al. revealed that the turbulence model characteristics had a greater effect on dispersion error than model resolution. Swall and Davis (2006) used a Bayesian hierarchical model to estimate sulfate concentrations for locations that lack monitoring data, and compared those values to predictions from the CMAQ Eulerian air quality model. These efforts support the use of Bayesian methods for reducing uncertainties of environmental model predictions.

In a related application, Thomson et al. (2007) located underground gas sources for purposes of extraction by assimilating above-ground gas concentration and wind data. They frame the system they study as an inverse problem and estimate the unknown source location using simulated annealing. While their application is related to the application here, their methods and objectives are different. The objective in this dissertation is not only to identify the release parameters, but also, to characterize the uncertainties associated with those estimates. Throughout this dissertation, the reduction



of uncertainties is a common theme. Furthermore, simulated annealing is likely to be too computationally demanding for real-time implementation.

## **2.2 Indoor contaminant fate and transport**

The physical and chemical processes that determine the fate and transport of indoor airborne contaminants are numerous and complex. The individual species' properties, such as particle size, diffusivity, chemical reactivity and sorption properties, are important. Several system-level processes, such as the building's ventilation rates, are relevant to the fate and transport of a contaminant. Many factors affect the system-level processes. Examples of these factors are the architectural features of a building, thermal-mechanical system design and operation, weather conditions, interior temperatures and occupancy behavior. For example, solar radiation, building orientation and dimensions determine the radiative heat transfer to the building envelope. Heat transfer through the building envelope affects interior natural convection and mechanical cooling processes that in turn influence the contaminant movement through the building.

A thorough understanding of environmental contaminant fate and transport requires knowledge of several subjects that include contaminant chemistry and physics, heat and mass transfer, fluid mechanics and mechanical system design and control. The subject is vast. For the purposes of this dissertation, it is important to understand the main processes that influence the fate and transport of non-reactive gaseous and particle phase contaminants. I describe system level and contaminant scale processes. These processes collectively determine the concentrations that may be observed temporally and spatially during a release event.

## **2.2.1 System level processes**

I use three categories to discuss system level contaminant fate and transport processes. The first category describes interactions between the building and the environment and focuses on building-to-environment air exchange and contaminant transport. The second category considers the airflow and mixing processes among the different zones in the building. Third, I consider processes at the room scale.

### **2.2.1.1 Building and environment interactions**

Contaminants are transported into and out of a building by advective transport. Air exchange between the building is an important system-level contaminant transport process. Three processes account for interior to exterior air exchange: thermal buoyancy (known as the stack effect), wind and mechanical ventilation.

Airflow induced by thermal buoyancy, commonly referred to as the stack effect, is pressure-driven flow caused by density differences between two adjacent columns of air (Feustel, 1999). Wind action generates a velocity and pressure profile around the building. Compared to the static pressure of the undisturbed wind-velocity pattern, the pressure field is characterized by higher static pressure on the windward side (reduced velocity) and lower pressures on the parallel and leeward sides. The windward side is characterized typically by exterior-to-interior flow and the leeward side by interior-to-exterior flow (Feustel, 1999).

Wind driven and thermal-buoyancy induced airflow collectively contribute to natural ventilation and leakage (or infiltration). The distinction between natural ventilation and leakage is that the former is an intended airflow process, and is enhanced through

architectural features such as operable windows. Linden (1999) gives a description of the fluid mechanic fundamentals of natural ventilation. Although wind may appear to be the main driving mechanism, thermal buoyancy is a significant driver and the dominant driver of natural ventilation in many circumstances (Linden, 1999).

Much of the work that investigates wind-driven natural ventilation is conducted using wind tunnel experiments (Carey and Etheridge, 1999; Jiang et al., 2003). Jiang et al. (2003) compared the experimental results with those from a large eddy simulation (LES) computational fluid dynamics (CFD) model. They found good agreement between the experiments and model predictions for the three ventilation cases that they studied.

Experiments have been conducted to study buoyancy-driven natural ventilation (Dascalaki et al., 1996; Jiang and Chen, 2003; Wilson and Kiel, 1990). Wilson and Kiel (1990) and Dascalaki et al. (1996) measured airflow for a test building with a large, single-sided opening under multiple thermal conditions. Jiang and Chen (2003) measured airflow for a single-sided opening and compared the airflows with two different CFD models. Other studies investigate natural ventilation with modeling (Allocca et al., 2003; Axley and Emmerich, 2002; Axley et al., 2002; Seifert et al., 2006).

Wilson and Kiel studied airflow through open doors in which they varied the duration that the door was kept open. Their experiments traversed an interior-to-exterior temperature difference of 0.5 °C – 45 °C and door opening durations of 0.5 s-120 s. The smallest temperature, 0.5 °C, exchanged ~ 10% of the room volume when the door was open for 120 s. Temperature differences of 4.7 °C and 17.9 °C resulted in 20% and 35% air volume exchanges, respectively, for a 120 s door opening duration.

Dascalaki et al. compared their measured airflow rates using tracer gas measurements with calculated airflow rates using hot wire anemometry. As the wind speed and temperature difference increased, the ratio between the “calculated” and “measured” airflow rates decreased. The increased temperature difference may result in more homogeneous flow; however, they had inadequate data to correlate the airflow differences with wind speed.

Natural ventilation processes are particularly relevant for residential buildings since many single-family homes are not equipped with a mechanical ventilation system. However, natural ventilation is becoming increasingly popular as a design tool in commercial buildings owing to energy and occupant comfort considerations. Research has shown that occupants can tolerate wider temperature ranges in buildings using natural ventilation compared to buildings exclusively using mechanical conditioning (Brager et al., 2004).

In practice, naturally ventilated commercial buildings may have a supplemental mechanical system. Such systems are termed mixed-mode or hybrid systems. Heinonen and Kosonen (2000) describe design and control methods for minimizing energy consumption of mixed-mode ventilation systems for northern climates.

Leakage through the envelope is an important system-level process and has consequences for indoor air quality and thermal conditioning energy requirements. Wind, thermal buoyancy, and mechanically induced pressure differences can induce leakage across the envelope through unintended pathways. Gaps in windows, beneath doorways, and openings around service penetrations are some examples of leakage pathways. Several factors influence the leakage rates of a building such as the

architectural design, mechanical system design and operation and weather conditions. Chan et al. (2005) compiled the results of a residential leakage database. They found that older and smaller buildings had significantly higher normalized leakage rates (normalized by floor area) compared to newer, larger residences. Price et al. (2006) studied the leakage characteristics of commercial buildings and apartments in the U.S. and in other developed countries. They found commercial buildings and apartments to be about twice as leaky as single-family houses, per unit envelope area.

Some energy codes, such as Title 24, require that buildings be tested for leakage. Building leakage can be investigated using fan pressurization tests, as described in ASTM standards (e.g., (ASTM, 2003)) and McWilliams (2002). The building is pressurized with a fan inserted in an opening (typically a doorway). The power supplied to the fan is varied to adjust the airflow until a target pressure differential,  $\Delta P$ , of 20-70 Pa across the building envelope is reached. Using the fan curves and the measured pressure differential, the airflow across the fan is known. Hence cumulative leakage from the building is also known (McWilliams, 2002; personal communication with author, March 31, 2004).

Most commercial buildings use mechanical systems to control air exchange between the building and the environment. A building may be designed with one or more mechanical air handling units (AHUs) that typically meet ventilation and thermal conditioning requirements simultaneously. Regardless of the use of the building, mechanical systems (and buildings) must be designed to provide a minimum level of ventilation to maintain acceptable interior conditions, often following ventilation standards such as the ASHRAE 62 ventilation standard or California's Title 24 energy

standard. For commercial buildings, ventilation rates are determined based on the type of occupancy. Title 24, for example, requires 7 (L/s)/person (15 (ft<sup>3</sup>/min)/person) for office environments. Today, most mechanical AHUs are equipped with economizers that can be modulated to actively meet ventilation needs and simultaneously minimize energy use by altering the ratio of fresh air to recirculated air.

The total ventilation rate in a building is the sum of natural ventilation, leakage, and forced ventilation. Differences in leakage characteristics and mechanical system design and operation, building type (i.e., residential or commercial), envelope construction, and age of the building has resulted in significant variability in ventilation and leakage rates in the commercial and residential building stocks. Persily et al. (2006) found median ventilation rates of 1 ACH (air change per hour) with 25<sup>th</sup> and 75<sup>th</sup> percentiles of 0.4 and 2 ACH, respectively, in U.S. office buildings. Murray and Burmaster (1995) analyzed residential data and reported a median, mean, and standard deviation air change rates of 0.5, 0.76, and 0.88 ACH, respectively.

Ventilation rates control the rate at which a contaminant, originating from an acute release, is completely removed from the building. In the case of a short-term acute release, the time required to remove a contaminant from the building is referred to as the removal time scale and is roughly equivalent to the inverse of the ventilation rate. Larger ventilation rates result in smaller removal time scales and expedite the rate at which indoor air quality returns to normal conditions. For a continual release, the ventilation rate determines the rate at which interior concentrations increase over the duration of the release and the concentration decay rate following the release.

### **2.2.1.2 Interzonal airflow and mixing processes**

Contaminants spread through the rooms of a building by advective transport and interzonal airflow. The transport of contaminants within the building affects the spatial distribution of concentrations that may be observed in the event of a contaminant release and this has consequences for occupant exposure. Compared with the ventilation literature, there are fewer publications that address interzonal airflow processes for both commercial and residential buildings. However, exposure considerations have motivated indoor air research that seeks to understand the complicated relationships between interzonal airflow and contaminant transport (Miller and Nazaroff, 2001; Ott et al., 2003). Recent studies also consider the influence of occupant behavior (Klepeis and Nazaroff, 2006a; Klepeis and Nazaroff, 2006b).

Separate interior sections of the building are often referred to as zones. Zones may include rooms or areas that have some physical barrier to distinguish them from adjacent sections of the building. Supply ducts, stairwells, or a ceiling plenum are other examples. Airflow between rooms, vertical airflow through a stairwell, and supply airflow from a duct are relevant and important interzonal airflow processes.

Thermal differences, wind, and mechanical systems drive air exchange within the building. These are the same processes that drive ventilation airflow. Thermal differences between rooms stimulate air exchange because of the buoyancy differences between the adjacent columns of air. In many residential buildings, natural convection drives interzonal airflow due to the absence of a mechanical system. Although their work did not quantify interzonal airflow, Wilson and Kiel showed that small temperature differences (1 °C) can generate significant buoyancy-driven airflow.

Mechanical system or air handling unit (AHU) operation often determines the airflow patterns in commercial buildings. In a typical commercial building design, air is circulated by the AHU through a system of overhead ducts. Overhead systems generally meet thermal and ventilation requirements simultaneously. The airflow from the ducted system to the rooms and from the rooms to the return space (which is often the plenum space above a drop ceiling) result in significant interzonal airflows. Mechanical systems alter internal pressure distributions that can also induce airflow between rooms.

The rates at which air is exchanged among zones determines the rate at which contaminants are transported through the building by advection. Consider a scenario in which a contaminant is released as a pulse in a room. For this scenario, the interzonal transport and mixing time scale is the time required for the contaminant to be transported out of the release room and to the remainder of the building, or to the outside. The interzonal transport time scale is inversely proportional to interzonal airflow rates. Uniform concentrations may be reached among a group of zones that share airflow communication paths within the interzonal transport time, depending on the relative rates of air exchange and ventilation rates.

Systems with higher interzonal airflows may exhibit more rapid whole-building mixing. Rapid interzonal transport tends to equalize contaminant exposures. Slower interzonal transport provides more opportunity to minimize occupant exposure to an unexpected contaminant release, in the event that a greater number of occupants are not in close proximity or in a room in which a release has occurred. Klepeis and Nazaroff (2006a) found that residential exposure to secondhand smoke could be reduced by isolating the smoker in a closed room with an open window.



Mechanical air distribution systems, especially overhead systems, are effective at inducing mixing on the whole-building scale. Consider that a mechanical system recirculates a portion of the air. In the event that a contaminant is released somewhere in the building, the transport of the contaminant to other locations in the building is accelerated. Mechanical systems generally increase the risk that an occupant anywhere in the building will be exposed to some level of contaminant more quickly, irrespective of their location in the building. The opposite effect occurs if a mechanical unit delivers 100% outdoor air. Instead, the contaminant air may be quickly expelled from the release room to the outdoors, preventing contamination of other zones.

Commercial buildings may alternatively be designed with radiant thermal systems and underfloor systems. In radiant systems, heated or cooled water may be supplied to a system of tubes embedded in a concrete slab or through ceiling panels. A small overhead mechanical unit may be installed for meeting ventilation requirements. Underfloor systems are similar to a traditional overhead systems except air is supplied to rooms by pressurizing the plenum space below a raised floor. Ideally, underfloor systems condition the occupied space instead of the whole room.

Buildings with radiant systems may exhibit different interzonal airflow processes than buildings using overhead systems. In radiant system, natural convection airflow within a room and across rooms may be enhanced by large heated or cooled surfaces. An overhead mechanical unit, if included in the design, may provide minimal air to meet ventilation requirements and is unlikely to recirculate air in the building.

Interzonal airflow can be determined experimentally using two methods. Fan pressurization techniques can be applied to determine airflow between specific rooms and

across individual components. In brief, a fan is inserted in an opening and the airflow is adjusted to reach a target pressure differential. Adjacent zones are pressurized (using additional fans) to suppress any leakage into the zone of interest, thus isolating it.

The second method uses tracer gas experiments to quantify interzonal airflow. Tracer gas experimental procedures are well documented and provide a non-intrusive method of studying airflow under typical operating conditions (McWilliams, 2002). A known amount of tracer is released and the concentration in different rooms is measured. Typically, contaminant mass balance equations are derived with the unknown quantities being the interzonal airflow rates. Usually, the concentrations are assumed to be uniform throughout a room. Inaccuracies can arise from the well-mixed assumption and from measurement error as is documented by Miller et al. (1997).

There are some published data on interzonal airflows. Miller et al. (1997) used tracer gas decay methods to estimate the interzonal airflows in a two-zone residential building. They measured interzonal airflow rates of  $60 \pm 2 \text{ m}^3/\text{h}$  and  $154 \pm 17 \text{ m}^3/\text{h}$  for two sets of experiments. The first was conducted with no active ventilation and the second with  $20 \text{ m}^3/\text{h}$  fresh air. Ott et al. (2003) also used tracer gas techniques to estimate airflow between two zones in a residence. They measured interzonal airflow rates of 102 and  $120 \text{ m}^3/\text{h}$ , and air change rates of 4 and 4.6 ACH.

### **2.2.1.3 Airflow and mixing processes at the room scale**

The rate at which contaminants are mixed within a room has consequences to occupant exposure and to how rapidly contaminants are transported to other zones and subsequently out of the building. Mixing within a room often occurs by turbulent diffusion. Turbulent diffusion processes occur on a much larger scale compared to

molecular diffusion and is characterized by irregular flow structures termed eddies (Csanady, 1973). Natural and forced convection can induce turbulent diffusion.

Natural convection airflow in a room is influenced by the temperature distribution of surfaces, room dimensions, and air temperature. Small differences in surface to air-stream temperatures can induce turbulent airflow and convection mixing because of the relatively large height dimension of rooms. Consider a wall height of 3.6 m (12 ft) and temperature difference of 2 °C between the wall and room air. For these conditions, the Rayleigh (Ra) number is  $\sim 10^{10}$  (assuming dynamic viscosity of  $16 \times 10^{-6} \text{ m}^2/\text{s}$ , expansion coefficient of  $3.4 \times 10^{-3}$ , and Prandtl number of 0.7). The Rayleigh number indicates the relative magnitude of buoyancy and viscous fluid motion. A transition from laminar to turbulent airflow occurs at  $Ra \sim 10^9$  (Mills, 1995), indicating that a temperature difference of 2 °C may generate turbulent flow and mixing in a room.

Baughman et al. (1994) studied the rates of contaminant mixing in rooms under natural convection conditions with tracer gas experiments. They observed mixing times of  $\sim 7$ -10 min, 13-15 min and 100 min for the following respective test cases: the room was subject to solar radiation, a 500 W heater was operated in the room, and the room was kept at nearly isothermal conditions.

Forced convection systems, when present, significantly affect the airflow and mixing processes in a room. Overhead mechanical air distribution systems are designed to supply air in a manner that promotes mixing within the room to achieve a uniform air temperature. Diffuser outlets are designed such that the air mixes with the ambient room air by entrainment, reducing the air velocity (compared to the diffuser exit velocity) and equalizing the air temperature (ASHRAE, 1993).

Drescher et al. (1995) investigated the mixing times within a room under forced convection conditions. The experiments were designed to simulate mixing in rooms served by overhead mechanical distribution systems. The mixing times ranged from 2 - 42 min with typical values  $\sim 7$  min. Some of the experimental results were later substantiated using CFD (Gadgil et al., 2003).

Underfloor systems are likely to exhibit different airflow patterns and mixing times compared to overhead systems. An underfloor system supplies air from below a raised floor and at lower velocities. A well designed system conditions the occupied space of a room by displacement of warm air close to the occupant, rather than by inducing turbulent mixing in the entire room. The room may be thermally stratified, if in cooling mode, and this has implications for the contaminant mixing. Theoretically, mixing processes will be slower for underfloor systems compared to overhead systems, when both are in cooling mode.

### **2.2.2 Contaminant scale processes**

Contaminants are affected by different removal and transformation processes. The extent to which the contaminant transport will deviate from the behavior of an ideal tracer gas is determined by the influence of the removal and transformation processes. The distinction between the behavior of tracer gases and other contaminants is important because this dissertation uses experimental data from tracer gas releases for exploring the sensor system design concepts with intended application to contaminants.

Airborne contaminants can be in gaseous or condensed phases. The term *aerosol* technically means a suspension of solid or liquid particles in a gas (Seinfeld and Pandis,

1998). I use the term *particle* to refer to any non-gas-phase airborne contaminant. All contaminants are influenced, to some extent, by species and class-specific transformation and removal processes, such as sorption for gases, or deposition for particles.

#### **2.2.2.1 Dynamics of gaseous species**

Concentrations of gaseous contaminants are influenced by transport and transformation processes. Sorption and chemical reactions are transformation processes. Advection, molecular diffusion, and turbulent mixing qualify as transport processes. Contaminant scale processes for non-reactive chemical species include molecular diffusion and sorption. The discussion focuses on non-reactive chemical species. However, it is worth noting that for chemically reactive species, homogeneous and heterogeneous chemical reactions, including redox processes, acid-base reactions, and hydrolysis reactions are important (Weschler, 2004).

Molecular diffusion is often negligible based on typical time scales. For example, using a diffusivity of  $0.1 \text{ cm}^2/\text{s}$  and length scale of 100 cm, the time scale for diffusive transport is  $\sim 10^5 \text{ s} \sim 1 \text{ day}$ . Therefore, diffusion plays a small role in transporting contaminants except over short distances, typically near surfaces.

Sorption is an important species-specific process for chemically nonreactive gaseous contaminants. Sorption in the indoor context refers to the partitioning of a species between indoor air and indoor materials (Cano-Ruiz et al., 1993; Nazaroff et al., 1993). If the partitioning of the species is weak then sorption can be neglected as a removal process. The rate of sorption is important in the context of real-time sensor systems. If the rate of sorption is slow, relative to the contaminant transport time scales, it can be

ignored. Sorption phenomena within the indoor environment is an active research area (Huang et al., 2006; Sakr et al., 2006; Singer et al., 2007; Won et al., 2001).

Singer et al. (2004) investigated the sorption of several different volatile organic compounds (VOCs) in a simulated residential space. The removal times by sorption were in the range 10 – 40 min. A subsequent study investigated the sorption characteristics of surrogates for sarin and nerve agents (Singer et al., 2005). The researchers performed chamber experiments, in which they measured the sorption and desorption characteristics based on fitting the data to a four-parameter sorption model. The initial sorption (to the surface) times ranged from ~ 10 min to 70 min, with variability reflecting the different surrogate compounds and differences in furnishings.

These results indicate that sorption time scales can be comparable to intrazonal removal times, which can range from ~ 10 min to 25 min, depending on mechanical air delivery design. They may also be insignificant, relative to the interzonal transport times, depending on the particular contaminant.

#### **2.2.2.2 Dynamics of particles**

Indoor particle concentrations and fate are influenced by removal, transport, and phase-change processes. Removal processes include ventilation, filtration, and deposition onto surfaces. Particle resuspension from surfaces may also impact concentrations. Resuspension is not yet well understood but is expected to occur on a time scale longer than deemed relevant to the operation of a real-time sensor system. Particle transport occurs by means of advection, settling and turbulent and Brownian diffusion. Removal and transport processes influence indoor particle concentrations.

Phase-change processes such as nucleation, evaporation and condensation may be important in some circumstances. Nazaroff (2004) summarizes indoor particle dynamics.

Particle dynamic processes are strongly influenced by particle diameter. The particle size spectrum is typically divided into three ranges based on particle diameter: ultrafine ( $\leq 0.1 \mu\text{m}$ ), accumulation ( $0.1 \mu\text{m} - 2 \mu\text{m}$ ) and coarse ( $> 2 \mu\text{m}$ ). Filtration can be effective for both ultrafine and coarse mode particles, with ultrafine particles removed by Brownian motion and larger particles by impaction. Deposition loss rates also follow a “U” shape, in which the ultrafine and coarse particles exhibit higher loss rates. Accumulation-mode particles experience lower filtration and deposition loss rates than do the other modes. If a contaminant behaves as an ideal tracer (i.e., transport occurs via advection and removal through ventilation), it may closely simulate the behavior of accumulation-mode particles, but not of ultrafine or coarse-mode particles, especially for buildings in which HVAC systems are equipped with efficient filters.

### **2.3 Indoor contaminant fate and transport modeling**

This dissertation uses an inverse problem approach to detect and characterize the properties of a contaminant release (Figure 1.3). The contaminant transport model is an integral component to this process.

Accuracy, computational expense, and user input requirements should be considered in selecting the model. A more detailed or finer grained model does not necessarily result in a more accurate model, as the performance of more detailed models are even more dependent on the definition of input parameters and boundary conditions (which are not always easily identifiable). Ultimately, the model should be selected based on its ability

to achieve a particular performance objective of the system. If, for example, the performance objective is to identify a release location as a particular group of rooms, or one particular room, the detail provided by a computational fluid dynamics (CFD) model may not be necessary.

In this section I give an overview of the main contaminant modeling approaches, ranging from CFD approaches to the lumped-parameter approach of the multizone model. The latter approach is used in this research, as CFD remains prohibitively expensive for even the simplest of buildings. Nevertheless, it is important to understand the context of the decisions made in this work.

### **2.3.1 Multizone models**

A multizone model describes a building as a network of well-mixed chambers, termed zones, that are connected by flow paths. The outdoor environment is represented as an additional zone with unbounded volume. Typical multizone models solve for the airflow between zones by adjusting the zone pressures in order to balance the mass flows through the flow paths, which have assigned pressure-flow characteristics.

Fan pressurization tests can be conducted to determine leakage parameters. Given measurements of the airflow,  $Q$ , and pressure drop,  $\Delta P$ , values of the airflow parameters  $n$  and  $C_Q$  of the power law (equation 2.1) can be estimated.

Multizone models can represent airflow induced by wind, thermal buoyancy, or mechanical forces. Most commercial systems contain one or more air handling units (AHUs) which provide adequate quantities of fresh air by mixing a fraction of the return air with outdoor air. This mixture is discharged to supply zones. Each AHU



compartment, such as the mixing box, return and supply may be represented as additional zones in the model. The fundamental physics, computation and limitations of multizone models are next discussed.

### **2.3.1.1 Multizone model approach**

Two multizone models are widely used, and are maintained by U.S. national laboratories: COMIS and CONTAM. While COMIS is used in this research and some of the references may refer to COMIS, none of the ensuing discussion is limited to this particular model. Feustel (1999) and Feustel and Rayner-Hooson (1990) describe the physical foundations of COMIS.

Multizone models require knowledge of the following boundary conditions: wind velocity and pressure fields around the building, internal and external temperatures, and fan characteristic operation curves that relate airflow to pressure differential. Using these conditions, the airflow between two zones is calculated using the pressure differential between the zones and a set of equations that relate the airflow rates to the pressure difference and flow resistance. The flow paths between two zones may include doors, windows, cracks, and ducts. I discuss the equations used to describe some flow components briefly. Additional details on flow components can be found in the Feustel references.

The crack component is commonly used to describe leakage between two zones. A power law equation is used to model airflow through a crack:

$$Q = C_Q (\Delta P)^n \tag{2.1}$$

where  $C_Q$  and  $n$  depend on the geometry of the particular airflow pathway. For rectangular building cracks,  $C_Q$  is based on a combination of geometry and fluid

properties, and  $n$  ranges from 0.5 and 1, depending on whether the airflow resistance is dominated by either inertial or viscous forces. When viscous forces dominate  $n \rightarrow 1$ , and  $n \rightarrow 0.5$  when inertial forces dominate (Liu and Nazaroff, 2001).

Rather than specifying sets of parameters for every possible flow pathway across a building envelope, a collection of pathways are often grouped into a single set of parameters ( $C_Q$  and  $n$ ) that represent the overall envelope leakage for a pair of zones. Fan pressurization tests can be conducted to calculate a building's leakage characteristics. In lieu of experimentation, leakage parameters may also be estimated using published estimates, such as from ASHRAE.

Most building systems, especially in commercial buildings, include networks of ducts; thus, the prediction of duct airflow is important. Airflow through a duct is modeled using this fluid flow equation:

$$\Delta P = \lambda \frac{L \rho v^2}{D^2} \quad (2.2)$$

where  $L$  is the duct length,  $D$  is the diameter,  $\rho$  is the air density,  $v$  is mean air speed, and  $\lambda$  is a function of the flow regime and characteristic fluid properties. This equation can be rearranged to fit the power law function (equation 2.1) where the exact value of  $C_Q$  and  $n$  will depend on the airflow conditions (i.e., whether the flow regime is laminar, transition or turbulent). Because  $C_Q$  and  $n$  depend on the airflow rate, the model must calculate these values iteratively.

The energy or pressure loss through duct fittings, such as through elbows can also be calculated using basic fluid dynamic equations and correlations. While this presentation implies a simplicity at a fundamental level, the challenge in airflow prediction arises from the complexity of interconnected flow paths and the uncertainty that describes the more

complex network. For example, while correlations are available to predict energy loss through individual fittings, the cumulative effect of closely connected fans, fittings, junctions, and elbows is not as easily quantifiable. These closely connected flow elements, especially when they are placed near fans, increase the overall pressure losses and is known as the system effect (Coward, 1990).

Modeling airflow through large openings such as open doors and windows is a challenging task. Airflow is influenced by a combination of steady-state buoyant flow, dynamic wind effects, and recirculating flows from thermal effects. Large openings may have zero or one neutral-pressure planes, which will result in one-way or two-way flows. Therefore, airflows through vertical openings are calculated in sections, using the pressure differential of each section. At any level,  $z$ , the velocity across the opening is related to the pressure differential using Bernoulli's equation.

$$v(z) = \left[ 2 \left( \frac{P_1(z) - P_2(z)}{\rho} \right) \right]^{1/2} \quad (2.3)$$

COMIS combines these pressure difference relationships and solves for the airflows at each node by enforcing mass balance. To solve the governing equations, the computational solver iterates on values of the zones' reference pressures, using a gradient based optimization routine, until mass balance is reached within each zone. Details of the computation can be found in Lorenzetti (2002b).

Contaminant transport is subsequently computed, for each time step, using the modeled airflows and applying the mass-balance relationship to the contaminant. In the situation where the contaminant is transported solely by advection, the mass-balance relationship for the contaminant in zone  $i$  is:

$$\frac{V_i dC_i}{dt} = E_i + \sum_{j=1}^{N_{in}} Q_{ji} C_j - \sum_{j=1}^{N_{out}} Q_{ij} C_i \quad (2.4)$$

where  $V_i$  is the volume of zone  $i$ ,  $C_i$  is the concentration of contaminant in zone  $i$ ,  $E_i$  is the emission source in zone  $i$  (which can vary with respect to time),  $N_{in}$  zones supply  $Q_{ji}$  airflow to zone  $i$ ,  $C_j$  is the concentration of contaminant in zone  $j$ , and zone  $i$  supplies  $Q_{ij}$  airflow to  $N_{out}$  zones.

Equation 2.4 assumes that the contaminant behaves as an ideal tracer. As noted in Section 2.2.2, gaseous species and particles may not behave as passive tracer gases and species-specific characteristics can affect the rates at which different contaminants are transported among building zones. The multizone modeling framework can be extended to reflect cases where a contaminant is significantly impacted by processes such as sorption (for gaseous species) or deposition (for particles). For example, Sohn et al. (2007) linked the MIAQ4 particle model with the multizone model COMIS to model the fate and transport of environmental tobacco smoke. Since this dissertation uses tracer gas experiments as the basis for the data, it treats the contaminant as a tracer.

### 2.3.1.2 Multizone model limitations

Multizone models simplify the estimation of airflow and contaminant transport by dividing a building into manageable zones. These zones are assumed to be independently well-mixed, having uniform concentrations throughout each zone at all times. Assuming instantaneous mixing within a zone can be problematic. Several researchers have investigated the extent to which the well-mixed assumption may be valid, including the early research efforts by Baughman et al. (1994) and Drescher et al. (1995). Their work identified mixing times for forced convection conditions and some natural convection

conditions (i.e., when the room was subjected to heating by solar radiation through a window) to be on the order of 10 min. In cases where the ventilation times are longer (reasonable values are  $\sim 10$ -20 min), the well-mixed approximation may be reasonable.

In all cases, the early predictions from a multizone model, however, are likely to be inaccurate because at these times the zone is unlikely to satisfy the well-mixed condition. The resulting inaccuracies are relatively larger for short-term releases (i.e., when comparing two unit releases of differing durations). Furthermore, errors due to the well-mixed assumption propagate through the model typically in the form of premature arrival times of contaminant to each zone. The implications of this condition are addressed in Chapter 5.

Following the early works of Baughman et al., and Drescher et al., many other researchers have challenged the applicability of the well-mixed assumption. In recent work, Richmond-Bryant et al. (2006) compared the time required to reach well-mixed conditions with and without furnishings in a room for aerosol releases. The inlet velocity and flow rate to the room were 1.4 m/s and  $1.13 \times 10^{-2} \text{ m}^3/\text{s}$ , which is just above the ASHRAE ventilation standard of  $9.44 \times 10^{-3} \text{ m}^3/\text{s}$  per person for office spaces. They found that well-mixed conditions were reached in 300 s without furnishings, and 600 s with furnishings.

In addition to the inaccuracies caused by limitations of the well-mixed assumption, multizone models may also miscalculate airflows. The airflow between zones is estimated mainly based on engineering correlations for the pressure-flow relations in the flow paths, and not by solving the governing fluid mechanics equations in the flow paths

or zones. Accurate characterization of the flow-path parameters is also important and is not always achieved.

Consider the following situation. A room has an open window in the north wall and two interior doors: one in the south wall, and one in the west wall. Imagine that the wind is blowing from the north. The airflow through the south door is likely to experience preferential flow; however, the multizone model does not consider the orientation of the doors, relative to the window. The multizone model bases its calculation of the airflow through the doors solely on the static pressure in the zone, rather than by solving for the air velocity profile within the room.

Fundamentally, the multizone model conserves mass but not momentum. Therefore, multizone models are limited in the types of spaces to which they can be applied. Large spaces, such as atria, may require a computational fluid dynamics modeling approach. The investigation of numerical techniques for predicting airflows and contaminant transport in large spaces is an active area of research (Jayaraman et al., 2006).

One additional limitation of multizone models is that they typically do not contain an energy model. Multizone models solve for the airflow between zones using the temperatures as model inputs and assume that the airflow is buoyancy induced. The airflow and temperature are tightly related and a better engineering approximation is to solve the energy equations and airflow equations simultaneously. Recent efforts have been made to couple building energy models with multizone models so that temperature and airflow are solved simultaneously (Axley et al., 2002; McDowell et al., 2003).

Multizone models are known to have additional problems. Large vertical openings, for example in stairwells and elevator shafts, present problems for a variety of reasons

related to the interaction between the (usually unmodeled) thermal system and the flow system. Current implementations do not properly handle duct junctions. Another simplification is that some multizone models (including COMIS) do not account for the transport time spent in a flow path, such as in a duct. Therefore, the model may underpredict the time required for contaminants to move within a building. More discussion of these issues can be found in Lorenzetti (2001; 2002a).

### **2.3.2 Computational fluid dynamics**

Computational fluid dynamics (CFD) models solve the governing equations of continuity, momentum, energy, and contaminant transport. Thus, they can provide a level of detail unavailable to models driven mainly by engineering approximations, such as multizone models. For example, CFD models can provide detailed within-room profiles of velocities and contaminant concentrations (Fan, 1995; Nielsen, 2004). Such information can be useful in large spaces where the well-mixed assumption and uniform temperature profiles are commonly inappropriate (Finlayson et al., 2004; Gan and Riffat, 2004; Jayaraman et al., 2006; Yang et al., 2001).

Two broad CFD modeling approaches have been considered in building contaminant and airflow estimation: the Reynolds Averaged Navier Stokes (RANS) method, and large eddy simulation (LES). Each addresses turbulent flow differently. Eddies of different sizes characterize turbulent airflow. LES models the larger eddies in detail, with relatively less effort given to the smaller eddies. LES smooths the seemingly chaotic behavior of the fluid but retains a time variation of the flow for a given location. The RANS approach averages out the unsteady turbulent behavior (Ferziger and Peric, 2002).

Even with the computationally less-expensive RANS method, hours of computer time may be required to calculate the airflows for a small room (Gadgil et al., 2000).

CFD is not practical for the application pursued in this dissertation because the statistical approach requires generating hundreds or thousands of model simulations. Nor is CFD practical for a whole (large) building. A CFD model may not be necessary unless the detailed airflow patterns within a room are required. This may be the case in a large space, where multiple sensors may be placed within the space, and occupants may be evacuated from that space via a strategic pathway. However, in a conventional building where the well-mixed assumption is approximately valid for individual rooms, an inverse solution that uses a multizone model may be able to locate the release within enough time and confidence, such that an optimal response strategy can still be executed.

It is worth noting that there are active research efforts to combine multizone models with CFD (Tan and Glicksman, 2005; Wang and Chen, 2007), which may result in a computationally affordable modeling approach. Future sensor system research methods may benefit from such models.

### **2.3.3 Zonal models**

Zonal models are a third class of physics-based models, in which zones that would otherwise be treated as well-mixed are further subdivided. The airflow and energy transfer between subzones are calculated using continuity, pressure differences and a thermal energy balance. In addition, the analyst must assign each subzone a model type, for example identifying thermal plumes and jets. Thus the details of flows within the room may be influenced by the analyst's craft as well as from the model calculations.



Wurtz et al. (1999) describe the construction of zonal models. Another group of researchers have converted the COMIS multizone model into a subzonal model (Stewart and Ren, 2006). The user effort and computational requirements of zonal models are less than for CFD models, but greater than for multizone models.

As with multizone models, zonal models do not account for within-room momentum. Therefore, the results obtained from a zonal model will differ from the results obtained by CFD, even when the analyst succeeds in identifying the gross flow features correctly. Mora et al. (2003) demonstrated that an extremely coarse-grid RANS-CFD model gives more accurate contaminant concentrations than a zonal model of comparable complexity. Therefore, it is not clear that zonal models can be beneficial to the design of sensor systems. Nevertheless, it may be worth investigating the use of zonal models, along with combined CFD / multizone models, in future research efforts.

### **2.3.4 Semi-physics based network models**

I use the term semi-physics network models to denote those that treat a building as being comprised of interconnected zones, but which solve for the airflows using a statistical approach such that the only physical law that is obeyed is mass balance.

The general process is as follows. The building is divided into zones (or nodes). Assumptions are made about the possible airflow pathways between the different zones. Known amounts of tracer gases are released in each zone (simultaneously if different gases are used, sequentially if the same gas is used). The concentrations in each zone are measured. A mass-balance equation is set up to represent the transport of the tracer from the release zone to the adjacent zones. The unknown airflow rates to and from each zone

are estimated using numerical optimization. This method circumvents the need to estimate leakage characteristics, as with a multizone model, or to define detailed boundary conditions, as with CFD.

Miller et al. (1997) describe a least squares approach to estimate the airflow rates for a two-zone building driven mainly by natural ventilation. Pulses of distinct tracer gases were released into each zone and concentrations of both tracers were monitored in both rooms over time. In analyzing the data, the researchers encountered estimation problems because of errors arising from the well-mixed assumption. In addition, they note that the method becomes more computationally expensive as the number of zones increases:  $n^2$  equations are needed to model a building with  $n$  zones.

Sohn and Small (1999) used “effective” mixing volumes within a zone, rather than assuming an entire zone is at a uniform concentration. The effective mixing volumes become additional unknown parameters. They compared the performance of two optimization algorithms, steepest-descent and simulated annealing, in estimating the effective mixing volumes and airflow exchange rates for a three-zone building. Steepest-descent worked well when the sets of equations were determinate (i.e., known mixing volumes), while simulated annealing could find the global optimal values of mixing-volumes and interzonal airflow rates. Their work used simulated data. Similar to the Sohn and Small approach, Nicas (1996) used a two-zone model to account for imperfect mixing in a room, and applied this method toward estimating exposure intensity. I refer the reader to Miller et al. (1997), Sohn and Small (1999), Nicas (1996) and their references for details on the tracer gas release methods and airflow exchange rate estimation methods.

Some complications to this general approach include that the airflow network structure should be known, and the airflow parameters are not linked in a ‘causal’ way to the operating conditions. Therefore, extrapolation of the model to non-experimental operating conditions is not straightforward.

## **2.4 Bayes Monte Carlo**

Imagine the following scenario. A contaminant is released over a short duration somewhere in or near a building. The building is equipped with a network of sensors that operate in real-time or near real-time, to detect the contaminant. The sensors may either report a concentration measurement, or, for alarm-type sensors, simply indicate whether the contaminant is above a certain threshold level. When any sensor measurement exceeds the sensor’s detection limit, or threshold level, it is inferred that a release may have occurred.

In such a situation, a Bayes Monte-Carlo (BMC) algorithm can extract more information from the sensors than the simple fact that a release has taken place. For example, this sensor interpretation algorithm might be used to determine the release location, duration, or amount. The algorithm also can determine building operating conditions, or other system parameters that are not directly measured but that may influence future concentrations during the event.

The performance of the sensor system is judged by its ability to characterize the contaminant release to meaningful levels of statistical confidence. A contaminant release can be characterized by its release conditions, such as the strength, location, and duration. Some knowledge may exist as to the likelihood of a release taking on different values of

these parameters. The role of the sensor system is to synthesize real-time measurements with this knowledge and generate updated estimates of these parameters.

This section describes Bayes' rule, the Bayes Monte Carlo algorithm, both generally and as it is implemented in a real-time sensor system, and alternative algorithms.

### 2.4.1 Bayes' rule

The following paragraphs introduce Bayes' rule and relevant probabilistic terms.

During a release, the algorithm compares the data streaming in from sensors against each realization in the library in order to assess the likely values of each model parameter. The goal is to be able to make statements such as, "The probability that the release occurred in the building lobby is 0.5 and that the release amount was greater than 100 g is 0.9."

These probability statements take into account the observations,  $O$ . Mathematically, the probability of some parameter,  $Y$ , taking on various values, given  $O$ , is written as  $p(Y|O)$ . In fact, the mathematical framework described below allows  $Y$  to represent not just model parameters, but more generally any hypothesis or event.

We term  $p(Y|O)$  the "posterior probability distribution function." It represents the probability of a hypothesis or event  $Y$ , conditioned on observations,  $O$ . (The descriptive "posterior" anticipates the application of Bayes' rule, which calculates  $p(Y|O)$  as a consequence of the observations and of the probabilities assumed prior to the release.)  $Y$  can be either a continuous or discrete variable. If  $p(Y|O)$  is used to represent a continuous variable, it is a probability density function and if  $Y$  is a discrete variable, then

$p(Y|O)$  is a probability mass function. In either event, the important point is that  $p(Y|O)$  is an assessment of  $Y$  based on observation.

To calculate  $p(Y|O)$ , the BMC algorithm applies Bayes' rule (Lee, 2004):

$$p(Y | O) = \frac{p(O | Y)p(Y)}{p(O)} \quad (2.5)$$

The factor  $p(O|Y)$ , is called the likelihood function. It represents a relative agreement of the observation  $O$ , given some known parameter value, or hypothesis,  $Y$ . In this application,  $p(O|Y)$  represents the likelihood of observing a set of measurements based on the modeled outcomes,  $Y$ .

The following example illustrates the distinction between the posterior probability,  $p(Y|O)$ , and the likelihood function,  $p(O|Y)$ . Let  $Y$  represent some disease, and  $O$  some symptom. Assuming a condition,  $Y$ , such as a flu, it is possible to know the probability of exhibiting certain symptoms,  $O$ , such as coughing. A valuable diagnostic is the probability that a patient has the flu if she exhibits coughing,  $p(Y|O)$ . While it is possible, given an assumed relationship between a hypothesis and its consequences, to calculate  $p(O|Y)$ , it is more challenging to calculate directly  $p(Y|O)$ . Bayes' rule provides a means to estimate the inverse probability,  $p(Y|O)$ .

The next factor in Bayes' rule,  $p(Y)$ , is called the prior probability. In the application in this dissertation,  $p(Y)$  represents an assessment of a particular type of release occurring, before the release actually occurs. The term "prior" accompanies this factor, because this probability is estimated before actual observations are available. It is used in Bayes' rule, in combination with the observations to estimate  $p(Y|O)$ .

The denominator of Bayes' rule,  $p(O)$ , is the probability of observing a particular outcome. With the cough and flu example, the probability of exhibiting a cough is

dependent on all of the possible ways by which a cough may be induced. It is calculated as follows:  $p(O) = \sum_Y p(Y)p(O|Y)$  if  $Y$  is discrete, and  $p(O) = \int p(Y)p(O|Y)dY$  if  $Y$  is continuous. Because  $p(O)$  is a constant with fixed  $O$  that normalizes the numerator, Bayes' rule is often written as a proportionality:

$$p(Y | O) \propto p(O | Y)p(Y) \tag{2.6}$$

Bayesian statistics introduces subjectivity through the prior probability,  $p(Y)$ , and is sometimes controversial for this reason. The prior probability is based on a prior belief of  $Y$  before interpreting  $O$  and can be based on expert judgment or empirical information. The value of  $p(Y|O)$  depends on  $p(Y)$ . Thus, if the prior beliefs between two analysts differ, the evaluation of  $p(Y|O)$  will also differ.

The determination of  $p(Y|O)$  is dependent on two components: the observations and the influence they have via  $p(O|Y)$  and the prior probability,  $p(Y)$ . As more data become available, the evaluation of  $p(O|Y)$  dominates the outcome  $p(Y|O)$ , and  $p(Y)$ , the prior probability, becomes less significant. In the application at hand,  $Y$  represents the occurrence of a particular contaminant release event (and its associated set of concentration outcomes) and  $O$  are the observations (measurements). At early times after the release, the prior probabilities of the concentration outcomes are highly influential. As more data are available, the data dominate the results from Bayes' rule.

## 2.4.2 Foundation of Bayes Monte Carlo

Consider a scenario where the objective is to use Bayes' rule to update an uncertainty distribution of a continuous model output,  $Y_i$ , using observations,  $O_i$ . Bayes' rule is restated in equation 2.7 with an elaboration on the denominator.

$$p(Y_i | O_i) = \frac{p(O_i | Y_i)p(Y_i)}{\int_{-\infty}^{+\infty} p(O_i | Y_i)p(Y_i)dY_i} \quad (2.7)$$

The evaluation of the denominator involves integrating the likelihood function over all possible values of the model output. Suppose that the prior uncertainty distribution of the model output,  $p(Y_i)$ , is normally distributed, and the likelihood function, which represents the model and measurement error,  $p(O_i|Y_i)$ , is also normally distributed. Then, the posterior uncertainty distribution of the model output,  $p(Y_i|O_i)$  can be solved analytically, given the variances and mean of  $p(Y_i)$  and  $p(O_i|Y_i)$ . For example, many analytical solutions exist when the distribution of  $p(Y_i|O_i)$  follows the same parametric form as  $p(Y_i)$  (Gelman et al., 2004).

A numerical solution is needed in the absence of an analytical solution to Bayes' rule. Markov Chain Monte Carlo integration (MCMC) is the common numerical approach to solving equation 2.7 (Gelman et al., 2004). MCMC, however, requires many iterations in which each new iteration involves generating a new set of model outcomes based on the statistical calculations from the preceding iteration. Thus, MCMC is not ideal for a real-time application.

Bayes Monte Carlo (BMC) is a numerical solution of Bayes' rule that provides a means to separate the time consuming modeling from the statistical calculations (Sohn et al., 2002a). Samples are drawn from the solution field so exhaustively that the integration in equation 2.7 is sufficiently approximated.

The key process in BMC is that samples are drawn from assumed prior distributions of the model inputs and are used to assemble a set of model outputs. A finite set of samples are drawn from the prior distributions of the model inputs. These samples

collectively constitute a probability mass function such that the probability of any one set of model input values is defined. Each set of model inputs generates values of model outputs, which are referred to as a model realization. Applying Bayes' rule to solve for the posterior probability of a set of discrete events is relatively straightforward.

Consider that  $N$  realizations are generated. The uncertainty distribution of the output  $Y_i$  is updated with corresponding observations  $O_i$  by calculating the posterior probability of each realization. The posterior probability of the inputs and outputs associated with the  $k^{th}$  realization is computed using equation 2.8

$$p(Y_{i,k} | O_i) = \frac{p(O_i | Y_{i,k})p(Y_{i,k})}{\sum_{u=1}^N p(O_i | Y_{i,u})p(Y_{i,u})} \quad (2.8)$$

where  $p(Y_{i,k}|O_i)$  is the posterior probability of the  $k^{th}$  model realization conditioned on the observations  $O_i$ ,  $p(O_i|Y_{i,k})$  is the likelihood of observing measurements  $O_i$  given model output  $Y_i$ ,  $p(Y_{i,k})$  is the prior probability of the  $k^{th}$  model realization,  $Y_{i,u}$  denotes the  $u^{th}$  realization of model output  $Y_i$ , and  $N$  is the number of model realizations. The updated uncertainty distributions of each model output can be used to update the uncertainty distributions of each model input. Additional details of BMC, such as those associated with generating the realizations and determining a likelihood function are discussed in the context of the two-stage BMC implementation.

A noteworthy advantage of a numerical solution, such as BMC, is that Bayes' rule can be applied to any modeling inference. All that is needed is a defined set of model input parameters, a coded relationship between the inputs and their uncertainty distributions, and an assumed relationship between the model input and outputs (Brand and Small, 1995). BMC imposes no restrictions on the shapes of the uncertainty distributions that describe the prior probabilities and likelihood function. Thus, the



domain of potential application of Bayes' rule is greatly widened, especially for complex systems.

### **2.4.3 Applications of Bayes Monte Carlo**

The BMC algorithm has been used in many environmental applications. Early application of the method includes that of Dilks et al. (1992) who quantified the uncertainties of water quality models. Freeze et al. (1990) also discussed the use of BMC for modeling the uncertainties of hydrogeological models with application to environmental design.

Following those early works, Dakins et al. (1996) presented a conceptual framework that uses Bayes Monte Carlo to assess the value of collecting future environmental samples for site characterization. Such samples help to reduce the uncertainty of an environmental fate-and-transport model. Different sampling schemes are compared and subsequently used to make environmental remediation decisions.

More recently, Sohn et al. (2000) used BMC methods to characterize parameters associated with groundwater contaminant transport. In their approach, they did not directly compare models and measurements in evaluating the likelihood function. Instead, they compared statistical metrics of concentration fields and pressure gradients. They used this approach because measurements and model predictions were not available at the same level of temporal and spatial aggregation.

Bayesian methods have been introduced for updating uncertainties of an integrated environmental health risk assessment model (Brand and Small, 1995). Health risk models often are composed of several interconnected submodels, each of which may be

characterized by large uncertainties. For example, a source model may be linked to a fate and transport model, which, in turn, is connected to an exposure model, and ultimately to a dose-response model.

If data are available at any stage of the overall model, they can be interpreted using a Bayesian Monte Carlo framework to reduce the uncertainty of the model predictions at that stage. Brand and Small (1995) illustrate how updating can be executed in a forward and backward sense. In the forward propagation method, data at a temporal or spatial measurement point,  $k$ , are used to update the uncertainty distributions of the model predictions at  $k+1$ ,  $k+2$ , and so on. That approach is used in this dissertation. For example, chemical sensor information at the present time may be used to better predict future concentrations. In backwards propagation, data at the  $k^{\text{th}}$  interval are used to update the uncertainty estimates at the  $k-1$ ,  $k-2$ , ... intervals. For example, in a real-time sensor system, information at the present time may be used to change the inferences that were made using previous temporal measurements.

In a recent publication, Lo et al. (2005) used BMC analysis to reduce the uncertainties of a lifecycle assessment for different waste treatment options. The global warming impacts estimated with the BMC approach revealed information that may be useful for policy decisions.

The final application I discuss is the problem addressed in this dissertation: the development of real-time sensor systems for detecting indoor contaminant releases. Sohn et al. (2002a) were the first to use the BMC algorithm for this application, by proposing a two-stage approach that separates the system modeling from the Bayesian updating step. Because the modeling may be performed before an event, the procedure allows real-time

estimation in cases where models cannot be run in real-time (for example, due to a large number of uncertain parameters, or even because the model itself takes longer than real-time to run). The two-stage approach is next discussed in detail.

#### **2.4.4 Two-stage Bayes Monte Carlo algorithm**

The BMC algorithm is incorporated into a real-time sensor system in two stages (Figure 2.1). The first stage assembles a library of simulated releases before any event takes place. The second stage compares the simulated results to sensor measurements, in real-time, during an event.

The release simulations for Stage 1 of the BMC algorithm are generated using a contaminant fate and transport model of the building in which the monitoring system is deployed. The contaminant fate and transport model simulates the concentrations that may be observed during a release event. Each simulation differs primarily in the values of the model input parameters, which are sampled from preselected distributions. The BMC approach is flexible enough to accommodate alternate models as well.

Stage 2, the sensor interpretation part of the algorithm, uses Bayesian updating to evaluate the agreement between each hypothetical simulated release and the sensor data. The probabilities of each simulated release are in turn used to estimate the uncertainty distributions of the release parameters and other information, such as the future concentrations and fates of the contaminant.

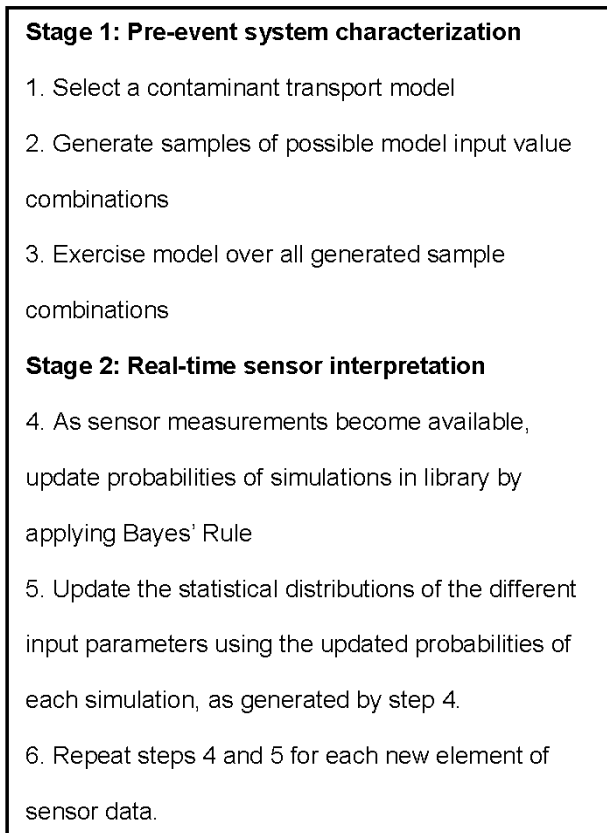


Figure 2.1. Two stage Bayes Monte Carlo framework

The primary advantage of this 2-stage approach becomes apparent when the model execution time prohibits simulating releases during the event. This may be the case when a single model run takes longer than real-time (as, for example, with computational fluid dynamics). More commonly, the model runs quickly, but so many runs are required to explore the parameter space that the simulations can't be performed in real-time. The two stages are next discussed in detail.

#### **2.4.4.1 Stage 1: Pre-event system characterization**

In the first stage, a contaminant fate and transport model is exercised to produce a library of simulations, or realizations, that represent the behavior of all possible

contaminant release scenarios. Because the inputs and outcomes are continuously variable, it is not possible to completely represent the outcome space,  $\Omega$ , with a finite set of simulations. Rather, it is necessary that the library, denoted  $\Omega'$ , must approximate the outcome space,  $\Omega$ , such that any contaminant release event is sufficiently close to a realization in  $\Omega'$ .

The library is generated by sampling from the prior distributions of each of the variable model input parameters and exercising a contaminant transport model for each set of parameter values. The prior distribution is a probability distribution of the respective parameter based on a best estimate (rather than using classical statistical methods, which exclusively use data). Expert judgment can be used to construct the prior distributions of each parameter. In this dissertation, log-uniform or uniform distributions are often used to represent the distributions of continuous model input parameters. The types of parameters may include the pollutant release location, pollutant mass, release duration, or inter-room door positions – i.e., any input to the model that may be unknown during a release event.

Individual sets of parameter values may be generated using Latin Hypercube sampling (LHS) (Iman et al., 1980). To generate  $m$  samples, the input parameter space is divided into  $m$  equally probable intervals and  $m$  samples are drawn from each interval in the parameter space. LHS reduces the number of samples needed to cover the range of the distribution (such as with a uniform, or log-uniform distribution) compared to random Monte Carlo sampling.

Determining the required number of parameter sets that are needed to approximate  $\Omega$  is a non-trivial step in generating the library. Consider that the library consists of  $n$

scenarios. If  $n$  is too small, there is a risk that a particular pollutant release event is not represented (or represented closely enough) in the library. In this case, the numerical approximation to  $\Omega$  would be inadequate. In such a situation, the updated uncertainty distributions of the model parameters, and any other unobserved variables, may not be accurate. For example, an incorrect release location may be identified as the true release location to a very high probability.

Theoretically,  $n$  can never be too large – since  $\Omega'$  is generated by selecting instances of continuously varying model input parameters. It is expected that, as  $n$  increases,  $\Omega'$  approaches  $\Omega$ . However, additional computational challenges are encountered as  $n$  increases. Each parameter set requires an execution of the contaminant fate and transport model in Stage 1. Stage 2 requires updating  $n$  probabilities in real-time, the evaluation of which is constrained by finite computing capacity and possibly by limited computer storage as well.

In addition to an inadequate number of parameters sets,  $\Omega'$  may poorly characterize a release event, also, owing to poor performance of the fate-and-transport model used to generate the library. Both decisions — the selection of an appropriate model and the number of samples analyzed — are important towards characterizing  $\Omega$ .

In practice, the completeness of  $\Omega'$  can be assessed using basic statistical measures. Given a fixed number of parameters and possible parameter ranges,  $n$  can be increased progressively until the selected statistical metrics (e.g., standard deviation, interquartile range, median) of  $\Omega'$  (i.e., based on modeled output variables) are insensitive to increased values of  $n$ . Assuming the metrics are meaningful, once the statistical properties of  $\Omega'$  are insensitive to increasing  $n$ , a sufficient value of  $n$  has been reached.

In this application  $\Omega'$  consists of time varying concentrations, and, thus, the statistical properties of  $\Omega'$  are evaluated on a temporal basis.

#### **2.4.4.2 Stage 2: Real-time sensor interpretation**

Stage 2 consists of real-time interpretation of sensor information using the updating procedures from the Bayes Monte Carlo method (referred to as Bayesian updating). As data become available to the sensor system, the algorithm estimates and updates the level of agreement between the observed data and each of the realizations in the library (i.e., the contaminant transport predictions). Each model realization in the library is compared to the data to assess the likelihood that the realization is consistent with the data. Models that fit the data more closely will result in higher likelihoods than others. The uncertainty distributions of the model input parameters, such as release location, or release mass, are then estimated using the posterior probabilities of each realization.

The probabilities of each model realization before and after assessing the agreement between the data and modeled concentrations are termed the prior,  $p(Y_k)$ , and posterior probabilities,  $p(Y_k|O)$ , respectively. Each model realization is assigned a prior probability. Because expert judgment was already used in Stage 1 to determine the prior uncertainty distributions of each of the model input parameters, it is often assumed that each realization is equally probable. The likelihood function,  $p(O|Y_k)$ , denotes the level of agreement between the model realization and observations. It is best understood in terms of measurement error. Given a particular modeled concentration, how likely are different measurement values? The likely deviation between modeled and measured values will increase with greater sensor error.

Consider a library that contains  $N$  realizations. The posterior probability of the  $k^{th}$  realization (making predictions  $Y_k$ ) given the sensor measurements  $O$  is denoted as  $p(Y_k|O)$  and is calculated with equation 2.9.

$$p(Y_k | O) = \frac{p(O | Y_k)p(Y_k)}{\sum_{i=1}^N p(O | Y_i)p(Y_i)} \quad (2.9)$$

where  $p(O|Y_k)$  is the likelihood of observing measurements  $O$  given model prediction  $Y_k$  and  $p(Y_k)$  is the prior probability of the  $k^{th}$  Monte Carlo realization. The updated statistical descriptions of each model input parameter and output are calculated from the ensemble of  $p(Y_k|O)$  results using standard methods.

While the BMC framework can update uncertainty distributions of multiple model outputs, it is mainly used in this dissertation to update the uncertainty distributions of contaminant concentrations. Consequently, the “ $i^{th}$ ” subscript used in equation 2.8 is not included in equation 2.9.

Equation 2.9 implies that each update uses all available data, but equation 2.9 can be rewritten so that the current update is determined from the most recent data plus results of the prior update. Equation 2.10 illustrates this alternative form, which is equivalent to equation 2.9, but also illustrates the sequential nature of the algorithm.

$$\begin{aligned} p(Y_k | O_{1:q+1}) &= \frac{p(O_{1:q+1} | Y_k)p(Y_k)}{p(O_{1:q+1})} \\ &= \frac{p(O_{1:q} | Y_k)p(O_{q+1} | Y_k)p(Y_k)}{p(O_{1:q+1})} \\ &= \frac{p(O_{1:q} | Y_k)p(Y_k)}{p(O_{1:q})} \cdot \frac{p(O_{q+1} | Y_k)}{p(O_{q+1})} \end{aligned}$$



$$\begin{aligned}
&= p(Y_k | O_{1:q}) \cdot \frac{p(O_{q+1} | Y_k)}{p(O_{q+1})} \\
&= p(Y_k | O_{1:q}) \cdot \frac{p(O_{q+1} | Y_k)}{\sum_{i=1}^N p(O_{q+1} | Y_i) p(Y_i)} \tag{2.10}
\end{aligned}$$

where  $q$  rounds of data (a round denotes that all sensors in the network have been polled for a particular time),  $O_{1:q}$ , have already been processed, and the most recent round of data, indicated by  $O_{q+1}$ , is being used to update the posterior probabilities. Either equation 2.9 or 2.10 could be used in an implementation.

The likelihood,  $p(O|Y_k)$ , represents the likelihood of a vector of measurements (because  $O$  comprises multiple measurements, possibly made at multiple times). It is evaluated by combining the likelihoods of each individual measurement. Consider that vector  $O$  consists of  $M$  total measurements, which differ in space (owing to the different locations of the sensors) and time. The likelihood of an individual measurement ( $O_i$ ), given a corresponding modeled concentration ( $Y_{k,i}$ ), i.e.,  $p(O_i|Y_{k,i})$ , is calculated using the probability distribution of the model-to-measurement errors ( $O_i - Y_{k,i}$ ). For example, if the errors follow a normal distribution, then  $p(O_i|Y_{k,i})$  is calculated as follows:

$$p(O_i | Y_{k,i}) = \frac{1}{\sqrt{2\pi}\sigma_{Ei}} \exp\left(-\frac{1}{2}\left[\frac{O_i - Y_{k,i}}{\sigma_{Ei}}\right]^2\right) \tag{2.11}$$

where  $\sigma_{Ei}^2$  is the variance of ( $O_i - Y_{k,i}$ ). (In the absence of model specification error,  $\sigma_{Ei}^2$  would be the variance of the measurement error.)

A typical evaluation of  $p(O|Y_k)$  assumes that the model-to-measurement errors are independent for different concentration determinations. This assumption is not always correct. For example, if the sensor contains calibration drift, then the errors from two measurements are correlated by this drift, and the independence assumption would be

violated. Similarly, imperfect model representation may also cause the errors to be consistently under- or overdetermined. In such cases, the information content from the observations is diminished and the likelihood function should reflect this. Sohn et al. (2000) describe likelihood function methods that account for correlated observations. However, assuming independence,  $p(O|Y_k)$  is calculated using equation 2.12:

$$p(O | Y_k) = \prod_{i=1}^M p(O_i | Y_{k,i}) \quad (2.12)$$

Constructing an appropriate likelihood function is an important and challenging task that is an important component of this dissertation. The likelihood function,  $p(O|Y_k)$ , quantifies the error structure of the data, which includes differences between the data and the model predictions resulting from any combination of measurement error, spatial and temporal averaging or correlations, and imperfect model representation. Error characteristics for a sensor operating under controlled laboratory conditions may be known, but the error characteristics of chemical sensors in field operation are often not well understood. Furthermore, depending on the type of model that is employed, model specification error may contribute to a large percentage of the overall uncertainty. If so, then it is important that this source of error be well represented in the likelihood function.

There are consequences to underestimating or overestimating the errors from measurement and model specification. When the errors are underestimated, then the posterior probabilities are overly sensitive to the data, which may result in unreliable statistical estimates. For example, an incorrect release location may be determined to be the true release location to a high level of confidence. This type of inaccurate inference could result in a disastrous response plan.

When errors are overestimated, the data are not used in an optimal sense. The release event could be characterized in greater detail and, thus, information is wasted. Small and Fischbeck (1999) describe the problems that are encountered in developing likelihood functions for risk-analysis and environmental problems.

#### **2.4.5 Limitations of Bayes Monte Carlo**

The BMC algorithm has characteristics that make it amenable for a real-time environmental monitoring system. However, other characteristics may present challenges for widespread implementation. I discuss two such characteristics here to support the analyses, discussion, and themes that arise in the succeeding chapters.

The first issue concerns how knowledge of contaminant behavior is generated and stored. Information pertaining to contaminant transport is generated prior to the occurrence of a release event. A reasonable concern is that the associated library may not adequately represent the true conditions of the release. As an example, the model may assume a certain set of HVAC operating conditions that are outdated.

One approach to prevent this situation is to carefully design and if necessary, augment the library to reflect all possible scenarios. However, it is impractical to incorporate or perhaps even anticipate all possible operating conditions, as well as all possible release conditions. As the library becomes larger, computational challenges may be encountered both in generation of the library, and in real-time sensor interpretation.

Rather than using a “brute-force” method to approach the issue of dynamic operating conditions, it may be possible to employ innovative methods in processing the library, such as identifying subsets within the library based on directly observed boundary

conditions (such as weather, operating conditions, etc.). This issue is addressed within the context of sensor fusion in Chapter 6. Alternatively, an algorithm that does not rely on a library may be considered. One possible algorithm is Markov Chain Monte-Carlo (MCMC), which is discussed in the next section.

The second issue concerns the efficiency with which the posterior uncertainty distributions of the release parameters are estimated. BMC can be classified within a broader class of statistical methods known as importance sampling (Reichert, 2002). These methods are known to be inefficient (Gelman, 2004). I briefly describe importance sampling and illustrate its relationship to the BMC algorithm. The concept of importance sampling is used in a later discussion on Sequential Monte Carlo algorithms.

In this discussion (and in the next section), the term “target distribution” refers to the statistical distribution of a parameter of interest. The goal is to describe the target distribution, which is unknown. The term sample distribution refers to a distribution from which samples can be drawn.

The idea of importance sampling is that samples are not drawn directly from the target distribution but from some sample distribution. The sample densities or probabilities are then weighted according to the relationship between the target distribution and sampled distribution. Statistical textbooks usually frame importance sampling around the process of evaluating an expectation. If we are interested in evaluating the expectation of some function,  $g(x)$ , we write:

$$E(g(x)) = \int g(x)f(x)dx \cong \frac{\sum_{i=1}^N g(X_i)}{N} \quad (2.13)$$

Consider that  $f(x)$  represents the probability distribution of a random variable  $X$  (The capitalization denotes that  $X$  is a random variable, versus  $x$  which is a dummy variable used to evaluate the probability of any value of  $X$ ). Also consider that  $N$  samples of  $X$  are drawn from the probability distribution,  $f(x)$ , using some sampling method such as Monte Carlo sampling. The expectation of  $g(x)$  is calculated in a straight-forward manner using equation 2.13.

Now, suppose that samples cannot be drawn directly from  $f(x)$  (i.e., the target distribution), but from some other distribution,  $h(x)$  (i.e., sample distribution). Then  $E(g(x))$  can be evaluated as:

$$E(g(x)) = \int \frac{g(x)f(x)}{h(x)} h(x)dx \cong \frac{\sum_{i=1}^N g(X_i) \frac{f(X_i)}{h(X_i)}}{N} = \frac{\sum_{i=1}^N g(X_i)w(X_i)}{N} \quad (2.14)$$

Here,  $N$  samples are now drawn from some distribution,  $h(x)$ , and  $w(X_i)$  is a ratio, often termed the importance weight, or importance ratio. It compensates for the fact that the instances of  $X$  were not drawn from  $f(x)$ , but rather  $h(x)$  (Lee, 2004).

The variability of the importance weights determines the success of importance sampling. Importance sampling is not useful if the importance ratios vary substantially (Gelman et al., 2004). (The usefulness is in relation to the ability of the sampling method to approximate the distribution with discretely drawn samples.) If  $g(x)f(x)/h(x)$  is roughly constant, then precise estimates of the integral can be obtained. If the importance ratios vary substantially, then more samples will be drawn from the “unimportant” regions of the target density, and the more relevant samples (i.e., those with higher probability density) will be represented with fewer sample draws. The sample distribution will not adequately resemble the target distribution in this case.

In the BMC approach, the posterior uncertainty distribution of the concentrations are desired. To approximate these distributions, samples are drawn from the prior uncertainty distributions of the model input parameters, which result in a prior uncertainty distribution of the concentrations. The posterior uncertainty distributions are generated by reweighting the prior uncertainty distributions by the likelihood function according to the ratios obtained from equation 2.9. BMC is a specialized version of importance sampling, in which the prior uncertainty distributions serve as the sampling distribution (Reichert et al., 2002). Equation 2.15 illustrates how Bayesian updating, equation 2.9, can be viewed as an importance sampling method.

$$p(Y_k | O) = p(Y_k) \frac{p(O | Y_k)}{\sum_{i=1}^N p(O | Y_i) p(Y_i)} = p(Y_k) w_k \quad (2.15)$$

where  $w_k$  is the importance “weight” according to the calculated likelihood for the particular set of observations and model realization.

Each model realization (total of  $N$ ) is reweighted according to its own evaluation of the likelihood function, relative to the denominator in equation 2.9, in proportion to its prior probability. The weight,  $w_k$ , for a particular realization will be small if that realization poorly reflects the measurements, relative to the other realizations. A small weight is indicative that the sample and respective concentrations in the library are not helpful towards representing the posterior uncertainty distributions of the concentrations. When the weights,  $w_k$ , vary significantly, the BMC algorithm is inefficient, as it is evaluating more realizations that are further from the observations.

The selection of the prior uncertainty distributions of the model input parameters, and sampling from these uncertainty distributions, strongly influences the degree to which any realization in the library contributes to the posterior uncertainty distributions.

BMC will perform more efficiently if the prior uncertainty distributions of the model input parameters are concentrated around the truth. This would result in more samples being drawn from more relevant areas of the prior uncertainty distribution and would result in a lower variance among  $w_k$ .

As a larger number of parameters are required to model the system uncertainties it may become important to implement steps that improve the efficiency of the BMC algorithm. For example, it may be possible to discard, in real-time, less relevant realizations and replace these with realizations from more relevant areas of the uncertainty distributions. Other methods based on importance sampling principles, such as sequential Monte Carlo methods, may offer additional opportunities for efficiency improvements.

#### **2.4.6 Alternative algorithms**

Two algorithms frequently appear in the literature in closely related domains: Markov Chain Monte Carlo (MCMC), which has been used in ecological modeling applications, and sequential Monte Carlo (SMC) methods, which have been used in target tracking applications. MCMC methods have been proposed in the literature as alternative approaches to BMC (Qian et al., 2003). I discuss the basic elements of MCMC and highlight issues surrounding its applicability to indoor contaminant real-time characterization. SMC methods are also discussed because of their suitability for real-time estimation and their resemblance to the BMC algorithm. Knowledge of both classes of algorithms may offer insight for possible future research directions.

### 2.4.6.1 Markov Chain Monte Carlo methods

Some characteristics of BMC (and more fundamentally, importance sampling) may pose challenges for the implementation of the BMC algorithm. The inefficiency of BMC may be potentially avoided by using Markov Chain Monte Carlo (MCMC) sampling methods. If a large number of simulations can be performed in real time, then MCMC is preferable to BMC or other importance sampling methods (Reichert et al., 2002).

From a historical perspective, the advent of MCMC techniques provided Bayesian analysts with a tool to tackle many problems that were inaccessible before (Gelfand, 2000). Often, Bayesian analysts would solve Bayes' rule to the proportionality constant (equation 2.6) but were unable to solve for the constant (i.e., the denominator) analytically, as it involves high-dimension integration (Lee, 2004). Numerical tools are thus needed. While the BMC method and importance sampling methods provide examples of such tools, MCMC is far more sophisticated. With MCMC, complex problems can be solved with just a specified likelihood function and a prior distribution.

MCMC's usefulness extends beyond Bayesian analysis and is a powerful sampling tool that can be used to perform high-dimension numerical integration. Consequently, the early part of this discussion is general and the latter part is related to the Bayesian application of this dissertation.

The idea of a Markov Chain Monte Carlo (MCMC) method is as follows. Consider the target distribution,  $f(x)$ , which represents the distribution of some variable  $X$ . We wish to draw samples from  $f(x)$ .  $X$  may be some unknown parameter or vector of parameters. The magic of MCMC is that if transition probabilities can be selected,  $p(y|x)$ , which represent the probability of a new value of  $X$ , based on a previous value of



$X$ , then, over time, an MCMC algorithm can be applied to generate a long series of samples that will eventually follow the target distribution,  $f(x)$ . The long series of samples is called a Markov Chain. The theory of MCMC is extensive and the reader is referred to Robert and Casella (2004) and Ross (2003) for details.

There are two commonly used MCMC algorithms: the Metropolis-Hastings (M-H) algorithm, and the Gibbs sampler, which is considered by many statisticians to be a special case of the M-H algorithm (Gelfand, 2000). Two tutorial style papers exist on both the M-H algorithm and Gibbs sampling (Casella and George, 1992; Chib and Greenberg, 1995).

I summarize a few key ideas of the Gibbs sampler, based on Casella et al. (1992), as it is the most widely used MCMC technique. Suppose we are interested in obtaining some characteristics of a distribution,  $f(x)$ . It is necessary to evaluate this distribution by taking into account the possible values that all other variables may take in combination with the possible values of  $X$  (i.e., the joint probability distribution needs to be considered):

$$f(x) = \int \dots \int f(x, y_1, \dots, y_p) dy_1 \dots dy_p \quad (2.16)$$

Using the Gibbs sampler, we avoid calculating  $f(x)$ , which involves multi-dimensional integration, and instead generate samples  $X_1, X_2, X_3, \dots, X_m \sim f(x)$  without requiring  $f(x)$ . This is achieved by composing conditional probability distributions for each random variable (e.g.,  $f(x|y_1, y_2, \dots, y_p)$ ) and sampling from these distributions.

The Gibbs sampler follows a set of rules to determine whether each new sample is accepted or rejected. In this way, new samples are drawn iteratively from the conditional

distributions (e.g.,  $f(x|y_1y_2\dots y_p)$ ) forming what is known as a Gibbs sequence. The early samples in the Gibbs sequence make up the “burn-in” period. Eventually, the Gibbs sequence reaches “stationarity” and all samples follow the distribution  $f(x)$ . Using the samples from the Gibbs sequence, statistical properties of  $f(x)$  can be computed.

Gibbs sampling is an active research area, both in terms of its use in applications, and in developing rules for proper implementation. There are different ways to construct the Gibbs sequence: starting with one initial condition and generating one sequence, or initiating multiple chains simultaneously. A significant issue associated with Gibbs sampling (and any MCMC technique) is assessing when a chain has converged to a stable point. Kass et al. (1998) document a discussion among well-known statisticians on this and other implementation issues associated with MCMC.

In a Bayesian analysis, the  $f(x)$  from equation 2.16 represents  $p(Y|O)$  from equation 2.9. The objective is the same: to draw samples from  $p(Y|O)$  without solving Bayes’ rule completely, which requires a solution of the normalizing constant or denominator. Because MCMC methods facilitate the solution of Bayes’ rule, they have been used extensively in environmental applications and may potentially offer an attractive alternative to the BMC algorithm. For the reasons discussed in Section 2.4.5, MCMC may utilize the observations more “efficiently,” and generate more accurate distributions of the unknown parameters. Qian et al. (2003) give a flavor of some of the potential problems of BMC, favoring MCMC.

Both Malve et al. (2005) and Harmon and Challenor (1997) used MCMC (in particular, the Metropolis-Hastings algorithm) for modeling aquatic ecosystems. Harmon and Challenor assert that MCMC methods are impractical for systems in which more than

~ 10 parameters are being estimated. However, Dowd and Meyer (2003) used the Gibbs sampler to estimate the distributions of 158 parameters of an aquatic ecosystem, some of which were time-varying. (Parameter values at new time steps were modeled as separate parameters.) They emphasized that if the model is chosen carefully, and is not unnecessarily complex, then MCMC can be applied to ecological inverse problems with more than 10 parameters.

Several papers in the literature, such as Kass et al. (1998), attest that successful implementation of the Gibbs sampler or M-H algorithm is not a simple task. However, if the particular application allows for an iterative computational approach, MCMC is a more efficient alternative to BMC, or any other importance sampling method.

Despite the benefits of MCMC, it is the iterative nature of the method that makes it unattractive for real-time applications. For this reason, MCMC methods have not commonly been used in controls engineering and signal processing problems, such as target tracking. However, in the case of a very important, high-profile building, there is a possibility that a large number of computers may be employed to perform real-time MCMC sensor interpretation.

In rare situations, MCMC techniques may be considered as a practical alternative to BMC. To execute an MCMC method, ranges and possible values of each parameter must be defined, as well as a likelihood function for comparing the measured concentrations to modeled concentrations. (Note, these two tasks are required for the BMC implementation as well.) Following the ideas outlined in the previous paragraphs, successive samples of parameters values would be drawn from the conditional distributions,  $p(O|Y)$ , as defined by the likelihood function. In essence, MCMC samples

strategically from among the possible combinations of parameter values and generates the posterior distributions,  $p(Y|O)$ , where  $Y$  reflects the parameter space and  $O$  the observations.

The likelihood function would rely on the predictions from a transport model, which could be made either prior to any event, or in real time. The advantage of executing the transport model in real time is that pre-event computation requirements are reduced. In addition, pre-event transport modeling of all possible parameter combinations may not be desirable. Avoiding pre-event modeling was an attractive feature of MCMC, as compared to BMC. However, given that the Gibbs sampler often requires many iterations (on the order of 1000's) to reach convergence, it may be impractical to perform contaminant transport modeling in real time.

In the case where the transport model is executed prior to the release, there are still noteworthy differences between MCMC and BMC. Specifically, the MCMC method selectively samples from the “library”, where BMC samples comprehensively. The significance here is that MCMC targets the more representative section of the library while BMC considers every sample in the library regardless of the relevance of the sample to the present conditions. The selectivity of MCMC renders it more efficient.

#### **2.4.6.2 Sequential Monte Carlo methods**

Sequential Monte Carlo methods (SMC) constitute a class of non-iterative parameter estimation methods in which observations arrive sequentially in time, and statistical inference is performed in real time. SMC methods have been used in process engineering, target tracking, signal processing, and financial modeling (Doucet et al., 2001). In target tracking, *dynamic* parameters are estimated in real time. For these

real-time applications, MCMC methods are widely viewed as inappropriate (Gilks and Berzuini, 2001), as they require a large number of simulations which cannot be practically implemented in real time in most cases.

As with BMC and MCMC, SMC methods combine observations with prior distributions to generate updated posterior distributions. While MCMC is iterative, SMC is non-iterative and, thus, more appropriate for real-time applications. Doucet et al. (2000) describe an SMC recursive equation by rearranging Bayes' rule.

$$p(x_{0:t+1} | y_{0:t+1}) = p(x_{0:t} | Y_{0:t}) \frac{p(y_{t+1} | x_{t+1})p(x_{t+1} | x_t)}{p(y_{t+1} | y_{0:t})} \quad (2.17)$$

where  $x$  and  $y$  are vectors; the subscript (e.g.,  $0:t+1$ ) represents the time index of the data; the most recent posterior distribution,  $p(x_{0:t+1}|y_{0:t+1})$ , is calculated using the posterior distribution from the previous time index,  $p(x_{0:t}|y_{0:t})$ , with the measurements (and model parameters) from the most recent time index,  $y_{t+1}(x_{t+1})$ , and preceding time index,  $y_t(x_t)$ .

Equation 2.17 resembles equation 2.10 because the problem described in this dissertation is also a sequential parameter estimation problem: concentration measurements arrive sequentially, and properties describing the contaminant release are estimated as new data become available. The BMC algorithm, as it is implemented in this dissertation, is a sequential type Monte Carlo method. It is relevant to understand the commonalities and differences between the BMC and SMC methods discussed in the literature. Neither method is described in a common literature since they are typically used in different application domains.

SMC methods typically use some form of importance sampling. Importance sampling is efficient when the sample distribution closely matches the target distribution. In the contrary case, the samples may not derive from relevant areas of the target

distribution. The problem of selecting an appropriate sample distribution goes to the heart of the problem of implementing importance sampling. As noted, the BMC algorithm is an importance sampling method in which the prior distribution is the sampled distribution. When applying the BMC algorithm, a poorly constructed prior will result in a situation where some elements in the library may result in very large weights, while others have insignificant weights and contribute very little to the posterior distribution.

While it is not possible to construct a sample distribution that equals the target distribution (otherwise, samples would be drawn from the target distribution directly), it is important to not generate sample distributions in an ad-hoc manner. Hence, the overall goal in improving SMC methods that derive from importance sampling is in selecting a sample distribution in such a way that the variance of the importance weights is small. The term “degeneracy” is often used in the SMC literature to describe the phenomenon where the variance of the weights becomes very large and only a few samples contribute in a significant way to the posterior distribution. Arulampalam et al. (2002) present a nice discussion of the degeneracy issue and SMC methods. To “measure” degeneracy, Liu (2001) proposes the following heuristic for calculating the effective sample size (ESS), based on the variance of the sample weights,  $w(x)$ :

$$ESS(m) = \frac{m}{1 + \text{var}(w(x))} \quad (2.18)$$

where  $m$  is the number of samples drawn from the sample distribution. Such heuristics are useful because there is no statistical definition of an “efficient” importance sampling method. In the BMC method, degeneracy would result in a very small number of library realizations contributing to the overall posterior distributions.

Liu (2001) discusses different approaches to improving the SMC algorithms, one of which is known as resampling. With resampling, samples that have very low weights (or low representation of the posterior distribution) are discarded and replaced with samples from regions with higher weights. This resampling can be executed at any time interval. In BMC, a procedure known as Bayesian windowing is somewhat similar. There, a subset of realizations that have very low posterior probabilities are discarded, and the remaining realizations are set to equal probabilities (Sohn et. al., 2000). Bayesian windowing differs from resampling because discarded realizations are not replaced with new realizations.

There is no one “optimal” resampling rule. Some of the methods described by Liu are inspired by the non-iterative sampling-importance resampling (SIR) algorithm by Rubin. This algorithm was originally presented in Rubin (1987) and has been described subsequently in many statistical texts such as Gelman et al. (2004).

Gilks and Berzuini (2001) describe an improved SMC method they term the resample-move algorithm for estimating dynamic parameters. It is inspired by the concepts introduced by resampling and MCMC sampling, but exhibits less degeneracy. In another hybrid approach, Godsill and Clapp (2001) combine importance sampling with MCMC resampling techniques to produce a new SMC method.

In summary, the class of SMC methods may offer some insight into measuring or improving the efficiency of the BMC algorithm, as the problems associated with the SMC methods (e.g., degeneracy) are analogous to the problems that may be encountered in implementing the BMC method.

## 2.5 Closing comments

While this background chapter describes a wide array of tools that are available, a select few are used in this research. This dissertation takes the perspective that simulation and statistical tools are to be used to support the system design research questions, rather seeking to demonstrate the use or development of tools and algorithms alone. However, this research constitutes early work that may be followed by subsequent research efforts in environmental sensor systems. I anticipate that hybrid or more sophisticated tools may be useful in future research efforts. It is my intention that this chapter provides a sufficiently thorough background to facilitate the next stage of research questions that move beyond this dissertation.



# Chapter 3

## 3 Case study description

The algorithms in this dissertation are evaluated using experimental data obtained in a real building. This case study should be viewed as the “experiment” in this research. A description of the buildings, experiments and contaminant transport models are provided in this chapter.

The experiments were conducted in one unit of a multi-unit building located at the Dugway Proving Grounds, Utah (Sextro et al., 1999). The study unit comprises 660 m<sup>3</sup> of interior volume and approximately 280 m<sup>2</sup> of floor area on three levels (Figure 3.1). For consistency, I retain the room names used by Sextro et al.

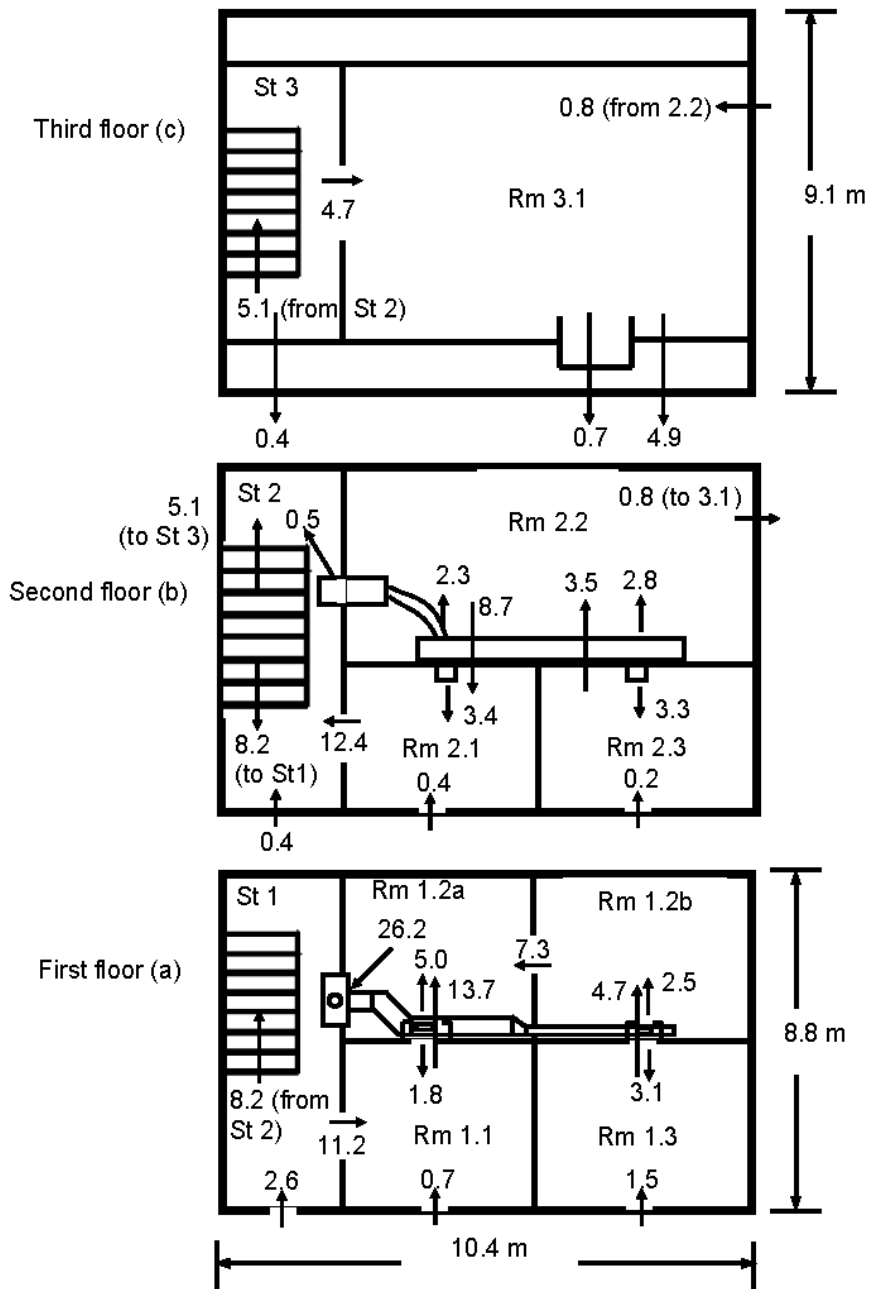


Figure 3.1. Plan view of case study unit. (a) Floor 1; (b) Floor 2; (c) Floor 3. Selected intrazonal and AHU airflow rates for the tracer release of Experiment 1 (as calculated by COMIS at the first time step) are shown, in units of  $\text{m}^3/\text{min}$ .

A mechanical air-handling unit (AHU) supplies air to the first and second floors, and the intake to the return duct is located on the 1<sup>st</sup> floor within Room 1.2a. The AHU is a 100% recirculating unit (i.e., there is no deliberate outside air intake). Air exchange between the inside and outside occurs by means of pressure-driven airflow through leaks in the building envelope. Airflow between interior zones occurs by mechanically induced flow through the system ducts, by pressure-driven or convective airflow through open doorways and wall grilles, and by pressure-driven airflow through interior leakage paths.

Fan pressurization testing was conducted to measure leakage characteristics for the purposes of calibrating the transport model. Subsequently, twelve tracer experiments were performed over a 5-day period. Sextro et al. developed a contaminant fate and transport model for these releases using the COMIS multizone model.

Both the experimental data and the original, base multizone models were used in this dissertation. Specifically, the experimental tracer data were used to simulate sensor network signals from hypothetical releases by selecting and modifying the data to represent different sensor placements and performance characteristics. The multizone models were used in the sensor interpretation algorithms, as described in Chapters 4-6.

The remainder of this chapter describes the experiments, airflow patterns and model generation. Additional details are presented in an appendix (§3.A).

### 3.1 Experiments

Sextro et al. (1999) conducted fan pressurization tests to determine the leakage characteristics of the building. From these fan pressurization tests, the leakage characteristics of the different flow paths in the building were determined.

Measurements of airflow,  $Q$ , and pressure drop,  $\Delta P$  were used to calculate values of the airflow parameters  $n$  and  $C_Q$  of the power law (equation 2.1). In the case study experiments, the parameter  $n$  was set to a fixed value of 0.61, and values of  $C_Q$  were determined. The leakage characteristics determined for all flow pathways are reported in the appendix (§3.A).

Measurements were also conducted to characterize the AHU airflows. The AHU air supply rates were measured with all doors open, including the entrance door into the unit, and pressure drops over the supply and return grilles were on the order of 20 Pa. These measurements are also reported in the appendix.

Subsequent to the fan pressurization tests, twelve tracer experiments were conducted at various times of day, spanning a period of 5 days. Approximately 20 g of propylene was released instantaneously at a single location in each experiment. Measurements were taken at 20 s intervals for the first 20 min of the release, and at 60 s intervals thereafter. Each room and staircase level contained a sensor that was centrally located. The propylene concentrations were measured using photo-ionization detection (PID) sensors.

The AHU was operating for 11 of these experiments, and was off for the twelfth. The interior door positions were varied across all experiments, using 3 different door

position combinations. Table 3.1 shows the operating conditions during the tracer experiments.

Table 3.1 Description of twelve tracer-gas experiments

Experiment number <sup>b</sup>	Release location	Door position	AHU status
1, 5, 10	Return duct	All doors open	On
3, 8, 12	Return duct	All doors closed	On
2, 6, 11	Return duct	Stairwell doors closed, other interior doors open	On
4, 9	Room 1.3	Stairwell doors closed, other interior doors open	On
13	Room 1.3	Stairwell doors closed, other interior doors open	Off

a There was no experiment numbered 7.

During the tracer experiments, interior temperatures were recorded every 30 s. Weather conditions (outside temperature, wind direction, wind velocity, and barometric pressure) were measured every 15 min. The appendix (§3.A) reports interior temperatures and weather conditions data for all the experiments.

### 3.2 Airflow patterns and tracer analyses

Airflows were not directly measured during the tracer experiments. Instead, they can be obtained from the transport model. Figure 3.1 displays the modeled airflows from Experiment 1. The overall airflow patterns in, out, and within the building are shown.

The airflow between the building and the surrounding environment, (i.e., infiltration and exfiltration) is dictated by the relative indoor to outdoor air temperatures, leakage characteristics of the envelope, and wind-induced pressures. The AHU recirculates

indoor air, and thus, does not actively play a role in ventilation (though it does in an indirect way as it alters the pressure distribution within the building). An important driving force is the stack effect. When the outdoor conditions are cooler than indoors, air enters through leaks in the 1<sup>st</sup> floor envelope, and exits from the 3<sup>rd</sup> floor. The staircase provides an easy pathway for this airflow to follow, and outside air will enter the interior rooms to a degree that depends on the positions of the staircase doors. When outdoor conditions are warmer than indoors, the stack flow is reversed, with air entering at the 3<sup>rd</sup> floor, and exiting at the 1<sup>st</sup> floor.

The airflow within the test unit is largely determined by the operation of the AHU, whenever it is on. The AHU contains only one ducted return pathway, located on the 1<sup>st</sup> floor. Because of this return, when the AHU is on, the 1<sup>st</sup> floor operates at a lower pressure than the exterior and 2<sup>nd</sup> floor. Hence, air flows from the 2<sup>nd</sup> floor into the 1<sup>st</sup> floor. The airflow to and from the 3<sup>rd</sup> floor depends on the stack effect. Within the 1<sup>st</sup> floor, owing to the location of the return intake, air flows from Rooms 1.1 and 1.3 to Rooms 1.2 and 1.2a. Furthermore, the diffusers supplying airflow to Rooms 1.1 and 1.3 are horizontally oriented, and located in the doorway, which may facilitate the airflow to the adjacent rooms, 1.2a and 1.2b. On the 2<sup>nd</sup> floor, the researchers noted that air flows from Room 2.1 into the staircase, therefore drawing air from Room 2.2, and also from 2.3 into 2.2.

Based on the fan pressurization tests, it was evident that leakage between the interior and exterior for the first floor was relatively small, but that it was comparatively large for the second and third floors. The large leakage rates between the rooms on the first floor

are due to the presence of the air return registers, while the large leakage rates between the rooms on the second floor are due to large cracks in the floor and the walls.

Figures 3.2 – 3.5 show the propylene measurement results for Experiments 1, 4, and 13. Data are shown for this subset of experiments because they represent the greatest variability in experimental conditions and also are representative of the concentrations observed in the other experiments. Among these three experiments, there are 2 different release locations (i.e., the return intake in Experiment 1, and Room 1.3 in Experiments 4 and 13), and the AHU was both on and off (i.e., Experiments 1, 4: on; Experiment 13: off).

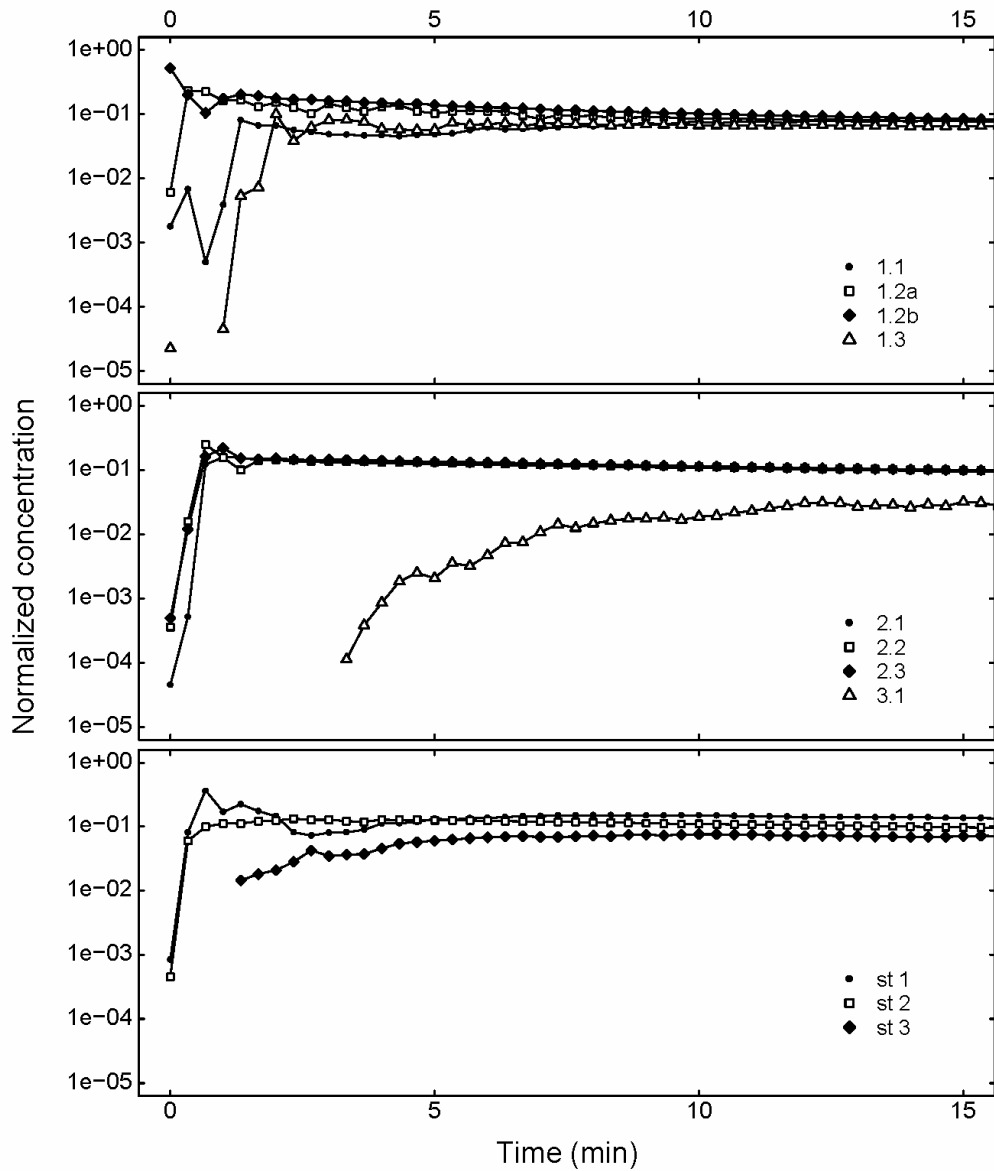


Figure 3.2. Time-dependent normalized tracer-gas concentrations for Experiment 1. Concentrations are normalized by theoretical peak concentration in the experiment; time is in reference to the time of the release; upper frame: first floor rooms; middle frame: second floor and third floor rooms; lower frame: staircases.

For Experiment 1, in which the release occurred at the return intake, the concentrations rise sharply in Rooms 1.2a, 1.2b, and all of the 2<sup>nd</sup> floor rooms, before decaying. This observation is intuitive, as these rooms are directly served by the AHU.



Relatively high concentrations also are observed in Stair 1, most likely owing to leakage from the AHU duct. The concentrations in Rooms 1.1 and 1.3 reach their peak values within a few minutes, but after the other rooms served by the AHU. An explanation is that a fraction of the supplied airflow to these rooms is diverted to the Rooms 1.2a and 1.2b. Therefore, less contaminant is transported to these rooms initially, and more time is required for the contaminant to reach the sensors, which were centrally located, in these rooms. The concentrations in Room 3.1, Stair 2, and Stair 3 increase toward their peak values more gradually, as these zones are not directly served by the AHU. The concentration profile is not smooth – that is, there are fluctuations, particularly for Rooms 1.2a, 1.2b, and Stair 1. These fluctuations indicate that well-mixed conditions are not instantaneously met within these zones. The fluctuations cease at ~ 5 minutes, which suggests that the mixing time for the respective zones is on this order.

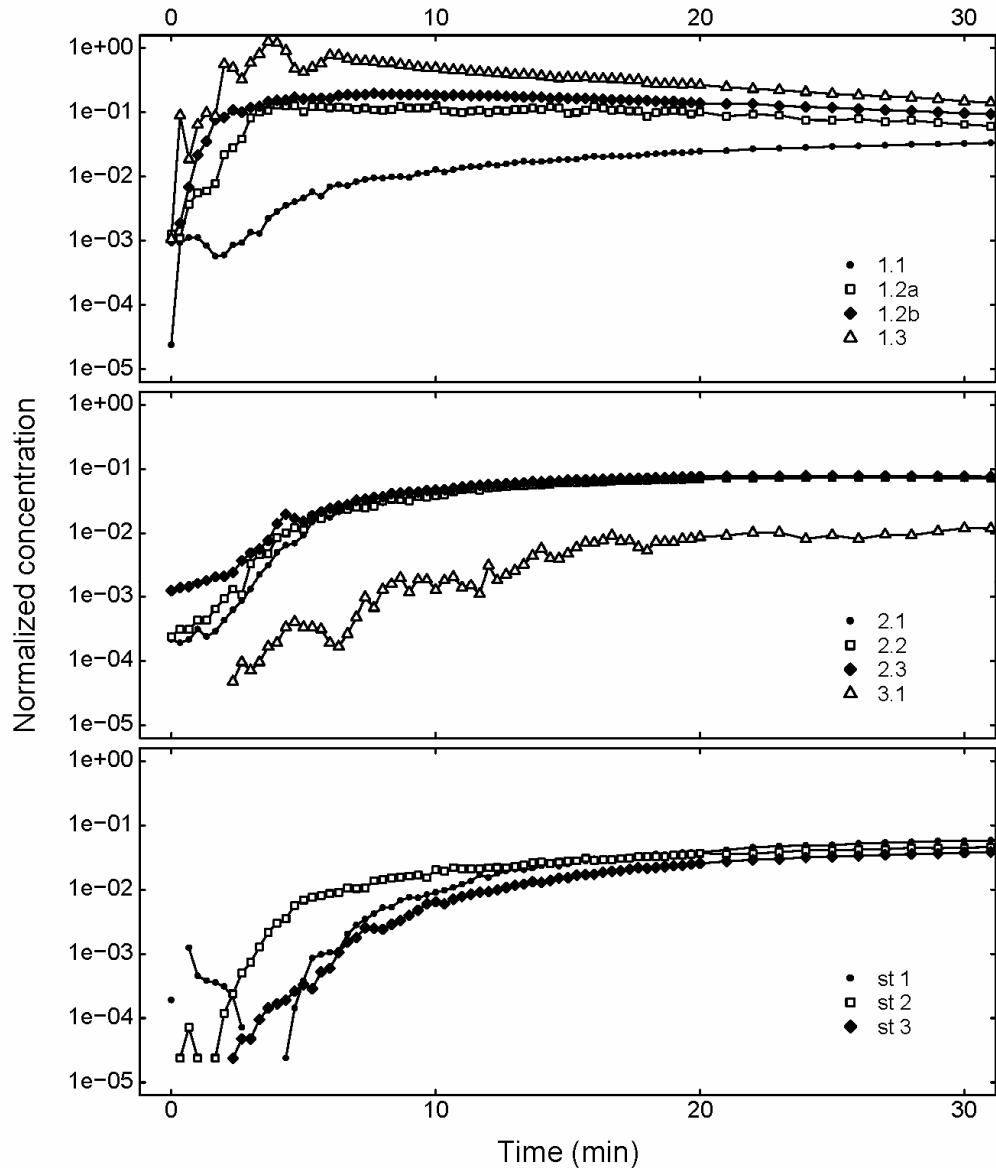


Figure 3.3. Time-dependent normalized tracer-gas concentrations for Experiment 4. Concentrations are normalized by theoretical peak concentration in the experiment; time is in reference to the time of the release; upper frame: first floor rooms; middle frame: second floor and third floor rooms; lower frame: staircases.

In Experiment 4, the contaminant was released into Room 1.3. The highest concentrations are observed in the release room; in fact, the peak concentration in this room exceeds the peak concentration of all other zones by almost an order of magnitude. Among the non-release zones, the highest concentrations are observed in Rooms 1.2a and

1.2b. For the contaminant to be circulated to the other zones in the building, it must first be transported to the adjacent Room 1.2b, and then to Room 1.2a, before entering the intake to the AHU. Concentration fluctuations for Room 1.3 cease by ~ 8 minutes.

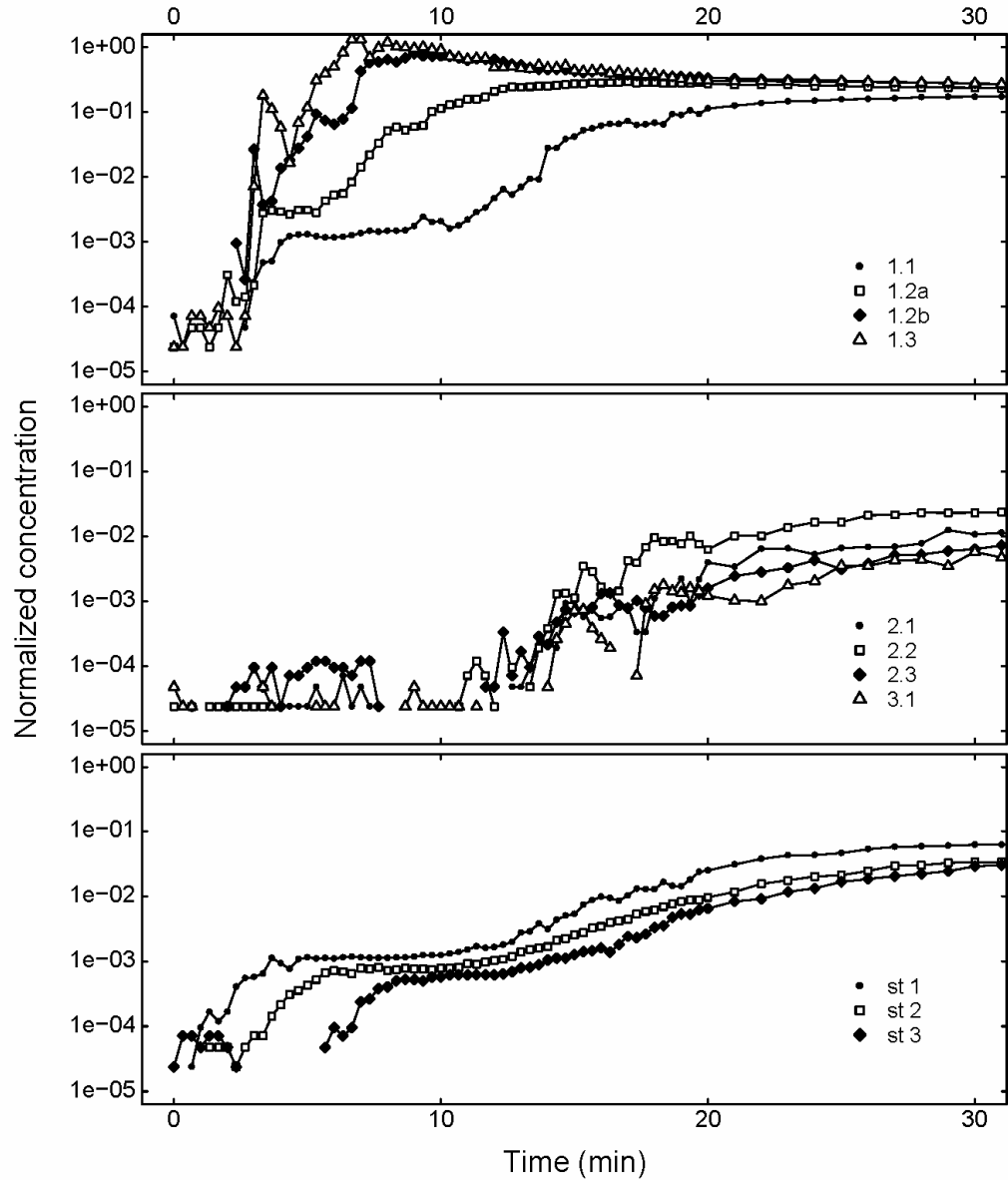


Figure 3.4. Time-dependent normalized tracer-gas concentrations for Experiment 13. Concentrations are normalized by theoretical peak concentration in the experiment; time is in reference to the time of the release; upper frame: first floor rooms; middle frame: second floor and third floor rooms; lower frame: staircases.

In Experiment 13, the AHU was off, and the contaminant was released in Room 1.3. The concentration profiles in this experiment are unique, mainly because the intrazonal transport is relatively slow, since all airflow occurred by nonmechanical means. The highest concentrations are observed in the release room, as expected. Since Room 1.3 communicates directly with Room 1.2b, this room experiences relatively high concentrations as well. Similarly, because Room 1.2b communicates with Room 1.2a (through an open pathway), the concentrations rise in Room 1.2a after several minutes. The contaminant eventually reaches Room 1.1, much later, and with a peak concentration an order of magnitude less than the peak concentration in the release room. In this experiment, the staircase doors were closed, inhibiting airflow from the main rooms to the stairwell. Because the stairwell is also the greatest pathway for vertical transport among floors, the concentrations in the 2<sup>nd</sup> and 3<sup>rd</sup> floors are significantly less than those on the first floor.

Figures representing the concentrations for the remaining experiments are included in the appendix (§3.A). The overall results do not differ significantly compared to the variability observed among Exps 1, 4, and 13. However there are some noticeable differences in the concentrations observed in the stairwell and on the 3<sup>rd</sup> floor. For example, if Experiment 2 is compared to Experiment 1, it is clear that the closed staircase doors inhibit contaminant transport to the staircase. Therefore, lower concentrations are observed in Experiment 2 in the stairwell. Because the stairwell is not actively served by the AHU, it is more greatly influenced by changing temperature and weather conditions and different interior door positions.

### **3.3 Multizone model development**

Sextro et al. (1999) used the COMIS multizone model to simulate the contaminant fate and transport in the test unit. The airflow characteristics (including the AHU characteristic fan curve), AHU operating status, door positions, temperature and weather conditions were incorporated into the model, and the model performance was evaluated for the twelve tracer experiments.

Each room in the test unit, each level of the staircase, the AHU return, and the supply were represented as individual zones. Figures 3.5-3.7 show the model-to-measurement fit for Experiments 1, 4, and 13. Figures for the other experiments are included in the appendix (§3.A).

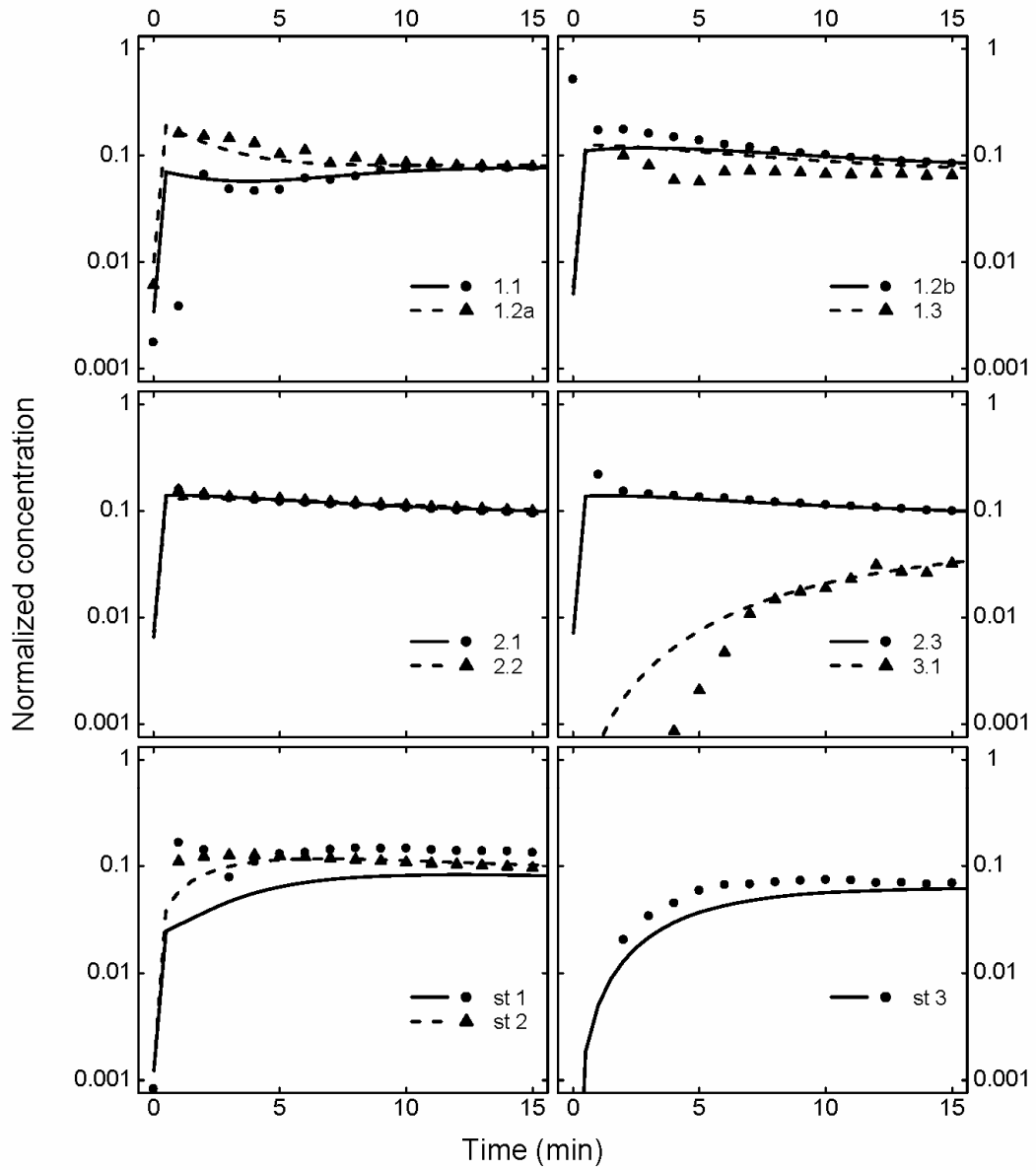


Figure 3.5. Measurements compared with multizone model predictions for Experiment 1. For clarity, every 3<sup>rd</sup> data point is plotted.

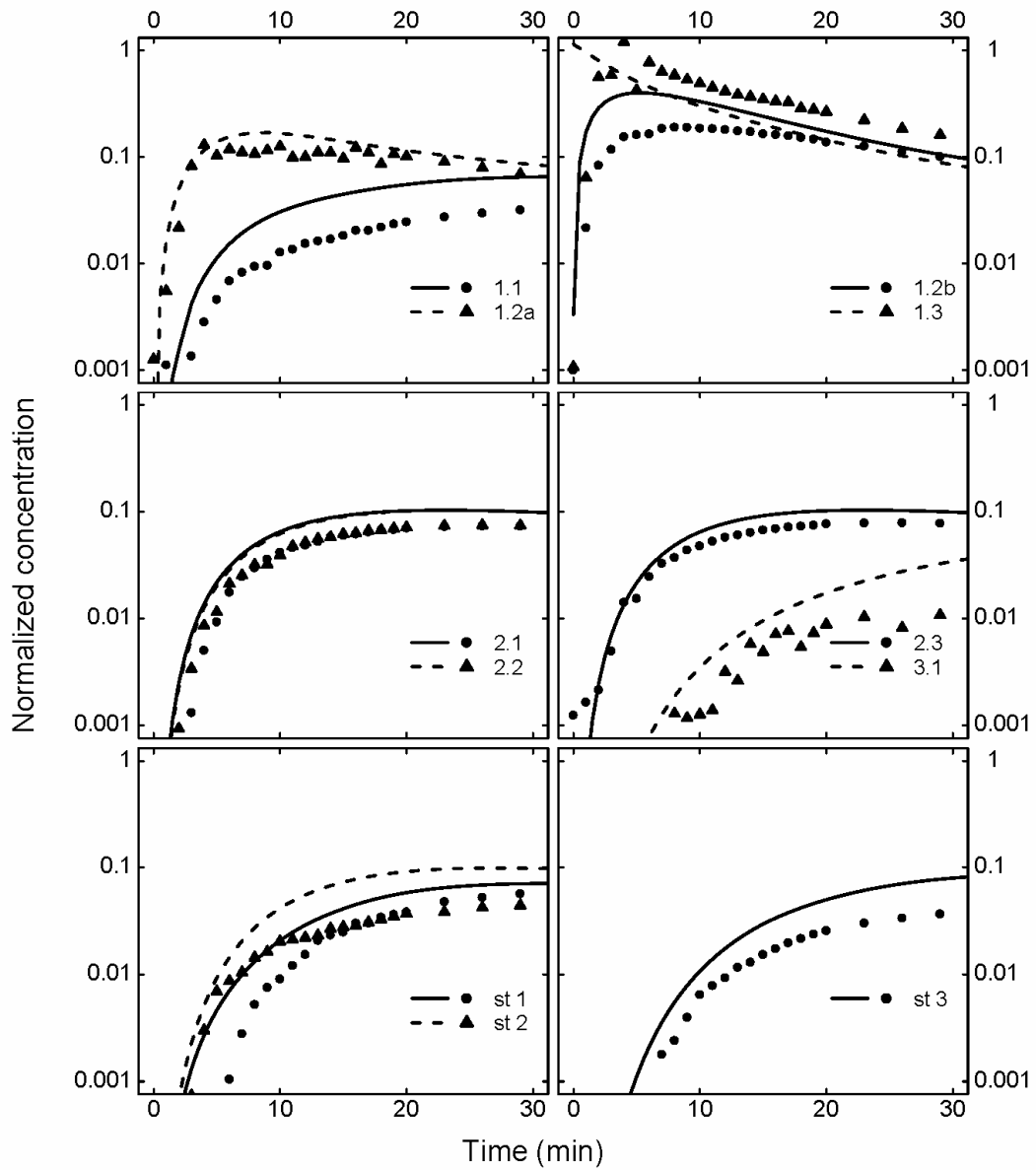


Figure 3.6. Measurements compared with multizone model predictions for Experiment 4. For clarity, every 3<sup>rd</sup> data point is plotted.

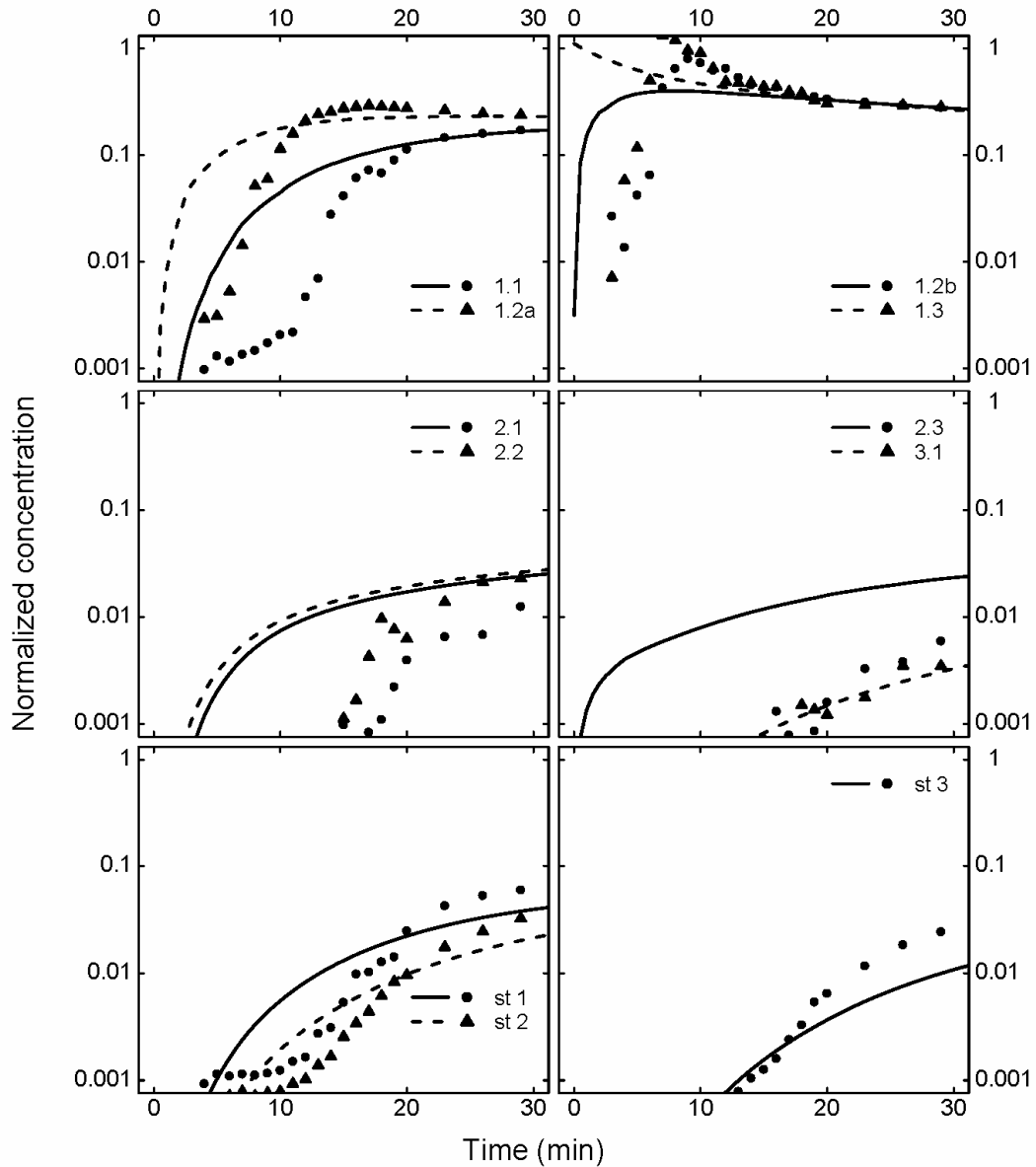


Figure 3.7. Measurements compared with multizone model predictions for Experiment 13. For clarity, every 3<sup>rd</sup> data point is plotted.

For Experiment 1 the model has difficulty matching the measured concentrations in some zones. The multizone model underpredicts the initial concentrations in Room 1.2b (see Figure 3.8). The orientation of the supply diffuser to Room 1.3 may be responsible for the higher concentrations in Room 1.2b. Specifically, the diffusers that supply air to



Rooms 1.1 and 1.3 are oriented horizontally in the “door nooks” (i.e., the diffuser plane is parallel to the floor), in contrast to a vertical orientation that characterizes the other supply diffusers in unit. Given that the inter-room airflow pathway is from Room 1.3 to Room 1.2b (when the AHU is operating) it is possible that the orientation in combination with the airflow pathway results in a “short circuiting” that prevents the contaminant from being adequately mixed in Room 1.3 before being transported to Room 1.2b. The model would be unable to account for this type of effect, as it assumes instantaneous perfect mixing.

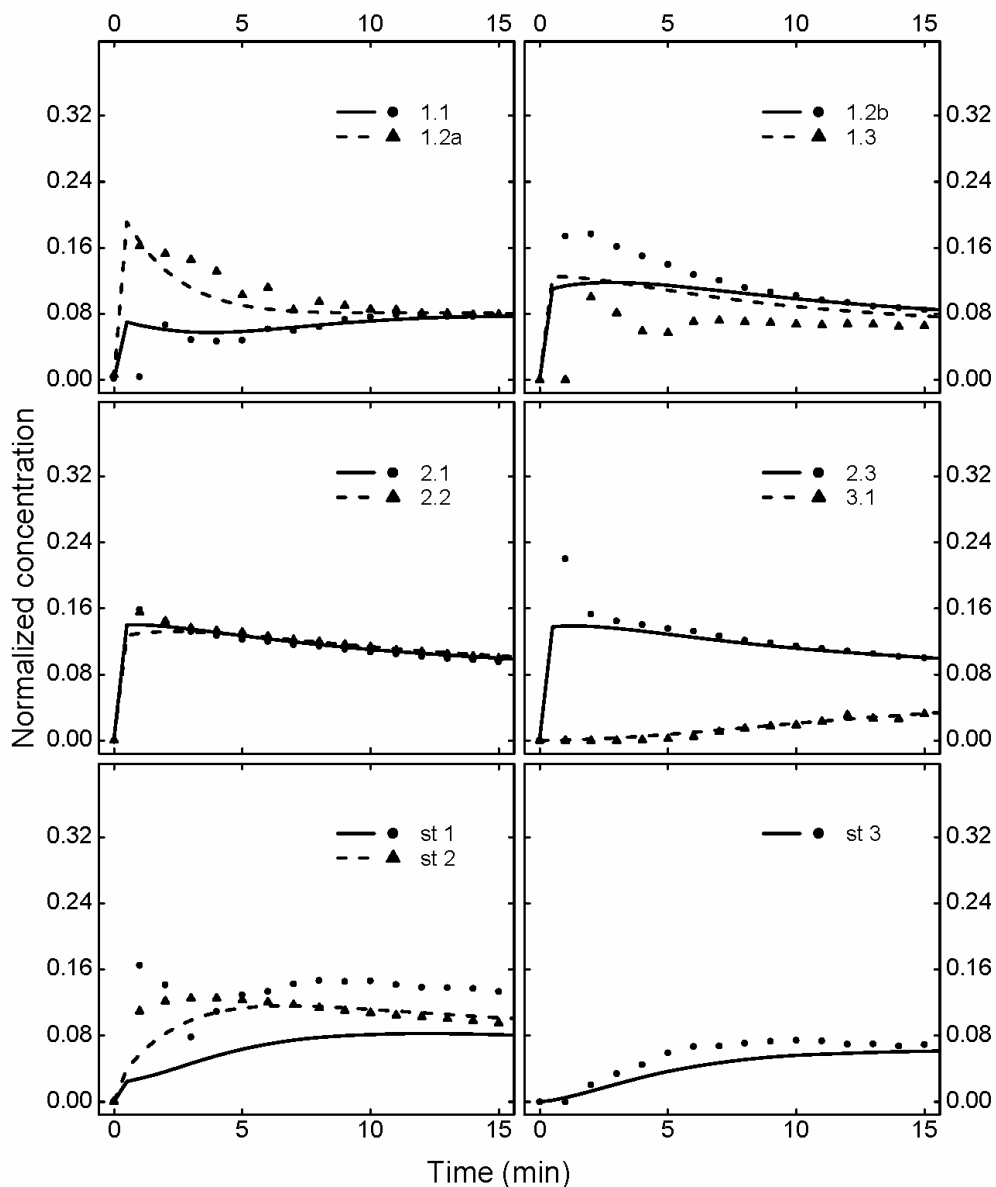


Figure 3.8. Linear scale plot comparing measurements and model predictions for Experiment 1. For clarity, every 3<sup>rd</sup> data point is plotted.

Concentrations in Room 1.3 are substantially overpredicted at the early part of the experiment. There are two possible explanations for this. (1) There is a lag time for the contaminant in the room to reach the sensor (as evidenced by the first several measurements) owing to the slow within-room mixing. (2) The horizontal orientation of

the diffusers may divert some airflow away from this room such that the contaminant is not adequately mixed within Room 1.3.

The concentrations in Stair 1 are underpredicted as well. Two possible explanations are that the leakage from the duct is not well represented and that the vertical transport through the staircase is not accurate (which is a recognized limitation of multizone models, as noted in Chapter 2). In addition, if the model miscalculates the stack effect, which introduces outside air into the stairwell, this would have ramifications for all other airflows into that zone. For example, if the leakage into the lower staircase level is overestimated, this could result in an underestimate of the airflow due to duct leakage, as the model has to balance the airflows.

Modeled concentrations in the 2<sup>nd</sup> and 3<sup>rd</sup> floor are relatively accurate for Experiment 1, except for the staircase. This is consistent with the observations in Stair 1, and substantiates the inference that the overall contaminant transport to the staircase is underpredicted.

In some zones, the measured data exhibit fluctuations with time, particularly at the earlier time periods, while the modeled concentrations are smooth. The initial fluctuations in measured concentrations are likely a consequence of the finite duration of the mixing process occurring within each room. At later times, the modeled concentrations may not match the measurements because of errors in the airflow calculations. The concentration profiles produced by the multizone model are smooth because the model assumes that the contaminant is always well-mixed within a zone.

The results for Experiment 4 (see Figure 3.9) exhibit reasonably good model-to-measurement fits for most rooms but for two: Rooms 1.3 and 1.2b. Room 1.3

is the release room and Room 1.2b is adjacent. The main exit pathway for contaminant from Room 1.3 is through Room 1.2b, before being transported to the remainder of the building. The model assumes instantaneous mixing of the contaminant within the release room, and consequently overpredicts transport to Room 1.2b. Viewing Figure 3.9, it may be inferred that the contaminant requires approximately 8 min to be well-mixed within Room 1.3. Prior to this time, the measured concentrations exhibit fluctuations, as expected for an incompletely mixed room.

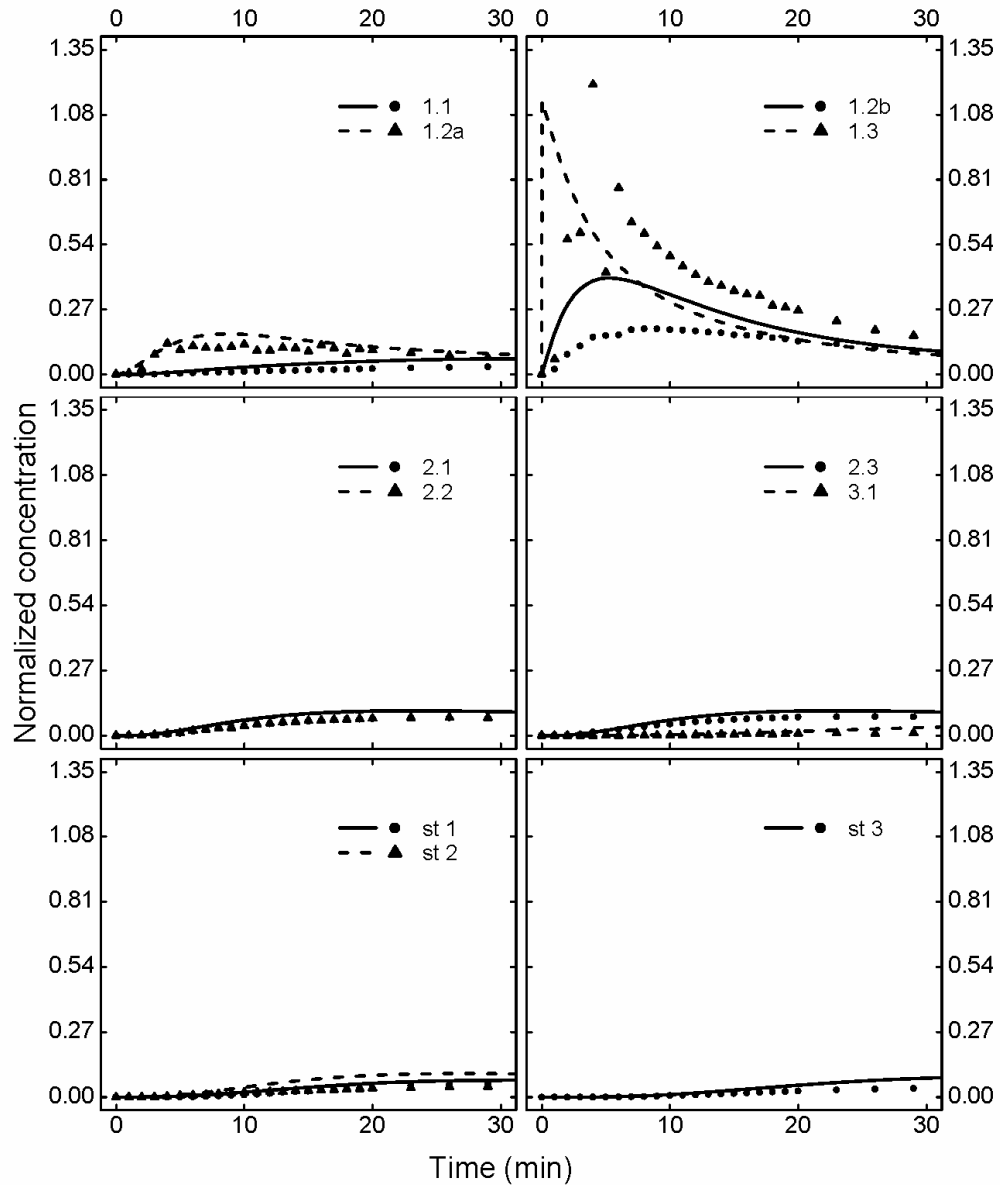


Figure 3.9. Linear scale plot comparing measurements and model predictions for Experiment 4. For clarity, every 3<sup>rd</sup> data point is plotted.

The early overestimation of the rate of contaminant transport to Room 1.2b appears to translate to a bias of model predictions in Room 1.2b that remains for the duration of the experiment. In turn, the rate of transport to the other rooms is also modeled to occur

faster than was observed. This effect, however, results in barely perceptible differences between model and measured values for the remaining rooms.

In Experiment 13, with the AHU off, there are persistent discrepancies between model predictions and measurement results in some rooms. (See Figure 3.10.) All transport is due to buoyancy and wind-induced flow. The model assumes that the contaminant instantaneously mixes in the release room, which is certainly not true. Because the room is not subject to any mechanically induced airflow, the mixing time within the release room is longer than in the other experiments. Several minutes are required for the sensor to detect significant concentrations. In all zones, especially as observed in Room 1.2b, the modeled transport rates are overestimated. Consequently, the model overestimates the initial concentrations in Room 1.2b especially, and to a lesser extent in other rooms.

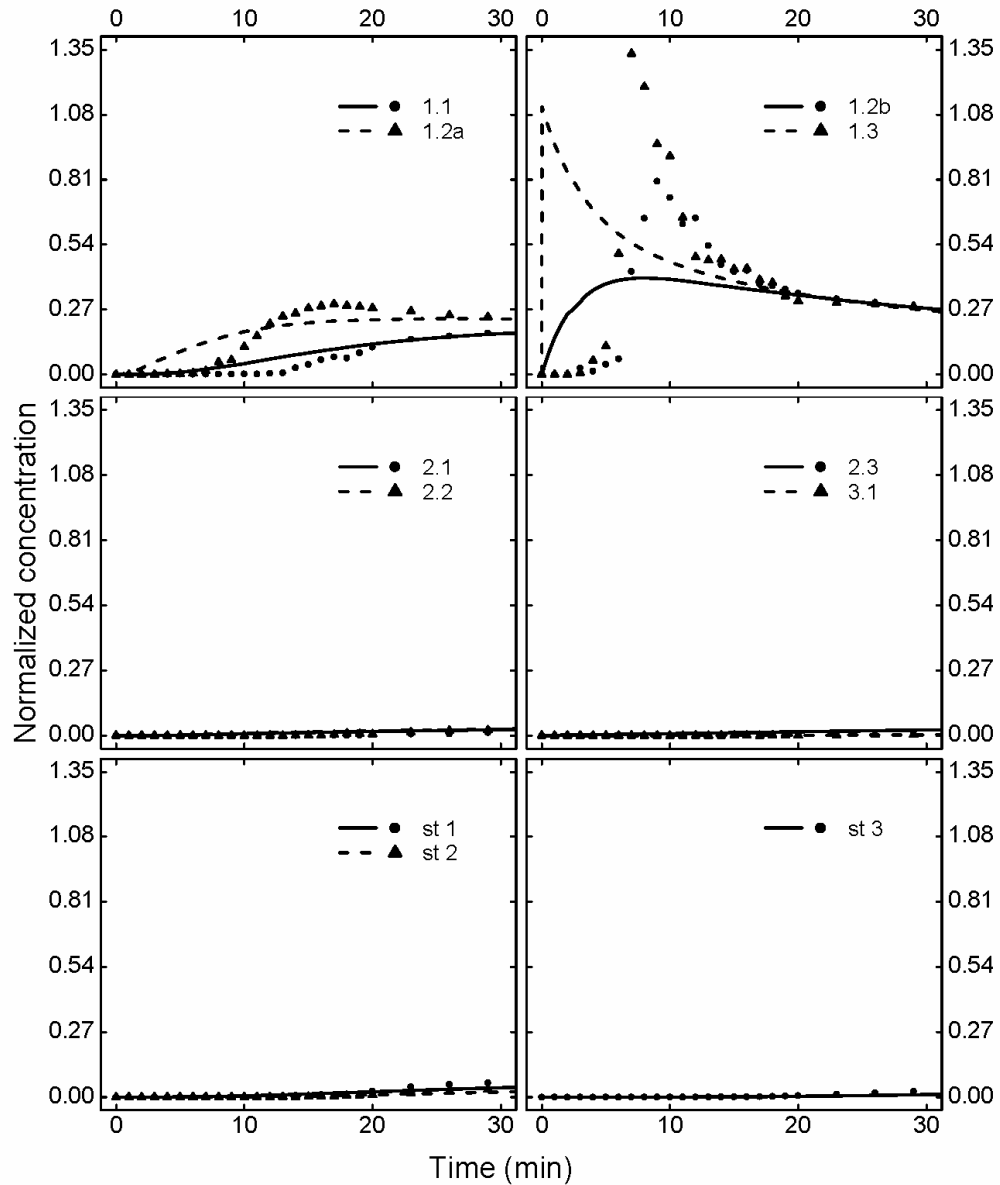


Figure 3.10. Linear scale plot comparing measurements and model predictions for Experiment 13. For clarity, every 3<sup>rd</sup> data point is plotted.

These model runs were used as the base simulations in this dissertation. Further “tuning” of the models was not performed. As discussed in Chapter 2, more detailed modeling approaches, such as CFD are impractical, and network models, such as

multizone models, come with their own set of limitations. The implications of model behavior on sensor system performance are considered in detail in Chapter 5.

### **3.4 Scope of the case study**

Much of the research reported in this dissertation relies on this case study. It is important to understand the unique characteristics of the case study, so as to assess the limitations and broader relevance of the dissertation's findings. I discuss here how the release conditions, building type and scale, HVAC system design, and operation differ from what may be encountered among commercial buildings.

First, the extent to which the findings are relevant to all possible release scenarios in the case study itself is not known. Only twelve tracer experiments were conducted, each of pulse duration. These encompassed only two release locations, two AHU operating conditions (i.e., on and off), with windows all closed, and with three door configurations.

Additional data are needed to ensure that the proposed interpretive algorithms are robust for releases of longer duration and in different locations. Also, while it is conceivable that a release event is likely to consist of a single short-term release, staged multiple releases, or an extended duration release may also occur.

The test space is small, with a floor area of only  $\sim 280 \text{ m}^2$ ; it goes without saying that most buildings of interest for this application are larger. However, the building did consist of an overhead air distribution system, with mechanically driven airflow. In fact, the case study resembles a mixed-mode ventilation system, which includes naturally driven (i.e., stack-driven) airflow and mechanically-driven airflow. Thus, though small,



the test space bears some resemblance, in process, to larger buildings that may be encountered in the field.

There are additional noteworthy characteristics. The case-study site had only one AHU, which was operated at a constant rate in a purely recirculating mode. Most ventilation systems are required by code to deliver a minimum amount of fresh air based on the occupancy of the space (e.g., the California energy standard, Title 24 ([www.energy.ca.gov](http://www.energy.ca.gov)), requires 7 l/s per person or 15 ft<sup>3</sup>/min per person). Economizer systems, which modulate the amount of fresh air to minimize energy use, are becoming increasingly prevalent for energy efficiency reasons. The presence of differing economizer modes for AHU operation may have design ramifications for sensor systems, as addressed in Chapters 5-6.

The presence of only one air delivery device separates this case study from what may be encountered in typical commercial buildings. Besides the existence of a greater number of AHUs, the control and operation of these units may also influence the generalizability of results from the case study. Notwithstanding these details, the design concepts and insights may still be relevant to buildings with multiple AHUs and economizers.

In addition to differences between the case study and practice with regard to the conventional overhead air distribution systems, some buildings may be thermally conditioned using underfloor air distribution systems or radiant systems. The characteristics of contaminant transport can differ markedly in these cases, particularly for underfloor air distribution systems. With alternative distribution systems, a different

modeling approach might be required to characterize contaminant transport, though the sensor interpretation framework may still prove useful.

Not only may the design differ from the case study, but the control and operation of real systems is likely to differ. The control mode in the case study is simple: the AHU was either on or off. Real systems are usually controlled by an EMCS. Commonly, a variable frequency drive controls the speed of the fan, or, in the case of a single speed fan, the fan is cycled on and off under thermostatic control. The EMCS also controls the economizer, if present, to achieve desired levels of temperature and or humidity of the mixed air stream. These are examples of operation that will introduce more complexity into the airflow time history, and hence increase the challenge of properly characterizing contaminant transport. Some of these issues are addressed in Chapter 5.

### 3.A. Appendix: Supplemental case study information

This appendix contains additional information and measurements for the case study. The reported measurements include those taken during the fan pressurization tests (and the derived parameters) and tracer experiments. These include the room volumes (Table 3.A.1), power-law air leakage coefficients (Table 3.A.2), airflow measurements (Table 3.A.3), concentration data from the additional tracer experiments (Figures 3.A.1-3.A.9), interior temperatures (Table 3.A.4), and weather measurements (Table 3.A.5). Figures 3.A.10-3.A.18 illustrate model-measurement comparisons for the tracer experiments displayed in Figures 3.A.1-3.A.9.

Table 3.A.1. Individual room volumes

Room	Volume (m <sup>3</sup> )
Room 1.1	45
Room 1.2a	45
Room 1.2b	45
Room 1.3	45
Room 2.1	45
Room 2.2	95
Room 2.3	45
Room 3.0	190
Stairwell 1	35
Stairwell 2	35
Stairwell 3	35

Table 3.A.2. Leakage characteristics for the test unit

Leakage path	Leakage coefficient, $C_Q$ (m <sup>3</sup> /h @ 1 Pa) <sup>a,b</sup>
Room 1.1 to outside	25
Room 1.1 to staircase	74
Room 1.1 to Room 1.2	202
Room 1.1 to Room 1.3	0
Room 1.1 to Room 2.1	3
Door Room 1.1 to staircase	25
Door Room 1.1 to Room 1.2	43
Room 1.2 to outside	4
Room 1.2 to staircase	28
Room 1.2 to Room 1.3	202
Room 1.2 to Room 2.2	14
Door Room 1.2 to Room 1.3	43
Room 1.3 to outside	58
Room 1.3 to Room 2.3	8
Room 2.1 to outside	41
Room 2.1 to staircase	71
Room 2.1 to Room 2.2	122
Room 2.1 to Room 2.3	65
Room 2.1 to Room 3.1	14
Door Room 2.1 to staircase	102
Door Room 2.1 to Room 2.2	16
Room 2.2 to outside	8
Room 2.2 to staircase	5
Room 2.2 to Room 2.3	136
Room 2.2 to Room 3.1	145
Door Room 2.2 to Room 2.3	117
Room 2.3 to outside	25
Room 2.3 to Room 3.1	8
Room 3.1 to outside	921
Room 3.1 to staircase	197
Room 3.1 door to staircase	30
Staircase to outside	105
Door staircase to outside	10
Duct system to staircase	10
Duct system to Room 1.2	30
Duct system to Room 2.2	12

<sup>a</sup> Exponent in the flow equation is assumed to be  $n = 0.61$  for all pathways.

<sup>b</sup> Measurements per Dugway project report, received from Richard Sextro, Lawrence Berkeley National Laboratory.

Table 3.A.3. Airflow measurements taken in the AHU at the test unit

Register ID	Register flow (m <sup>3</sup> /h) <sup>a</sup>
Supply to Room 1.1	126
Supply to Room 1.2a	170
Supply to Room 1.2b	176
Supply to Room 1.3	218
Supply to Room 2.1	244
Supply to Room 2.2 (above 1.2a)	161
Supply to Room 2.2 (above 1.2b)	199
Supply to Room 2.3	234
Total Supply	1528
Return	1846
Calculated duct leakage	318

<sup>a</sup> Measurements per Dugway project report, received from Richard Sextro, Lawrence Berkeley National Laboratory

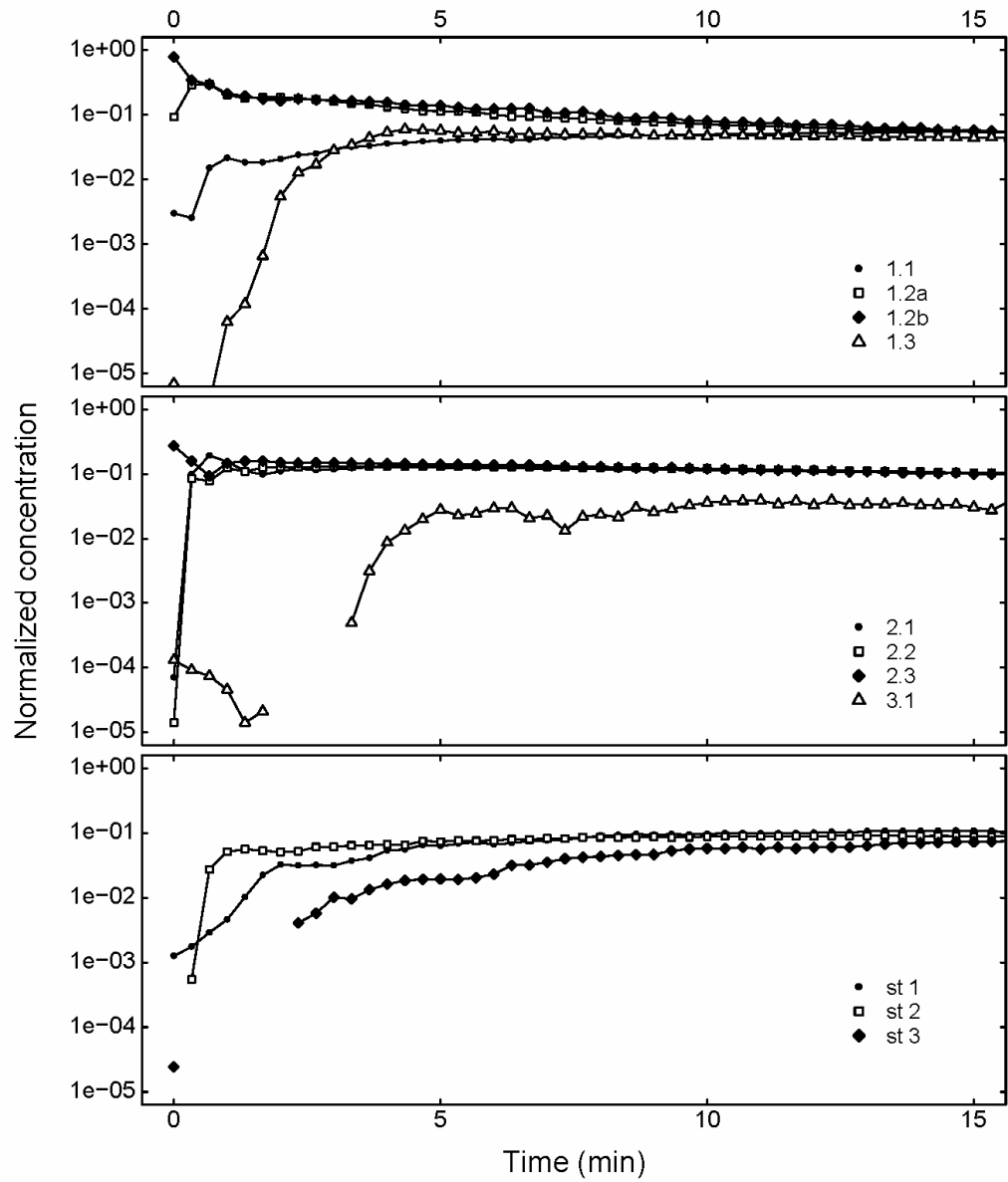


Figure 3.A.1. Time-dependent tracer gas concentrations for Experiment 2.

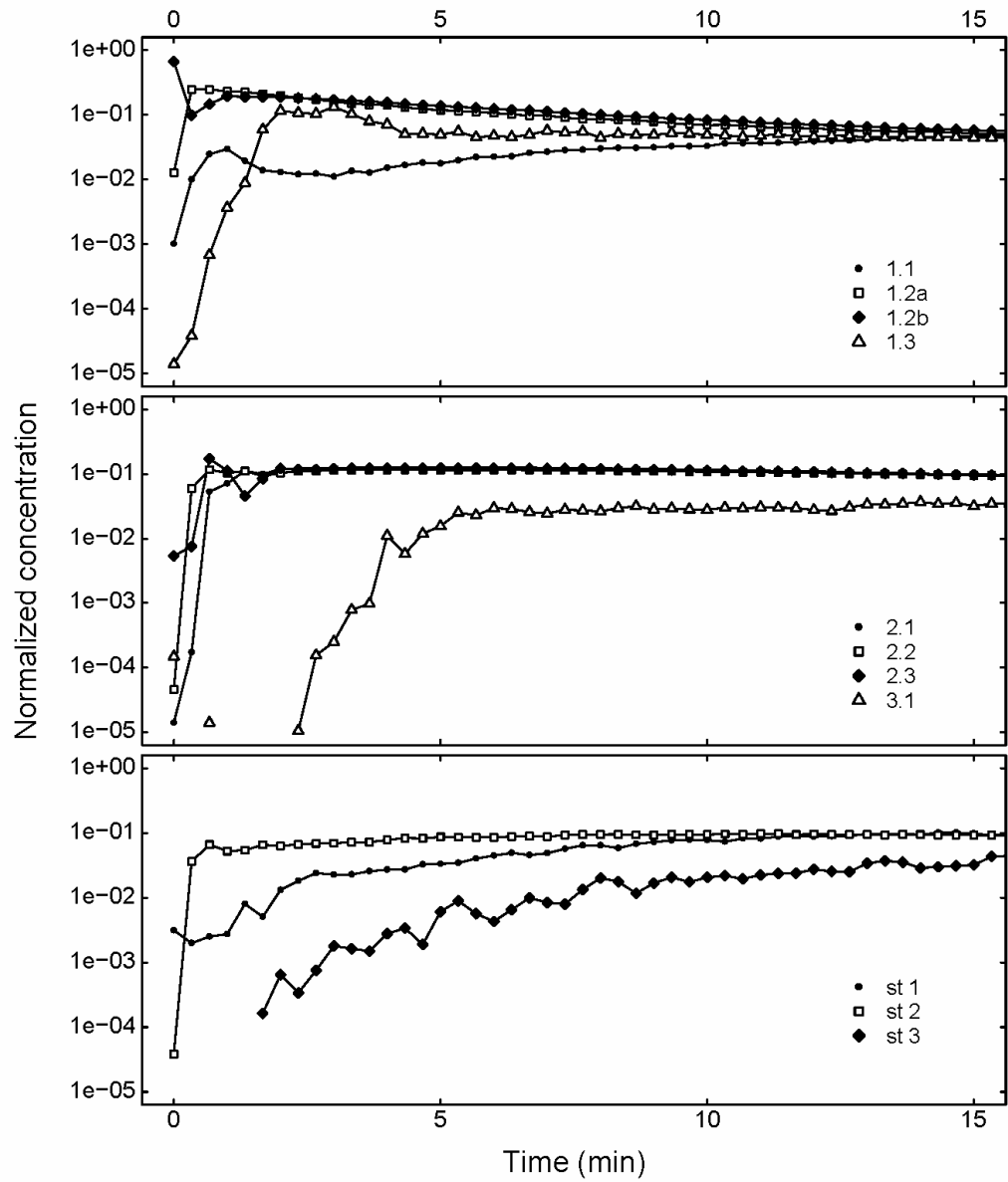


Figure 3.A.2. Time-dependent tracer gas concentrations for Experiment 3.

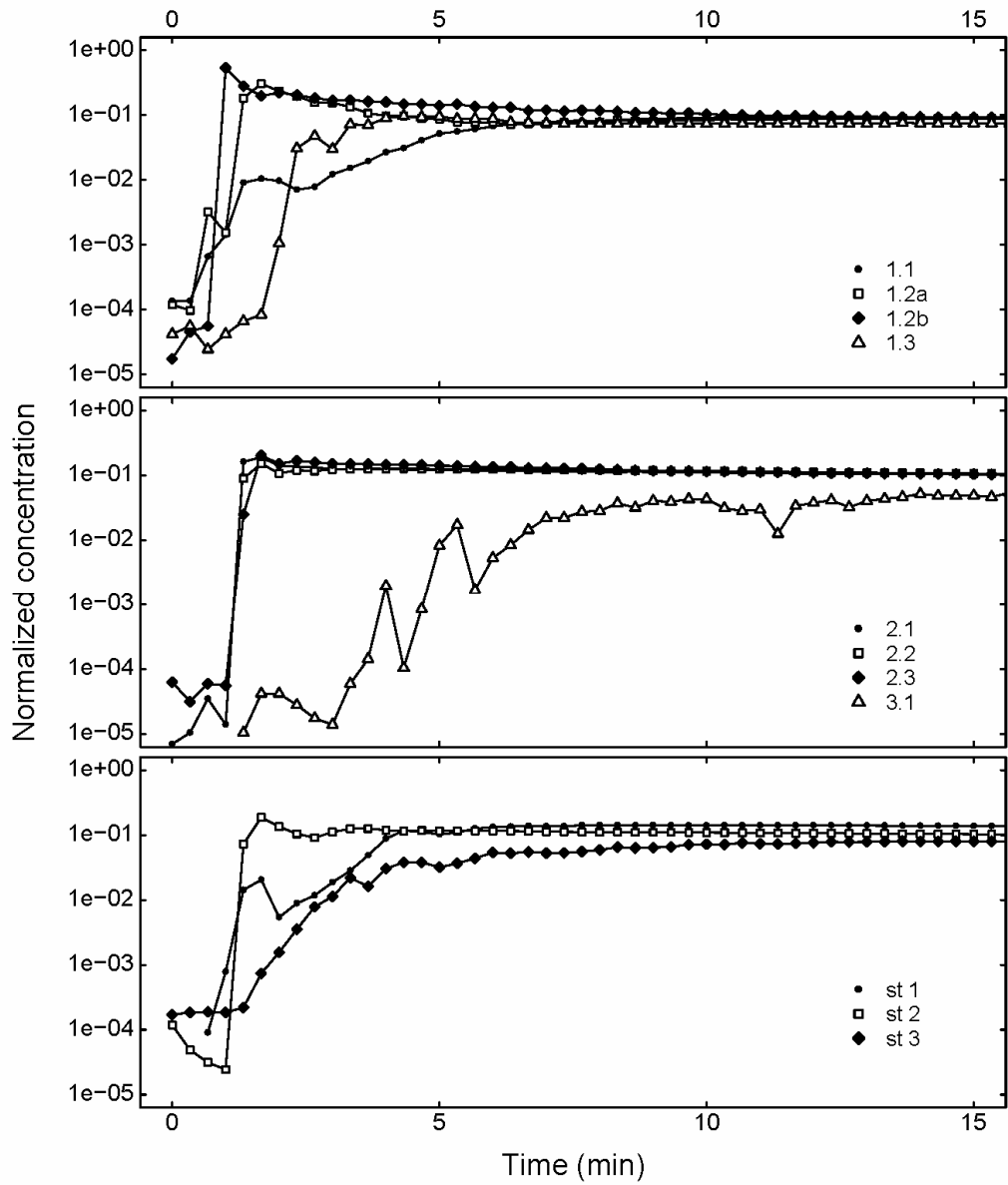


Figure 3.A.3. Time-dependent tracer gas concentrations for Experiment 5.



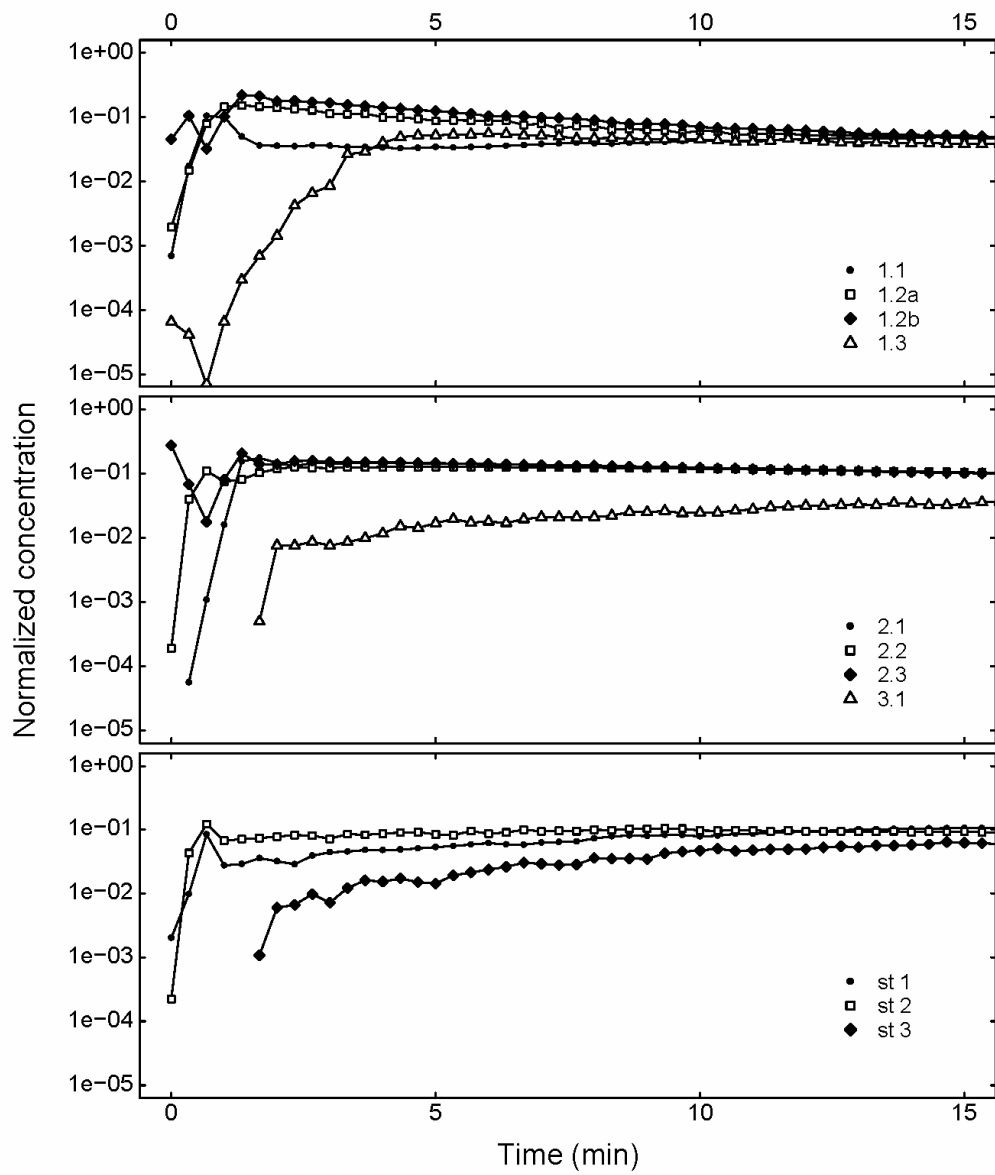


Figure 3.A.4. Time-dependent tracer gas concentrations for Experiment 6.

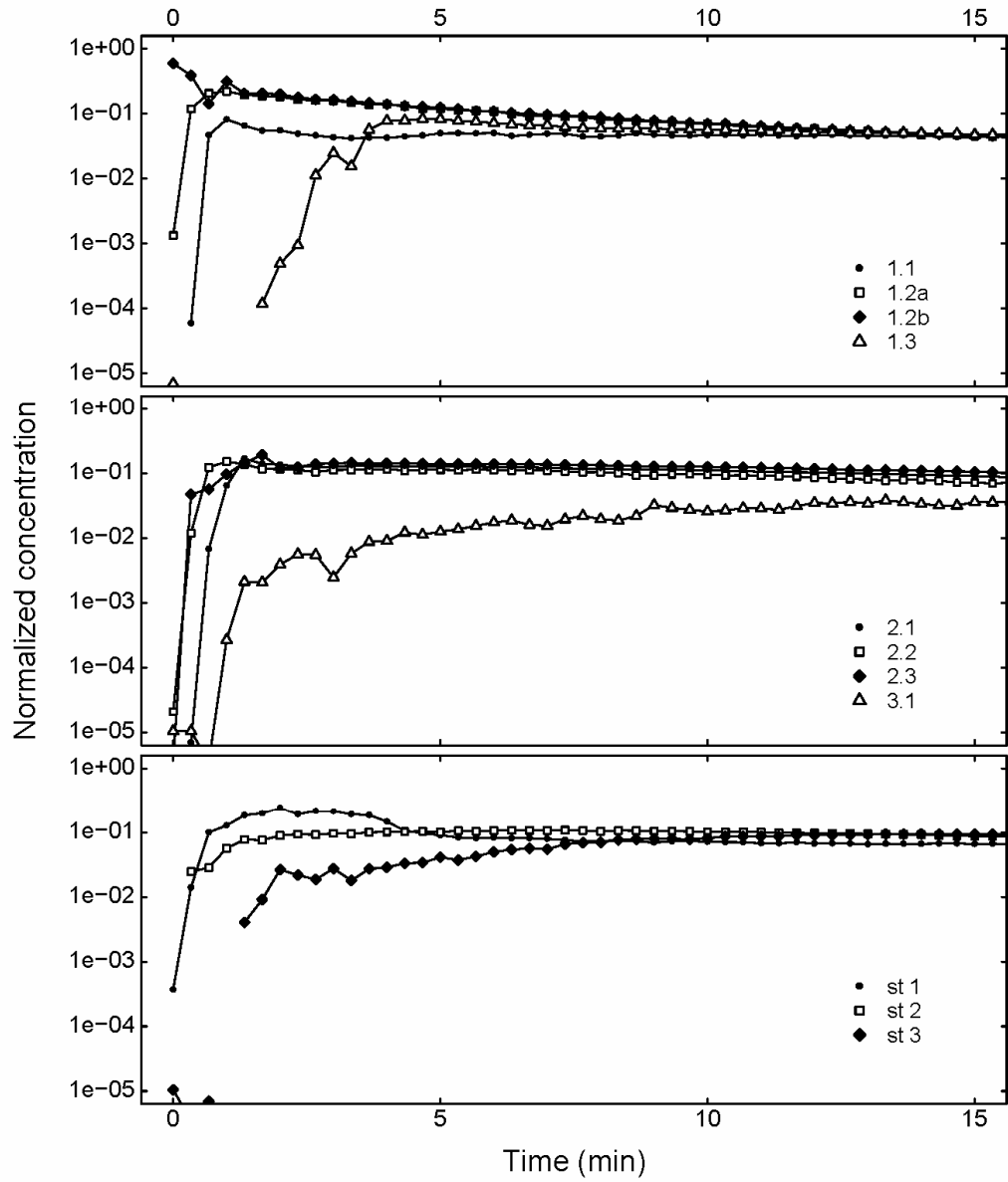


Figure 3.A.5. Time-dependent tracer gas concentrations for Experiment 8.

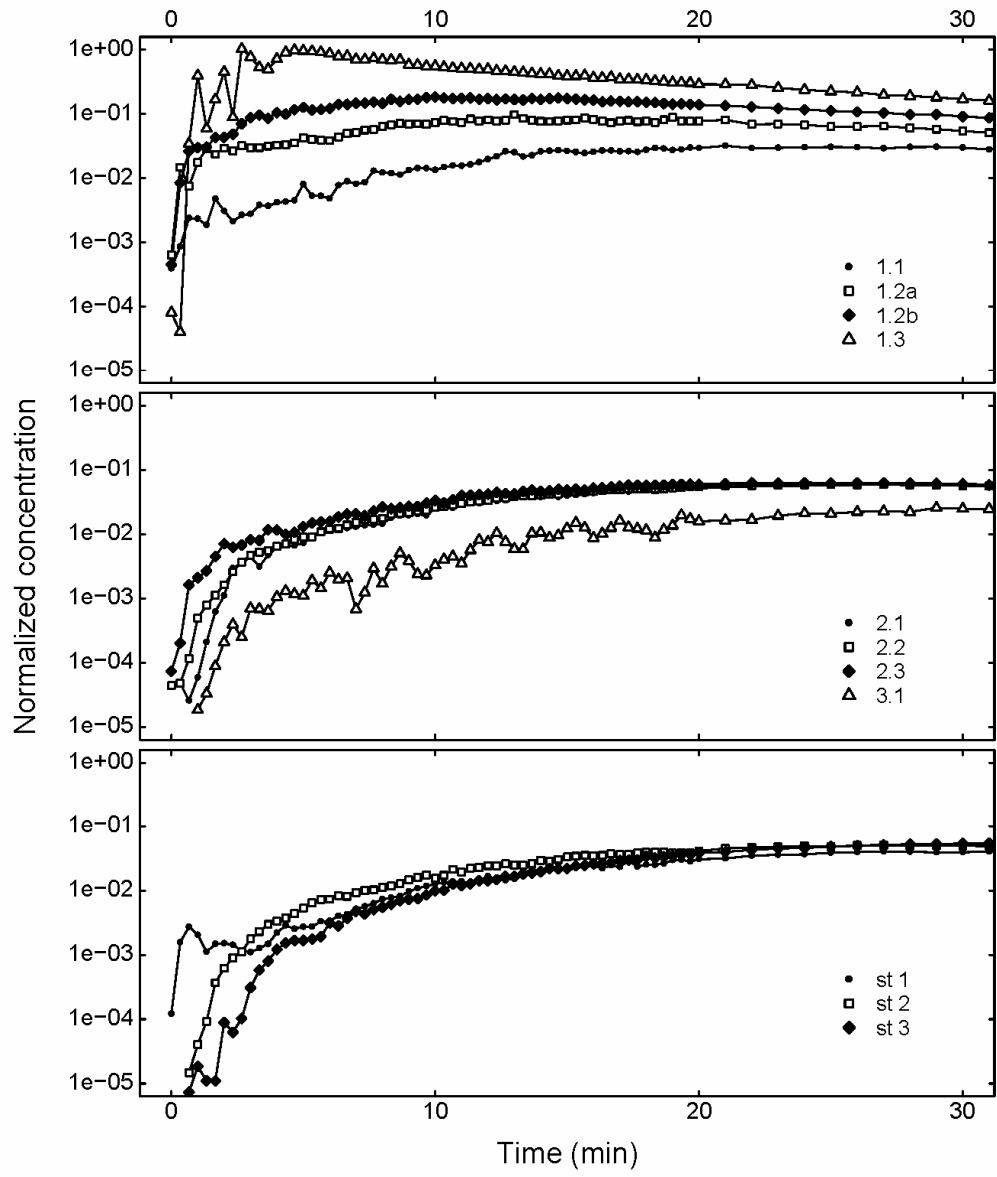


Figure 3.A.6. Time-dependent tracer gas concentrations for Experiment 9.

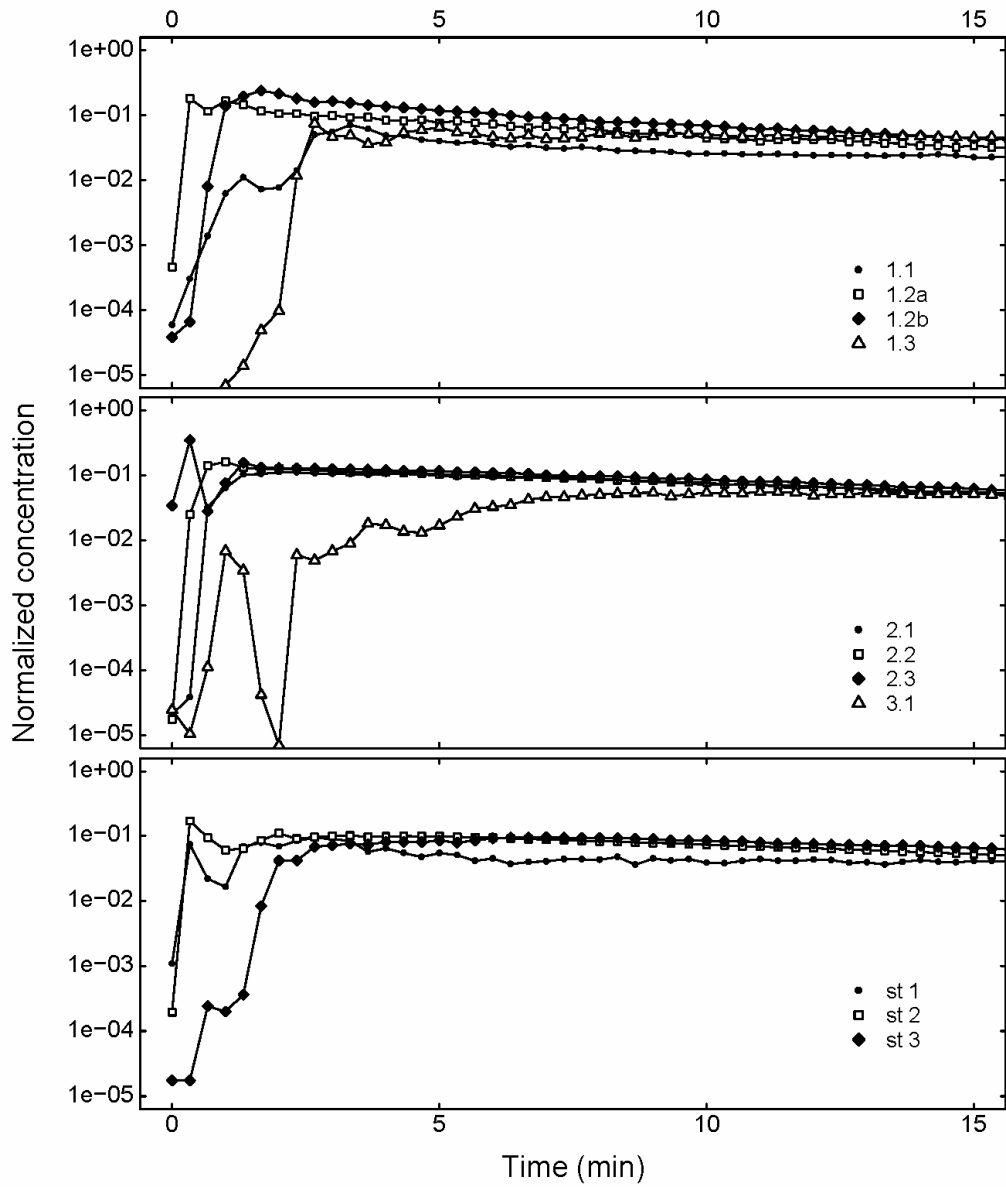


Figure 3.A.7. Time-dependent tracer gas concentrations for Experiment 10.

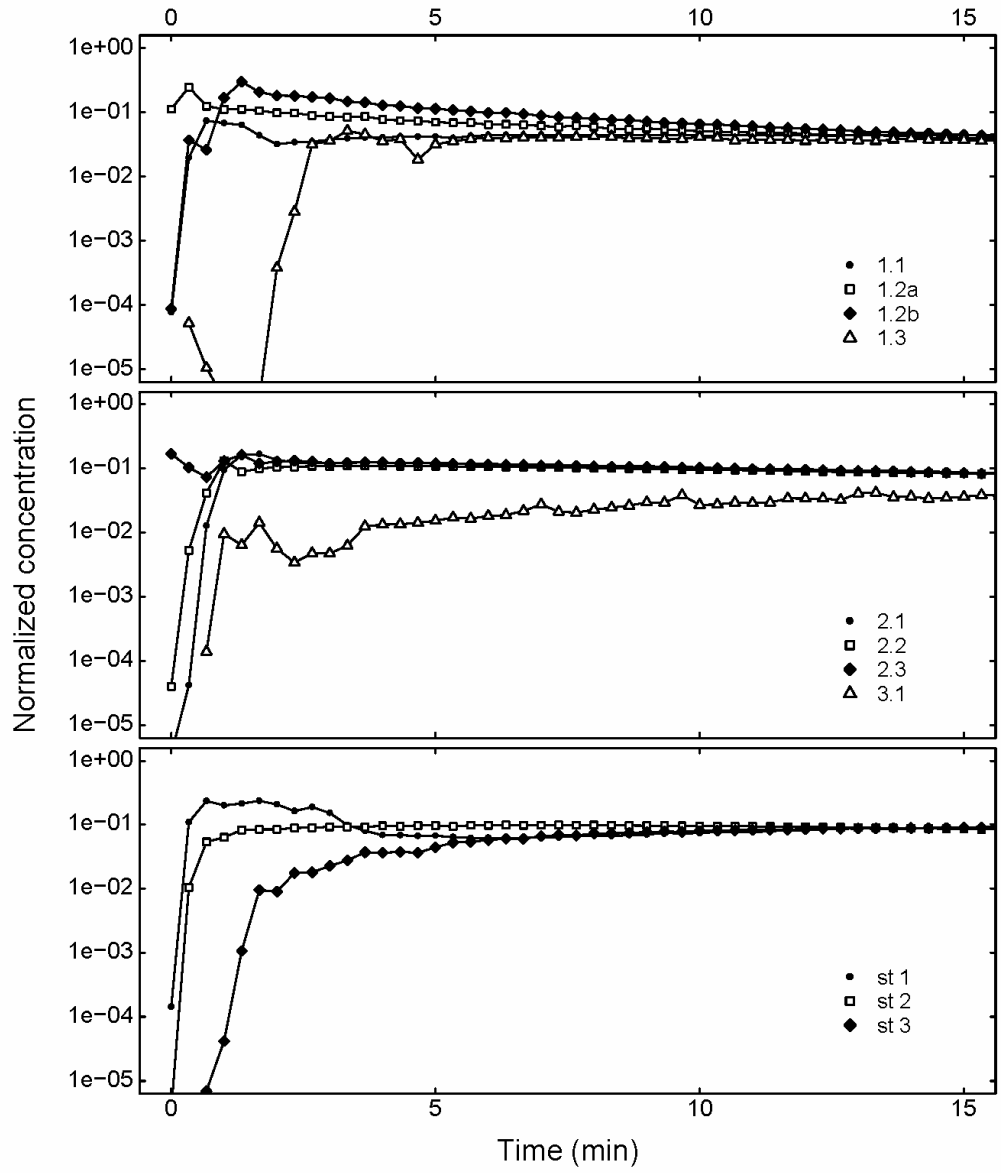


Figure 3.A.8. Time-dependent tracer gas concentrations for Experiment 11.

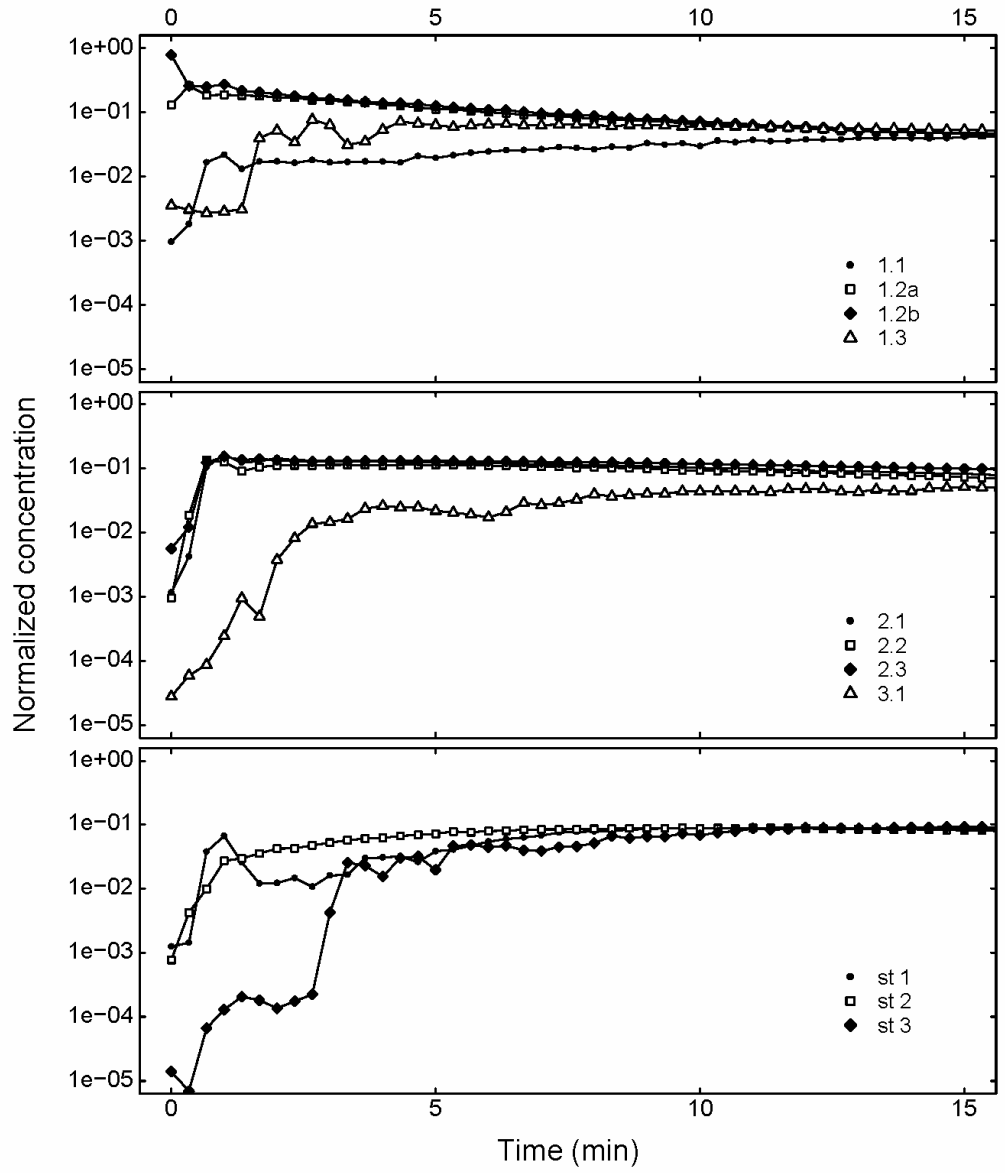


Figure 3.A.9. Time-dependent tracer gas concentrations for Experiment 12.

Table 3.A.4. Average interior temperatures (°C) recorded during tracer experiments <sup>a</sup>

Zone	Experiment Number									
	1	2	3	4	5	6	8	9	10	
Room 1.1	16.5	18.0	17.3	16.1	17.8	12.1	9.4	9.7	9.6	
Room 1.2a	16.2	17.0	16.6	15.8	17.3	12.9	10.9	10.8	10.6	
Room 1.2b	15.9	16.6	16.3	15.5	16.6	13.4	11.3	11.4	11.4	
Room 1.3	16.6	17.7	17.2	16.4	18.1	14.6	12.5	12.5	12.4	
Room 2.1	18.0	19.2	18.8	17.4	18.6	14.7	12.0	11.8	11.5	
Room 2.2	17.4	18.1	17.5	16.5	17.5	14.6	11.9	11.8	11.4	
Room 2.3	18.9	19.7	19.3	17.7	19.3	15.7	13.2	12.9	12.7	
Room 3.1	13.5	19.3	20.5	10.8	16.7	12.1	9.0	9.2	10.0	
Stair 1	17.1	17.9	17.7	15.9	17.7	13.2	9.9	9.9	10.0	
Stair 2	18.5	19.4	18.7	17.8	18.6	14.1	11.5	11.6	11.2	
Stair 3	15.8	18.8	19.4	14.4	16.6	13.9	11.2	11.4	11.1	

<sup>a</sup> Based on one hour of measurements immediately following the release; original measurements recorded at 30 second intervals.

Table 3.A.4. (continued)

Zone	Experiment Number		
	11	12	13
Room 1.1	9.0	9.9	9.5
Room 1.2a	9.0	9.2	9.2
Room 1.2b	9.1	9.6	9.5
Room 1.3	10.5	11.7	10.8
Room 2.1	10.3	11.7	10.6
Room 2.2	9.5	9.8	10.3
Room 2.3	10.9	12.4	11.5
Room 3.1	8.9	9.0	10.1
Stair 1	8.9	8.9	8.9
Stair 2	11.5	11.1	9.6
Stair 3	9.0	10.4	10.1

<sup>a</sup> Based on 1 hour of measurements immediately following the release; Original measurements recorded at 30 second intervals.

Table 3.A.5. Weather conditions recorded during tracer experiments <sup>a, b</sup>

Exp No.	Wind speed (m/s)				Temperature (°C)			
1	0.0	0.0	0.0	0.0	1.9	4.0	6.1	8.0
2	2.8	3.0	3.4	3.9	15.3	15.6	16.2	16.5
3	2.8	2.6	3.0	3.6	18.2	18.3	18.2	18.1
4	0.7	1.2	1.3	1.2	1.7	2.8	4.3	6.1
5	0.4	0.2	0.7	0.5	16.8	17.6	18.3	18.4
6	1.6	1.9	2.2	0.0	8.4	8.8	9.3	8.6
8	7.9	7.8	8.7	9.3	2.1	2.3	2.5	2.7
9	8.0	7.8	8.7	9.2	2.6	2.6	3.2	3.6
10	10.5	10.5	10.5	10.5	4.5	4.5	4.5	4.7
11	4.5	4.8	2.8	2.2	-0.4	-0.3	0.4	1.1
12	7.5	8.1	7.6	7.3	5.7	5.9	6.3	6.6
13	5.9	5.9	6.1	5.9	9.1	9.1	9.2	9.0

<sup>a</sup> Based on 15-minute average measurements recorded over the first hour following the release.

<sup>b</sup> The 4 columns represent 4 distinct measurements taken in sequential 15-minute periods.



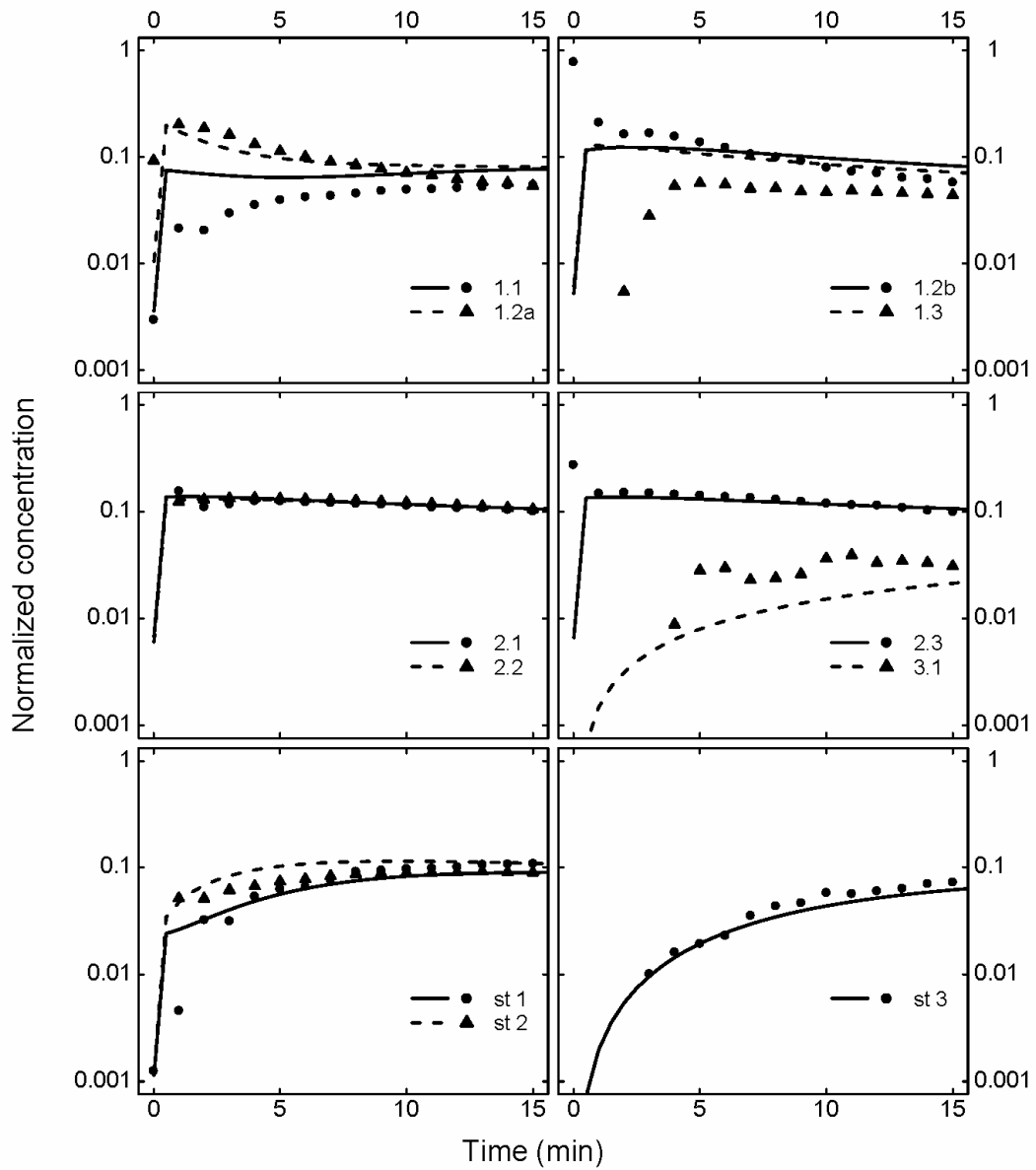


Figure 3.A.10. Time dependent modeled and measured concentrations for Experiment 2. For clarity, every 3<sup>rd</sup> data point is plotted.

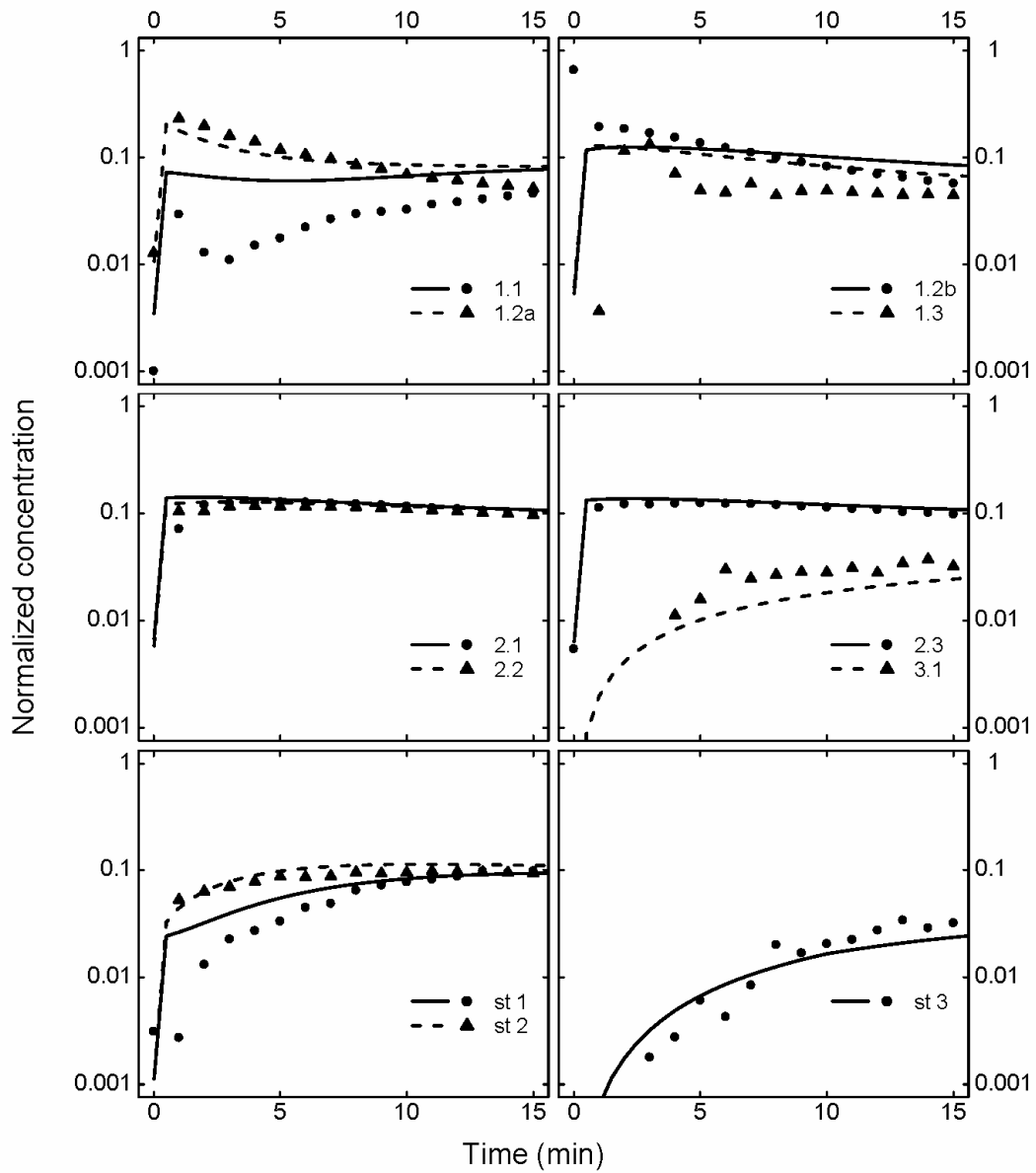


Figure 3.A.11. Time dependent modeled and measured concentrations for Experiment 3. For clarity, every 3<sup>rd</sup> data point is plotted.

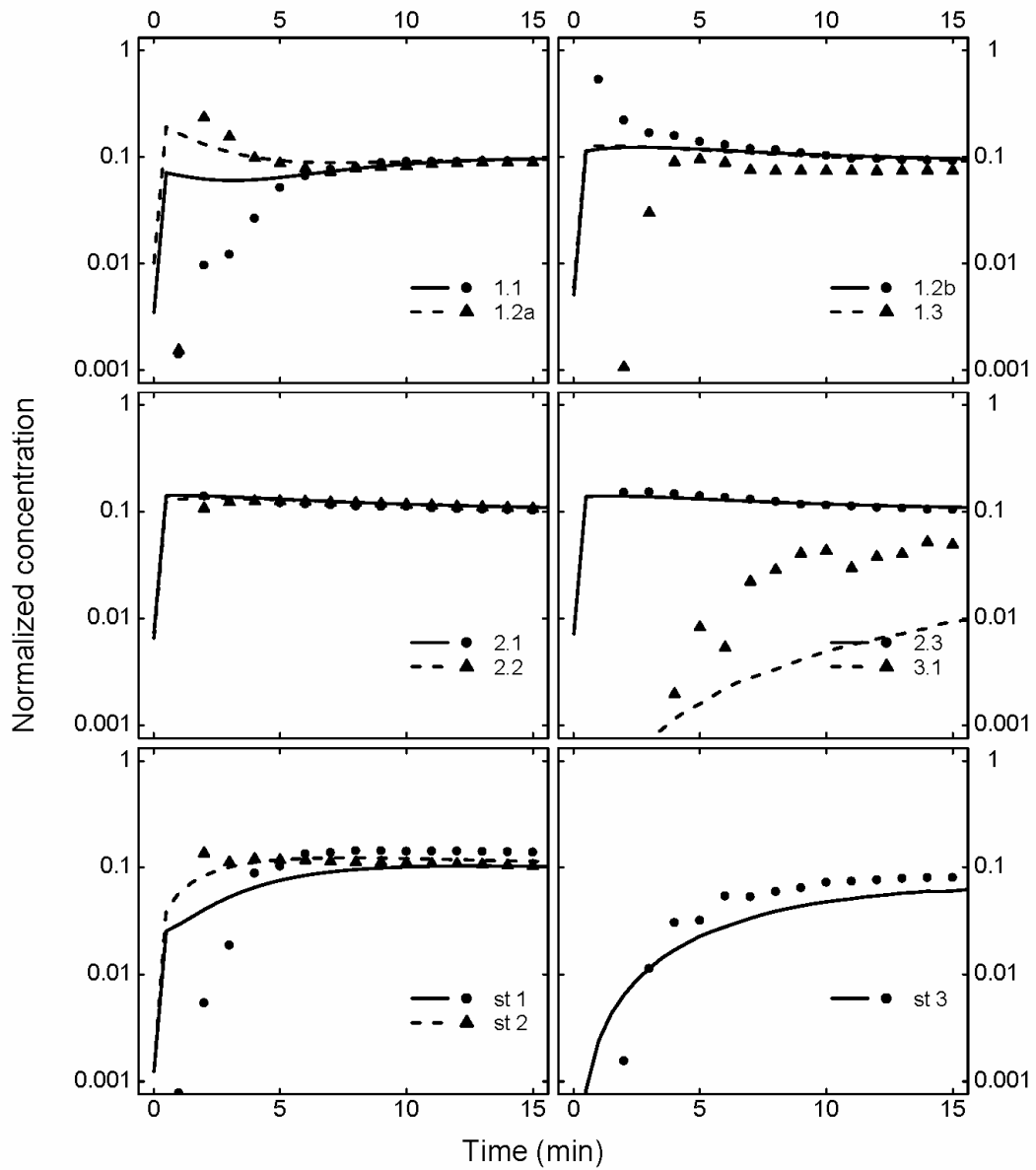


Figure 3.A.12. Modeled and measured concentrations for Experiment 5. For clarity, every 3<sup>rd</sup> data point is plotted.

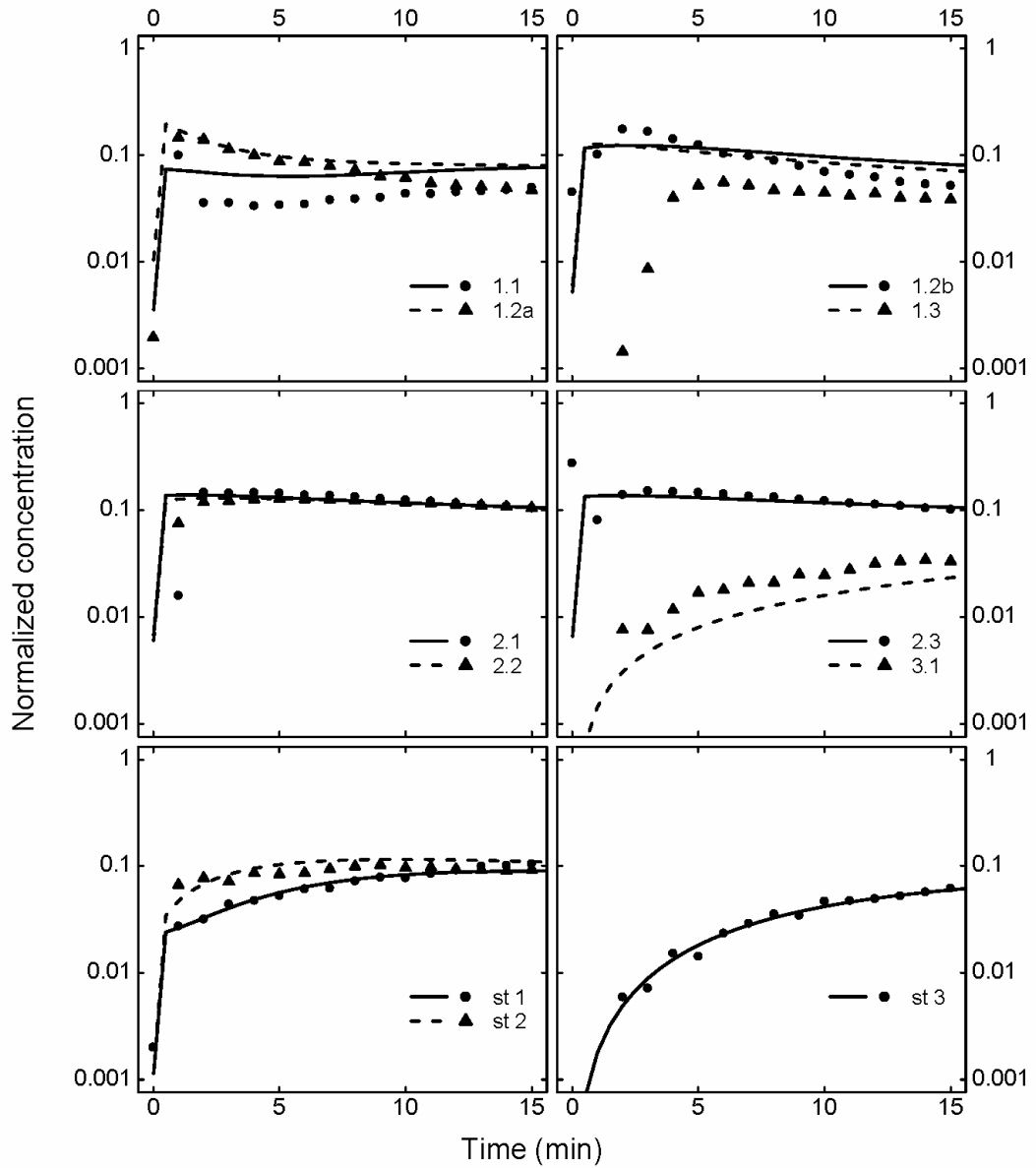


Figure 3.A.13. Time dependent modeled and measured concentrations for Experiment 6. For clarity, every 3<sup>rd</sup> data point is plotted.

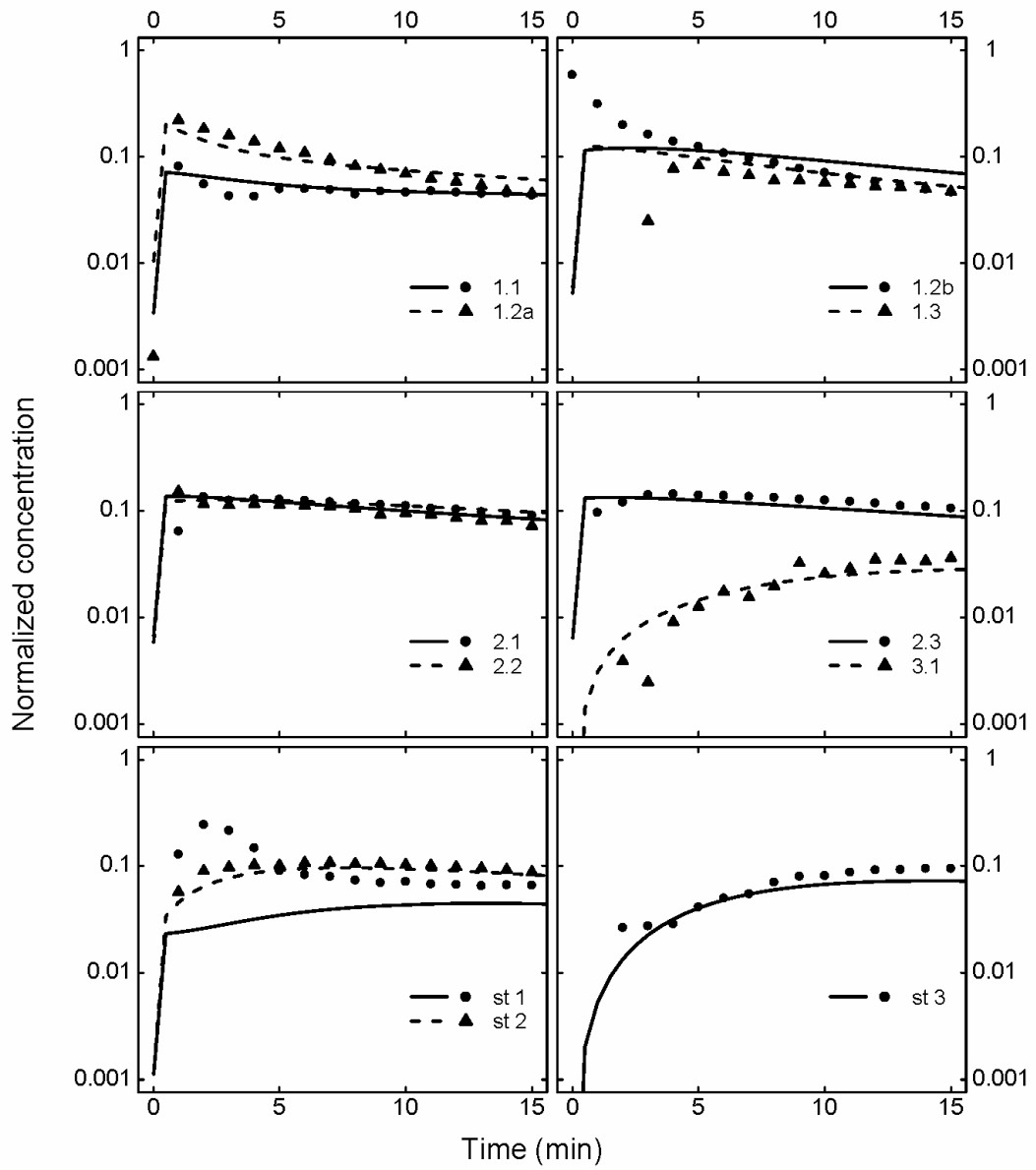


Figure 3.A.14. Time dependent modeled and measured concentrations for Experiment 8. For clarity, every 3<sup>rd</sup> data point is plotted.

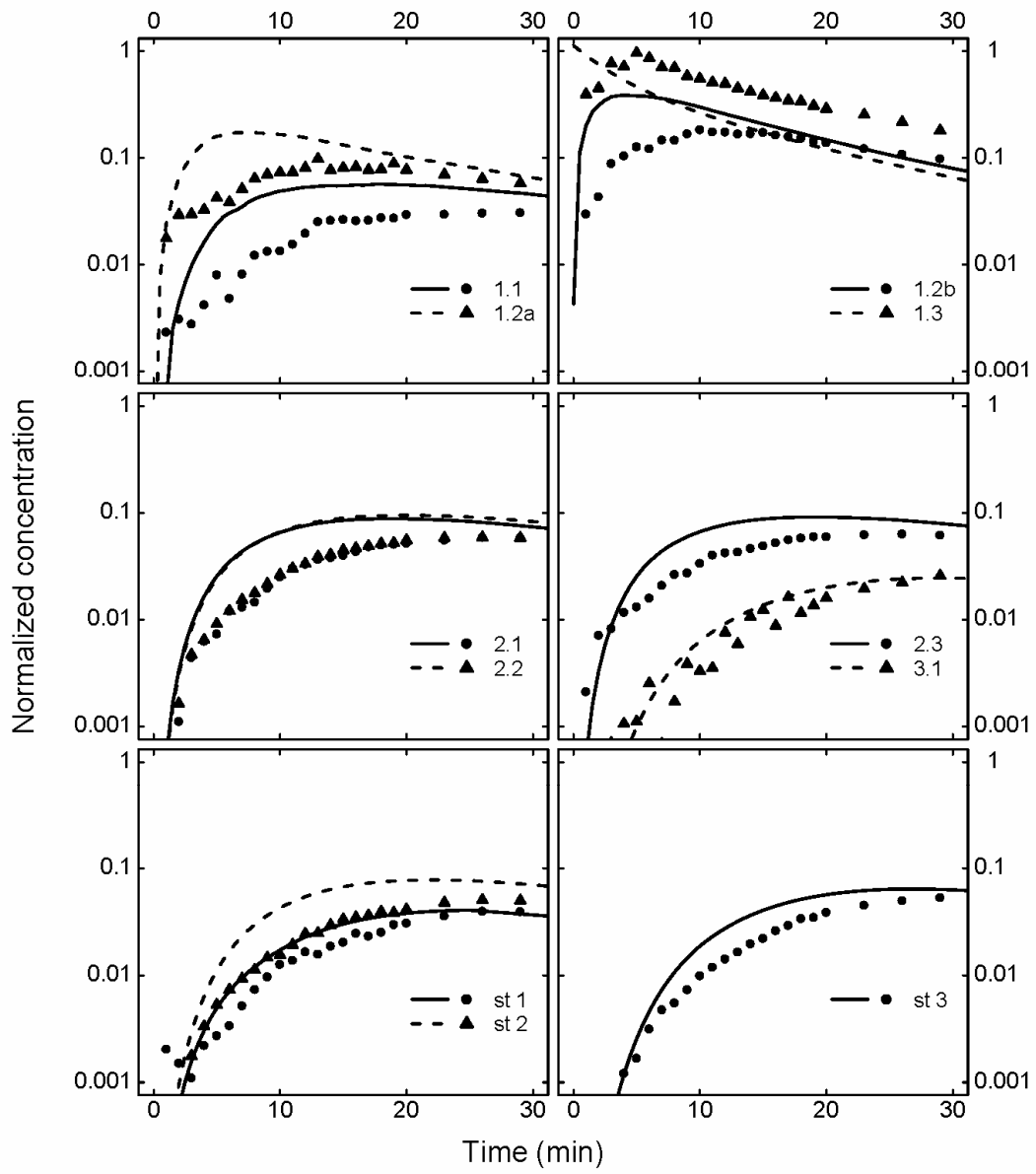


Figure 3.A.15. Time dependent modeled and measured concentrations for Experiment 9. For clarity, every 3<sup>rd</sup> data point is plotted.

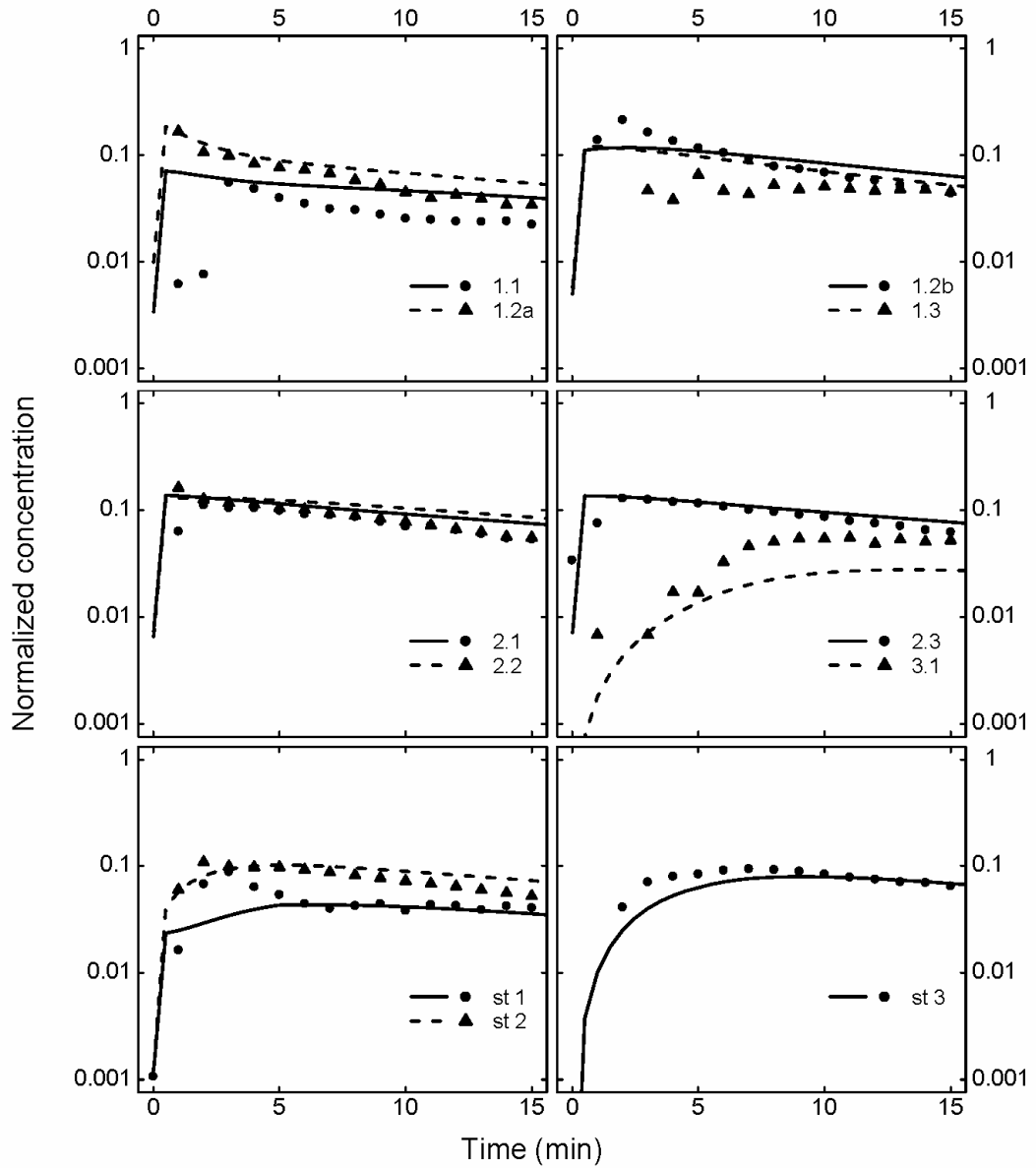


Figure 3.A.16. Time dependent modeled and measured concentrations for Experiment 10. For clarity, every 3<sup>rd</sup> data point is plotted.

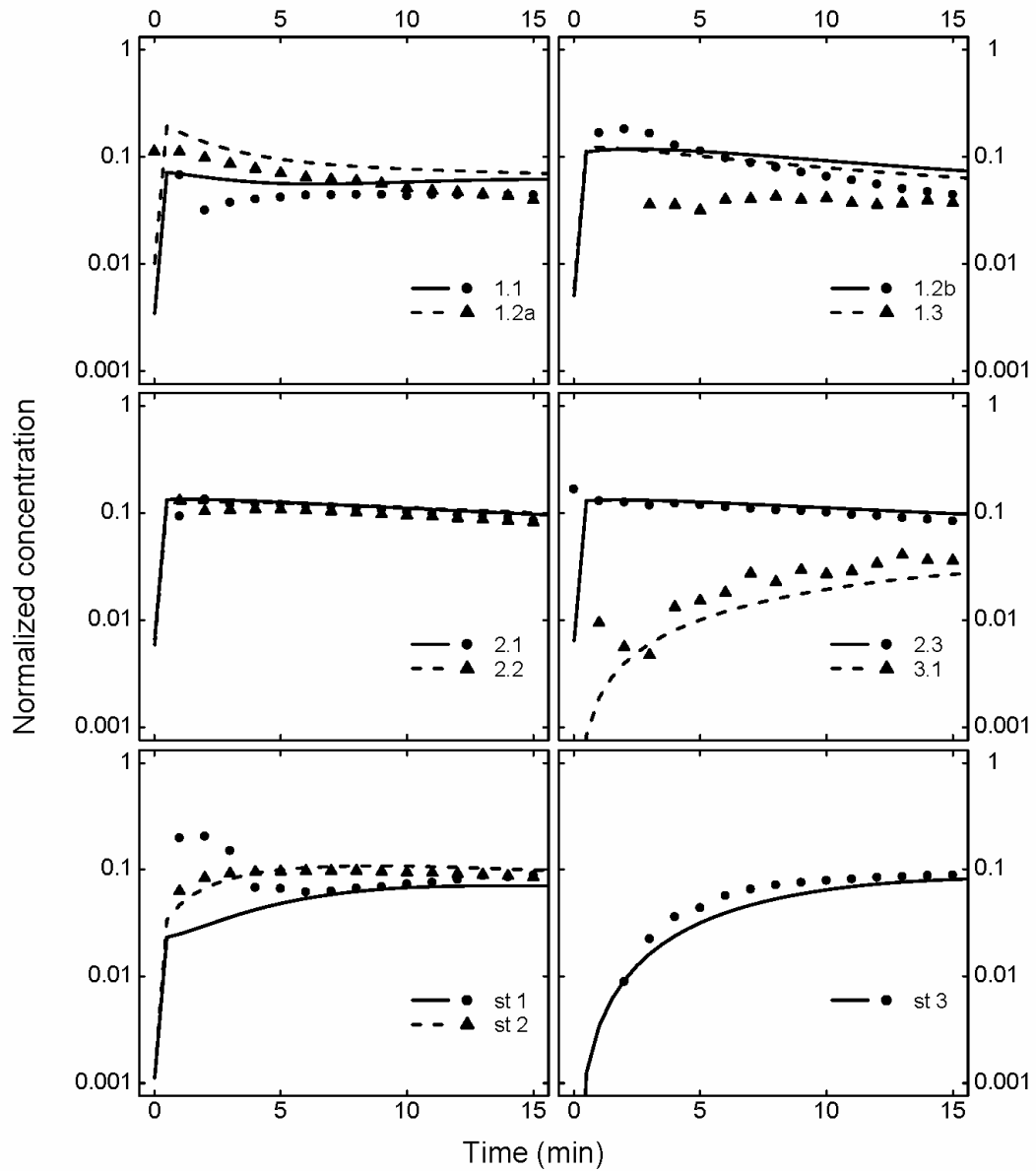


Figure 3.A.17. Time dependent modeled and measured concentrations for Experiment 11. For clarity, every 3<sup>rd</sup> data point is plotted.



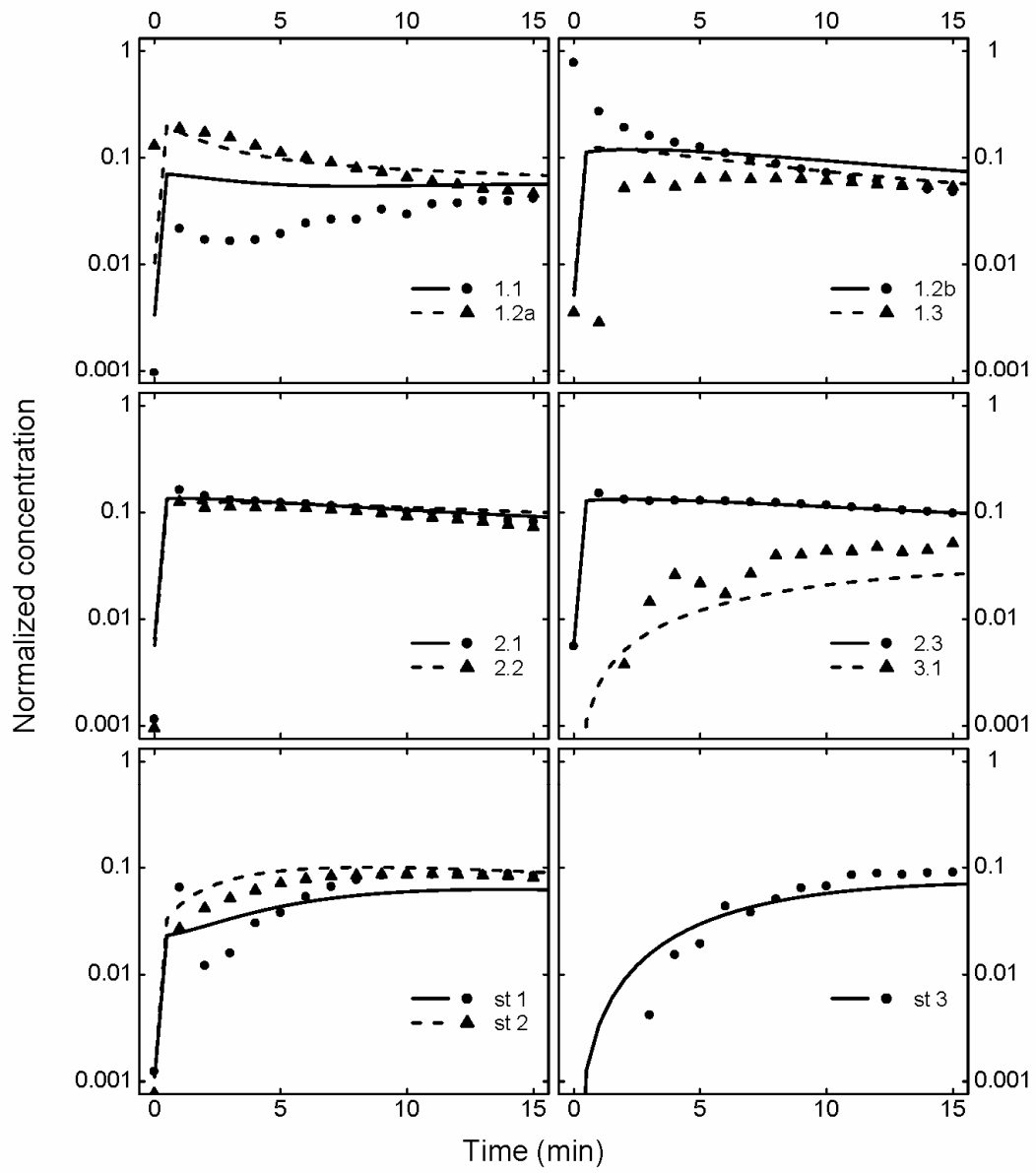


Figure 3.A.18. Time dependent modeled and measured concentrations for Experiment 12. For clarity, every 3<sup>rd</sup> data point is plotted.

## Chapter 4

# 4 Evaluating sensor characteristics in the case of threshold sensors

Reproduced in part from research published in Atmospheric Environment.<sup>1</sup>

### 4.1 Introduction

Developing robust indoor contaminant sensor systems is a challenging design task that must consider the interactions and influence of several factors. Selecting the types and characteristics of individual sensors, placing them into networks and devising interpretation algorithms are all part of the design problem. Limitations inherent to the sensors, such as high rates of false positive or negative measurements, and other characteristics are likely to influence the performance of sensor systems. An important design goal is that the sensor interpretation algorithms perform well when faced with such challenges. In this chapter, I consider the influence of sensor characteristics in developing robust sensor systems, focusing specifically on threshold- or alarm-type sensors.

---

<sup>1</sup> Sreedharan, P., Sohn, M.D., Gadgil, A.J., Nazaroff, W.W, 2006. Systems approach to evaluating sensor characteristics for real-time monitoring of high-risk indoor contaminant releases. Atmospheric Environment 40, 3490-3502. Copyright 2006, Elsevier, Inc.

Strategic placement and selection of sensors is necessary to ensure that the release event is detected and characterized. However, to select an appropriate mix of sensors, one must understand what the performance trade-offs are between different sensor characteristics. Such characteristics may include detection limits, response times, accuracy, or threshold levels. A systems level analysis may be helpful for identifying and selecting an optimal set of sensors given a mix of available sensor characteristics. This kind of analysis is arguably necessary since the benefits of a particular sensor characteristic ought to be evaluated against an overall system goal.

Much effort is being directed to the development of toxic chemical sensors. However, relatively little attention has been devoted to identifying and selecting sensors characteristics that will optimize the performance of sensor systems designed to protect occupants in indoor environments. Moreover, journal publications on improving chemical protection within indoor environments almost exclusively focus on optimizing the performance of sensors individually (e.g., electronic noses and robotic systems for plume tracking). For example, the proceedings of the 2004 IEEE Sensors conference contained no papers on the topic of system-level design.

As noted in Chapter 2, the statistical approach used in this research, the Bayes Monte Carlo (BMC) algorithm, has been exercised using synthetic data and concentration type data. A few published papers focused on demonstrating the potential benefit of the BMC algorithm for performing real-time reconstruction of an indoor contaminant release (Sohn et al., 2002a; Sohn et al., 2002b). In both of these investigations, the sensors were assumed to be capable of reporting continuous concentration measurements.

This chapter advances earlier work by investigating the performance of a threshold- or alarm-type sensors, rather than concentration-output devices. Threshold sensors constitute a class of chemical sensors that is being considered for use in indoor sensor systems. There is significantly less information in a binary signal compared to a concentration signal. In the former, all that is known at any time is whether the concentration is above or below a certain target level. At the outset of this work, it is not clear if a network using this type of sensor is capable of *characterizing* the release in any meaningful manner (as contrasted with simply *detecting* that a release has occurred). However, efforts in other sensor network applications, such as in target tracking, have shown that binary sensor signals can provide useful information for modeling physical systems (Aslam et al., 2003).

Through a series of examples, I examine how well various sensor systems can reconstruct the release event, with each system consisting of sensors with different characteristics, focusing on threshold level, response time, and accuracy. Adhering to a case study approach, each example sensor network is evaluated using data obtained from the tracer gas experiments described in Chapter 3. In this chapter (and throughout the dissertation), the success of a sensor system is characterized in terms of by the rate at which the sensor system characterizes the release to meaningful levels of confidence. These examples demonstrate the importance of a systems perspective in selecting sensors with desirable characteristics for overall sensor-system performance.

## 4.2 Background

This chapter seeks to understand the performance of sensor networks composed of alarm-type sensors. In an operational system, when any sensor indicates a measurement above the detection limit, a release event may have occurred, and the BMC interpretation algorithm is initiated. The BMC algorithm is briefly reviewed.

### 4.2.1 BMC Review

A two-stage Bayes Monte-Carlo algorithm is the approach used to first represent prior assumptions of the possible release characteristics and to then interpret sensor information in real-time. As they become available, the data augment and eventually displace the prior assumptions about the physical system as they become available. The two-stage approach is described in detail in Chapter 2 with key points reiterated here.

The first stage consists of developing a library of hypothetical contaminant transport simulations spanning the set of all plausible pollutant release and internal airflow conditions. In the second stage, the sensor data are interpreted by estimating their degree of statistical agreement with predictions in the simulation library using Bayes' rule.

The posterior probability of the  $k^{\text{th}}$  simulation best representing the true release, having made prediction  $Y_k$ , and given sensor measurements  $O$ , is denoted as  $p(Y_k|O)$ , and is evaluated as follows:

$$p(Y_k | O) = \frac{p(O | Y_k)p(Y_k)}{\sum_{i=1}^N p(O | Y_i)p(Y_i)} \quad (4.1)$$

where  $p(O|Y_k)$  is the likelihood of observing measurements  $O$  given that simulation prediction  $Y_k$  is true,  $p(Y_k)$  is the prior probability that the  $k^{\text{th}}$  simulation is true,  $N$  is the total number of simulations in the library, and  $p(O|Y_k)$  is the likelihood function, reflecting the error structure of the data. Thus, larger values of  $p(O|Y_k)$  indicate a better match of the data,  $O$ , to the modeled concentrations of  $Y_k$ .

#### 4.2.2 Case study: Experiments and model development

This section summarizes key elements of the case study experiments used in the current assessment. Additional details are presented in Chapter 3. The case study unit is a three-floor multi-unit building comprising 660 m<sup>3</sup> interior volume and 280 m<sup>2</sup> of floor area (Figure 3.1). A 100% recirculation mechanical air-handling unit (AHU) supplies air to the first and second floors with its return on the first floor.

Sextro et al. (1999) conducted fan pressurization tests to characterize the airflows, followed by tracer gas experiments. In each tracer experiment, 20 g of propylene was instantaneously released at a specific interior location. Concentrations in each room and the staircase were recorded at 20 s intervals (for the first 20 min) and 60 s intervals (after 20 min). The leakage and airflow characteristics were used to develop a multizone airflow and pollutant dispersion model which was used as the base model in this chapter.

I selected the data from one tracer-gas experiment (Experiment 1) for the investigation in this chapter. In this case, the release occurred outside the return grill of Room 1.2a (first floor). The AHU was operating and all interior doors were open. Figure 4.1 displays concentration data collected from the sensors.

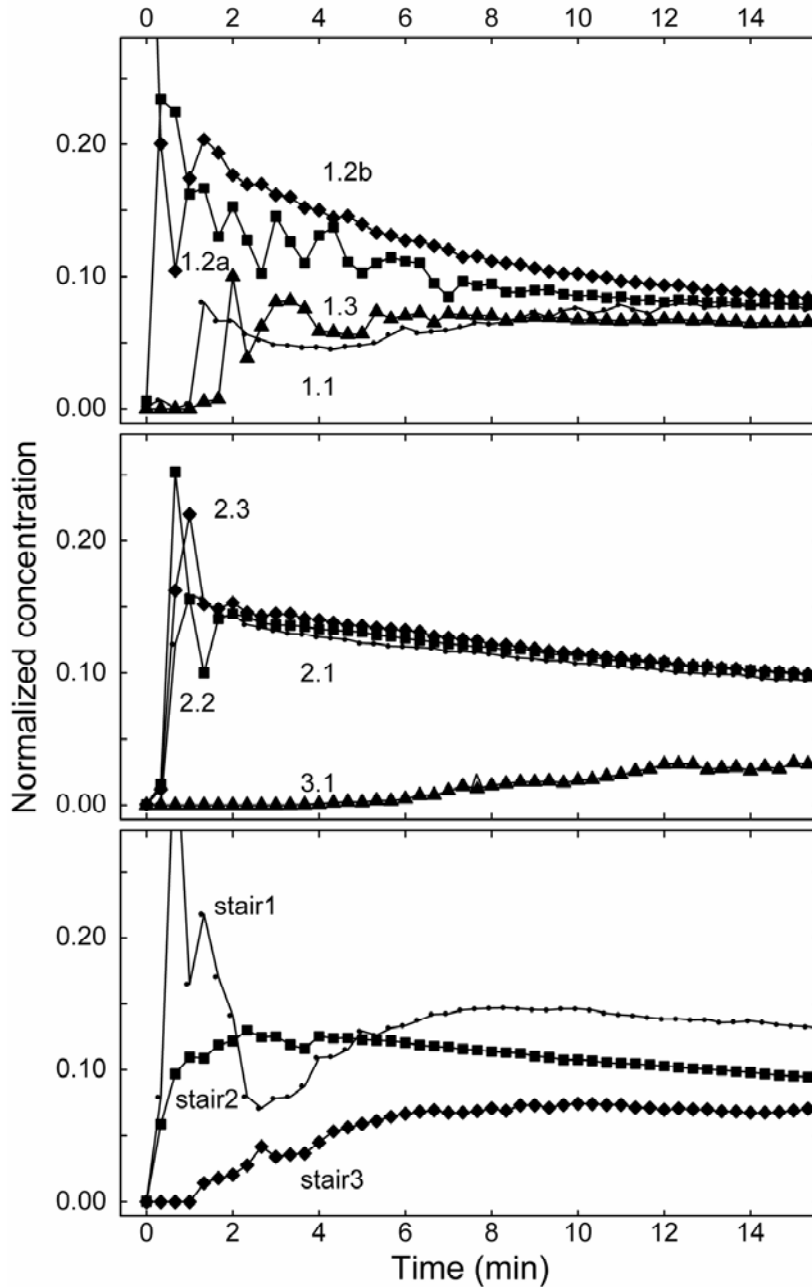


Figure 4.1. Time-dependent normalized tracer gas concentrations for Experiment 1. Concentrations are normalized by the theoretical peak concentration in the experiment (i.e., the maximum concentration that would be observed if the contaminant was instantaneously and uniformly mixed in the release room). (a) first floor rooms; (b) second floor and third floor rooms; (c) staircases.

### 4.3 Approach

This section explores the potential utility of the two-stage Bayesian interpretation algorithm for designing a sensor system that consists of single-level threshold sensors. Several examples are constructed using data from a single case-study experiment. The objectives are: (1) to test how well Bayesian interpretation works with threshold sensors; and (2) to explore how the system's performance varies with the sensor threshold level, response time, and error. The emphasis here is not on optimizing the Bayesian algorithm for this case study, but to demonstrate how a Bayesian methodology can be used to assess sensor characteristics for any sensor type, such as single-level threshold sensors. Actual implementation for a specific sensor system may require further refinement of the algorithm.

For the purposes of this chapter, I assumed that each zone is equipped with one sensor, and that each sensor has a single threshold, meaning that the output is either "below threshold" or "above threshold." I define several possible threshold levels that are at or above the minimum detection limit of the sensor. Three important parameters characterize sensor performance in this study: threshold level; response time (also called "integration time"); and accuracy.

A sufficiently large release will trigger "above threshold" responses from one or more sensors. A Bayes Monte Carlo algorithm is then initiated to determine key information about the release event, including the location of the release and the mass of contaminant discharged.



Each network is evaluated in terms of how accurately and quickly the system estimates the contaminant release parameters. In practice, different performance criteria may be set depending on the objectives of the particular monitoring system.

### 4.3.1 Library generation

Stage 1 of the BMC algorithm involves generating a library of simulations that represent the behavior of all possible contaminant releases. Using the transport model for Experiment 1, a library of 5000 simulated contaminant releases was generated. Each parameter set was obtained by sampling from statistical distributions of a set of key input parameters. The variable input parameters are summarized in Table 4.1.

Table 4.1 Parameter values of the library of 5000 simulated contaminant releases

Parameter	Values
Source location	Ten locations: any room plus stairwell. Each location is equally probable.
Source duration	1 s to 5 min; log-uniform distribution.
Source amount	10 to 100 g; log-uniform distribution.
Door position	Three equally probable scenarios: (1) all interior doors open; (2) all interior doors closed; (3) doors between staircase landings and adjacent rooms closed, all others open.

### 4.3.2 Hypothetical threshold sensor data

I generated hypothetical threshold sensor data by interpreting the tracer data from the experiment as if they were concentrations to which surface acoustic wave (SAW) sensors were exposed. SAW sensors are piezoelectric devices, often configured to provide alarms based on whether the incoming concentrations are above or below a predefined trigger or threshold level. False positive and false negative alarms may occur, according to the

performance characteristics of each sensor, or the inability of the sensor to distinguish between the contaminant of interest and interfering chemicals that also may be present in the air.

Using the case study data from Experiment 1, I generated hypothetical alarm data for sensors with different performance characteristics, based on discussions with several developers and users of SAW sensors. Three sensor attributes were varied: threshold level, response time, and error.

The threshold levels were chosen relative to the measured concentrations during the first 120 minutes of the release event. The lowest selected threshold would cause 98% of the data to trigger the alarm, while the highest threshold would trigger an alarm for only 1% of the data. However, for presentation purposes, I normalize both threshold levels and concentration data by the concentration that would be found in the release zone if the entire release amount instantaneously mixed throughout only that zone. That is, thresholds and concentrations are reported in terms of the theoretical maximum peak concentration that could be measured in the system under the perfectly well-mixed assumption for an isolated release room. With this normalization, the lowest threshold level was set to  $2 \times 10^{-4}$  and the highest threshold was set at 0.16. (These are equivalent to 0.02% and 16% of the theoretical peak concentration.)

Sensor response times were fixed at values that ranged from 20 seconds to 180 seconds. In the simulations, concentrations are averaged over the response time and then compared to the appropriate threshold level. Note that averaging over the response time corresponds to an assumption that the SAW desorption cycle is brief relative to its adsorption cycle. In these simulations I ignored the duration of the desorption cycle (i.e.,

each sensor started integrating the next cycle of data as soon as it reported an alarm or no-alarm condition).

Simulations were run using data with and without synthetically added error. For simulations with added error, I generated sensor signals according to the following assumptions: (1) if the actual concentration was within 25% of the sensor threshold level, the signal would be false 50% of the time; and (2) for concentrations outside of this range, the signal would be false either 10% or 30% of the time, depending on the assumed sensor error. In the implementation of Bayes' rule utilized here, the likelihood function is based on the probability used to generate the false positives and negatives. For example, for data generated with a 30% error, the likelihood is 0.3 when the modeled concentration is more than 25% above the threshold level and the sensor has not signaled on; conversely, the likelihood is 0.7 if the sensor has signaled on. For the simulations using data without synthetically added error, I assume 5% error for all data. I did not assign 100% confidence to the data because of inherent uncertainty and variability. In practice, the designer of the sensor system should have reliable information on the sensor's actual rate of false positives and false negatives.

Figure 4.2 illustrates the conversion of measured concentration data to simulated threshold data. Figure 4.2(a) shows normalized time-averaged concentration data, with the threshold level indicated. Figure 4.2(b) shows the threshold data that would result from a threshold sensor with no error and an instantaneous response. Here, "1" signals that the concentration exceeds the threshold. Figure 4.2(c) shows the threshold data, corrupted with false negatives or positives. Because the false readings are generated

stochastically, different realizations of the data in Figure 4.2(c) would exhibit different patterns of output signals.

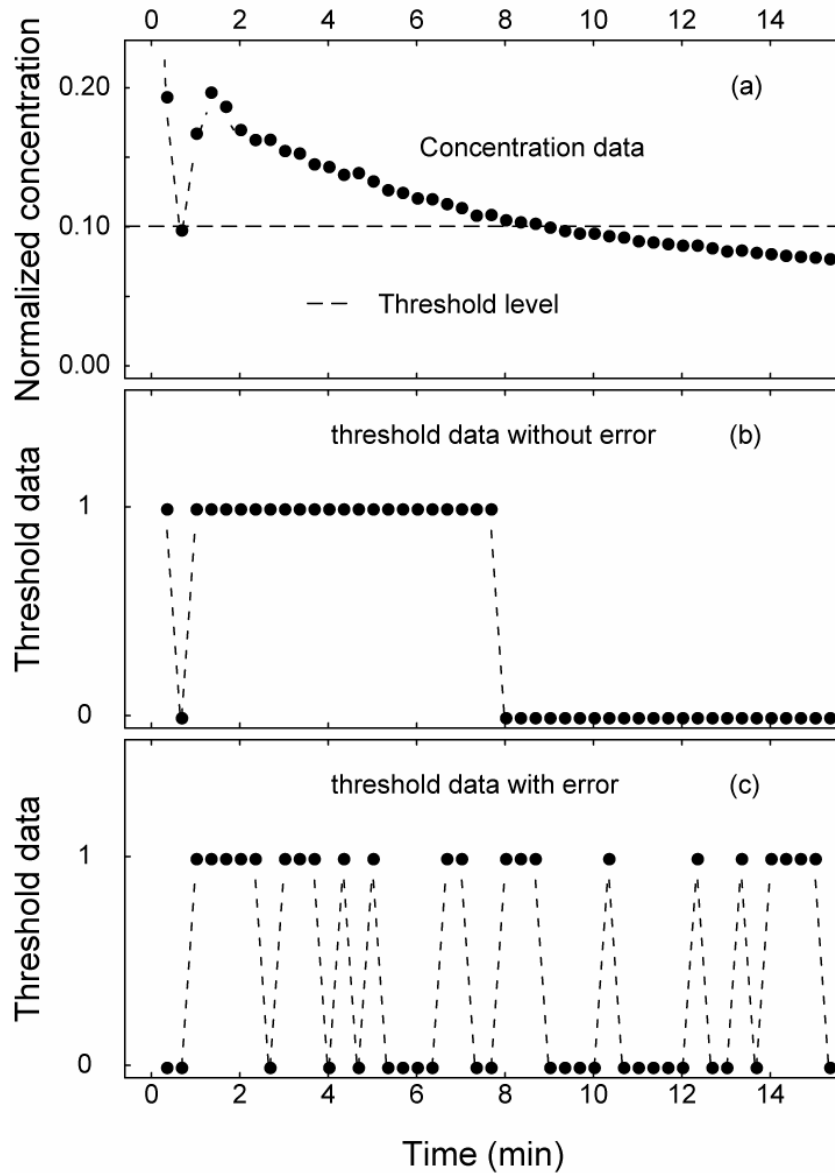


Figure 4.2. Sample illustrating conversion of tracer gas concentration to threshold data: (a) concentration data; (b) threshold data without simulated error added; (c) threshold data with simulated error added.

For the case-study examples, I systematically varied the sensor characteristics. In cases where error was specifically investigated, for each sensor attribute, I generated 50

sets of error-added threshold data, analogous to those displayed in Figure 4.2(c), for each sensor in the system. Each combination of threshold level, response time, and error produced a data stream against which to apply the Bayes Monte Carlo algorithm. The algorithm was used to determine the release location and release magnitude; the time of release was assumed to be accurately known.

## **4.4 Results and Discussion**

To demonstrate data interpretation using threshold sensor data, I first investigate the ability of the sensor system to identify the release location, mass, and duration. Next, I investigate the effect of changing the threshold level and response time characteristics, and lastly, the effect of changing the sensor error in conjunction with these characteristics.

### **4.4.1 Estimating release characteristics with threshold data**

The information content in threshold sensor data is significantly less than that in direct concentration measurements. Nevertheless, the sensor system can successfully reconstruct the source, at least in some circumstances. I demonstrate this with an example in which the concentration data have been converted to threshold data using a threshold level of 0.023, a sensor response time of 20 seconds, and without added error. The likelihood function was evaluated assuming a 5% error rate. I judge the sensor system performance by its ability to reduce the uncertainty of the probabilistic outcomes that describe the release location, mass, and duration parameters, and by the time required to do so.

Figure 4.3 depicts the time required to identify the release location. In this chapter, localization to either the intake of the return duct or to the room containing the return, Room 1.2a, is considered equally acceptable. At time zero, each zone is assumed to be equally likely as the release location. As sensor data arrive, the Bayes algorithm adjusts these probabilities, identifying the correct release location with greater than 90% probability within one minute. If rapid response hinges on locating a source quickly, this example suggests that threshold sensors under this network configuration and data quality might be acceptable for real-time monitoring.

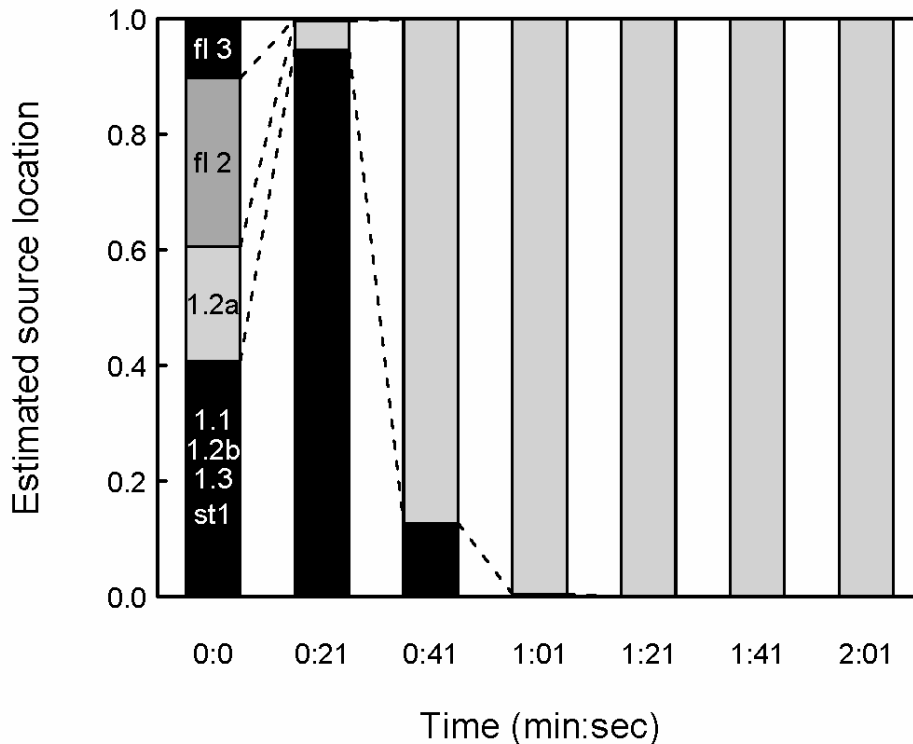


Figure 4.3. Probability of source being in location indicated using threshold data. Sensor characteristics: response time of 20 s, threshold level of 0.023, and without added error. The actual release location is Room 1.2a. Time is referenced to the instantaneous release event.

Figure 4.4 shows the time-resolved estimates of the release mass and duration parameters. These parameters are accurately estimated within tight uncertainty limits after ten minutes. Rather than emphasizing the accuracy at which these parameters are estimated, a more important observation is that threshold type data are capable of reducing the uncertainties that describe the release and thus, this type of sensor may contain information that is relevant to a system-level analysis and sensor-system objective.

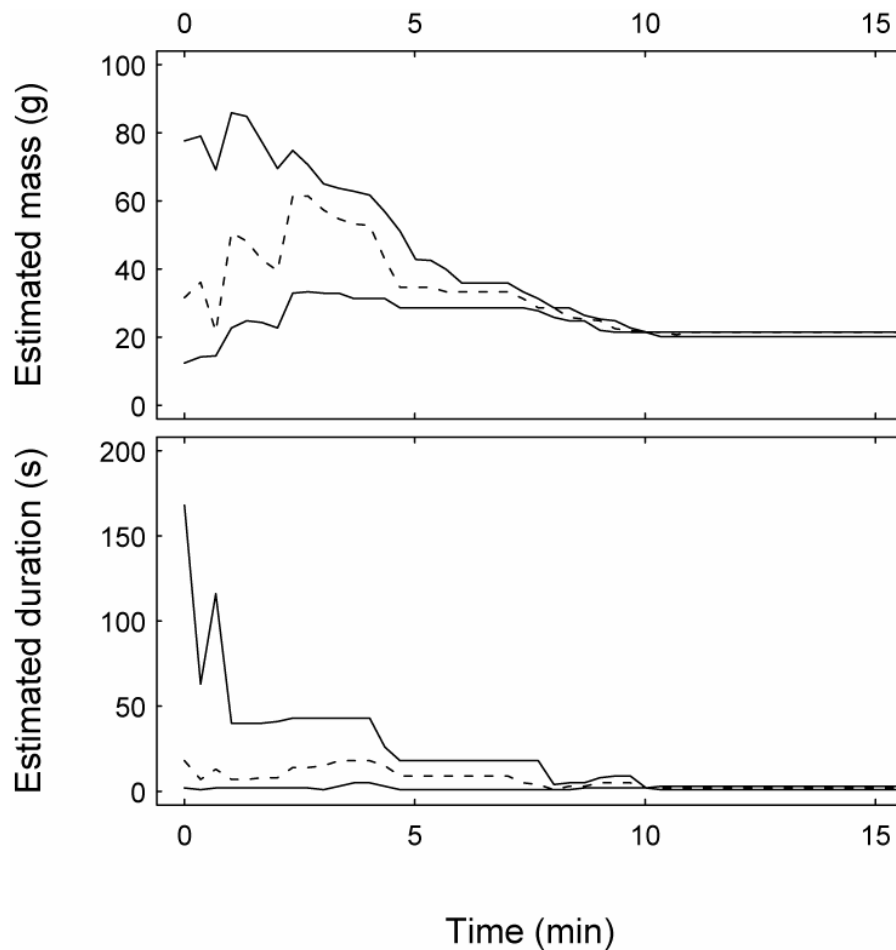


Figure 4.4. Release mass and release duration estimated using threshold data. Sensor characteristics: response time of 20 s, threshold level of 0.023, and without added error. The solid lines indicate 80 percent confidence intervals; dashed lines indicate medians. (Actual release mass was 20 g, released as an instantaneous puff.)

In addition to estimating key release parameters, it is useful to know the past and future uncertainty distributions of concentrations in different zones. When the sensor signals are alarm type, rather than concentration signals, one may expect that the uncertainty descriptions of concentrations may not be significantly reduced. In contrast, if concentration data were available to the sensor system, it is more likely that the description of concentration distributions at current and future times will be represented by comparably more narrow confidence intervals. (Note that even in the case where sensors produce concentration signals, there is still uncertainty surrounding the true concentrations because of measurement error.)

Figures 4.5-4.7 show the uncertainty distributions for the concentrations in three zones based on three different conditions of data: no data (Figure 4.5), after analyzing 5 min of data (Figure 4.6), and after 10 min of data (Figure 4.7). One zone from each floor was selected: Rooms 1.2a and 2.2 and Stair 3. Figures 4.6-4.7 are both based on the use of threshold-type sensors, with one sensor per zone, and with sensor characteristics as described in the figure caption.



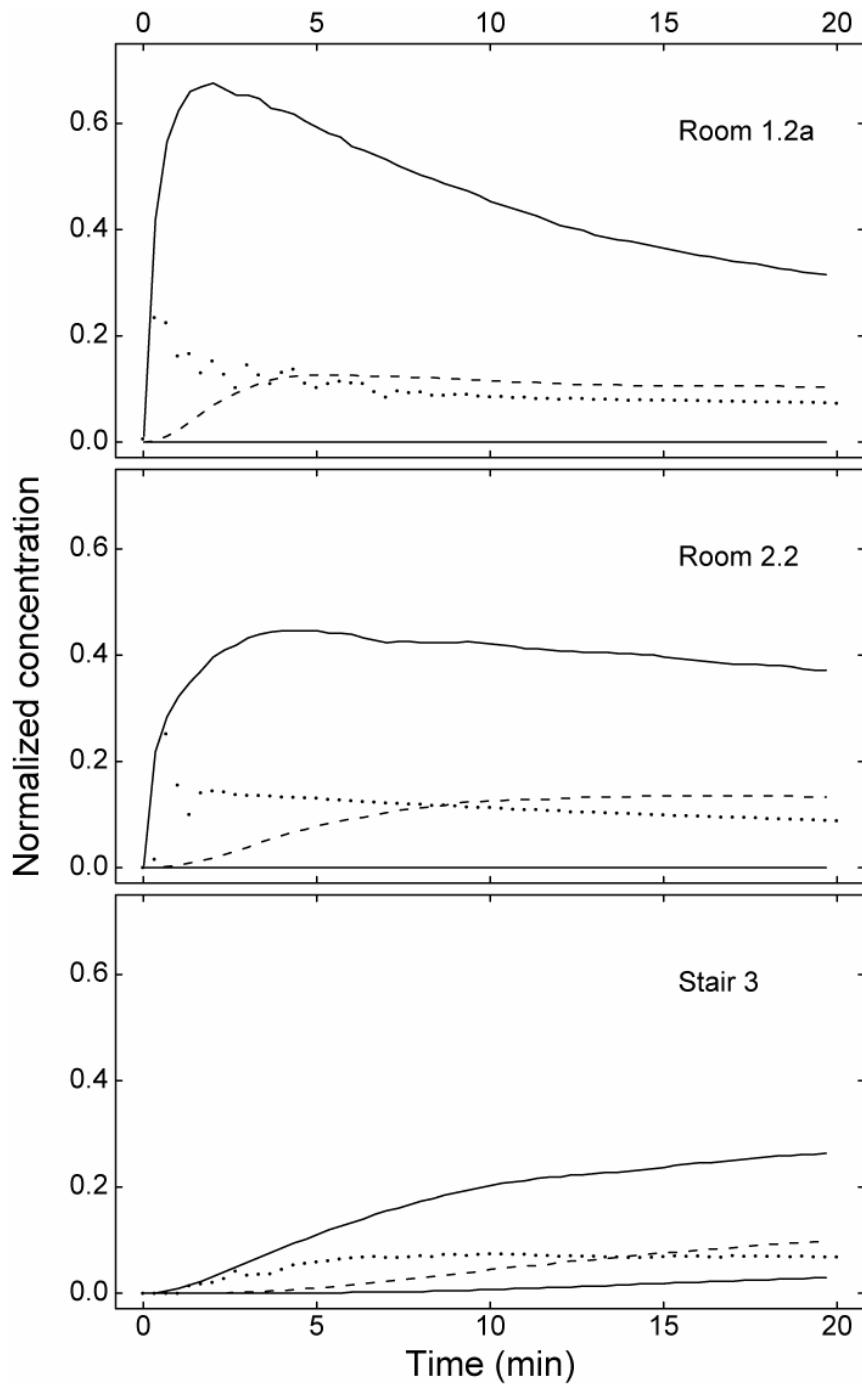


Figure 4.5. Prior confidence intervals of the concentration profiles for three zones with concentration data. Median (dashed line), 10% and 90% percentiles (solid lines), and concentration data from Experiment 1 (dotted line).

Figure 4.5 represents the best assessment using prior knowledge of the distribution of concentrations that would be observed in these zones, if a release were to occur at time

zero. The actual concentrations measured by in Experiment 1 are also shown in the figure. The measured concentrations lie well within the prior distributions for each zone.

After 5 min of threshold data have been assimilated by the sensor system, an updated assessment of the uncertainty distributions of the concentrations in these zones is obtained (Figure 4.6). The original concentration data obtained from Experiment 1 are shown as well. Two types of information are obtained from this figure: the updated concentration distributions prior to the actual time (i.e., 5 minutes) and the updated concentration distributions for future time. The BMC algorithm substantially reduces the uncertainty with respect to both time periods, relative to the prior distributions. It is also worth noting that the original concentration data lie outside the 80% confidence intervals for the zones. This suggests that the confidence assigned to the measurements may not reflect the true model-to-measurement fit for all cases.

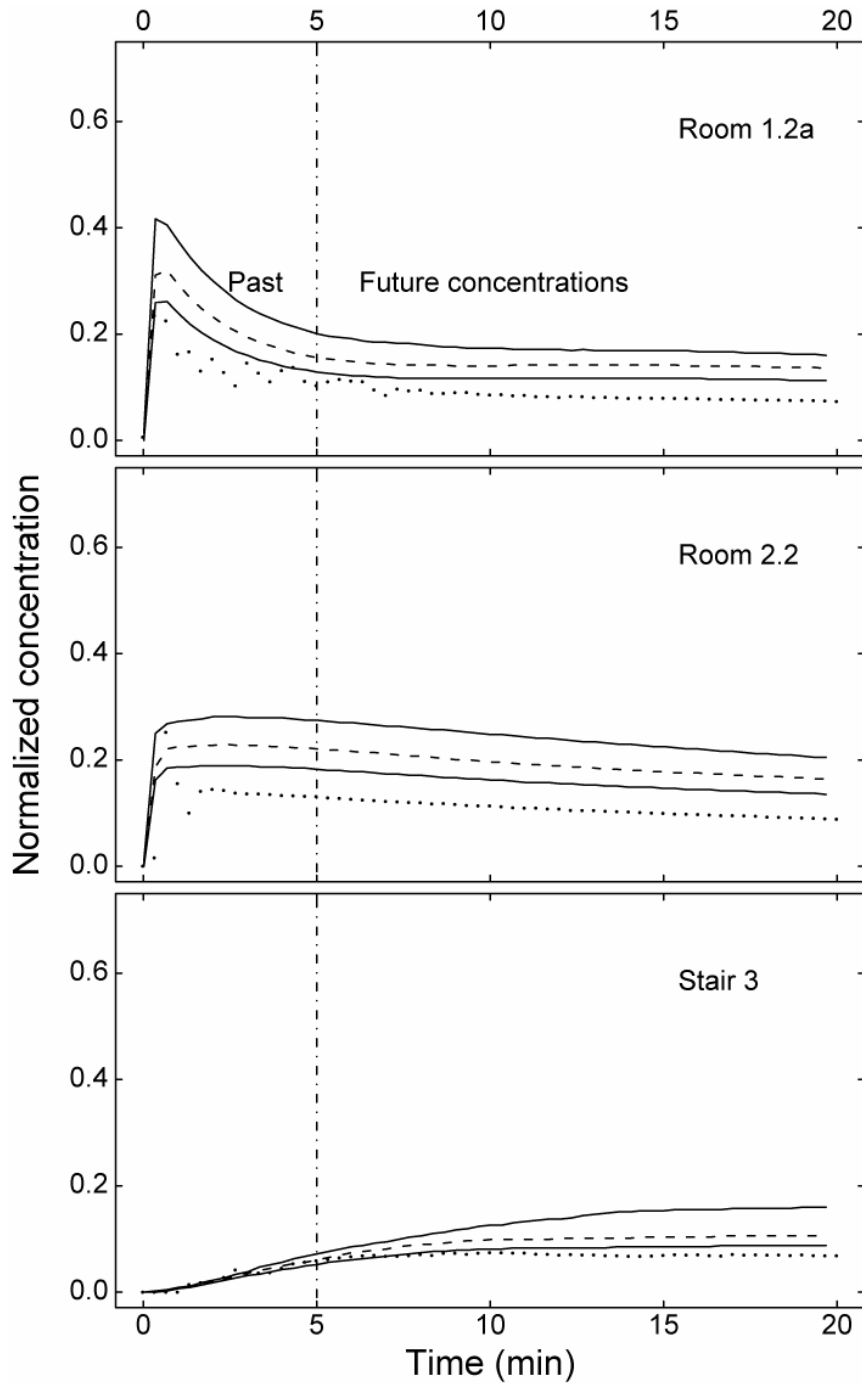


Figure 4.6. Confidence intervals of the concentration profiles for three zones after performing Bayesian updating using 5 min of alarm-type sensor data. Median (dashed line), 10th and 90th percentiles (solid lines), and concentration data (dotted line). Sensor signals were generated using a response time of 20 s and threshold level of 0.023.

Figure 4.7 shows the updated concentration distributions after assimilating 10 min of threshold data. The uncertainty regarding the concentrations is further reduced. The confidence intervals are now quite narrow. Two possible explanations for these narrow intervals exist. First, very high confidence is assigned to the model-to-measurement comparisons, resulting in very high posterior probabilities for a small number of realizations. In a real situation, these uncertainties are likely to be larger than the assumed 5% error in these simulations. Second, an improperly sampled library could result in negligible posterior probabilities for select realizations, leaving certain realizations, by default, with relatively high posterior probabilities. In this example, the first explanation is true. (Avoiding the second situation is the reason for a large library sampling size of 5000 simulations.) The purpose of these figures is to demonstrate the ability of the sensor system, using threshold signals, to reduce the uncertainty of the interpreted and predicted concentration distributions, and not to generate an optimal algorithm for this particular data set.

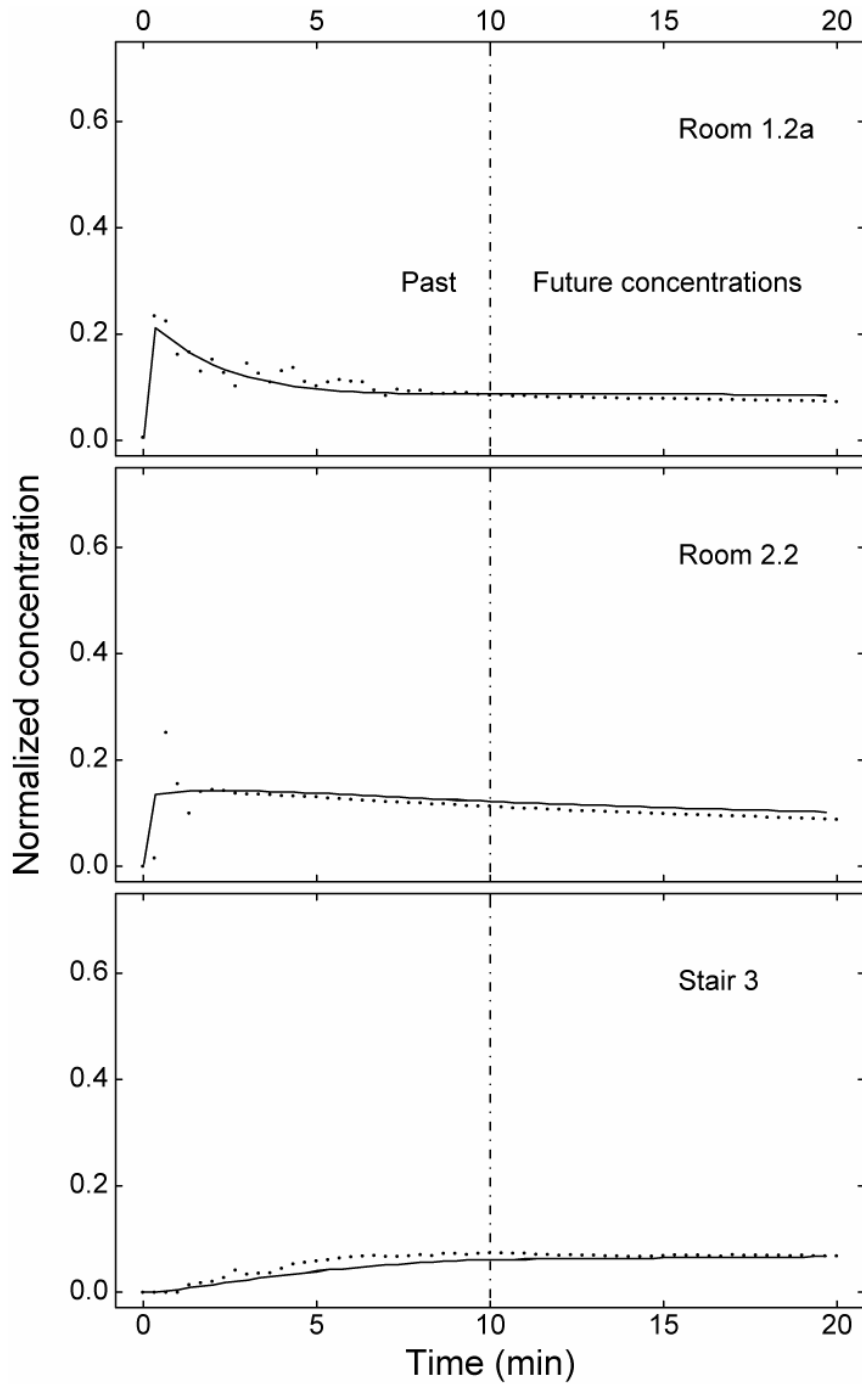


Figure 4.7. Confidence intervals of the concentration profiles for three zones after performing Bayesian updating using 10 min of data. Median, 10th and 90th percentiles of the estimated concentration profiles (solid line), which are coincident; concentration data (dotted line). Sensor signals were generated using a response time of 20 s and threshold level of 0.023.

#### 4.4.2 Effects of sensor threshold level and response time

In this section, the trade-offs between sensor threshold level and response time are explored. The threshold level was varied between  $2 \times 10^{-4}$  and 0.16 (equivalent to a range of 0.02% to 16% of the theoretical peak concentration) and the response time was varied between 20 seconds and 180 seconds. No added error was included in these data sets, to isolate the effects of threshold level and response time. Figure 4.8 shows the time required to identify the correct release location with 90% confidence for these cases. The lowest threshold levels ( $2 \times 10^{-4}$  and  $2.3 \times 10^{-3}$ ) and highest threshold levels (0.14 and 0.16) are often unable to identify the release location because they yield data that varies little among zones. Figure 4.8 also shows that for the intermediate threshold levels there is little difference in the time required to locate the source, particularly for sensor response times between 20 and 120 seconds.

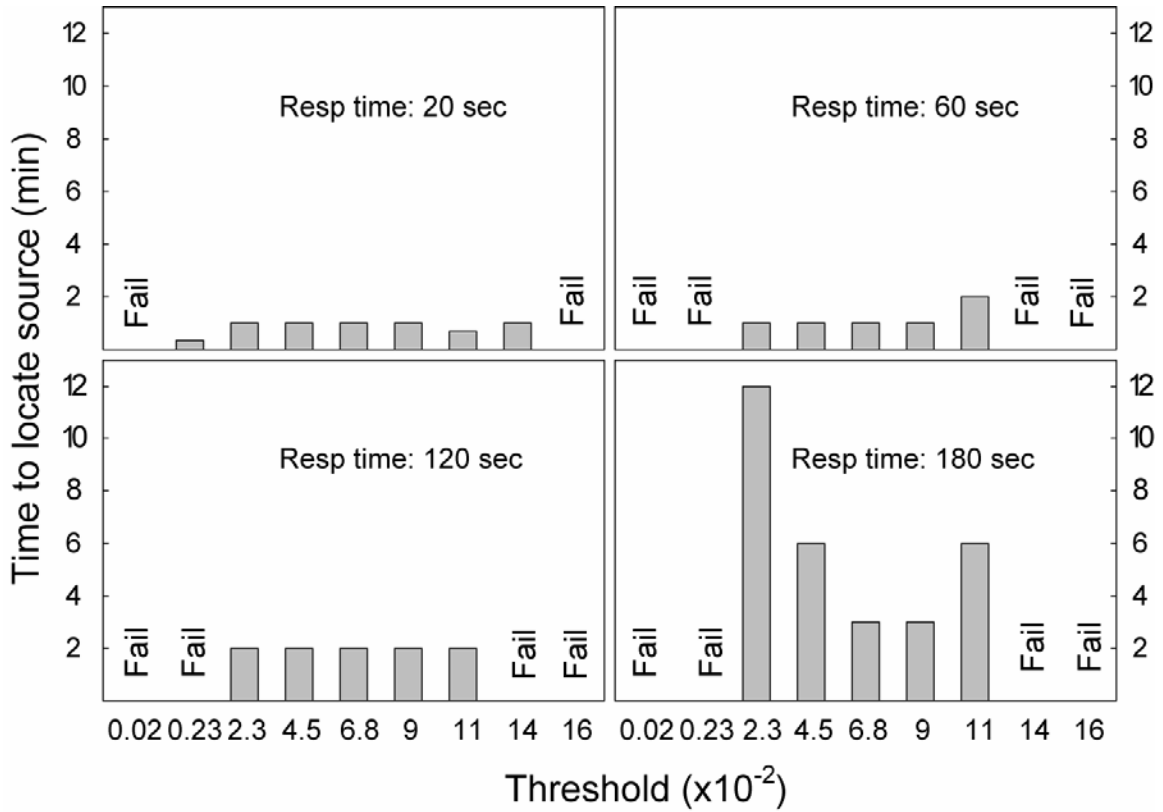


Figure 4.8. Time required to locate the release location with varying threshold level and sensor response time.

Figure 4.9 shows determinations of the release amount for four combinations of response times (60 seconds and 120 seconds) and threshold levels (0.023 and 0.09). With higher threshold level sensors, the algorithm estimates the release mass with a more narrow confidence interval more quickly. The sensor system with the high threshold level sensors reduces the posterior probabilities of many more realizations than the sensor system with the low threshold level sensors. Therefore, the calculated parameter uncertainty bounds for release mass are narrower. This relationship was observed consistently across all response times and thresholds between 0.023 and 0.11.

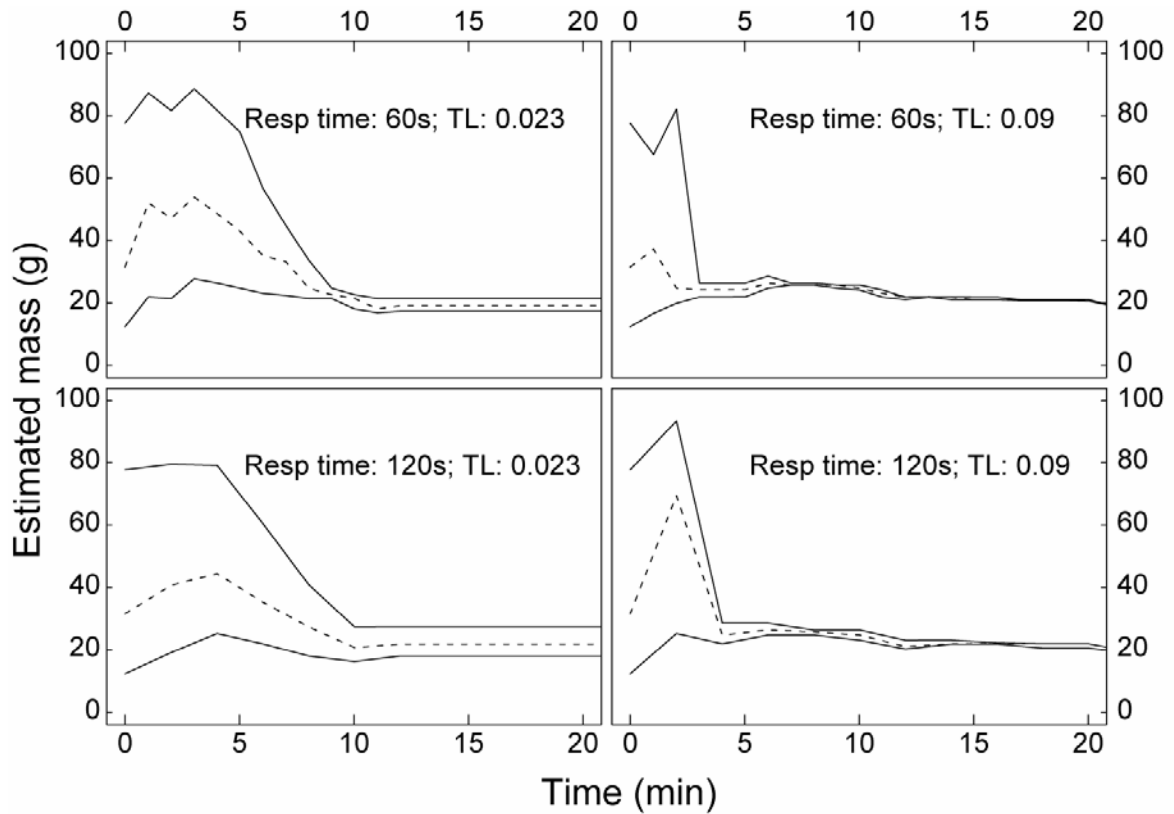


Figure 4.9. Estimated mass released as a function of threshold level and response time. Median (dashed line), and 10th and 90th percentiles (solid lines) of the estimated release mass are shown. (True release mass was 20 g.)

Over the range of conditions shown in Figure 4.9, the effect on the results of sensor response time is weaker than the effect of threshold level. As the response time increases, the sensor system receives information less frequently, and the confidence intervals broaden somewhat. In comparing Figures 4.8 and 4.9, one can observe that more time is needed to estimate with high confidence the release mass than the release location.

Summarizing the results depicted in Figures 4.8 and 4.9, the performance of the sensor system depends significantly on threshold levels relative to actual concentrations, and less on sensor response time, at least over the range of conditions studied. If airborne



concentrations in the actual building can vary over a broad range, a sensor system with multi-threshold sensors should be considered, since they may better cover the range of airborne concentrations.

#### **4.4.3 Effects of sensor error**

In this example, I include error in the threshold sensors. As expected, the results show that the algorithm requires more data (and, thus, more time) to identify the source zone with high confidence when error is present in the sensor signals.

Figure 4.10 shows the time required to locate the release location for the case of fixed sensor response time (20 s) and error (30%), but with varying threshold level. Each of the three cases is based on 50 data realizations with randomized error applied to the threshold signals. Each frame in the figure shows the time-dependent probability of correctly identifying the release location.

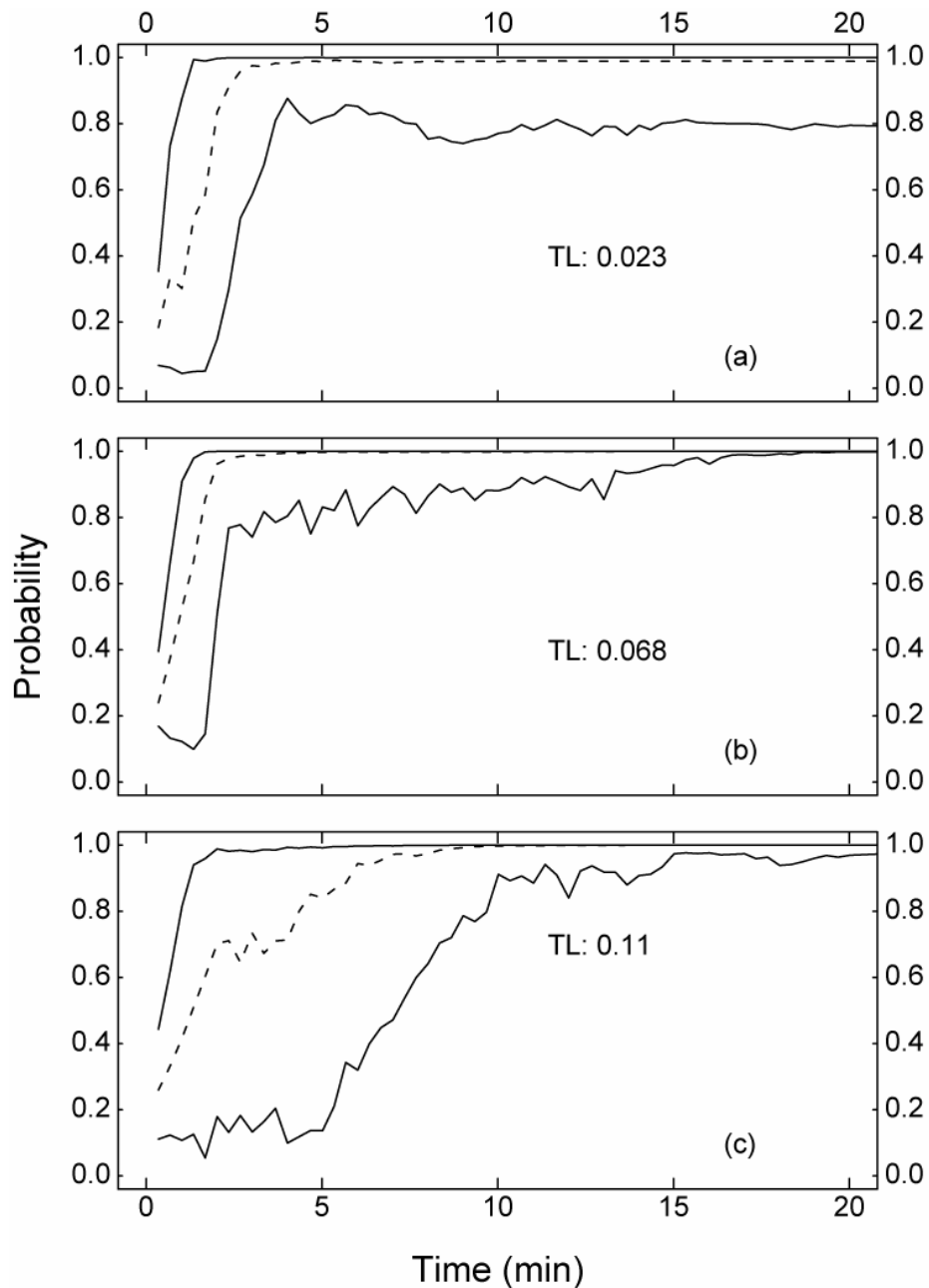


Figure 4.10. Probability of correctly identifying the release location versus time for release at time = 0 using threshold data generated with 30% error, variable threshold level (TL), and fixed response time (20 s). Median (dashed line), 10th and 90th percentiles (solid lines) for 50 data realizations. Threshold level: (a) 0.023, (b) 0.068, (c) 0.11. Each zone in the building has one sensor (11 total).

Even with relatively high error, the system containing sensors with low (0.023) and medium (0.068) threshold levels rapidly identifies the source location with high

confidence in most realizations. The sensor system with medium threshold sensors performs best under these conditions, demonstrating the ability to locate the source with 90% confidence more rapidly in most cases than the low threshold level.

The medium-threshold sensor performs better than the low-threshold sensor in this case for two likely reasons. (1) The medium threshold level reduces the posterior probabilities of a larger subset of model realizations than the low-threshold level. (2) While a low-threshold sensor receives signaling-on information (i.e., when the concentration first exceeds the threshold level) before the medium-threshold sensor, a medium-threshold sensor receives the signaling-off information (i.e., when the concentration subsequently falls below the threshold level) sooner. If both signaling-on and signaling-off information is important for data interpretation, the sensor system based on medium-threshold sensors may provide important information more rapidly.

In contrast to the sensor system with low and medium threshold sensors, the sensor system with high threshold (0.11) sensors requires significantly more data to estimate the release location with high confidence, for a majority of the data realizations (Figure 4.10(c)). The wide confidence intervals in Figure 4.10(c) prior to seven minutes are caused by inconclusive or incorrect estimates of the release location for several data realizations.

Overall, the results obtained by investigating data without adding simulated error (i.e., false positives and negatives) suggest that certain parameters are more easily estimated than others; namely, the release location was determined more easily than the release mass. A sensor system should be designed with consideration of the specific performance objectives of the overall system, given the many possible goals of a

monitoring system. For example, accurate predictions of evolving concentrations may rely on accurate estimates of both release location and mass.

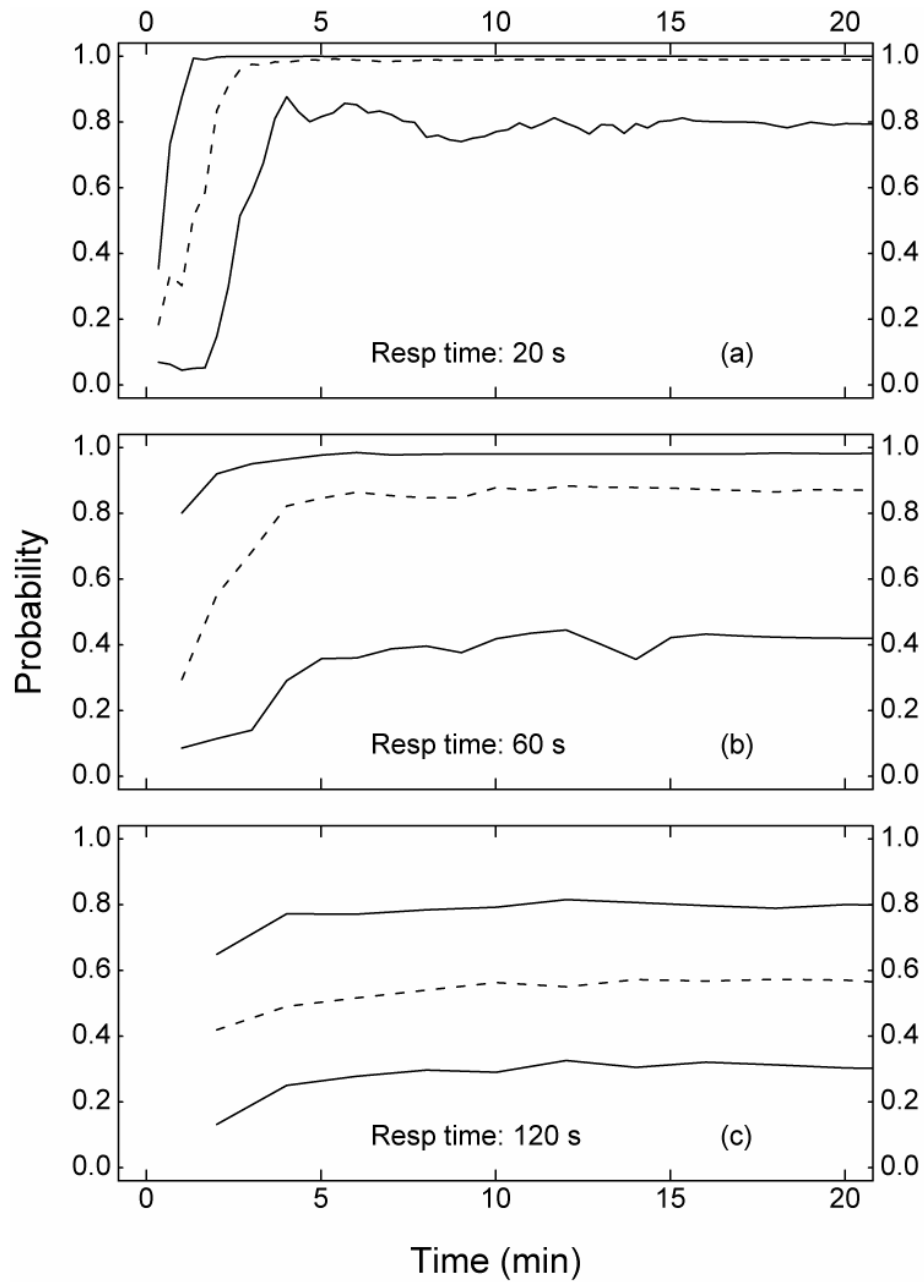


Figure 4.11. Probability of correctly identifying the release location using threshold data generated with 30% error, variable sensor response time, and threshold level of 0.023. Median (dashed line), 10th and 90th percentiles (solid lines) for 50 data realizations. Sensor response time: (a) 20 s, (b) 60 s, (c) 120 s.

Figure 4.11 shows how the time required to identify the release location varies with sensor response time. In these simulations, I used sensors with a 0.023 threshold level and 30% error. As the response time increases from 20 s to 120 s, the ability of the system to determine the release location and to estimate the released mass degrades. These findings are similar to those discussed in Section 4.4.2. For sensors with a fixed threshold level, a longer response time reduces the amount of information, and, therefore, reduces the certainty with which release parameters can be estimated.

To investigate tradeoffs between sensor speed and accuracy, I compare system performance with more rapid but less accurate sensors against system performance with slower but more accurate sensors (Figure 4.12). System 1 consists of sensors with a 20-s response time and 30% error (Figure 4.12(a)); System 2 consists of sensors with a 60-s response time and 10% error (Figure 4.12(b)); and System 3 consists of sensors with a 120-s response time and 10% error (Figure 4.12(c)). All three systems utilize sensors with a threshold level of 0.023. System 1 receives data at a higher rate, but lower accuracy than do the other systems. System 2 receives data at a higher rate than System 3, but with the same accuracy. I observe that System 2 clearly outperforms the other two systems, and System 3 performs slightly better than System 1. Therefore, in this example, a sensor system that is based on slower but more accurate sensors identifies the release location with higher confidence, more quickly, than a sensor system using faster, but less accurate sensors. This result was consistently observed in the simulations across most threshold levels, suggesting that slower, but more accurate threshold sensors may be more informative and lead to faster definitive interpretation than rapid, less accurate, threshold sensors when deployed in a building-monitoring system.

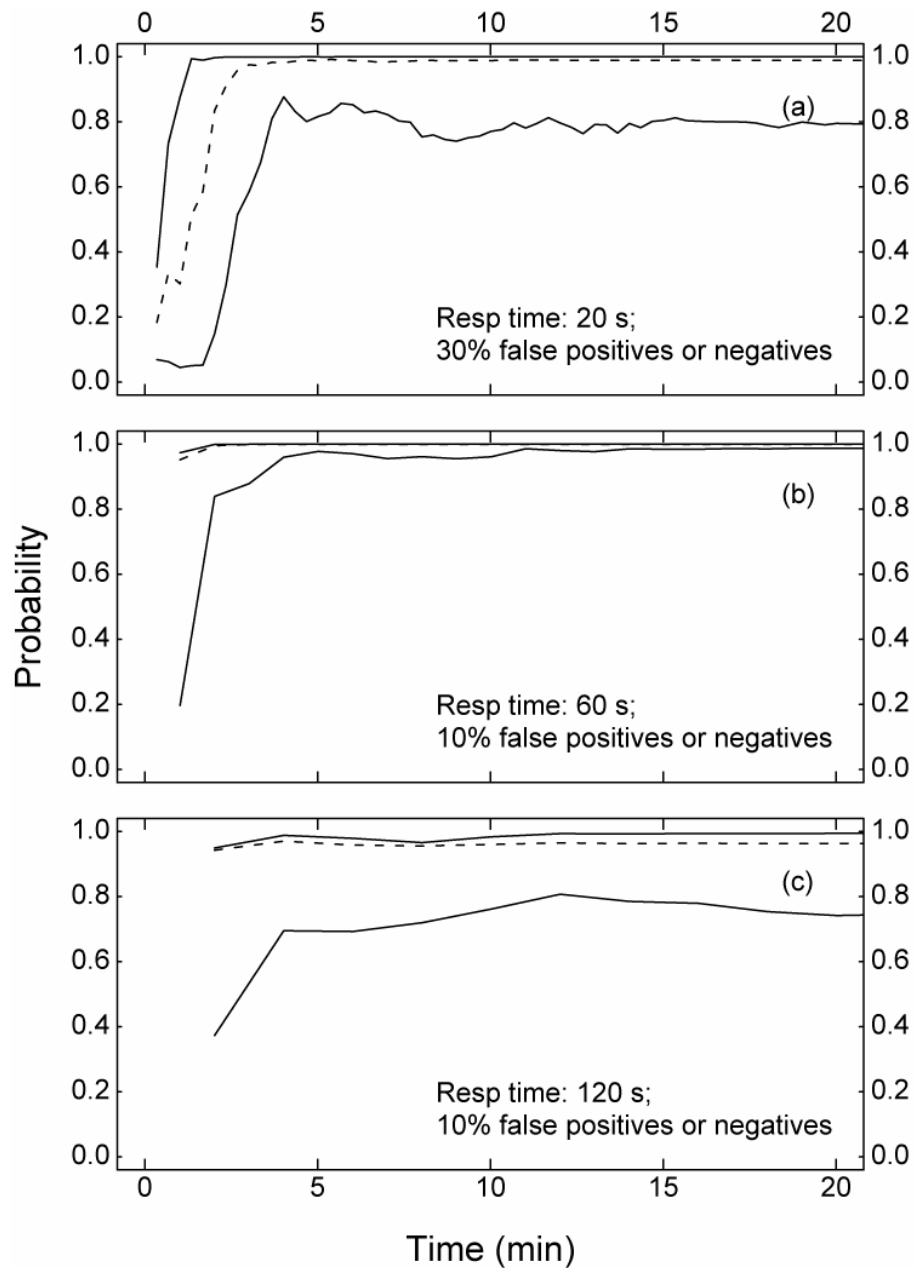


Figure 4.12. Probability of correctly identifying release location versus time since release using sensors with varying sensor response time and error and at a threshold level of 0.023. Median (dashed line), 10th and 90th percentiles (solid lines) for 50 data realizations. (a) Response time 20 s, error 30%; (b) response time 60 s, error 10%; and (c) response time 120 s, error 10%.

#### **4.4.4 Evaluating additional release scenarios**

The investigations in this chapter discussed so far were conducted using tracer data obtained from Experiment 1. To ensure that the findings were not idiosyncratic, owing to some anomalous features of Experiment 1, hypothetical threshold data were generated from the other experiments using a subset of the total sensor characteristics investigated.

In these explorations, threshold data were generated using a low threshold level of 0.023 and a sensor response time of 20 s. Randomly generated error was not added to the signals and the library was not altered. From this exercise it was found that the results from all experiments exhibited some common behaviors. In almost all cases, sensor systems using threshold signals were eventually capable of identifying the source location to a relatively high probability. The 80% 2-sided confidence intervals of the mass and duration follow a similar pattern as exhibited by Experiment 1. More data are required to generate narrow confidence intervals for emitted mass, relative to the data required to identify the release location. For illustration purposes, the probabilities of accurately identifying the release location and the estimates of the released mass are shown for Experiments 2-4 in Figures 4.13-4.15. The experimental conditions differed mainly by the door positions. In Experiments 1 and 2 all doors were open, while in Experiments 3 and 4, the staircase doors were closed.

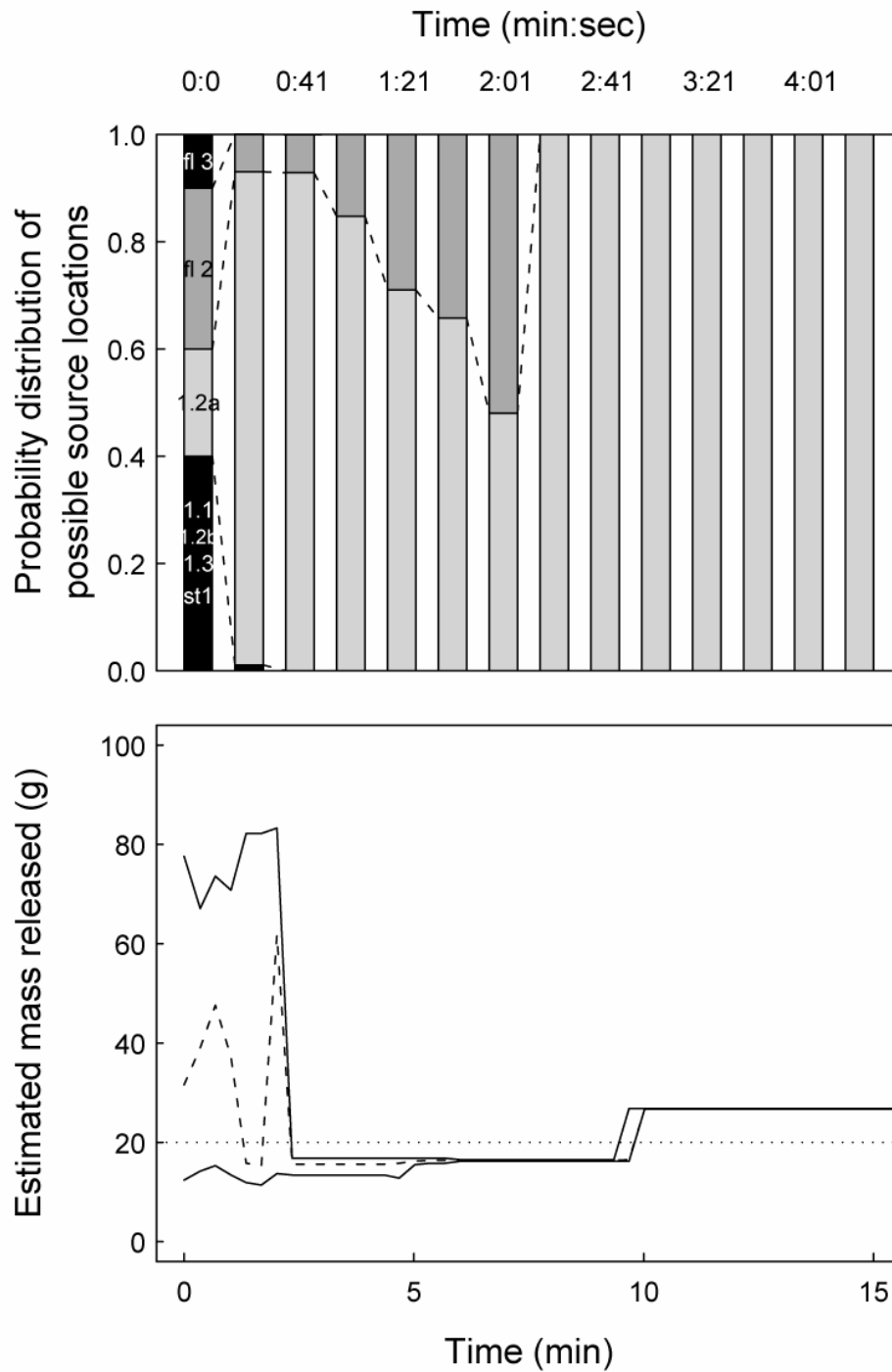


Figure 4.13. Determination of source location and released mass using threshold signals generated from Experiment 2 data, plotted versus time since release. Mass is shown with the 80% 2-sided confidence interval and median. Signals were generated with a sensor response time of 20 s, threshold level of 0.023, and without added error. The actual release location was the return intake in Room 1.2a, and release mass was 20 g.



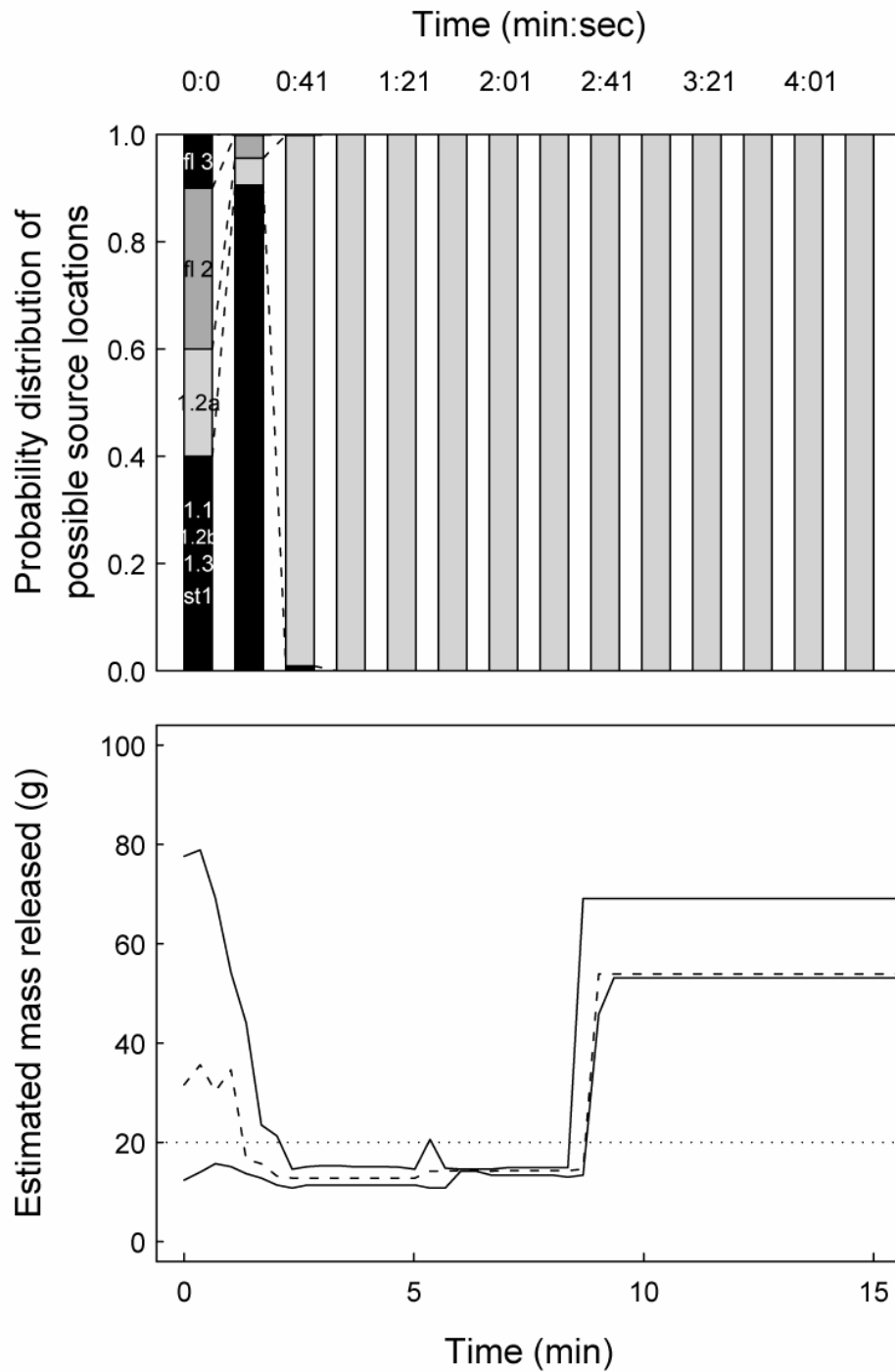


Figure 4.14. Determination of source and mass using threshold signals generated from Experiment 3 data. Mass is shown with the 80% 2-sided confidence interval and median; Signals were generated with a sensor response time of 20 s, threshold level of 0.023, and without added error. The actual release location was the return intake in room 1.2a, and release mass was 20 g.

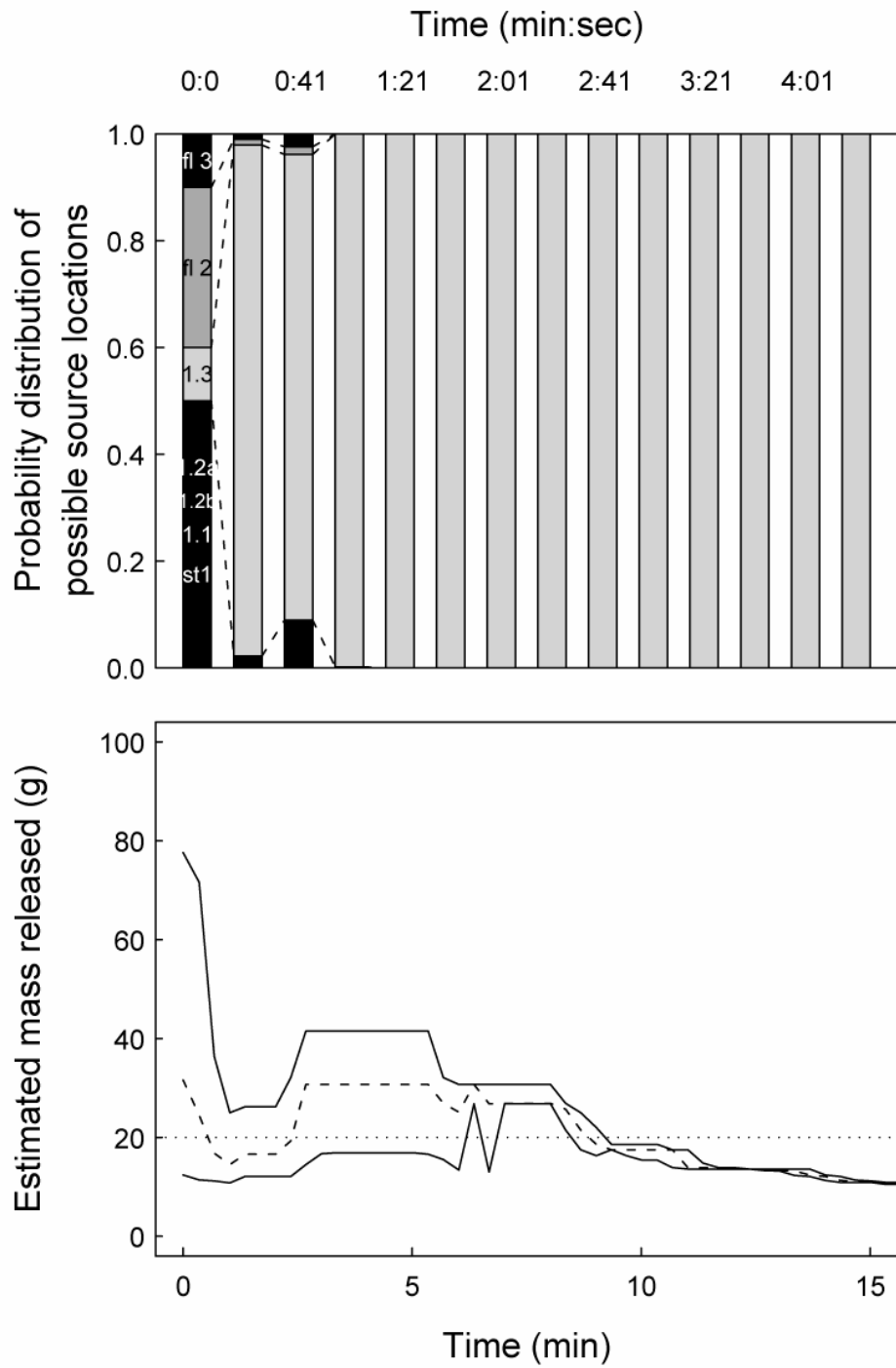


Figure 4.15. Determination of source and mass using threshold signals generated from Experiment 4 data. Mass is shown with the 80% 2-sided confidence interval and median); Signals were generated with a sensor response time of 20 s, threshold level of 0.023, and without added error. The actual release location was Room 1.3, and release mass was 20 g.

In Experiments 3 and 4, the algorithm eventually estimates a median mass release value that is either substantially higher (Experiment 3) or lower (Experiment 4) than the true quantity. These results suggest that the likelihood function assigns too much confidence to certain sensor signals. The time at which the signals exceed the threshold in certain zones may not be modeled accurately.

In the case of Experiment 2, the algorithm estimates consecutively lower probabilities for the true release room during the first two minutes after the release, suggesting that there may be model to signal mismatches at the early time periods that are not adequately accounted for by the likelihood function. In practice, early mismatches may arise from false positive or negative measurements and also from modeling errors. A series of false positive or negative measurements are likely to occur when the true concentration is close to the threshold level. I discuss the statistical implications of this behavior in the next section. Chapter 5 considers model specification error in detail and efforts to improve the likelihood function.

#### **4.4.5 Statistical considerations for interpreting threshold data**

A key objective of this chapter is to investigate how a system-level analysis may help identify an optimal mix of sensor characteristics for populating a sensor network. The sensor type selected for this illustration is alarm-type or binary signal sensors. While the objective of this chapter was not to develop an optimal algorithm for interpreting binary type data, it is worth discussing some related points that may in turn provide insight if further enhancement of the approach is desired.

Threshold sensors produce a binary output, only indicating when a contaminant's concentration is above or below a threshold. The likelihood function compares modeled concentrations and sensor signals at each time to the respective sensor threshold level. Hence, the only information that is needed — and used — from the library is the knowledge of whether the modeled concentration is above or below the threshold level. Therefore the information content for each sensor in each realization in the library can be described with parameters that indicate the time at which a threshold level has been crossed. Most sensors may experience two threshold level crossings: when the concentration is first exceeded as concentrations rise and the time that concentrations fall below the threshold level. For sensors with high threshold levels the threshold level may never be exceeded; and for others, oscillatory behavior may cause a threshold level to be crossed multiple times. In each case the amount of data that describes the threshold level crossing events is less than the concentration time series data.

Rather than implementing a likelihood function based on comparing thresholds to modeled concentrations, it may be possible to compare each sensor signal to the respective time that the threshold level is crossed for each realization in the library. Two possible advantages may result from a time-based likelihood function: (1) the library information for each realization can be consolidated from large sets of data to fewer parameters, which has implications for computational efficiency; and (2) the uncertainty at which the threshold level is crossed can be directly treated by the likelihood function.

Turning to another issue, the concentration data shown for Experiment 1 indicates that the concentration profiles does not consist of smooth curves, particularly, at early times following the release (Figure 4.1). Oscillatory behavior of the contaminant

concentration can result from spatial variability of the contaminant concentrations within each respective room. After some time these oscillations decay, and the concentration follows a well-mixed profile. The current statistical treatment does not account for oscillations about a threshold level in any special way. Because the realizations are generated using a multizone model that assumes the contaminant is well-mixed within a room, the realizations do not replicate any kind of oscillatory behavior.

In its current form, the presence of oscillations is a liability for the BMC algorithm. However, oscillatory behavior is potentially rich in information, irrespective of the cause for the oscillations, which can be error, or variability of the within-room concentrations. Oscillations about a threshold level suggest that the true concentration, which is unknown, is close to the threshold level. It is unlikely that any modeling approach can capture these oscillations with accuracy – even computational fluid dynamics is unlikely to capture the frequency and time of oscillations exactly, unless a large eddy simulator is used.

Hence, the burden is on the algorithm to capture the information-richness exhibited by oscillations. One approach may be to implement backwards Bayesian updating. In backwards Bayesian updating, previous rounds of data can be recycled to re-estimate current posterior probabilities of the realizations based on current information. Thus, while oscillations may be a liability as they are being interpreted in real-time by the algorithm, they become useful once they cease, and negative effects can be transformed through backwards updating.

## 4.5 Conclusion

The premise of this chapter is that the selection of sensor characteristics is best performed from a systems perspective. Hence, the primary purpose of this chapter is to illustrate the relationship between sensor characteristics and sensor system performance. Here, I have demonstrated — albeit for a limited set of circumstances — that a network of single-level threshold sensors can be used to determine the location and magnitude of a short-term release quickly and accurately using a Bayes Monte Carlo framework.

Sensor networks with sensor response times ranging from 20 s to 120 s and threshold levels ranging from 0.0023 to 0.11 could identify the release location to a probability exceeding 90% within 2 min. Sensors with very high and low-threshold levels have little discriminating power and could not characterize the release to useful levels. The BMC algorithm could estimate the release mass to a narrow confidence interval within 4 to 10 min across different threshold levels and response times. Higher threshold levels reduced the uncertainty of the released mass more quickly than lower threshold levels. Sensor response time had comparably less influence on the estimation of the mass.

When the sensor signals were spiked with random error of 10% and 30%, the systems were able to identify the release location to a 90% probability, for a majority of networks. Higher threshold levels, however, increased the risk of failure because they reduce the posterior probabilities of a greater number of realizations in the library as compared to lower threshold levels. Networks with more accurate but slower sensors resulted in more reliable predictions of the release location as compared to networks with

faster but less accurate sensors. This result suggests that the sensor response times evaluated are adequate for capturing the dynamics of the contaminant transport.

More important than the specific results, this chapter demonstrates that treating the network as a system may lead to better choices for sensor characteristics, like response time and error, than might be the case when considering sensors individually.

With respect to the specific investigations, the actual rate of false readings (i.e., false positives or negatives) for real sensors is likely to be less than 30%. Probabilistic results reached by the Bayesian algorithm based on the assigned confidence in this chapter are therefore likely to be conservative for release conditions considered in the library.

Several questions emerge from the research reported in chapter. Would networks consisting of fewer sensors have success at characterizing the release? And, if fewer sensors are adequate, where should they be placed? If only a few chemical sensors are available, and inadequate for achieving system performance goals, can other types of sensors be incorporated into the system to help achieve those goals in a more cost-effective manner? This chapter only investigated an instantaneous release. How well would the algorithm work against a full array of release conditions, including slow, steady releases? Currently, the algorithm does not advantageously use sensor oscillations, which are potentially information rich. Such oscillations are not easily accounted for in the multizone transport and fate model. To what extent does this behavior, or other unmodeled contaminant transport behavior present challenges, or opportunities, for developing more robust sensor systems? These questions inspire the investigations reported in Chapters 5 and 6.

## Chapter 5

# 5 Influence of transport and mixing time scales on sensor system performance

Reproduced in part from research published in Atmospheric Environment.<sup>2</sup>

### 5.1 Introduction

The release of airborne contaminants, whether accidental or by malicious intent, can cause acute harm to those exposed. Because of the risks posed, efforts are under way to develop sensors capable of rapidly detecting toxic airborne chemicals. Using such sensors effectively requires integrating them into a system that ideally should characterize the release in addition to detecting its occurrence, both in real-time. An integrated system design approach that collectively considers the contaminant transport physics, sensor characteristics, and sensor interpretation algorithms can help to achieve these goals.

In Chapter 4, the system-level performance trade-offs among individual sensor characteristics were investigated. It was shown that the Bayesian approach could successfully characterize releases when interpreting data from alarm-type sensors in a

---

<sup>2</sup> Sreedharan, P., Sohn, M.D., Nazaroff, W.W., Gadgil, A.J., 2007. Influence of indoor transport and mixing time scales on the performance of sensor systems for characterizing contaminant releases. Atmospheric Environment 41, 9530–9542. Copyright 2007, Elsevier, Inc.



real building. This effort underscored the importance of a systems perspective for selecting the optimal mix of sensor performance characteristics, such as response time and accuracy. Furthermore, the relationships between individual sensor characteristics and system level performance could be explained through the contaminant dispersion properties, and was not solely an algorithmic consequence. Thus, there may be conditions in which further improvement in sensor performance, or network design may be unfruitful owing to some characteristic of contaminant transport. In this chapter, I consider factors for sensor systems associated with the contaminant transport physics.

Recall, the sensor system operates on a two-stage Bayes Monte Carlo (BMC) algorithm. The first stage, which occurs before monitoring, involves generating a database of realizations with time-dependent chemical concentrations that reflect possible release scenarios. In this research, a multizone indoor air pollutant transport and fate model executes this step. The second stage occurs during monitoring. Once an event is detected, Bayesian inference is applied to estimate the probability of each library simulation having occurred. This step provides statistical information about the model inputs and future concentrations.

Chapter 4 demonstrated that a system-level analysis may help select sensors that better explain the physics. In this chapter, I reverse this design question and begin with the physics. How might the contaminant fate and transport influence the ability of a sensor system to characterize the release? How do the shortcomings in the model's representation of the transport physics, and used to generate the library, affect the BMC algorithm's ability to characterize the release? In turn, how can an understanding of these issues inform an improved sensor system design?

This chapter provides partial answers to these important questions. I discuss the time scales that characterize contaminant transport in buildings and how the information contained in sensor data degrades over time relative to these scales. The transport time scales provide an intuitive vehicle for exploring the role of transport physics in sensor system performance and are particularly pertinent to this dissertation because the sensor system processes information sequentially in time. This discussion is used to explore how limitations of a multizone transport model may affect the estimation of release parameters. I investigate the relevance of the time scales to the design challenge, first by blinding portions of the data to the algorithm, and then by evaluating the performance of several sensor systems differing in sensor number, location, and response time. To pursue the investigation, I first develop a likelihood function, using empirical data, that aims to characterize model specification error.

## **5.2 Background**

Two key issues in a Bayes Monte Carlo based sensor system are the selection of the model used to generate the simulation library and the likelihood function that quantifies the relationship between the modeled and measured values. These two subjects are reviewed in the remainder of the background section. The case study is also briefly reviewed.

### **5.2.1 Multizone model; likelihood function**

A multizone model is used to generate the simulation library, which is a network type model. The building is described as a network of well-mixed chambers or zones,

connected by flow paths. The outdoors and air handling unit (AHU) devices are represented as additional zones. Airflow between zones is induced by wind, buoyancy, or mechanical means, and is assumed to depend on the pressure difference between zones and the resistance of the flow paths that connect them. There are limitations to the multizone modeling approach. The assumption that the contaminant instantaneously mixes within a particular zone is a cause for errors particularly at early time periods immediately following the release. These errors may affect the algorithm's ability to extract information from the data.

Recall, the BMC monitoring scheme involves two stages. In the second stage, Bayes' rule is applied to compare the library of simulations to the observations and to update the probabilities that individual simulations accurately describe the situation as data arrive sequentially in time.

Given a vector of sensor measurements,  $O$ , the probability that these correspond to simulation  $k$ , with its corresponding concentration values  $Y_k$ , is denoted as  $p(Y_k|O)$  and is calculated as follows:

$$p(Y_k | O) = \frac{p(O | Y_k)p(Y_k)}{\sum_{i=1}^N p(O | Y_i)p(Y_i)} \quad (5.1)$$

where  $p(O|Y_k)$  is the likelihood function,  $p(Y_k)$  is the prior probability of  $Y_k$ , and  $N$  is the number of simulations in the library. The likelihood function describes the mismatch between measurements and model predictions, which are due to either model or measurement error.

## 5.2.2 Case study review and library generation

The case study consists of one unit in a 3-floor building at the Dugway Proving Grounds, Utah (Sextro et al., 1999). The interior volume is 660 m<sup>3</sup>, with approximately 280 m<sup>2</sup> of total floor area on three levels (Figure 3.1). An AHU supplied 100% recirculated air to the lower two floors.

Sextro et al. (1999) conducted fan pressurization tests to determine leakage rates and performed 12 tracer-gas experiments over a period of five days (Table 3.1). They used the measured air-leakage characteristics to generate a multizone contaminant transport model. I used the same model to generate the library for the first stage of the Bayesian algorithm, with one simplification: the average weather and temperature conditions from all tracer-gas experiments were treated as fixed input conditions. I generated variations on this basic model by sampling from a distribution of possible release locations, release masses, release durations, and door configurations, as summarized in Table 5.1. The resulting library comprised 5000 possible release scenarios, each with a simulated set of time-dependent concentrations for each sensor location.

Table 5.1 Parameter values used to generate the library of 5000 simulated contaminant releases

Parameter	Values
Source location	Twelve locations: any room, or stairwell, and return duct. Each location is equally probable.
Source duration	1 s to 5 min; log-uniform distribution.
Source amount	10 to 100 g; log-uniform distribution.
Door position	32 possible door positions, based on interior doors on each level and stairwell doors acting independently

### 5.3 Transport time scales in buildings

I seek to understand how knowledge of contaminant transport physics can help diagnose BMC algorithm results and thereby improve sensor system design. Because the BMC algorithm processes information sequentially, I describe three building transport time scales and explore how the information content of the data changes with time. An additional reason for studying these time scales is that they are intrinsically connected to important parameters that characterize the release.

The following discussion assumes that the monitored building is equipped with a conventional commercial building ventilation system that discharges conditioned air from overhead registers to promote rapid mixing in a room. I define time scales associated with three physical processes. The first, denoted  $\tau_1$ , represents the time required for mixing within a room. Imagine that a contaminant is suddenly released at time  $t = 0$  at a point within a room. Then  $\tau_1$  is defined such that for times  $t \ll \tau_1$  the contaminant concentrations vary strongly with position, whereas for times  $t > \tau_1$ , the concentrations are approximately uniform. For mechanically-induced airflow, Drescher et al. (1995) found that  $\tau_1 \sim 18 \rho^{1/3} V^{5/9} P^{-1/3}$ , where  $\rho$  is the air density,  $V$  is the room volume, and  $P$  is the mechanical power input of the induced airflow. For flow generated only by the AHU,  $P \sim \frac{1}{2} \rho A v^3$ , where  $A$  is the supply area of the discharge register and  $v$  is the discharge air speed. For example, taking  $V = 100 \text{ m}^3$ ,  $v = 2 \text{ m s}^{-1}$ , and  $A = 0.13 \text{ m}^2$  gives  $P \sim 0.7 \text{ W}$  and  $\tau_1 \sim 5 \text{ min}$ . The Drescher et al. experiments were performed in a room without any air exchange outside of the room, and with mixing induced by internal blowers. To my

knowledge, there are no analogous studies of mixing time when the room is subject to substantial air exchange.

A second time scale is that required for a local release to disperse uniformly through many rooms of a building, such as all rooms served by a single AHU. This time scale, denoted  $\tau_2$ , can be defined as follows. Consider a situation in which air exchange from outside the building is suppressed. A contaminant is suddenly released into one room at time  $t = 0$ . For times  $t \ll \tau_2$ , there is substantial room-to-room variation in the contaminant concentration. For times  $t > \tau_2$ , little room-to-room variation exists. Empirical data on  $\tau_2$  for buildings is lacking. A magnitude estimate is the total interior volume divided by the total rate of interzonal airflow, with the latter represented by the sum of the magnitudes of the incoming ventilation and open doorway airflows. By applying a rough AHU sizing rule (18 m<sup>3</sup>/h per square meter of floor area, or 1 ft<sup>3</sup>/ft<sup>2</sup> min<sup>-1</sup>) to a building with room heights of 3.7 m, and neglecting non-mechanically induced airflows,  $\tau_2$  for a group of rooms served by a single AHU is ~12 min. From this line of reasoning,  $\tau_2$  may be longer than, but is expected to be within an order of magnitude of  $\tau_1$ .

The third time scale,  $\tau_3$ , is the time required for the contaminant to be removed from the building and can be estimated as the building volume divided by the outside-air ventilation rate. Persily et al. (2006) reported a median value of 1 h, with 25<sup>th</sup> and 75<sup>th</sup> percentiles of 0.5 and 2.6 h for U.S. office buildings.

The extent to which concentration data contains information about a release may be explained using these time scales. Consider the typical situation, in which  $\tau_1 < \tau_2 < \tau_3$ , and that a multizone model is used to generate the simulation library.

For times  $t < \tau_1$ , mismatch between the modeled, well-mixed concentrations and sensor measurements are likely, because the model does not account for within-room variation. It is important to account for this mismatch statistically, to ensure that the interpretation algorithm is not confounded by the within-room variations. One method to do so is to assign less relevance to the measurements taken during this earliest time period.

A consequence of model-to-measurement errors for times  $t < \tau_1$  is that it may be challenging to extract useful information from the measurements within this time period. This characteristic may have consequences for estimating any parameter during this early time period, since all parameters are estimated from the posterior probabilities of the library simulations. For example, if the release duration is shorter than  $\tau_1$ , then an algorithm using a multizone model may be unable to estimate the release duration to a confidence interval that is less than  $\tau_1$ .

For times  $\tau_1 < t < \tau_2$ , conditions are likely to be ideal for determining the release location. Concentration differences among the rooms still exist, and greater model-to-measurement agreement is expected because of thorough intrazonal mixing, as compared to times  $t < \tau_1$ . Consequently, there are no special statistical challenges that must be overcome in comparing model with measured concentrations during this period.

For times  $\tau_2 < t < \tau_3$ , information about the release location is substantially lost, because the concentrations are similar among the rooms, and, therefore, there is little connection between the release location and the measured concentrations anywhere in the building. However, during this time period, an overall contaminant mass balance suggests that measured concentrations may still provide high-quality information about

the release quantity. The ability to estimate this mass will depend on the accuracy of the modeled indoor-to-outdoor air-exchange rates.

Finally, for times  $t > \tau_3$ , almost all information about the release is lost, since for  $t \gg \tau_3$  the contaminant is no longer present in the building.

Conditions may be more complex than represented above. For example, the transport and mixing time scales may be ordered differently, or may be sufficiently close as to be indistinguishable. As ventilation systems supply more outside air, with little or no recirculated air (e.g., to save energy), the interzonal transport time scale,  $\tau_2$ , will lengthen, and may exceed the ventilation time scale,  $\tau_3$ . A building with multiple AHUs may have several interzonal transport and ventilation time scales that vary with the characteristics of each AHU and with the airflow rates between zones. For a displacement ventilation system, a within-room transport time scale other than  $\tau_1$  may be relevant.

Despite these complications, identifying and quantifying transport and mixing time scales in relation to sensor performance and simulation model assumptions may be valuable for designing sensor systems of the type described in this dissertation. At a minimum, these concepts help frame the important issues for pursuing sensor system design for more complex systems.

## **5.4 The likelihood function**

The time scales help to explain how model-to-measurement mismatches may occur. Such mismatches must be accounted for in the likelihood function that is used in the



second stage of the BMC sensor system. I discuss here a method for reconciling model specification errors for the illustrative case study.

#### **5.4.1 Likelihood function development**

I considered all of the tracer gas data from the eleven experiments in which the AHU was operating (Table 3.1). Data from Experiments 5, 6 & 8-12 (“development experiments”) were used to produce the likelihood function, and data from Experiments 1-4 were reserved for subsequent testing.

For the seven development experiments, I ran multizone model simulations for the known experimental release conditions and subtracted the modeled concentrations at each monitoring location from the measurements. These errors were normalized by the hypothetical initial concentration of an instantaneously well-mixed building – that is, by the mass of the released tracer divided by the total interior volume of the test space. The smaller points plotted in Figure 5.1 show the normalized errors. They decay with time, as the tracer disperses through the building, but do not converge to zero. The large errors at early times are likely contributed to by the multizone model’s inaccurate assumption of instantaneously realized well-mixed conditions. At later times, errors arise in part from the model misspecification of interzonal airflow rates.

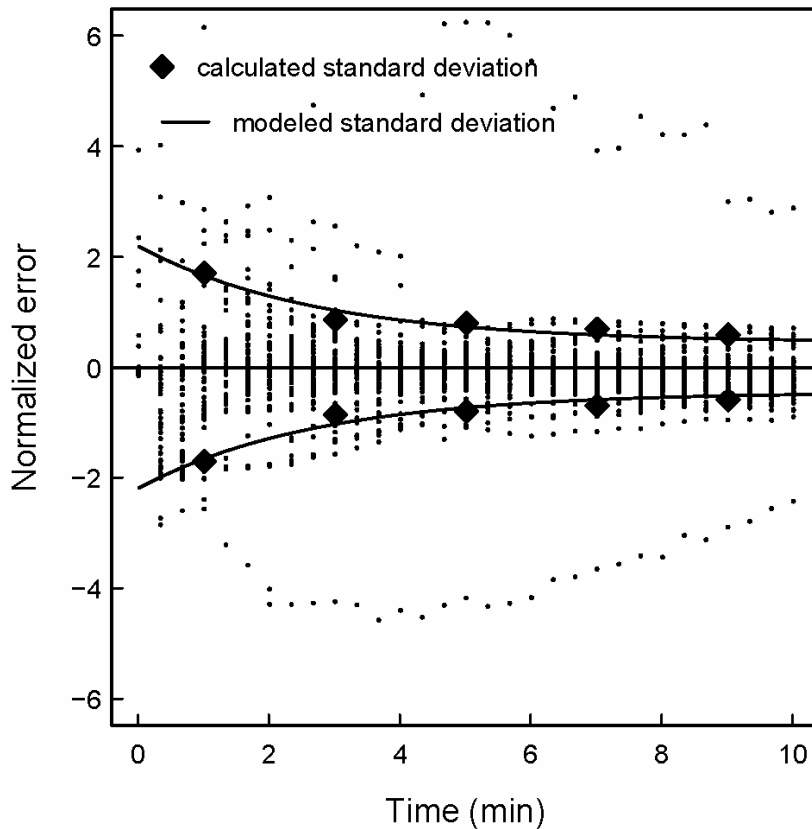


Figure 5.1. Normalized model-to-measurement error for seven tracer gas experiments. (small points, Experiments 5,6 & 8-12). Absolute errors are normalized by the instantaneous concentration of a well-mixed building. Also shown are calculated standard deviations for errors grouped into two-minute intervals and modeled standard deviation of normalized errors using an exponential fit, equation 5.2.

The data suggest a need for a time-varying error function. For the seven development experiments, I grouped the data into 2-min intervals, computed the standard deviation,  $\sigma$ , of the normalized errors in each group, and fit a three-parameter exponential function to the standard deviations:

$$\sigma(t) = Ae^{-Ct} + B \quad (5.2)$$

where  $A = 1.74$ ,  $B = 0.46$ ,  $C = 0.37 \text{ min}^{-1}$ . Figure 5.1 shows the modeled  $\sigma(t)$  in comparison to the empirical standard deviation of the normalized data-to-model errors. Since equation 5.2 describes the normalized errors, it can be scaled for different release masses by multiplying by the release mass and dividing by the interior volume of the entire building.

Equation 5.2 was developed using data from instantaneous releases. To populate a realistic simulation library, I require a likelihood function that includes the effect of finite release duration. Although I do not have experimental data for longer duration releases, I synthesized a likelihood function for longer duration releases. I next describe this process. It is important to note that this investigation is preliminary and further investigation is necessary, preferably using experimental data.

I generated hypothetical data for finite-duration releases, but within the limit of the durations considered in the generation of the library (i.e., 5 min), using linear superposition as follows. The original data, which were collected at 20 s intervals (for  $0 < t < 20 \text{ min}$ ) and 60 s intervals (for  $t > 20 \text{ minutes}$ ), were converted into 1 s data sets (denoted  $c_{inst}$ ) by linear interpolation. Hypothetical release data were constructed by summing 1 second time-shifted series of  $c_{inst}$  and dividing by the hypothetical release duration,  $\tau_{release, sec}$ . Values of 120 s (2 min) and 300 s (5 min) were assigned to  $\tau_{release, sec}$ . The concentration data at  $t$  seconds,  $c[t]$ , for a hypothetical release of finite duration is:

$$c[t] = \frac{1}{\tau_{release, sec}} \sum_{i=0}^{i \leq t} c_{inst}[i] \quad (5.3)$$

where  $c_{inst}[i]$  is the observed concentration at  $i$  seconds for the original data, as obtained by linear interpolation. Linear superposition will not model the turbulent flow characteristics accurately, especially that associated with large-scale eddies. However, in

lieu of experimental data, it provided a vehicle to begin exploring the relationship between release duration and model-to-measurement errors.

The concentrations obtained by equation 5.3 were compared to the predictions from a multizone model to generate errors, and these errors normalized using to the hypothetical release mass and building volume. To derive an empirical error function, which is required for the likelihood function, an assumption was made that the model-to-measurement errors of the longer duration releases at time  $t$  are linearly related to the fraction of the unit mass released at  $t$ . Then, the normalized errors for a release of duration  $\tau_{release}$  are as follows:

$$e(t) = e_{inst}(t) \times \frac{t}{\tau_{release}}, t \leq \tau_{release} \quad (5.4a)$$

$$e(t) = e_{inst}(t), t > \tau_{release} \quad (5.4b)$$

where  $e_{inst}$  is the error for a instantaneous unit release. Figure 5.2 shows the normalized errors from the hypothetical release data and empirical error function based on equation 5.4 for a 2-min release. The normalized errors for the 2-min release are observed to be smaller in magnitude at the beginning of the release, and increase to a peak value, before decreasing and reaching constant values. The early phenomenon occurs because only a fraction of the unit mass has been released. This suggests that the initial errors owing to poor mixing would be relatively small, compared to those associated with an instantaneous release.

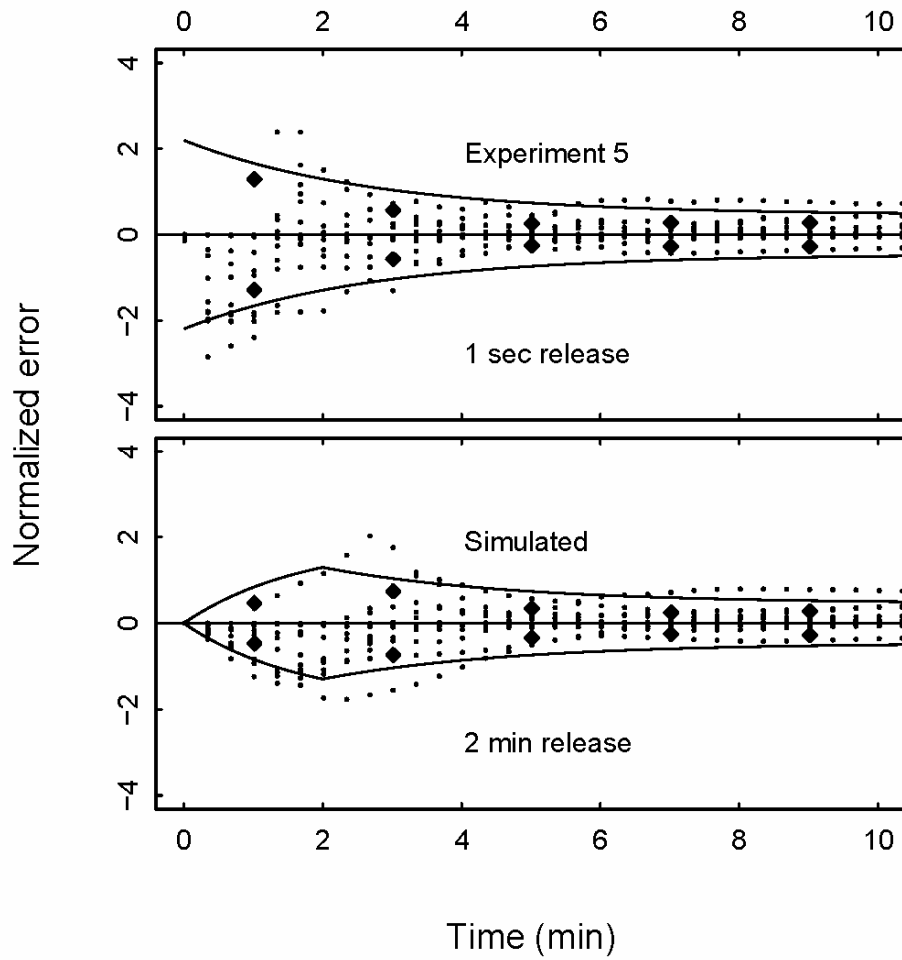


Figure 5.2. Normalized model-to-measurement error as a function of release duration. Black dots show errors based on tracer gas experiment 5; calculated standard deviations for errors grouped into two-minute intervals (black diamonds); modeled standard deviation of normalized errors using exponential fit (solid black line). Note, the agreement here differs from that in Figure 5.1 because this figure shows Experiment 5 results only.

The above procedure resulted in a reasonable fit, and equation 5.2 was extended to equations 5.5(a) and 5.5(b).

$$\sigma(t) = (Ae^{-Ct} + B) \times \frac{t}{\tau_{release}}, t \leq \tau_{release} \quad (5.5a)$$

$$\sigma(t) = Ae^{-Ct} + B, t > \tau_{release} \quad (5.5b)$$

where  $\tau_{release}$  is the release duration in minutes. Equation 5.5 represents only a preliminary attempt to incorporate longer release durations into the likelihood function. Experimental data are necessary to thoroughly investigate the effects of duration on sensor-system performance.

## 5.4.2 Implementation of likelihood function

If the release mass, time of occurrence, and duration are known, then the standard deviation of the absolute error can be estimated using equation 5.2 or 5.5, and the likelihood of observing a particular set of observations can be determined. However, during an event, these values are unknown. Because these variables influence the level of confidence that is assigned to the interpreted data, via the likelihood function,  $p(O|Y_k)$ , they in turn affect the probability that equation 5.1 assigns to each scenario in the library of possible releases. Misquantification of the model-to-measurement errors will lead to unreliable inferences about the release.

To estimate the model-measurement errors, I use the median mass and duration from the most recent round of updates of the Bayesian posterior probabilities to scale the likelihood function for the next update. An update occurs each time the system receives new data from a sensor and a full round is achieved after receiving data from all the sensors in the network. The initial mass is set to a large value, and duration to a small value, to place less emphasis on the initial data. Equations 5.6 and 5.7 are used to estimate the normalized ( $\sigma[t_i]$ ) and absolute errors ( $\sigma_{abs}[t_i]$  in  $\text{g/m}^3$ ), respectively, for an unknown release mass at time  $t_i$ . Note that the square brackets in equations 5.6 and 5.7

denote that these values (e.g.,  $\sigma$ ,  $\tau_{release}$  and  $mass_{rel}$ ) are to be evaluated as functions of time  $t_i$  or  $t_{i-1}$ .

$$\sigma[t_i] = (Ae^{-Ct_i} + B) \times \frac{t_i}{\tau_{release}[t_{i-1}]}, t_i \leq \tau_{release}[t_{i-1}] \quad (5.6a)$$

$$\sigma[t_i] = Ae^{-Ct_i} + B, t_i > \tau_{release}[t_{i-1}] \quad (5.6b)$$

$$\sigma_{abs}[t_i] = \sigma[t_i] \times \frac{mass_{rel}[t_{i-1}]}{Vol} \quad (5.7)$$

where  $i$  denotes the current round of updating,  $i-1$  is the previous round of updating,  $t_i$  is in minutes,  $\tau_{release}[t_{i-1}]$  and  $mass_{rel}[t_{i-1}]$  are the estimated release duration and mass, respectively, as assessed at time  $t_{i-1}$  (i.e., based on the previous round of updates), and  $Vol$  is the building's interior volume in  $m^3$ .

I address the problem of an unknown release time by assuming that the release occurred at discrete instants within a finite bound prior to the first non-zero concentration measurement. The unknown release time is incorporated into the library as time-shifted simulations, which results in a substantially larger library of simulations. When the first non-zero concentration is received, the BMC algorithm is initiated for the extended library. This first round of updates is performed by applying equation 5.5, with  $t = 0$ . The expected release time is estimated using the updated probabilities of each simulation in the extended library and is assumed to be the release time for the next round of updating.

## 5.5 Investigation of time scales by blinding data

To explore how time scales affect the interpretation of sensor data, I applied the Bayesian algorithm to Experiments 1 and 4, retaining data from all 11 sensors. I assessed the importance of early data by blinding the BMC algorithm to measurements taken at the beginning of the release, and then evaluating system performance for different blinding periods.

For each run of the Bayesian algorithm, an input data set was constructed by removing, from the experimental measurements, the first 4 min to the first 160 min of data. In this exploration, the release time was assumed to be known to explore the time-dependent utility of the data, without introducing the complication of an unknown release time.

In the data blinding process,  $0 < t < T'$  data were removed where  $T'$  was varied between 4 min and 160 min. The 4- to 160-minute windows of missing data straddle the expected  $\tau_1 - \tau_3$  time scales for the case study. Consider the within-room mixing time. Using the correlation from Drescher et al. (1995),  $\tau_1$  is estimated to be 12 min and 10 min for Experiments 1 and 4, respectively. It should be noted that this correlation may not be accurate for the release conditions of Experiment 1, since the release occurred adjacent to the return duct intake and the tracer may have entered the ventilation system without significant mixing in the room. The experimental data indicate that the concentration fluctuations observed by the sensors effectively cease by  $\sim 5$  min after the release, suggesting a lower value of  $\tau_1$ . The mixing time for Experiment 4 may have been somewhat longer than this estimate implies, since the supply register for this room was



oriented such that air was partly diverted into an adjacent room. Consequently, the effective volume may be larger than that used to estimate  $\tau_1 \sim 10$  min. Nevertheless, all indicators suggest intrazonal mixing time scales that are on the order of  $\sim 10$  min.

For both experiments, I estimate an interzonal mixing time  $\tau_2 \sim 17$  min, using the ratio of the volume served to the mechanical ventilation flow rates. Finally, I estimate a removal time scale of  $\tau_3 \sim 2$  h, based on the rate of air exchange with outdoors as given by the multizone model.

Figure 5.3 illustrates the ability of the BMC algorithm to accurately identify the release location. The plot shows the probability that the algorithm assigns to the actual release location as of the fourth round of data as a function of the time that the first round of data was received after the release (i.e., as a function of the blinding period). For example, withholding the first four minutes of data from Experiment 1, then presenting the Bayesian estimator with the next four sets of measured room concentrations, the interpretation algorithm identifies the true release location with better than 90% probability. Withholding the first 12 min of data, the probability falls to 70%. More extensive blinding (i.e., beyond 12 min) prevents the algorithm from inferring the release location with confidence.

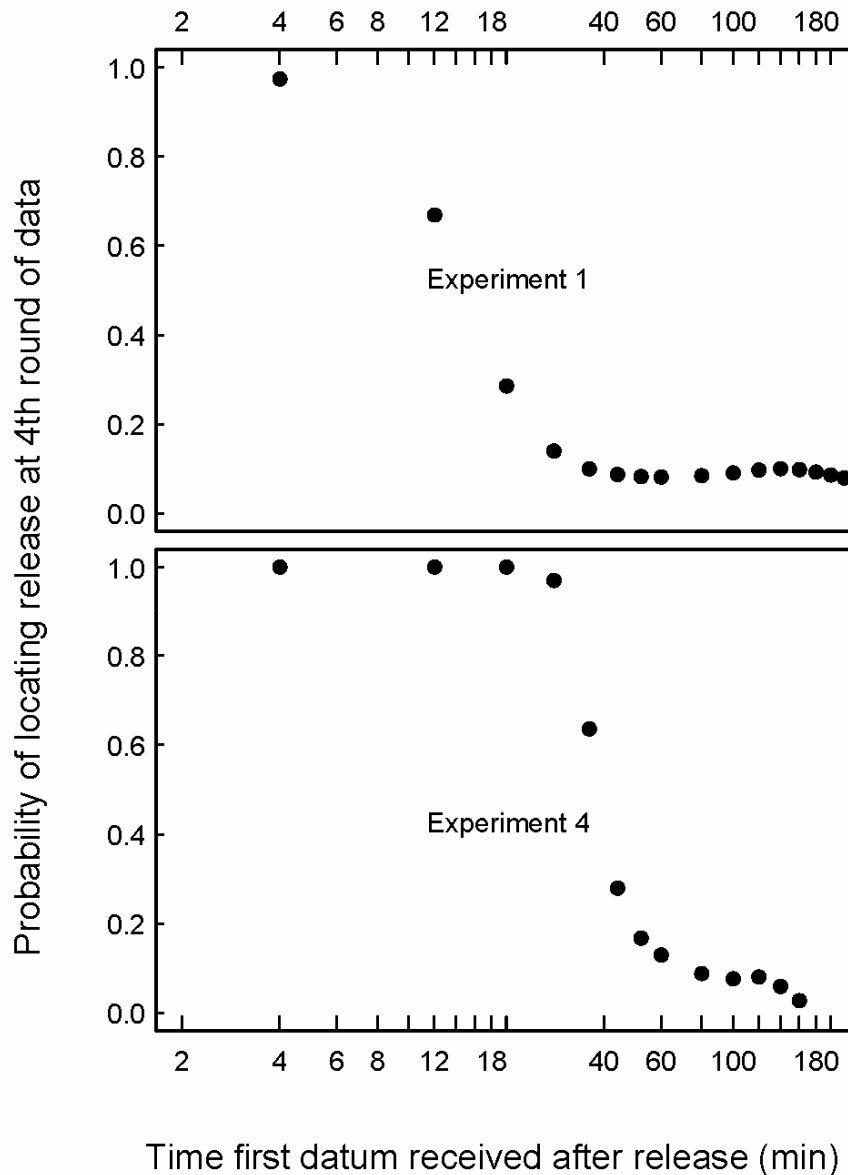


Figure 5.3. Probability of correctly identifying release location with blinded data sets. In each case, the algorithm has updated on four rounds of data and the sensor network is blinded to the data from the time of release (assumed to be known) until some later time, as represented by the x-axis. Results are shown for both Experiments 1 and 4.

Clearly, this example shows that the data available at the earlier part of the release is more rich in the information needed to identify the source location than that occurring later. I refer to this information as the “location fingerprint” and relate its loss to the  $\tau_1$  -

$\tau_3$  time scales. Blinding the interpretive algorithm to four minutes of data — within the lowest estimate of the mixing time for a single zone — hardly affects its ability to identify the release location. Because the tracer in Experiment 1 was released into the ventilation system (see Table 3.1), the strong location fingerprint at 4 min after the release does not result simply from the tracer not yet having left the release zone. Rather, it results from the fact that interzonal mixing processes, identified with  $\tau_2$ , have not yet had a chance to distribute the tracer evenly throughout the building. However, at 12 min after the release — when interzonal mixing has substantially occurred — the location fingerprint begins to fade.

Experiment 4 shows a similar pattern; the algorithm is able to accurately identify the release location up to 30 min after the release. The location fingerprint of Experiment 4 decays more slowly than that of Experiment 1 because the release occurred in a room, rather than at the intake of the ventilation system. This finding is consistent with interpreting the interzonal mixing time,  $\tau_2$ , as characterizing movement of a contaminant from zone to zone. As with Experiment 1, by the time that  $\tau_3$  has elapsed, mixing between the building and its environment has effectively erased the location fingerprint from the data of Experiment 4.

The Bayesian algorithm also estimates the release mass. Figure 5.4 depicts the estimated release mass, along with confidence bounds, for partially blinded data sets. In these figures, the “true value” is the amount intended to be released during the experiments, which may differ from the actual amount released owing to experimental error.

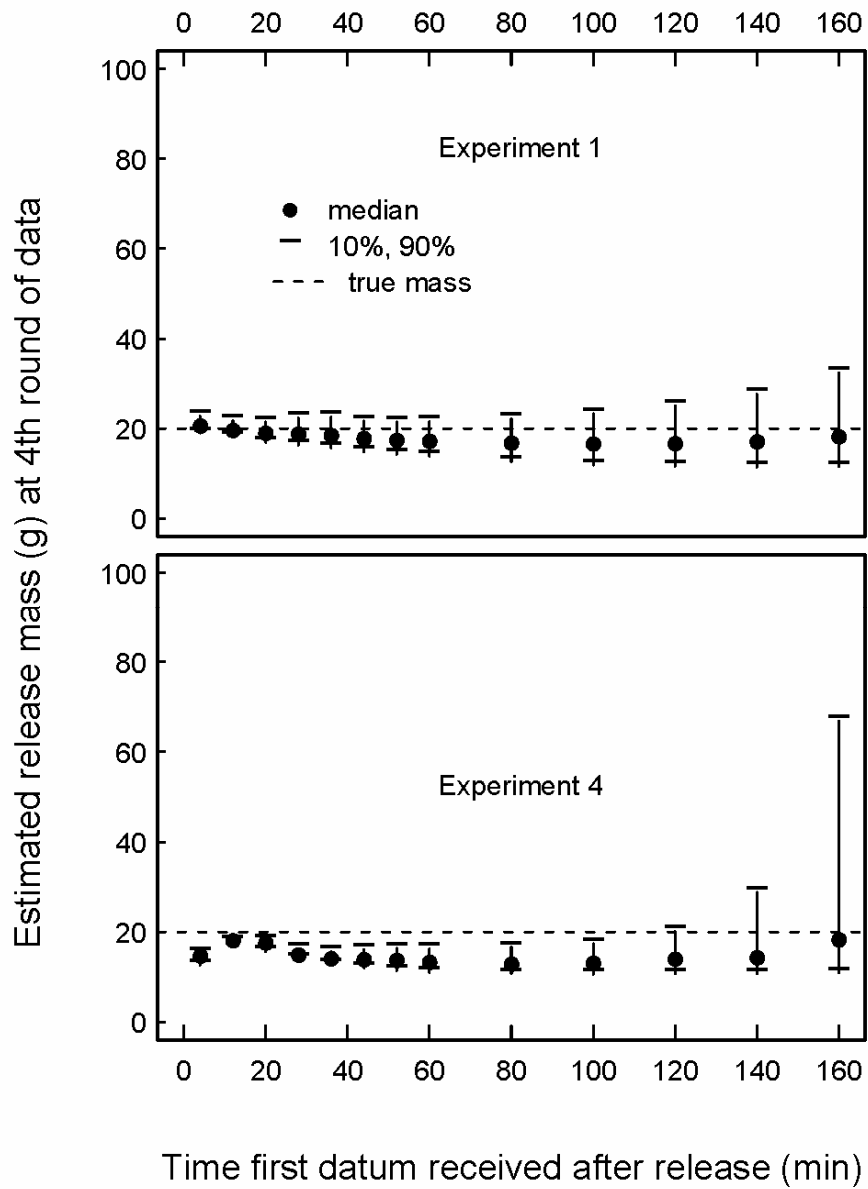


Figure 5.4. Estimated mass released with blinded data sets. In each case, the algorithm has updated on four rounds of data and the sensor network is blinded to the data from the time of release (assumed to be known) until some later time, as represented by the x-axis. Results are shown for both Experiments 1 and 4.

In the case of Experiment 1, the mass is correctly estimated to a narrow confidence interval even for data sets where the first 100 min of data have been removed. Unlike the

location fingerprint, the mass fingerprint does not significantly decay over time scales on the order of  $\tau_2$ . This observation reflects the fact that interzonal mixing simply redistributes contaminant mass throughout the building. Because the sensors continue to send this information to the Bayesian algorithm, the mass fingerprint does not significantly fade until time scales at least on the order of  $\tau_3$  have elapsed, i.e. when tracer removal from the building becomes important.

For Experiment 4, the median value of the estimated release mass is 30% less than the true value. The algorithm underestimates the mass, with a narrow confidence interval, for simulations in which as much as the first 120 min of the data have been removed. In this case, the mass is underestimated mainly because equation 5.5 underestimates the model-to-measurement error for one room in the building. Additional simulations (not presented here) show that the mass is more accurately determined for networks in which the sensor in this one room is absent. An optimized likelihood function would need to account for this type of error.

Overall, the time scales help to explain the success or failure of the Bayesian algorithm at interpreting the experimental data. In other words, the time scales help to identify those cases where the failure of the BMC method is attributable to the loss of information in the incoming data, and hence cannot be entirely overcome by increased algorithmic sophistication. In summary, a sensor system may require data from early sample times to identify a release location, but can estimate the release mass even with poor initial data. This suggests that sensors may be selected and optimized for either purpose and combined in the final system. For example, a system could be designed with a network of fast threshold-level sensors to quickly identify a release and its location, and

with a supplemental network of slower, finer-grained sensors that can be brought online to determine release mass at a later time during the event.

## 5.6 Evaluating different sensor systems

In this section, I describe the performance of networks that vary in the number of sensors, their placement, and in sensor response time. The relevance of the time scales in the context of these system design parameters is further explored. For example, can a network without a sensor in each zone correctly and confidently identify the release location? Can a network with few sensors identify the release mass to a high confidence, even if it fails to identify the location?

I investigated the performance of all possible sensor networks composed of from 1 to 11 sensors, for Experiments 1 and 4. Each network has a maximum of one sensor per room (8) plus one in each of the three levels of the staircase (3), for a maximum total of 11 sensors. The release time is treated as unknown, but assumed to have occurred either 1 s, 61 s, 121 s, or 181 s prior to the first indication from any sensor of a nonzero concentration.

Figures 5.5 and 5.6 show the probability of identifying the correct release location at three different times — 1 min, 5 min, and 10 min after the release. In Experiment 1, more than half the networks containing four or more sensors can correctly locate the source with 90% probability, using only 10 min of data. If the goal is to localize the release to either the return duct intake (its true location), or to the room that contains the intake of the return duct, more than 90% of the networks with three or more sensors correctly identify the release location to a probability of at least 90%.

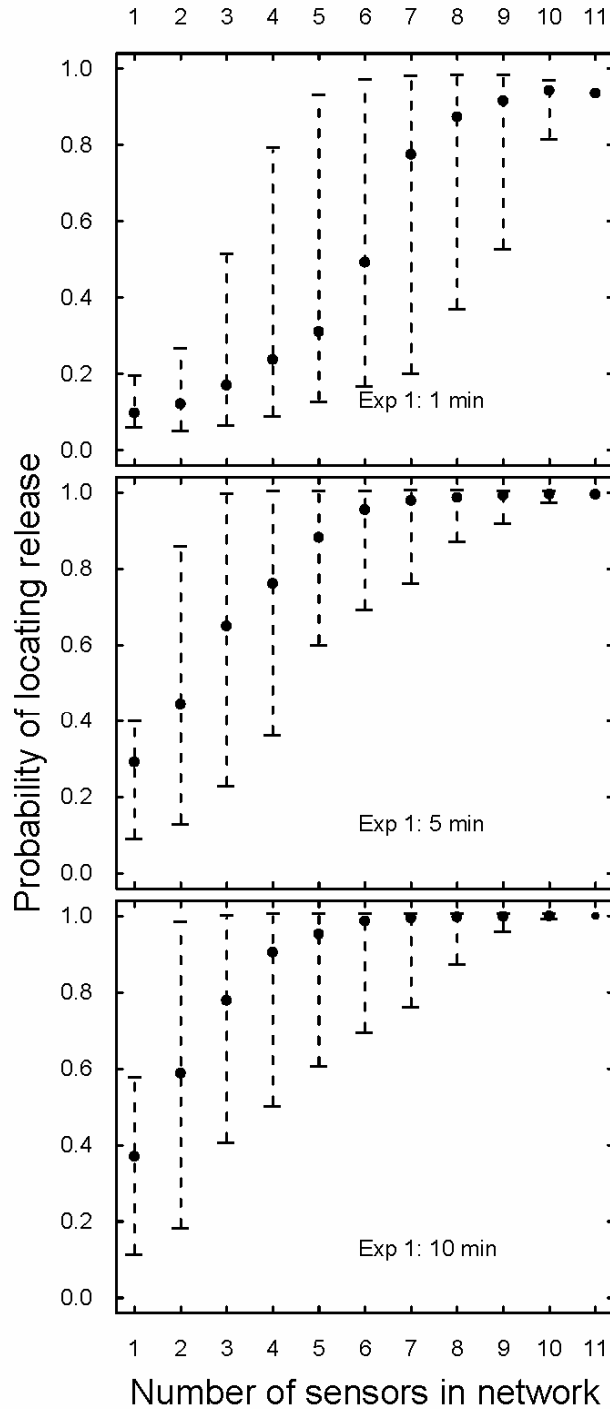


Figure 5.5. Probability of correctly identifying release location (Experiment 1). Results are shown at 1 min, 5 min, and 10 min after the release occurred; the 10th and 90th percentiles (dashed bars) and median (bullets) indicate the spread and central tendency of the estimated probabilities among the different network configurations for a particular number of sensors.

In this example, when the network contains many sensors, specific sensor placements matter little, suggesting that the sensors record overlapping information. For systems with few sensors, specific sensor placement becomes more important. This is indicated by the wider spread of probabilities shown by the 80% confidence intervals as the number of sensors in the network decreases.

Note the differences in the patterns between 1 min and 5 min, as compared to the more similar patterns for 5 min and 10 min. The larger differences in the former pair indicates that the system learns more about the release location during early periods ( $1 \text{ min} < t < 5 \text{ min}$ ) than later ( $5 \text{ min} < t < 10 \text{ min}$ ). This result suggests — as expected — that the location fingerprint loses definition soon after the release. The fingerprint can be seen better at earlier times even though the model-to-measurement errors are larger then. Relevant information can still be extracted from the data even when the model-to-measurement errors are relatively high, if appropriate confidence is assigned to these data.



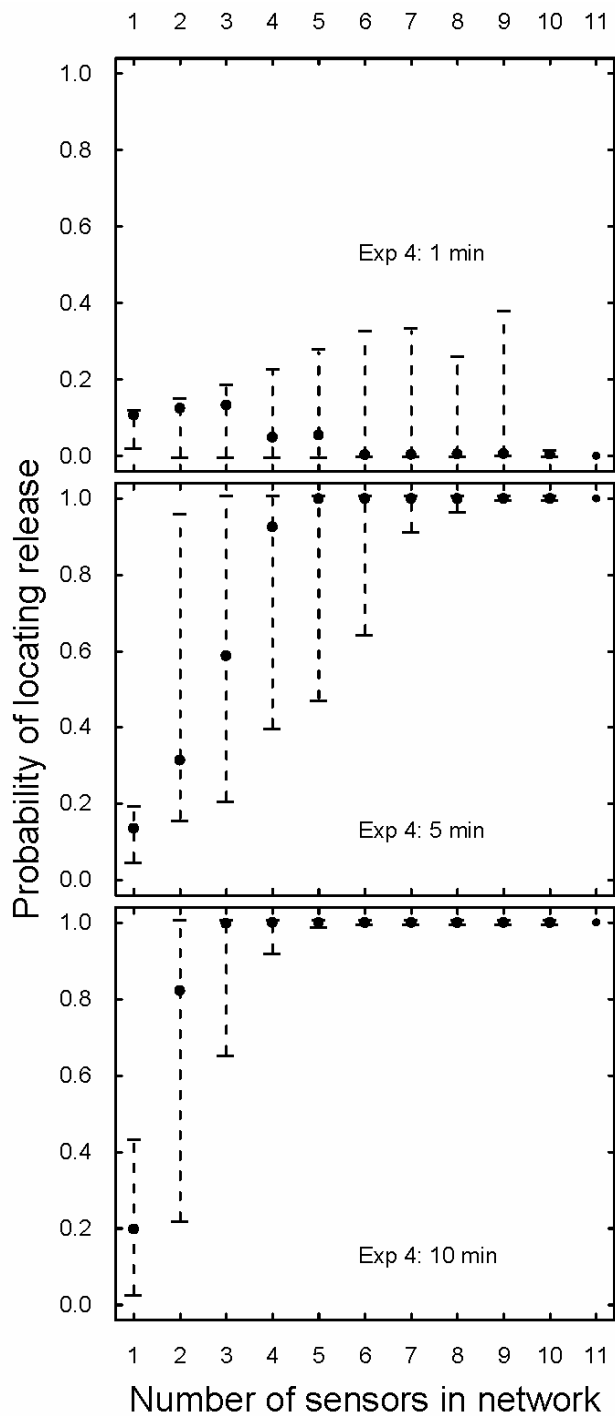


Figure 5.6. Probability of correctly identifying release location (Experiment 4). Results are shown at 1 min, 5 min, and 10 min after the release occurred; the 10th and 90th percentiles (dashed bars) and median (bullets) indicate the spread and central tendency of the estimated probabilities among the different network configurations for a particular number of sensors.

Turning to Experiment 4, Figure 5.6 shows that by  $t = 5$  min, networks with four or more sensors have a high probability of correctly identifying the release location. By  $t = 10$  min, the performance has little dependence on sensor number and placement, for networks ranging from 4 to 11 sensors. Comparing these results to those of Experiment 1, we see that a larger proportion of networks can correctly locate the release to a high confidence in Experiment 4. This finding is consistent with earlier results suggesting that it is easier to locate a release that occurs in a room than one that occurs in a return duct, again owing to the difference in how quickly the ventilation system spreads the contaminant (and hence washes out the location fingerprint). The good performance of networks with only 4-7 sensors suggest that the source may be correctly located with high confidence without a sensor in the true release location, at least in airflow conditions similar to this case study. More generally, it may be possible to use fewer sensors for locating releases whose fingerprints decay slowly.

A smaller number of sensors may be able to estimate the mass to a high confidence, as compared to the number that is required to identify the release location. I measure the confidence of the mass estimate at time  $t$  by calculating the uncertainty reduction of the 80% two-sided confidence interval, relative to the corresponding interval based on the prior mass estimate, as follows:

$$UR(t) = \left( 1 - \frac{q_{Mass,90\%}(t) - q_{Mass,10\%}(t)}{q_{Mass,90\%,prior} - q_{Mass,10\%,prior}} \right) \times 100\% \quad (5.8)$$

where  $q$  represents the quantile of the mass indicated by the subscript. That is,  $UR(t)$  represents the relative reduction in uncertainty of the mass based on the processing of data up to time  $t$ .

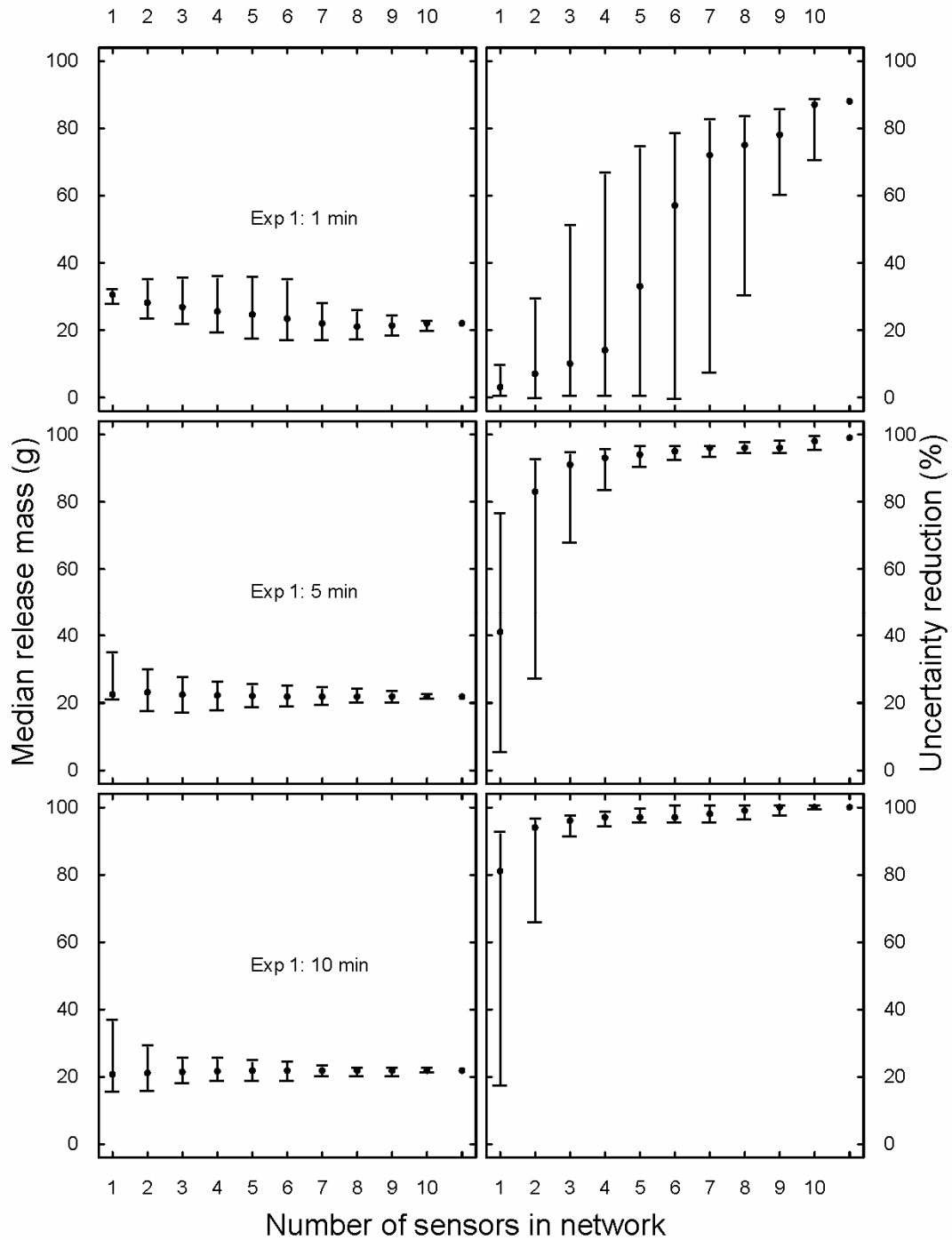


Figure 5.7. Estimates of the median release mass and uncertainty reduction of the 80% confidence interval (Experiment 1). Results are shown at 1 min, 5 min, and 10 min after the release occurred; the 10<sup>th</sup> and 90<sup>th</sup> percentiles (error bars) and median (bullets) indicate the spread and central tendency of the median mass and uncertainty reduction among the different network configurations for a particular number of sensors.

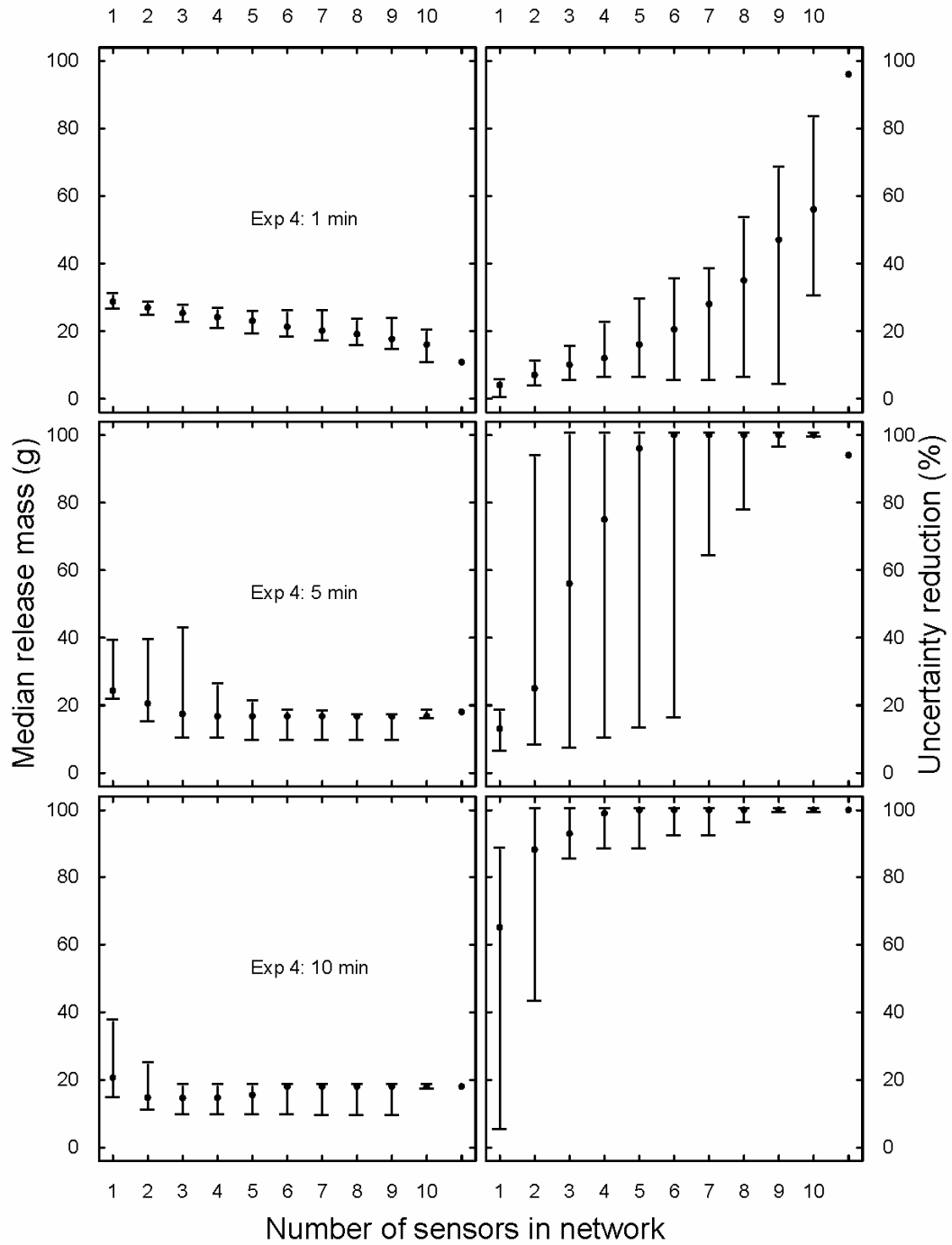


Figure 5.8. Estimates of the median release mass and uncertainty reduction of the 80% confidence interval (Experiment 4). Results are shown at 1 min, 5 min, and 10 min after the release occurred; the 10th and 90th percentiles (error bars) and median (bullets) indicate the spread and central tendency of the median mass and uncertainty reduction among the different network configurations for a particular number of sensors.

Figures 5.7 and 5.8 show the estimates of the median release mass and the uncertainty reduction of the 80% two-sided confidence interval for the same networks considered in Figures 5.5 and 5.6.

For Experiment 1, the released mass is estimated accurately to a high confidence for 90% of all networks with 3 or more sensors for  $t \geq 5$  min. The sensor system learns more about the released mass during the 1 min to 5 min monitoring interval than during the 5 min to 10 min period. Comparing Figures 5.5 and 5.6 with 5.7 and 5.8, fewer sensors are required to correctly determine the mass to a high confidence by 10 min, than are required to estimate the release location to a high confidence. At  $t = 10$  min, the release mass is estimated with high confidence for networks with three or more sensors for both experiments. In this case, few sensors are needed to limit the uncertainty of the estimated mass. This appears to be true for times much less than the pollutant removal time ( $t \ll \tau_3$ ), which suggests that perhaps even fewer sensors may be able to estimate the mass to a high confidence, if one can believe that there is mass-relevant information until  $t \sim \tau_3$ .

The relationship between time scales and performance of sensor systems with varied sensor response time was also explored. Figures 5.9 - 5.12 compare the performance of sensor systems with different response times for all networks containing an odd number of sensors. I generated time-averaged measurements for sensor response times of 60 s and 120 s, using the original data, which were collected at 20 s intervals.

One would expect that the ability of a sensor system to identify the release location is less sensitive to longer sensor response times for cases in which the location fingerprint decays slowly. Figures 5.9 and 5.10 show the probability of correctly identifying the

release location at  $t = 10$  min for Experiments 1 and 4. For Experiment 1, there is a significant deterioration in the ability of the system to locate the release to a high confidence using slower sensors. In this case, networks with fewer fast sensors outperform those with more slow sensors.

By contrast, for Experiment 4, the system performance is relatively insensitive to sensor response time. The contaminant requires more time to mix with other zones in Experiment 4 and, therefore, leaves a more persistent location fingerprint. In conditions characterized by slow dispersion, a sensor system may learn more by sampling densely in a spatial sense, rather than sampling at a high temporal frequency.

Based on the time scales assessment presented in this chapter, one should expect that the release mass estimation is relatively insensitive to sensor response time, but that longer response times will yield wider confidence intervals. Figures 5.11 and 5.12 depicts the median release mass, and corresponding uncertainty reduction of the 80% two-sided confidence interval, as defined by equation 5.8.

For Experiment 1, the median determination of released mass is observed to be relatively insensitive to sensor response time, but the final confidence intervals are somewhat larger for the longer response times, as exhibited by the lower values of uncertainty reduction. An interesting consequence is that networks with fewer, but faster, sensors are able to accurately estimate the released mass to a slightly higher confidence than networks with a greater number of slower sensors. In Experiment 4, it is seen that the median mass varies with response time, exhibiting greater sensitivity to sensor placement for slower sensors. This dependence could be due to the longer mixing times for Experiment 4. However, because the ventilation time scale exceeds 10 min, I

anticipate that as time progresses, the results produced by slower sensors will approach the results for networks with faster sensors.

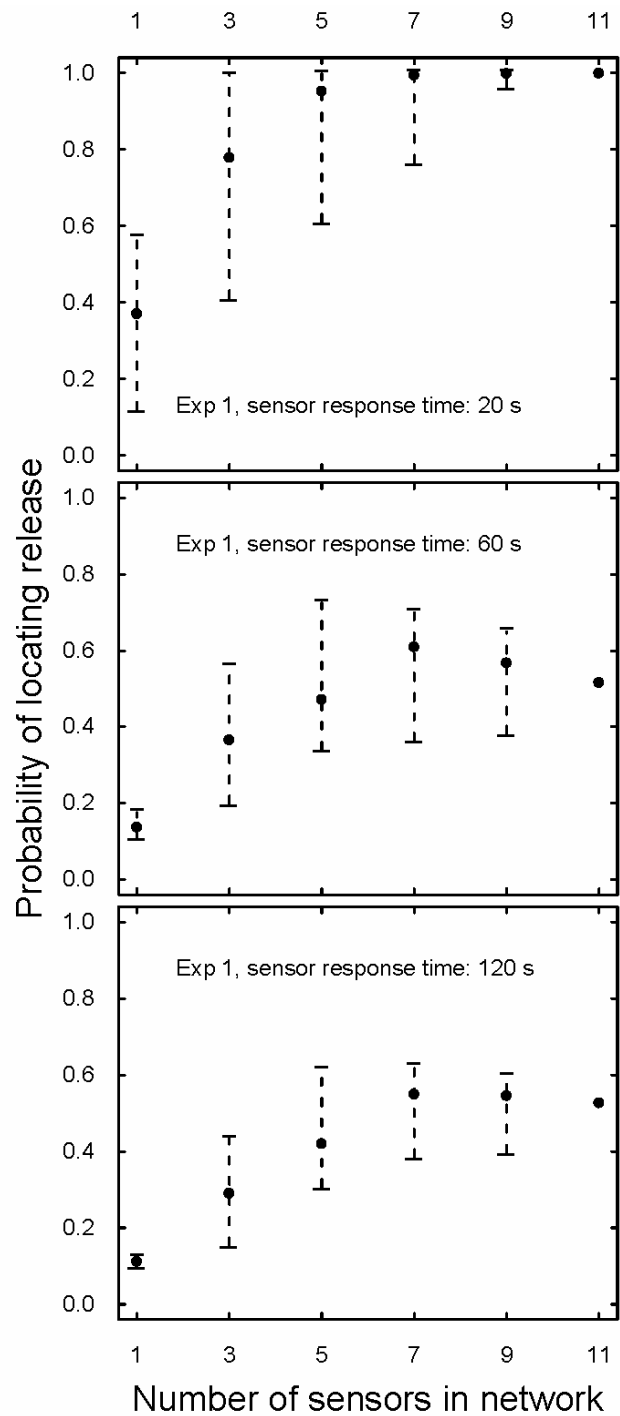


Figure 5.9. Probability of correctly identifying release location for 3 different sensor response times (Experiment 1). Shown at 10 min after release event for sensor response times of 20 s, 60 s, and 120 s; 10th and 90th percentiles (error bars) and median (bullets).



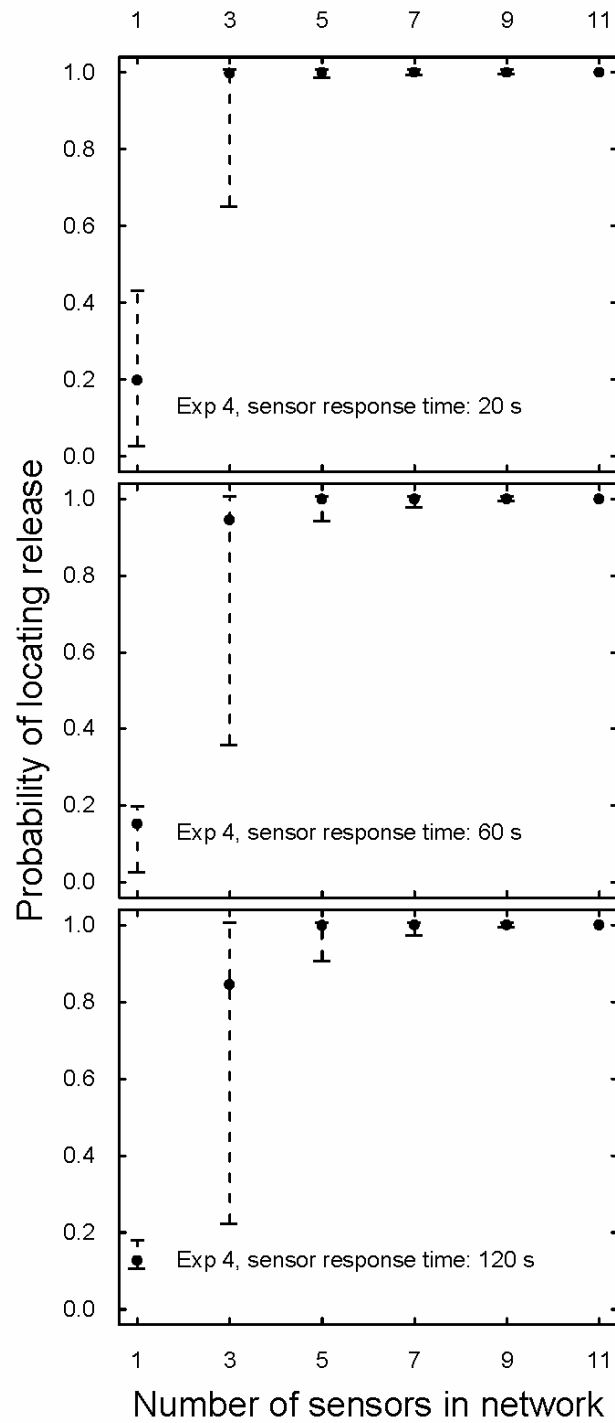


Figure 5.10. Probability of correctly identifying release location for 3 different sensor response times (Experiment 4). Shown at 10 min after release event for sensor response times of 20 s, 60 s, and 120 s; 10th and 90th percentiles (error bars) and median (bullets).

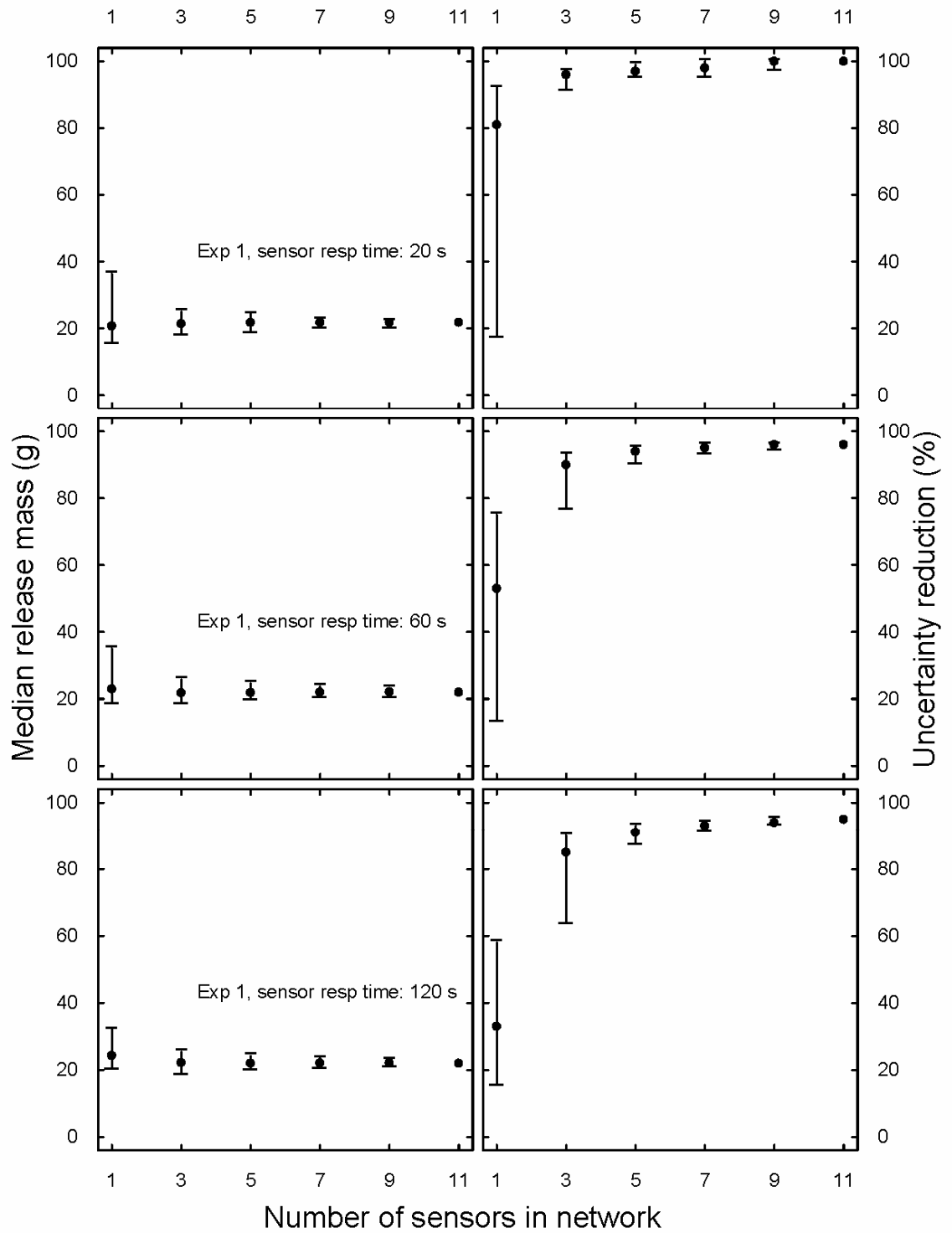


Figure 5.11. Estimates of median release mass and uncertainty reduction of the 80% confidence interval for 3 sensor response times (Experiment 1). Shown at 10 min after the release event for sensor response times of 20 s, 60 s, and 120 s; 10th and 90th percentiles (error bars) and median (bullets) of the median release mass.

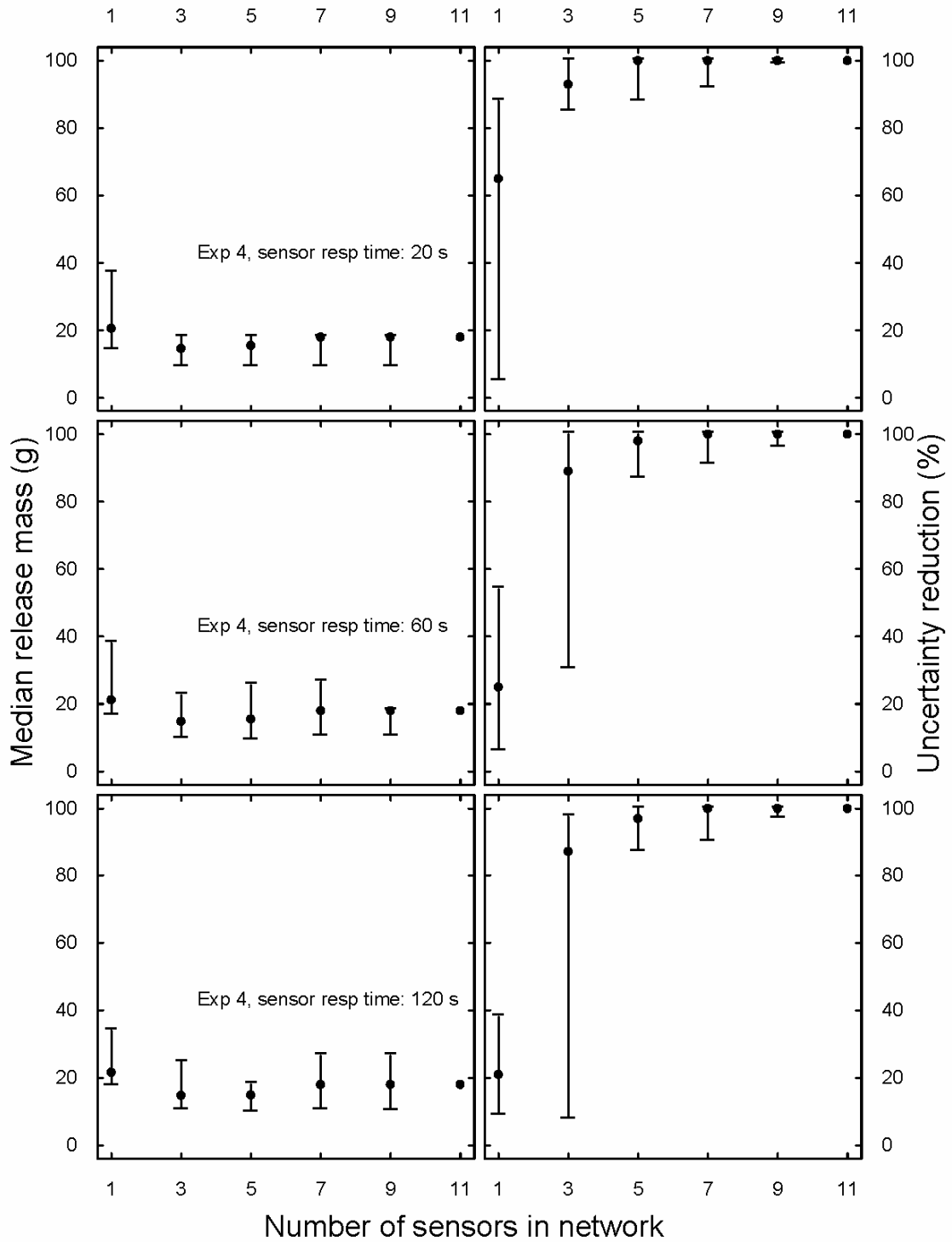


Figure 5.12. Estimates of median release mass and uncertainty reduction of the 80% confidence interval for 3 sensor response times (Experiment 4). Shown at 10 min after the release event for sensor response times of 20 s, 60 s, and 120 s; 10th and 90th percentiles (error bars) and median (bullets) of the median release mass.

## 5.7 Statistical considerations

The focus of this chapter is to understand the role of the contaminant transport time scales in the performance of a Bayesian-based sensor system. To pursue the investigation, a likelihood function that reconciled the model specification errors was required. While the likelihood function presented in Section 5.4 offers a promising step toward achieving this goal, the implementation in this chapter also reveals opportunities for further improvement. In particular, the results for Experiment 4 suggest that some aspects of model performance were not adequately captured by the likelihood function. This observation raises some interesting statistical issues, which may in turn provide insight towards a more robust sensor-system design.

Analysis of the results for Experiment 4 showed that the errors for two particular rooms — 1.2b and 1.3 — were not adequately treated, which resulted in the following problems. In some cases, an incorrect release location was identified with high probability, before identifying the correct release location. Figure 5.13 shows an example of this behavior for the fully instrumented network. In other cases, the release mass is estimated to a value below the true release mass (20 g), to a narrow confidence interval, as shown in Figure 5.14.

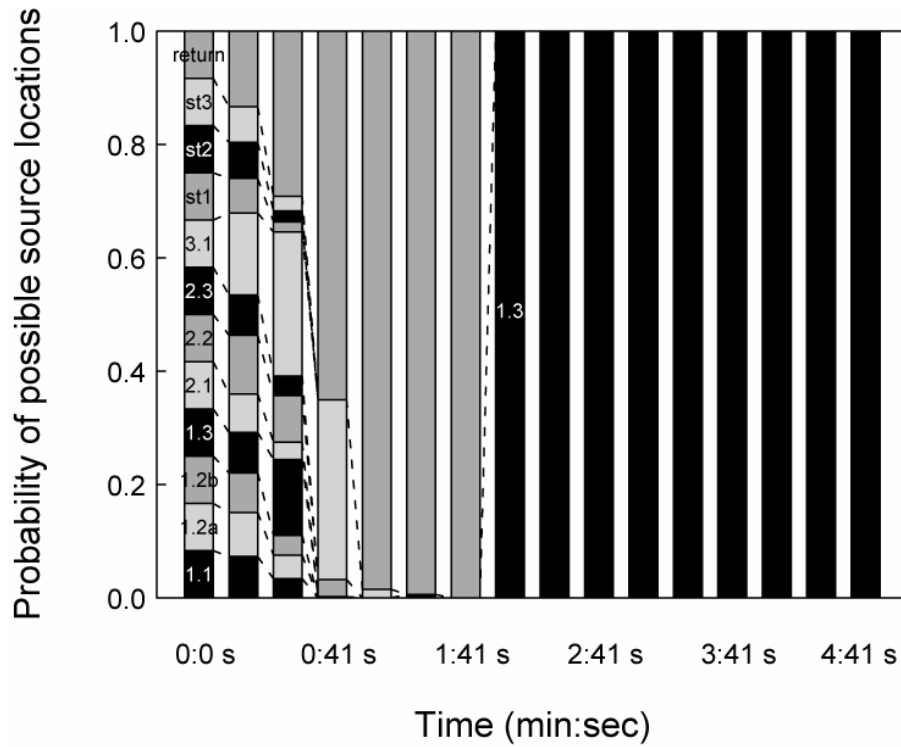


Figure 5.13. Probability of source being in location indicated for Experiment 4 data. The actual release location is Room 1.3, which is correctly identified as the most probable release location only after 2 minutes. Time is referenced to the instantaneous release event.

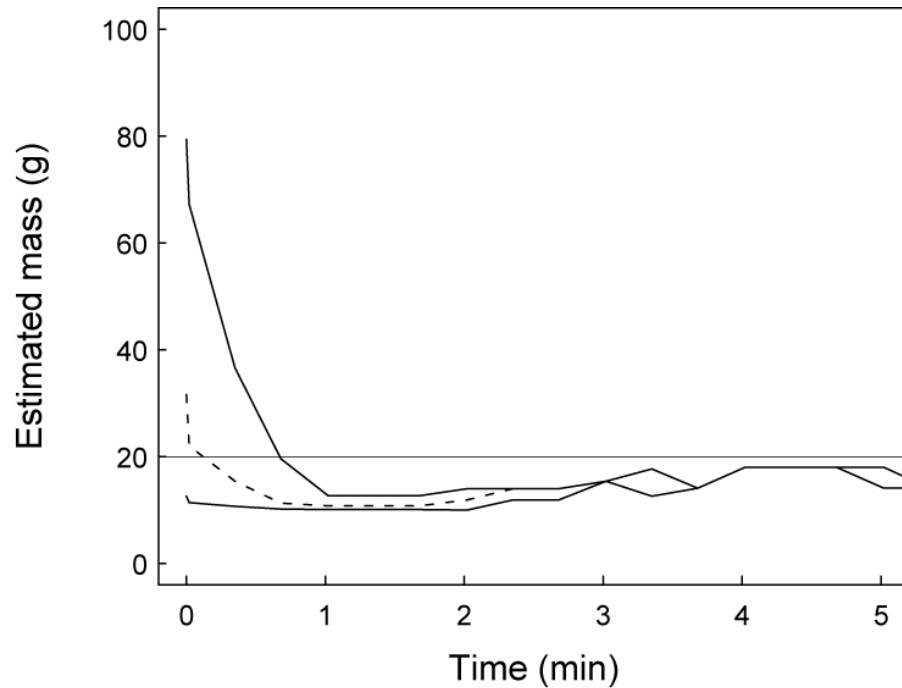


Figure 5.14. Release mass estimated for Experiment 4 data. The solid lines indicate 80 percent confidence intervals; dashed lines indicate medians. (Actual release mass was ~ 20 g.)

The confidence interval is so narrow, that only a small number of library realizations are likely to exhibit any significant influence on the parameter distributions. Behavior such as these do not necessarily indicate a “failure” of the sensor system, but have implications for response planning, and in addition, may offer insight into further improvement of the statistical algorithm. To better understand the idiosyncrasies that arose in this experiment, I next consider the effect of each key room individually.

The base multizone model significantly and consistently overestimates the concentrations in Room 1.2b. This overestimation is likely due to an inaccurate estimate of the true airflow conditions. In a real implementation, there are many scenarios that could result in this kind of phenomenon. For example, a damper in a duct system could

fail, or become stuck in a particular position. A window that is normally closed could be open, which might not be considered as a possibility in the library-generation stage. Or, a sensor may fail and exhibit significant bias. A statistical approach that is able to identify and properly respond to fault modes such as these, which are not adequately described in the library, is desirable. Solutions may include simple adjustments to the likelihood function, such as flattening the tails of the Gaussian distribution, which would limit the lowest possible likelihood function value. A more sophisticated approach may identify that there is persistent behavior that should be treated as a “fault,” and process the measurements accordingly.

In the case of Room 1.3, which is the actual release room for Experiments 1 and 4, significant errors result owing to the finite time required to mix the contaminant within the room. While the time-varying error function is a reasonable form for representing this error, the parameters obtained from the collective experiments do not fully capture the error describing the concentrations in Room 1.3. Of the seven “development” experiments, which were used to develop the likelihood function, only one experiment had a release occurring in Room 1.3. The remaining releases occurred at the intake to the return duct. Furthermore, the data from Experiments 4 and 9 suggest that, during these releases, the concentrations in the remaining nonrelease rooms exhibit few fluctuations, indicating that the contaminant is consistently well mixed in these zones. Differentiation between the degree of mixing within individual rooms is not considered in the proposed likelihood function. Therefore, a likelihood function that captures the effect of the location of the release may be required to describe the behavior of all possible releases.

In lieu of an error function that adequately links the errors with physical causes, such as the release location, a “black-box” approach may be desirable. Ideally, a well-designed likelihood function characterizes the error structure between the modeled and measured concentrations, such that the maximum utility of the data is realized. A proper likelihood function neither underestimates nor overestimates the value of the data. The negative results from either situation becomes evident from the statistics generated by the algorithm. A black-box approach recognizes that it may not be possible to develop an optimal or adequate likelihood function by linking the errors to the physical causes.

A black-box likelihood function may begin with some initial error structure and form, such as the one proposed in Section 5.4, and then adjust the level of confidence that is assigned to the data, based on some performance metric that indicates whether the assumed level of confidence is appropriate. For example, the range and disparity of posterior probabilities among the library realizations are indicative of the algorithm performance and are determined by the level of confidence that is assigned to the data. As less (more) confidence is assigned to the data, the range and disparity of the posterior probabilities decreases (increases). The statistical properties of the posterior probabilities may be used to determine an appropriate level of confidence to assign to the data in real time. For example, one method may be to limit the derivatives of the median and variance of the posterior probabilities in the library, with respect to time. This kind of approach moves towards an iterative method, and careful implementation would be required to preserve the real-time operational capability of the algorithm. It is worth noting that using the statistical properties of the posterior probabilities in real time to



improve the algorithm is analogous to using the variance of the “weights” in Sequential Monte Carlo methods (Section 2.4.6.2) to improve the algorithm performance.

## 5.8 Conclusions

This chapter shows how an analysis of the relevant time scales for contaminant transport may help to explain the performance of sensor systems that use Bayesian interpretation to estimate release conditions, based on concentration measurements. I show that the sensor system design should consider the physics governing contaminant transport, as well as the limitations that arise from the transport model.

I discussed three relevant time scales, the intrazonal mixing time ( $\tau_1$ ), interzonal mixing time ( $\tau_2$ ) and removal time ( $\tau_3$ ), and demonstrated how estimating different release parameters requires different kinds of information. Information relevant for finding the release location degrades beyond  $\tau_2$ , which can range from  $\sim 10$  to 30 min, for forced convection systems. The blinding analysis showed that removing data within  $\tau_2$  did not adversely affect the ability of the algorithm to identify the release location. For Experiment 1, which was characterized by a shorter  $\tau_2$ , the location “fingerprint” began to decay after removing 12 min of data. For Experiment 4, which had a longer  $\tau_2$ , the location fingerprint was detectable with as much as 22 min of the early data removed. Information useful for estimating the released mass remains in the system up to  $\tau_3$ . The blinding experiments showed that the release mass could still be estimated narrowly after removing 140 to 180 min of the early data, which is consistent with  $\tau_3$ .

Networks with varying number of sensors, placement, and sensor response time were evaluated. For networks based on Experiment 1’s data, a majority of networks with 4 or

more sensors could identify the release location to a 90% probability within 10 min. Networks based on Experiment 4 exhibited less sensitivity to sensor placement and a majority of networks with 3 or more sensors could identify the release location to a 90% probability. The effect of sensor response time was also evaluated. Networks with longer sensor response times based on Experiment 1 suffered a significant performance loss in the ability to identify the release location. In contrast, a majority of networks based on Experiment 4's data with sensor response times of 2 min could still identify the release location to a probability of 90%. The location results are consistent with the notion that the location fingerprint degrades more slowly for systems with longer  $\tau_2$ .

In contrast to the location results, the released mass predictions exhibited little sensitivity to the sensor placement for a majority of the networks with 4 or more sensors. The mass was estimated with an uncertainty reduction of 90%, for a majority of networks, with as few as 3 sensors. The influence of sensor response time on estimating the release mass was comparable to the effect on finding the release origin. Networks based on Experiment 1's data suffered a greater performance loss compared to networks based on Experiment 4's data because it is characterized by longer  $\tau_1$  and  $\tau_2$ .

Knowledge of the important time scales may provide insight into how to select sensor characteristics, where to place them, and how many sensors are needed to achieve a particular performance objective. Time scales may also be used to understand the results from the interpretation algorithm. While the particular values of the time scales explored were specific to this case study, I expect that the concepts are more broadly relevant. These concepts should be explored within the context of more complex

buildings, consisting of a greater number of zones, multiple air handling systems, and diverse operating modes.

To explore the relationship between time scales and system performance, I developed a likelihood function that quantifies model uncertainty, which is often a challenge for modeling contaminant dispersion. The development of this likelihood function was limited by the available experimental data and by the specific release conditions of the tracer-gas experiments. Further investigation of the effects of release duration and of the release location may help build upon this initial effort.

Although I performed all of these investigations using the BMC algorithm, the basic questions posed here are relevant to any statistical approach. Furthermore, because the time-scale arguments help one connect an algorithm's performance directly to the governing transport physics, the conclusions reached here may apply to other, non-Bayesian, sensor interpretation algorithms.

Future research efforts may usefully include the investigation of alternate models, such as multizone models with greater discretization, and the incorporation of heterogeneous sensors. It is important to understand the effect of model type more deeply; additional sensors may not be useful if the model is unable to use that additional information. It is also possible that additional sensors may compensate for model limitations. For example, multiple sensors within a room are likely to better represent the ideal well-mixed concentration compared to a single, centrally placed sensor, and thus reduce model specification errors.

In the limit where additional chemical sensors cannot describe the release, or where additional sensors are not available, non-chemical sensors (such as door-position sensors)

may further improve the sensor-system performance or reduce the overall system costs. Such systems may be of particular value to buildings that are more complex than the one considered in this case study. This chapter and Chapter 4 highlight the potential to select optimal sensor-system designs when the algorithm performance is interpreted through the contaminant transport physics. The purpose of Chapter 6 is to continue to investigate this theme in the context of networks comprised of heterogeneous sensors.

# Chapter 6

## 6 Fusing information from heterogeneous sensors

### 6.1 Introduction

Selecting and placing sensors into a sensor system for detecting toxic airborne contaminants is a challenging task for many reasons. The sensor-system performance should be robust under different release conditions, building operating conditions, and other uncertainties, which may not be anticipated during the design phase. As the complexity of the contaminant transport increases, for example, owing to larger building scale or more variable building operation, it becomes increasingly difficult to account for the uncertainties in the sensor-system design process.

An integrated design approach that collectively considers the contaminant transport physics, sensor characteristics, and sensor interpretation algorithms may help address this challenge. Chapters 4 and 5 explored the relationships among these factors and demonstrated how the task of selecting and placing sensors into a sensor system can benefit from an understanding of these relationships. Chapter 4 explored the influence of individual sensor characteristics, and Chapter 5 investigated how to use the time value of the data.

A contaminant detection system is typically composed of sensors that measure either chemical or biological contaminants. If the only goal of the sensor system is to detect the

presence of a contaminant, then incorporating noncontaminant sensors into the system may be of little value. The goal of sensor systems addressed in this dissertation, however, is to characterize an emission source (and to predict future concentrations), rather than simply detecting the source's existence. In this case, heterogeneous sensors (i.e., nonchemical or nonbiological species sensors) may help to either reduce the overall system cost by effectively displacing contaminant sensors or improve the performance for a given contaminant sensor network. Chapter 5 illustrated that the system performance is strongly dependent on sensor placement. Can heterogeneous sensors, for example, offset the penalty that results from poor sensor placement?

In addition to enabling less expensive and more informative networks, a mix of sensor types may help manage complexity by offsetting the increased computational burden needed to process information from more complex buildings.

Also referred to as disparate or multisensor fusion, heterogeneous sensor fusion is a mature concept. Over time, humans and animals have evolved their abilities to combine different sensory information. Dogs, for example, use scent along with sight and sound for basic survival. Heterogeneous sensors have been used in other information fusion applications. In the last decade, multisensor fusion has received significant attention in both military and civilian applications. Military applications include target recognition, autonomous vehicle guidance, and battlefield surveillance; civilian applications include process monitoring, robotics control, smart buildings and medical diagnostics (Hall and Llinas, 1997).

In contrast, the field of microscale sensors and wireless networks is less mature and has yet to explore heterogeneous sensor fusion. The development and application of

these relatively young sensor technologies to environmental and energy applications is relatively new. Multisensor fusion may have as-yet-unrealized benefits for environmentally related sensor-based technologies. Recently, Sanderson et al. (2006) implemented an oceanic network of sensors measuring temperature, conductivity and depth, and used fusion algorithms to characterize the temporal and spatial variations. Their research demonstrates the ability to deploy a real-time heterogeneous sensor network for an environmental application.

In the risk analysis literature, Brand et al. (1995) illustrate how the Bayes Monte Carlo (BMC) algorithm can be used to reduce uncertainties of linked environmental risk models. Drawing from this research, Sohn et al. (2000) used two types of sensors, hydraulic head and chemical sensors, to reduce the uncertainties of coupled groundwater flow and contaminant transport models. The flexibility of the BMC framework will be exploited in this chapter for exploring multisensor fusion.

This chapter explores heterogeneous sensor fusion in the following ways. Two multisensor fusion frameworks, physics- and algorithm-based, are described. This discussion builds on Chapter 5, where the relationships among the contaminant transport physics, model and algorithm were explored. I use the frameworks to discuss how specific sensors may be incorporated into the sensor system. Then, I illustrate the heterogeneous sensor fusion concepts by evaluating networks comprised of door position sensors and contaminant sensors based on the case study data that has been used throughout this dissertation.

## 6.2 Background

The background section begins with a brief review of the Bayes Monte Carlo algorithm. Next, pertinent aspects of the design and control of HVAC systems are discussed to provide insight into how existing building information systems may facilitate heterogeneous sensor fusion. Finally, the case study is reviewed.

### 6.2.1 Algorithm review

A two-stage Bayes Monte Carlo (BMC) algorithm is the computational engine of the sensor system. In the first stage, prior knowledge of possible release events is applied to generate a library of hypothetical release scenarios and concentration uncertainty distributions. The second stage updates the prior beliefs of the release conditions and concentrations by applying a sequential Bayesian updating technique to the contaminant measurements as they become available (Section 2.4.4).

I emphasize the methods by which information is incorporated into the algorithm for the later purposes of this chapter. Information regarding the system can be incorporated in either stage of the algorithm. The hypothetical release scenarios are generated by sampling from prior distributions of variable model parameters. Thus, knowledge of any model parameter, either fixed or variable, can be used towards the generation of the library of release scenarios. In Chapters 4 and 5, for example, leakage characteristics, weather conditions and temperatures, all based on sensor measurements, were used to generate the simulation library.



Bayesian updating provides a framework to incorporate dynamic sensor measurements. In Chapters 4 and 5, contaminant measurements were compared to the predictions from a contaminant fate-and-transport model. The fate-and-transport model is a combination of a serially connected airflow model and contaminant mass-balance transport model, in which the airflow predictions are inputs to the transport model. Thus, the estimates from the airflow model provide an additional opportunity for real-time sensor fusion. Brand et al. (1995) describe a general framework in which the BMC algorithm can be applied towards reducing the uncertainties of interconnected risk models. When available, multiple types of measurements can be used to reduce the uncertainties of interconnected models.

## **6.2.2 HVAC operation and control**

Existing building information systems, such as HVAC control systems, may contain resources that can be exploited for heterogeneous sensor fusion. This section describes basic design and control of common commercial air distribution systems. I later use this section to support the proposed heterogeneous sensor fusion methods.

In most commercial buildings, thermal and ventilation requirements are managed simultaneously using overhead air distribution systems. Appropriate quantities of outdoor and recirculated air are filtered, thermally conditioned, and redistributed through a ducted delivery system. Underfloor systems also provide thermally conditioned air but through a pressurized plenum beneath a raised floor. Radiant heating and cooling systems, in contrast, are designed to meet thermal requirements while outdoor air requirements are met separately through an air handling device. Currently, radiant

systems and underfloor systems are not common in U.S. commercial buildings. The remainder of this discussion focuses on overhead air distribution systems.

### **6.2.2.1 Overhead air distribution systems**

Overhead air distribution systems usually consist of common components that include an air delivery device, ducts, diffusers, and sensors. The air delivery device is typically referred to as a package unit or an air handling unit (AHU). A package unit is a self-contained system that includes the heating and cooling equipment, such as gas heaters, compressors, and condenser fans. In contrast, the AHU uses heat exchangers to condition the air, where the primary thermal-mechanical equipment is housed separately (such as the chillers, cooling towers, etc.).

Package units tend to be installed in smaller buildings, whereas AHUs are usually found in larger buildings. Package units commonly have less sophisticated controls than AHUs. I will disregard the differences in the ensuing discussion and, unless otherwise indicated, will refer to both types of systems as air handling units (AHUs).

### **6.2.2.2 Air management at the air handling unit**

The control system at the AHU level serves three main functions: manage the quantity of outdoor air needed for ventilation, condition the supply air to the desired supply temperature, and control the quantity of air supplied to the building. The economizer, heat exchanger, and supply fan, respectively, achieve these functions, and in that order.

Economizers are used to vary the amount of ventilation. Not all AHUs are equipped with economizers, although all commercial units must provide a minimal level of outdoor

air. Some package units are designed to allow a fixed, continuous quantity of fresh air. In contrast, an economizer reduces HVAC energy consumption by modulating the amount of outdoor air (which is mixed with some recirculated air) based on the temperature, and, in some climates, the humidity of the outside air. The control sequence makes use of the temperatures (or enthalpy) of the outside and return air and determines the operation of the economizer dampers. An example control sequence for the economizer outdoor air damper is illustrated in Figure 6.1.

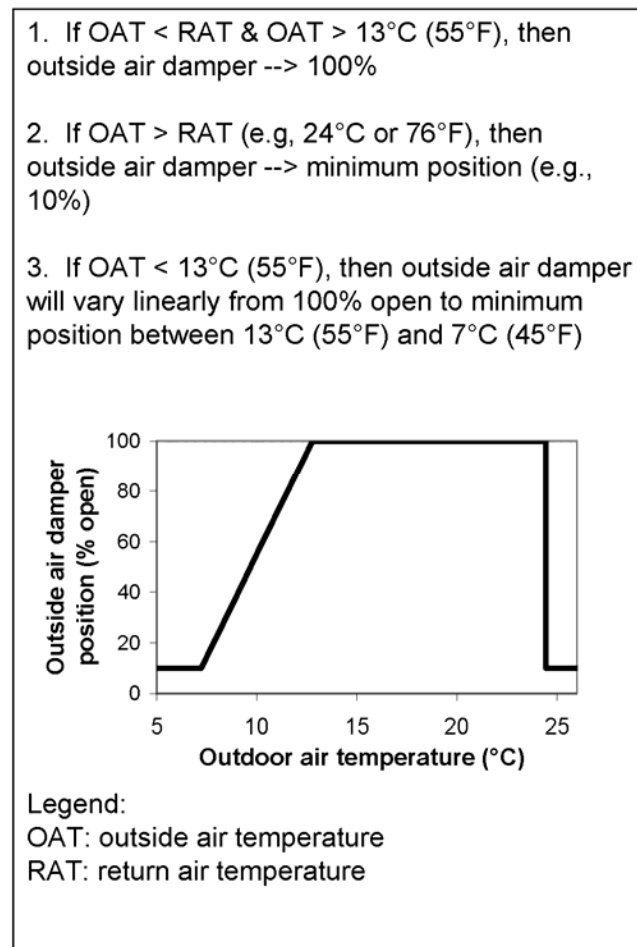


Figure 6.1. Sample economizer control sequence for economizer outside air damper. Percentage of outside air is varied based on the outside air temperature to minimize energy use.

Economizers are controlled through an energy management control system (EMCS), which provides access to several measurements such as the temperatures (and possibly humidities) of the outdoor air stream, the mixed air stream, the return air stream, and the supply air streams, and also the airflow rates. (In lieu of airflow measurements, temperatures of mixed air, return and outdoor air may be used to determine the proportions of recirculated and outdoor air.) The economizer damper-position signals are also accessible through the EMCS. In the event that other sensors fail, damper signals can provide information vital to knowing the ventilation condition.

In more sophisticated systems, the economizer may be controlled to not only minimize energy, but also to maintain a minimal amount of outdoor air that is required based on occupancy. Recently developed control sequences, promoted by the LEED standard ([www.usgbc.org](http://www.usgbc.org)), require that the interior carbon dioxide (CO<sub>2</sub>) levels be maintained within a fixed difference relative to the outdoor CO<sub>2</sub> levels. CO<sub>2</sub> measurements would be available through the EMCS if such a control sequence was implemented.

The supply fan controls the quantities of air distributed throughout the building. In many commercial systems, the supply fan speed is controlled using a variable frequency drive (VFD) that is modulated to maintain a desired static pressure downstream of the AHU (Figure 6.2). This strategy is used in what are called variable air volume (VAV) systems. In a VAV system, the quantity of air is modulated, rather than the temperature of the air, to meet thermal requirements of the zones that are served by an AHU. The information available from supply fan control loops may include the VFD control signal, duct static pressure, fan speed and airflow rate.

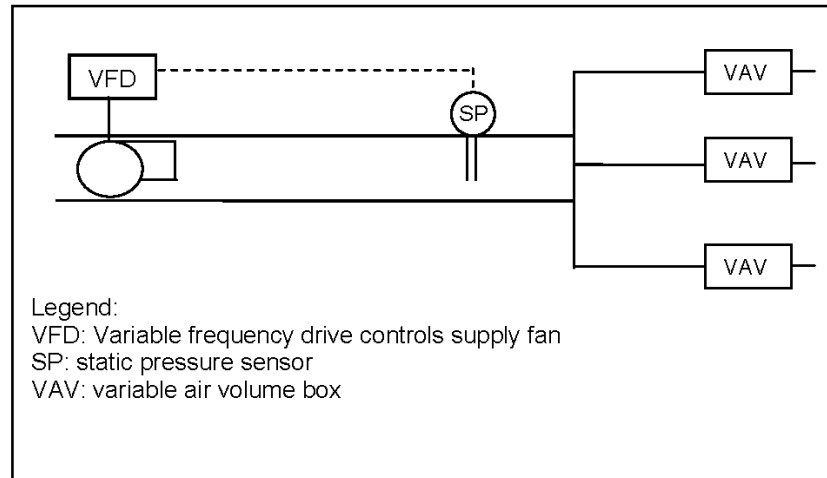


Figure 6.2. Supply fan diagram. Variable frequency drive controls the supply fan speed, based on the static pressure measured in the duct downstream of the fan. Variable air volume boxes supply air to individual zones.

In less sophisticated systems, particularly package units, the supply fan may operate at one speed and fluctuate between “on” and “off” states, to maintain the occupied zone’s temperature close to the desired value. Or, the supply fan may have two-speed control and modulate between these two speeds and the “off” position. In these types of control strategies, all that may be known is if the fan is on or off, or in the case of a two-speed fan, if the fan is on low or high speed, or off.

Relief systems allow the building to maintain a desired level of internal pressurization, which can vary based on the economizer function. There are different types of relief-system design and control. Although some systems use return fans, the use of return fans is often discouraged, because of the difficulty in implementing an effective control strategy, owing to the dependence on airflow measurements that are often unreliable (Taylor, 2000). A more common design is to use a motorized relief

damper that is adjusted to maintain a constant pressure difference between the inside and outside of the building.

### **6.2.2.3 Air management downstream of the air handling unit**

In most commercial systems using digital controls, variable air volume (VAV) systems manage the air distribution downstream of the AHU. The thermal needs of individual zones are satisfied by modulating the airflow to these zones. Both package units and conventional AHUs can be designed with a VAV system.

The control scheme is as follows (Figure 6.3). A box containing a damper (i.e., a VAV box) is located upstream of the zone, where the zone typically constitutes a group of rooms or areas that have similar thermal-conditioning needs. The VAV box controller rotates a damper according to the zone temperature. As the VAV box dampers close, the static pressure upstream of the VAV box increases and the supply fan speed is reduced to maintain constant static pressure.

VAV boxes often operate in cooling mode to compensate for internal heat loads. However, a VAV box may be designed such that it can deliver warm air when needed. These boxes are equipped with a hot water heat exchanger and sometimes a supplemental fan. A VAV box will operate by a different set of instructions when in heating mode. The sequence may instruct the damper to open fully and turn on the supplemental fan.

VAV boxes can be either pressure-dependent (PD) or pressure-independent (PI). In PD boxes, the damper position is controlled directly to meet the desired temperature (i.e., the temperature setpoint). The amount of airflow for a given damper position will vary depending on the static pressure upstream in the duct. In PI boxes, the VAV box controller calculates the airflow required to achieve the temperature setpoint, and this

airflow is the input signal to the control loop. The damper is adjusted to achieve the desired airflow through the box which is often measured using a pitot tube sensor. An advantage of PI boxes is that the airflow to the zone is independent of the duct static pressure upstream of the VAV box.

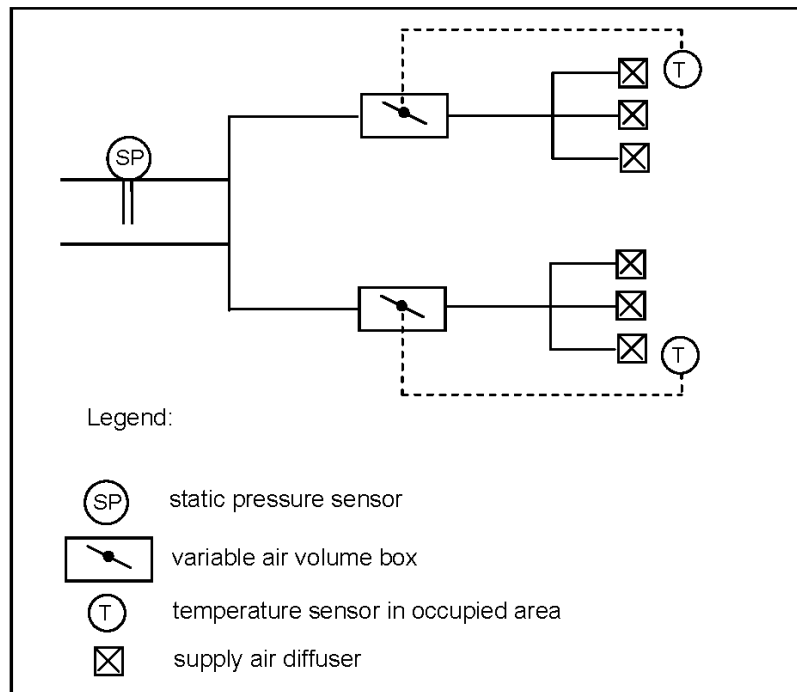


Figure 6.3. Variable air volume (VAV) units supply air to the individual zones based on the actual and desired temperature of the zones; zones in HVAC terminology consist of groups of rooms or areas that have similar thermal conditioning needs.

Most VAV box controllers are networked to the EMCS. Thus, information about the zone temperatures, VAV box airflows (for PI boxes), and damper positions is available to a sensor system. These measurements, especially of airflow rates, may be useful towards characterizing contaminant dispersion. Section 6.3.3 discusses how these measurements might be incorporated into a sensor system.

### 6.2.3 Case study

Essential elements of the case study are restated here. The test unit has 660 m<sup>3</sup> interior volume and 280 m<sup>2</sup> of floor area spanning three levels. One mechanical air-handling unit (AHU) recirculates the air to the first and second floors with one return on the first floor. Inside-to-outside air exchange occurs by infiltration, and intrazonal flows within the building result from a combination of mechanically induced, pressure-driven flow and convective airflow.

Fan pressurization tests were conducted to characterize the airflows and twelve tracer gas experiments were conducted over a five-day period (Sextro et al., 1999). In each tracer experiment, 20 g of propylene were instantaneously released at a specific interior location. Concentrations in each room and the staircase were recorded at 20 s intervals (for the first 20 min) and 60 s intervals (after 20 min). The operating conditions varied for each experiment (Table 3.1). Interior temperatures were recorded every 30 s, and weather conditions were recorded every 15 min. The leakage and airflow characteristics were used to develop a multizone airflow and pollutant dispersion model for the unit, based on the contaminant transport model COMIS (Feustel, 1999). This chapter uses data from Experiments 1 and 4.

Here, I discuss the relationship between the case study and typical HVAC design for commercial buildings. It is important to note that this case study may not be representative of typical design and control scenarios. For example, the AHU in the case study is either on or off. Many commercial systems are designed with VAV systems, and thus, the fan speed is modulated. On/off control is found more often in smaller buildings that use package units. An additional characteristic of the test unit is that the AHU is not



controlled to maintain thermal comfort, as is common in commercial systems. (This is because the test unit was built for experimental purposes.)

A typical commercial control system may use several sensors, such as airflow, temperature, or differential pressure, to inform their air distribution control algorithms. In contrast, the information types available in this case study are limited. However, the operating conditions may be used to simulate possible sensor types. For example, the known door positions during the tracer experiments could be used to produce simulated door-position sensor signals and the known AHU operating status could be used to generate a simulated control-system input signal. Temperature measurements and other weather conditions, which are commonly available through an EMCS, were collected during the case-study experiments and could also be considered as input information for heterogeneous sensor fusion.

### **6.3 Heterogeneous sensor fusion concepts**

I discuss two distinct perspectives that may be useful towards considering sensors for heterogeneous fusion. The first perspective begins with an understanding of the physical system (i.e., contaminant transport); I refer to this as the physics perspective. The second perspective is influenced by the sensor fusion algorithms and limitations associated with modeling and computation. I will refer to this latter perspective as the algorithm perspective. Both of these perspectives may be helpful towards identifying useful sensors. For example, an attractive candidate sensor may be one that helps to understand the physics (including characterizing the source) and also can be easily incorporated into an algorithm. The ensuing discussion attempts to identify examples of such sensors.

### 6.3.1 Physics perspective

The goal of characterizing the release can be framed as an inverse problem. Figure 6.4 reproduces the state diagrams from Chapter 1 with some modifications.

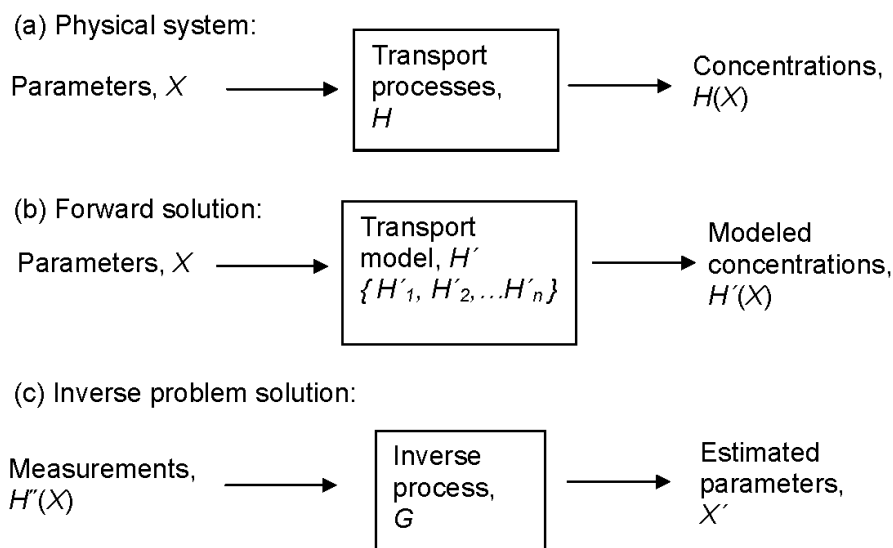


Figure 6.4. Physical system (a), forward solution (b) and inverse problem solution (c)

The inverse process,  $G$ , must estimate  $X$  using measurements  $H'(X)$ . In this dissertation, the BMC algorithm exercises the model  $H'$  prior to a release event over a range of possible parameters  $X$  to generate uncertainty distributions of  $H'(X)$ . In real-time,  $G$  compares  $H'(X)$  to multiple families of  $H'(X)$  and generates updated uncertainty distributions estimates of  $X$  (denoted  $X'$ ).

The model  $H'$  represents a complicated physical system ( $H$ ) that ideally encapsulates all the chemical and physical processes that govern the fate and transport of a contaminant species.  $H'$  may consist of different submodels, which at a minimum will include a bulk airflow model and contaminant mass-balance model.  $X$  consists of all parameters that significantly influence the contaminant concentration in the event of a

release. These include parameters that describe the contaminant release such as the location and quantity.  $X$  also includes the physical attributes of building that influence transport, such as operating conditions of the HVAC system, door positions, or occupancy.

In an actual implementation, sensors may exist, or could be deployed, to report information associated with  $H'$  (or any subset of  $H'$ ) such as the outputs from a submodel, which may include pressure differences or airflow. Sensors could also track characteristics of  $X$ , such as door positions. This chapter considers how information relating to  $H'$  and  $X$  can be incorporated in a multisensor system. I address this question by considering first the *real system*  $H$ , rather than the model,  $H'$ . By doing so, this exposition is not influenced by any particular model or algorithm limitation.

There are many factors that influence either airflow or contaminant transport and that may be useful in a multisensor system. Measurements can be made available to track these factors. Examples include HVAC system operating conditions, static and dynamic pressures, temperatures, elevator operation, door positions, wind direction and speed, and even the number and movement of occupants.

A useful concept for viewing such sensors is the degree to which the measured parameter influences contaminant transport. Referring to Figure 6.4, the more any parameter  $X$ , or outcome of subprocess  $H$  influences  $H(X)$ , the greater the degree of influence will be. Pressure differences between rooms, for example, play a large role in determining bulk airflow between rooms, which in turn influences contaminant transport between rooms. Door positions and occupancy may play a large role in room-to-room transport. For commercial systems using VAV technology, the damper position and

upstream duct static pressure will determine the overhead airflow rate to a zone. A poorly tuned or unstable control loop on a VAV box will alter the airflow to rooms.

The design and operation of the building determine the major airflow pathways and hence provide some insight into which measurements may exhibit a greater degree of influence. Design features and parameters include the architectural features of the building envelope, room occupancies and uses, and HVAC systems. For example, in a well-sealed building with uniform temperature setpoints throughout the building, infiltration and internal airflow driven by natural convection are likely to be small. Assuming the windows are closed, exterior temperature measurements, wind direction and velocity measurements may not be helpful, since the airflow is mainly controlled by the HVAC system. In contrast, consider a mixed-mode ventilation building where windows can be opened intentionally to provide natural ventilation. Here, door and window positions may be strongly relevant to characterizing the airflow pathways and resulting contaminant transport.

An interdependency may exist among different influencing factors and, if so, measurements of each may not be necessary. For example, the HVAC system operation is influenced by thermal loads, which in turn are influenced by real-time occupancies and weather conditions. Measurements of the HVAC system operation may be sufficient for achieving the purposes of the sensor system. The economizer function may also be determined by occupancy, if demand-controlled ventilation is employed. Here again, knowledge of the economizer function may be sufficient to achieve a sensor system goal, although information regarding occupancy could be used in parallel to possibly diagnose faulty operation of the economizer or HVAC system and to establish sensor redundancy.

The actual or *realized* degree of influence of any factor during a contaminant release event may be strongly related to the release parameters – especially the magnitude and site of the release. For example, the airflow through a doorway may be significant to the transport between two particular zones, but if the release originates in a far removed location, the door positions downstream along the flow path may be unimportant in affecting the final contaminant concentrations. An extreme scenario would be a case where a subset of zones are completely isolated from another set of zones in which the contaminant originates. In such a condition, the airflow in the isolated zones becomes irrelevant and the door positions in the isolated zones have no influence. The fact that the realized degree of influence is not known in advance poses an additional design challenge.

There is another class of measurements that do not influence contaminant transport, but may contribute to a heterogeneous sensor system: passive chemical measurements. I describe these measurements as passive, because they do not participate in overall contaminant transport. Measurements of non-contaminant species, such as carbon dioxide, or volatile organic compounds, may provide some useful information towards characterizing the overall airflow and transport pathways. These measurements are referred to as passive measurements in the later sections.

### **6.3.2 Algorithm perspective**

Information may be meaningful physically, but the ability to use that information may be limited by the transport model and algorithm used to solve the inverse problem. I first discuss general model and algorithm sensor fusion issues, and then issues specific to

the BMC algorithm and the multizone contaminant-transport model, both of which are used in the illustrative example to follow.

### **6.3.2.1 General model and algorithm considerations**

I classify sensors in two ways to address model and algorithm issues. The first classification considers whether the measurement represents a model input or model output. The second classification considers how the measurement informs the dynamic state of the system. I discuss algorithm implications for fusing information from sensors within each category and use this taxonomy to elucidate the fusion of data from specific sensor types.

#### *Model input / output classification*

The first classification addresses the role of the measurement in the model (or sets of models) used to solve the inverse problem: Does the model include the measurement as a parameter? Is the measurement a model input or output, or both, as can be the case in interconnected submodels? Airflow, for example, is an output from the airflow model and an input to the contaminant transport model. In such situations, should the information be treated as a model input or model output? This last question is important because information should be used only once to avoid double counting, which can result in overly confident determinations of the unknown parameters.

I first consider measurements that are classified as model inputs. Examples of measurements that fall under the model input classification include temperature, fan operating speeds, and door positions. The degree of influence concept was introduced and refers to the extent to which the parameter influences the final contaminant

concentrations. Except for the final contaminant concentrations, virtually all parameters that “influence” the final concentrations behave as model inputs. The degree of influence can be estimated by exercising the model (or models) over ranges of these model inputs. The physical conditions of the system can be used to further inform the potential usefulness of a measurement that is classified as a model input. The measurement accuracy and frequency should also be considered when selecting a sensor.

A measurement may be classified as a model output. Measurements classified as model outputs include static pressure and airflow, since these quantities can be estimated by multizone models. The issues to consider in interpreting such sensors are similar to the issues considered in interpreting concentration sensors: How accurately does the model predict this quantity? What is the accuracy of the measurement? Both these issues will have implications for constructing an algorithm to solve the inverse problem.

Excluding contaminant concentrations, all measurements classified as model outputs behave also as model inputs. This situation arises when the overall transport model is composed of a subset of linked models that are connected in parallel or serially. In the illustrative examples shown in this dissertation, airflow is calculated from the multizone airflow model and is an input to the contaminant mass-balance model. Thus, airflow is an output from the first model and an input to the second. Static pressure difference is another example of a variable that serves both as a model input and output. Pressure is estimated as an intermediate output by a multizone model (using temperatures as model inputs) that is subsequently used to estimate airflow. If pressure measurements were available, airflows could be estimated (or verified) directly using the flowpath’s leakage characteristics.

When sensors provide data that serve simultaneously as a model input and output, a decision has to be made on how to treat that information. In the case of an accurate sensor, it may be desirable to use the measurement as a model input. Consider static pressure measurements. These measurements can be made fairly accurately. In such a case, it may not be necessary to use temperature measurements and a multizone model to estimate pressure differences across zones. Instead, the pressure measurements could be directly used to calculate the airflows among zones.

On the other hand, if a sensor generates data that is characterized by large errors, two approaches could be considered. In the first approach, the measurement could be treated as a model input with an associated uncertainty distribution to represent the error. For example, in the BMC algorithm, a library could be generated such that multiple realizations represent the uncertainty of the measurement. Alternatively, the measurement could be treated as a model output and the sensor error could be formally treated when interpreting the information in real time. Airflow measurements, for example, may be abundant but are often inaccurate. It may not be desirable to abandon an airflow prediction model and feed inaccurate airflow sensors directly into a contaminant mass-balance model without incorporating the error characteristics of the measurement. Alternatively, airflow measurements could be used to validate the predictions from an airflow prediction model and reduce the uncertainty distributions of the airflow in a manner similar to how information from contaminant sensors is fused.

### *Dynamical state of the measurement*

A second classification within the algorithm framework considers the time variation of the measurement. I define three dynamical states — static, pseudo-static and dynamic



— and discuss algorithm implications for each. The extent to which a measurement varies with time has implications for information fusion from these sensors.

Static quantities are ones that rarely or never change. An example of a static measurement is the speed of a fan that operates at constant speed. Another example is the leakage characteristics of the building envelope. Static measurements are treated as constants in the model and algorithm. As such, repeat or real-time measurements of a static quantity are neither necessary, nor useful.

Pseudo-static quantities may change over time, but remain constant over the time scale of a release event (i.e., on the order of hours). Pseudo-static quantities can be treated as constants in the model and inverse algorithm. However, because they may change with time, updated measurements are needed for monitoring beyond the scale of release events. In addition, though the algorithm can treat pseudo-static quantities as constants, the algorithm must consider multiple possible values. Examples of pseudo-static quantities are the speed of a modulated supply fan and the position of a window.

Dynamic quantities change at a frequency that is comparable to or faster than the timescale of a release event. Examples of dynamic information include contaminant concentrations, elevator positions, or a damper position in a VAV box encountering unstable operation (i.e., owing to a poorly tuned control loop). Interpretation of dynamic model outputs, such as contaminant concentrations, does not require further algorithm discussion. However, the interpretation of information from sensors that qualify as dynamic model inputs is more challenging. The algorithm must either process the model inputs at a rate comparable to the rate at which the variable changes, or anticipate the

values prior to an event occurring. If the model input changes in real time, it is not clear which algorithm should be used to solve the inverse problem as both Markov Chain Monte Carlo (MCMC) and Bayes Monte Carlo (BMC) have limitations.

The issues associated with the dynamical state of the sensor should be considered collectively with its degree of influence. If a parameter is a highly dynamic quantity but exhibits little influence on contaminant transport, then its treatment will not be important in solving the inverse problem.

### *Implications for model selection*

While this dissertation uses only a multizone contaminant model, other modeling approaches are possible, and may be required to fuse information from different sensors. For example, occupancy influences the thermal loading of the building and airflow patterns within a room. Characterizing the influence of occupant behavior on room airflow patterns would require a CFD model. However, the influence of occupancy on thermal loading could be incorporated into a sensor system that couples a whole building energy model to a multizone airflow and transport model (Axley et al., 2002; McDowell et al., 2003).

The two classifications of sensor information, within the algorithm framework, can be used to evaluate the suitability of the contaminant transport model and algorithm for multisensor fusion. Conversely, sensors may be selected based on the ability of the modeling tools to accommodate such information.

### 6.3.2.2 Multisensor fusion with Bayes Monte Carlo

Sensor information can be fused in both stages of the BMC algorithm. In the first stage, both fixed and variable model inputs are used to generate the contaminant transport library. The fixed parameters are usually static variables. Variable parameters, which are generated from prior uncertainty distributions, are pseudo-static variables. Active measurements may be available for both categories of variables. In the second stage, measurements of model outputs, such as contaminant concentrations or airflow measurements, are available to be fused in real time.

When incorporating a model input into the library generation stage, a wide range of possible values must be considered. In real time, actual measurements of model inputs can be used to sort among the prior parameter values used to generate the library. If a sensor measures a discrete variable, such as a door position, this measurement can be used to effectively select from the library a subset that is representative of current conditions. If the sensor measures a continuous variable, then the library can be sorted in real time to retain a range of values that encompasses the measurement. In either case, a prior-sorting algorithm should be devised that considers the uncertainties of the measurement.

If a model input is pseudo-static, but follows a predictable pattern, it may be computationally efficient to use this measurement to generate a dynamic library, rather than attempting to generate a more comprehensive library at an earlier time. Dynamic generation of the library may be one approach to compensate for the increased complexity as more variables are considered. One example of a pseudo-static variable is

outdoor temperature, which changes over time, but usually not significantly over the timescale of a release event.

Sensors representing dynamic model inputs are not easily incorporated into the BMC algorithm, since typical implementation of the BMC algorithm assumes fixed parameters for the duration of the release. (Generating a library in advance by sampling over dynamic state variables, though not impossible, may be computationally impractical.) Dynamic state estimation may require a more sophisticated algorithm, such as a Sequential Monte Carlo method, which is used in target tracking applications. However, because the building system is not a highly dynamic system (on a release-event time scale), fusion of dynamic model inputs may not be necessary.

Some measurements fall under the classification of both model input and model output and are also pseudo-static. Examples include pressure and airflow measurements. If such measurements are available, it is necessary to determine whether the measurement will be used as model input or as model output. This decision may be influenced by the quality of the measurement. Pressure measurements could be used directly to estimate zonal airflows. In such a case, the airflow and contaminant transport library could be generated for BMC processing on a range of differential pressure measurements, a subset of which could be selected, using real-time pressure measurements.

Airflow measurements fall under the dual classification of model input and model output and have a pseudo-static dynamical nature. Compared to pressure measurements, they are likely to be less accurate and characterized by larger uncertainties. It is possible to represent an inaccurate measurement as a model input by generating a library based on the uncertainty distribution of the measured input. However, it may be more

computationally efficient to process such a measurement as a model output. In this scheme, airflow measurements would be interpreted in a manner similar to how contaminant concentrations are interpreted. Airflow measurements could be compared to modeled airflows, updating the posterior probabilities of each airflow realization in the library. Then, contaminant concentrations could be interpreted to further update the posterior probabilities of the simulations in the library. Brand et al. (1995) proposed the use of BMC for reducing the uncertainties of linked exposure assessment models.

### **6.3.3 Information fusion from specific sensors**

I use the physics and algorithm sensor fusion frameworks outlined in the previous sections to discuss the fusion of specific sensors. A variety of sensors are listed, some of which may be available as part of an Energy Management Control System (EMCS). Table 6.1 lists several specific sensors that could be used in a heterogeneous sensor network. Each sensor listed in Table 6.1 was selected because it measures an attribute that exhibits some degree of influence on contaminant transport. The sensors considered in this section are included for illustrative purposes and may not be exhaustive.

Table 6.1 Sensors for consideration in heterogeneous sensor networks

Measurement	Anticipated conditions for high degree of influence <sup>b</sup>	Model input output classification	Dynamical classification
Static pressure (e.g., between rooms, supply duct <sup>a</sup> , outdoor to inside)	Static pressure always exhibits a high degree of influence: static pressure between zones determines mixing between zones; supply duct static pressure determines airflow through HVAC system; outside to inside pressure is often used to control relief systems, and also determines airflow through exterior doors and windows	Both <sup>c</sup>	Pseudo-static
Temperature (e.g., zones, outdoor, AHU return air, supply and mixed air) <sup>a</sup>	When buoyancy induced airflow is significant and where temperatures are used to determine airflow supplied by the HVAC system	Input	Pseudo-static
Wind speed & direction	For buildings with leaky envelopes or operable windows	Input	Pseudo-static
Airflow (e.g., VAV box, AHU supply and outdoor air) <sup>a</sup>	Airflow influences contaminant transport in all conditions: VAV boxes determine airflow to zones, and supply and outdoor airflows determine overall ventilation time scales	Both <sup>d</sup>	Pseudo-static
Control system signals (e.g., economizer and VAV box damper positions, VFD signal) <sup>a</sup>	These signals are used to control overall airflow or ventilation and have a high degree of influence	Input	Pseudo-static
Door	Important when buoyancy induced airflow is significant; also important for zones not served by mechanical airflow	Input	Pseudo-static
Window position	Important for leaky envelopes and when windows are operable	Input	Pseudo-static
Occupancy <sup>c</sup>	Important when buildings are densely occupied and determine the supply from forced air convection systems; occupancy can also influence temperature (and airflow) in naturally ventilated buildings	Input	Pseudo-static
Elevator position	Important if leakage between the elevator shaft and building is significant	Input	Pseudo-static to dynamic

<sup>a</sup> Indicates that the measurement may be available from an EMCS.

<sup>b</sup> The degree of influence on contaminant transport is a criterion that applies only to measurements that influence transport; contaminant concentrations, for example, do not exhibit any degree of influence.

<sup>c</sup> Usually an output from an airflow model, but if available can be used as a model input to airflow model.

<sup>d</sup> Output from airflow model and input to contaminant transport model.

### **6.3.3.1 Temperature and window position**

Temperature measurements and window position are likely to be useful for buildings that are cooled by natural ventilation or mixed-mode ventilation. In these buildings, windows are manually controlled according to occupant preferences and offset the cooling that would otherwise be provided by an economizer. Temperature and window position are both pseudo-static model inputs. (Temperatures are used by the multizone model, in lieu of pressure measurements, to determine buoyancy induced airflow.)

The multizone model incorporates both sensors as model inputs. To include these measurements in the BMC framework, reasonable combinations of temperature and window positions must be anticipated during the generation of the library. While it is possible to sample from a wide range of temperature and window conditions, a more practical implementation might involve generating a more representative library based on occupant behavior and weather conditions. Monitoring may reveal correlations between typical occupant behavior and values of these two variables. Monitoring information may inform more realistic library generation.

### **6.3.3.2 Static pressure**

Static pressure measurements are attractive because they are accurate and strongly influence airflow throughout the building. Pressure differences between zones (along with the leakage characteristics) govern the magnitude of the airflow between the zones. Static pressure in ducts determines the overall airflow through diffusers. Interior to exterior pressure differences govern leakage.

Control systems are often configured to maintain specified pressure differences. Some AHU relief systems are controlled such that a damper is opened to maintain a desired interior-to-exterior pressure difference. Supply fans are often controlled to maintain a constant downstream duct static pressure, in VAV systems.

From a measurement perspective, a pressure sensor is relatively inexpensive and accurate. When pressure is controlled, it is unlikely to vary if the control system is properly tuned, which has positive algorithm consequences. In terms of sensor fusion, there are different ways to incorporate pressure measurements. In one method, real-time measurements of pressure measurements could be used directly by an airflow model to estimate zonal airflows. The library would then be generated to represent contaminant release scenarios. In another approach, the library could be generated based on possible ranges of differential pressure measurements. Then, during a release event, the library could be sorted using actual pressure measurements.

### **6.3.3.3 Variable air volume (VAV) information**

In many commercial buildings, airflow is dominated by HVAC function, and airflow is often distributed using a VAV system. In such a case, airflow information from VAV boxes is likely to be useful. VAV boxes represent the airflow to a group of zones.

In one approach, the VAV box airflows could be used directly as an input to a multizone model. The multizone model would calculate the airflow to the individual zones downstream of the VAV box. If the BMC algorithm is used, it would be necessary to generate a library of all possible VAV box airflows in advance of the release. This task is challenging but reasonably achievable. VAV box dampers are controlled in real-time, based on thermal loads. It is reasonable to approximate damper positions and



airflows as pseudo-static measurements over the duration of a release event. The occupancy and thermal loading of a building are relatively predictable, given historical data. Thus, a typical range of VAV box airflow conditions could be anticipated in advance and used to generate the multizone models that populate the simulation library.

In another approach, airflow measurements could be treated as model outputs, rather than inputs. Commonly, airflow measurements are characterized by large uncertainties. In such a case, a realistic library would still be generated; however, the VAV box airflows would be treated as model outputs, and the real-time measurements would be used to update the uncertainties of this airflow. The posterior probability of each possible airflow scenario in the library would be updated using real-time measurements of VAV-box airflow.

#### **6.3.3.4 Economizer information**

Incorporating realistic behavior of the economizer into a sensor system is likely to be critical to the overall success of a sensor system in a realistic commercial building. Along with infiltration, the economizer controls the indoor-to-outdoor contaminant transport by introducing fresh air. The introduction of fresh air determines the removal time scale of the contaminant, and thus limits the overall time available for characterizing the release.

The economizer is typically controlled to match either occupancy or climate conditions (i.e., temperature or enthalpy), which are unlikely to change over the duration of a release. Economizer-related variables, such as outdoor air fraction or airflow, and recirculated air fraction or airflow, can be treated as pseudo-static data. The damper positions or economizer airflows could be incorporated into an airflow model. As

proposed with temperature and VAV box airflows, it may be possible to capture realistic behavior by relating the economizer behavior to other variables, thus reducing the computational burden of generating a library of airflow and release scenarios.

The potential benefit of the measurements discussed in this section is not restricted to a BMC-based solution. Other inverse problem approaches would benefit from these measurements, since they provide information that facilitates the estimation of the airflow paths.

## **6.4 Illustrative example**

This section illustrates the fusion of information from heterogeneous sensors using a mix of door position and chemical concentration sensors. The purpose is to substantially explore the frameworks introduced in the preceding section. Door position sensors were selected for two reasons. The first is that the doorways in the case study experience significant airflows. (Figure 3.1 shows the modeled airflows for Experiment 1.) The second reason for choosing door position sensors is that the case study tracer experiments were conducted with different door positions and AHU operating conditions. Thus, the case study lends itself to an exploration of the different conditions under which door position sensors may be beneficial towards improving the overall sensor system performance.

Sensor fusion using door sensors is easily investigated with the BMC algorithm. Using the algorithm frameworks, door sensors are active measurements and can be treated as pseudo-static model inputs. Door sensor data can be fused by adjusting the

prior probabilities of the library realizations. The next section describes an algorithm that was developed for this purpose.

### 6.4.1 Door position fusion algorithm

In this case study, door position is treated as a binary variable (i.e., either fully open or closed). Real-time information on actual door positions is used to adjust the prior probabilities of each realization in the library, based on the degree of fit between the door position sensors and the assumed input values of each realization.

A prior filtering algorithm was developed based on three factors: (1) the degree of fit between the actual door positions and the door positions assumed in generating the realization ( $d_i$ ); (2) a threshold that determines the minimal degree of fit required to place value on the match ( $\delta$ ); and (3) the confidence that is assigned to the sensor-to-model fit, which determines how much the prior probability of each realization is adjusted ( $\gamma$ ).

The probability assigned to each simulation is adjusted based on  $d_i$ ,  $\delta$ , and  $\gamma$ , with intermediate factors that ensure that the final sum of each probability,  $p_i$ , is 1. The algorithm assumes that the initial probability of each realization is uniform. Figure 6.5 describes the prior filtering algorithm. The derivations of some steps are described briefly. For example, the upper limit,  $\theta_{max}$ , was derived by setting  $\sum p_i$  for all realizations with a favorable match equal to 1.

1. Calculate the degree of match,  $d_i$ , for each simulation:

$$d_i = \frac{n_{match,i}}{n_{sensors}}$$

where  $n_{match}$  is the number of sensors in the network that match the positions of realization  $i$ .

2. Calculate the maximum value of the factor used to adjust the prior probability of each realization with a favorable match:

$$\theta_{max} = \frac{1 - \frac{N_{d_i \geq \delta}}{N}}{\sum_{i=1, d_i \geq \delta}^N d_i}$$

where  $N_{d_i \geq \delta}$  is the number of realizations with  $d_i \geq \delta$  and  $0 \leq \delta \leq 1$  ( $\theta_{max}$  was derived by setting  $\sum p_i$  (for  $d_i \geq \delta$ ) = 1 based on step 4.)

3. Calculate the factor used to adjust each prior probability based on the confidence assigned to the sensors:

$$\theta = \theta_{max} \times \gamma \text{ where } 0 \leq \gamma \leq 1$$

4. Adjust the prior probabilities of each realization with a favorable match:

$$d_i \geq \delta : p_i = \theta \times d_i + \frac{1}{N}$$

5. Adjust the prior probabilities of all other realizations:

$$d_i < \delta : p_i = \frac{1 - \sum_{i=1, d_i \geq \delta}^N p_i}{N - N_{d_i \geq \delta}}$$

Step 5 was derived by setting  $\sum p_i = 1$  for all realizations in the library.

Figure 6.5. Prior probability algorithm for use with door position sensors.

This algorithm can be used to adjust the probabilities of each simulation based on the presence of binary sensors that represent active measurements, such as door position sensors.

At the maximum value of  $d_i$ , all sensors present in the system would match the values of the  $i^{th}$  realization. As  $\theta \rightarrow \theta_{max}$  and  $\delta \rightarrow 1$ , the probabilities of all realizations with exact matches are maximally weighted and the prior probabilities of all other

realizations approach 0. By adjusting  $\gamma$ , the degree to which the prior probabilities ( $p_i$ ) are adjusted, based on the sensor values, is varied.

The existence of two user-defined parameters,  $\delta$ , and  $\gamma$ , provides flexibility in treating the sensor measurements. The range of admissible values for  $\delta$  and  $\gamma$  is 0 - 1. Together, these parameters can be used to account for sensor and model uncertainty. Model uncertainty arises from the inability of the model to capture the physical effect of the parameter accurately (such as the effect of door position). There are different ways in which  $\delta$  and  $\gamma$  can be exercised to adjust the priors. Ultimately, the concentration measurements provide the “evidence” in Bayesian updating. The construction and subsequent adjustment of the priors is a subjective one, which, eventually becomes irrelevant with more evidence. However, the adjustment of the priors can be important for real-time estimation because the priors influences the efficiency with which the Bayesian updating scheme reduces the uncertainty distributions that describe the release.

A typical implementation may set  $\delta = 1$ , such that the priors of only those realizations with exact matches to the actual door positions will be increased. The extent to which these realization’s priors are increased can be tempered by setting  $\gamma < 1$ . Sensor error could result in some realizations not having an exact match to the sensor signals. By setting  $\delta < 1$ , it is possible to increase the prior probabilities of a larger subset of library realizations. In general,  $\delta$  could be interpreted as a parameter that compensates solely for sensor error and  $\gamma$  is a parameter that compensates simultaneously for sensor error and model uncertainty.

Furthermore, by setting  $\gamma < 1$ , it is possible to compensate for sparse sampling. The library may be adequately sampled if the entire library is used; however, if many

realizations are eliminated altogether through this fusion algorithm, it is possible that the resulting library may be insufficient for characterizing the release.

#### **6.4.2 Description of investigations**

The investigation makes use of two experiments plus simulated data. Collectively, these data sets address release conditions that include varying door positions, air handling unit (AHU) operation, source release location, and duration. The analyses are reported in two sections: mechanically driven airflow conditions (AHU is operating) and naturally driven airflow conditions (AHU off).

The mechanically driven airflow section evaluates hypothetical sensor networks constructed with data from Experiments 1 and 4, plus additional, simulated data sets. Comparing Experiments 1 and 4, door sensors were anticipated to be potentially useful for Experiment 4 but less useful for Experiment 1. The contaminant was released directly into the return intake for Experiment 1, but in the center of Room 1.3 for Experiment 4. With Experiment 1, the AHU distributed the contaminant rapidly to most of the rooms, excluding the stairwell and third floor. In Experiment 4, the contaminant is transported through a series of doors before being distributed widely by the AHU.

The second stage analyses of mechanically driven airflow conditions use simulated data. Simulated data were constructed to compensate for the limited scope of the case-study tracer experiments and to explore further conditions under which door sensors may be useful. A simulated data set based on Experiment 4 was constructed to provide a comparison case without the idiosyncracies of the experimental data set. A second

release originating in Stairwell 2 was simulated and investigated. The stairwell was chosen as the source location because it is not actively fed by mechanical airflow.

The usefulness of door sensors for conditions without mechanical airflow is evaluated in Section 6.4.4. Door positions are expected to bear a higher degree of influence on contaminant transport when the AHU is off. Thus, door sensors are expected to be more useful in such conditions. Only one empirical data set (Experiment 13) was available with the AHU not operating. Because one data set is inadequate to generate a statistical description of the model-to-measurement errors, simulated data are used to study this condition.

### **6.4.3 Mechanically driven airflow conditions**

This section investigates sensor networks comprised of combinations of door position and contaminant concentration sensors for conditions in which airflow is induced by both mechanical and natural means. Experiments 1 and 4 serve as the basis for the experimental component of the investigation. Subsequent data sets were generated to replicate Experiment 4 and to simulate a release originating in Stairwell 2.

#### **6.4.3.1 Library, likelihood function, and door sensor fusion**

The library and likelihood function from Chapter 5 are used in this section. To review, the base COMIS model (as described in Chapter 3) was used with the following simplification: the average weather and temperature conditions from all tracer-gas experiments were treated as fixed input conditions. Variations on this basic model were generated by sampling from a distribution of possible release locations, release masses, release durations, and door configurations, as summarized in Table 6.2.

Table 6.2 Parameter values used to generate the library of 5000 simulated contaminant releases

Parameter	Values
Source location	Twelve locations: any room, or stairwell and return duct. Each location is equally probable.
Source duration	1 s to 5 min; log-uniform distribution.
Source amount	10 to 100 g; log-uniform distribution.
Door position	32 possible door positions, based on interior doors on each level and stairwell doors acting independently

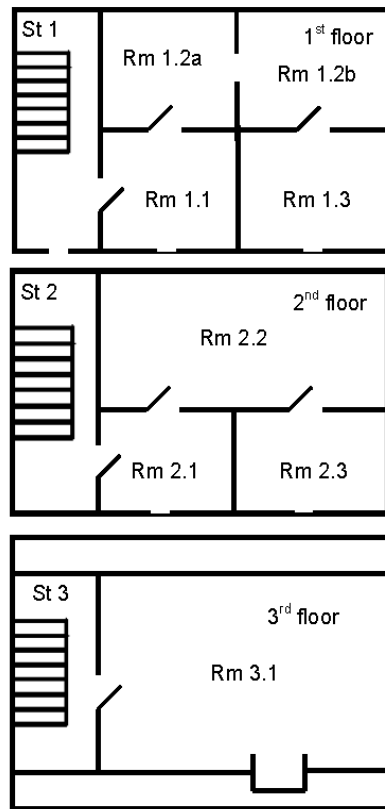


Figure 6.6. Layout indicating location of interior doors for the case study.

Although there are seven doors, the two interior doors separating the rooms on a given level were assumed to be in the same position. This reduced the number of



independent door settings to five, with  $2^5 = 32$  possible configurations. Figure 6.6 shows the locations of the interior doors.

The likelihood function was derived using the experimental data and model for Experiments 5-12 (AHU on). Figure 6.7 shows the normalized model-to-measurement error for these experiments. The errors were normalized by the release conditions. Also shown are the standard deviation of the binned errors (measurement – modeled concentration) and an exponential function fit to the standard deviations. Figure 6.7 is identical to Figure 5.1 and is described in more detail in Chapter 5.

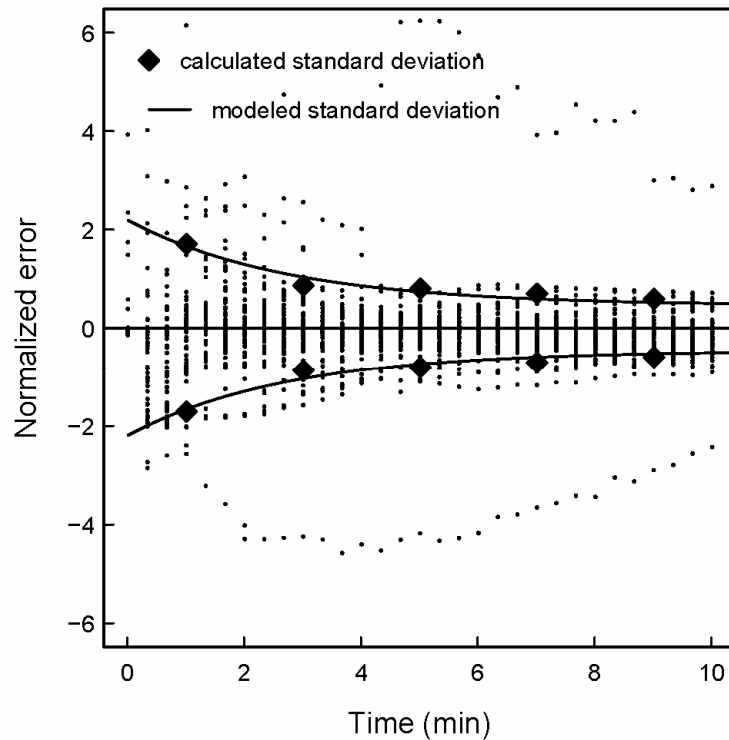


Figure 6.7. Normalized model-to-measurement error for seven tracer gas experiments (small points, Experiments 5, 6 & 8-12). Absolute errors are normalized by the instantaneous concentration of a well-mixed building. Also shown are calculated standard deviations for errors grouped into two-minute intervals and modeled standard deviation of normalized errors using exponential fit.

The exponential fit represents a time-based likelihood function for instantaneous releases. This equation was extended to very short duration releases through a preliminary investigation of the effects of duration. Equation 6.1 is the time-based likelihood function for the normalized errors for instantaneous or short-duration releases. (See Chapter 5 for more details.) The actual errors (units: g/m<sup>3</sup>) are obtained by multiplying the normalized parameters from equation 6.1 by the idealized well-mixed concentration assuming that the contaminant was instantaneous mixed throughout the entire building.

$$\sigma(t) = (Ae^{-Ct} + B) \times \frac{t}{\tau_{release}}, t \leq \tau_{release} \quad (6.1a)$$

$$\sigma(t) = Ae^{-Ct} + B, t > \tau_{release} \quad (6.1b)$$

where  $A = 1.74$ ,  $B = 0.46$ ,  $C = 0.37 \text{ min}^{-1}$ , and  $\tau_{release}$  is the release duration in minutes.

In this chapter, as in Chapter 5, the actual release time is treated as an unknown quantity. To summarize the procedure, the release time is treated as an additional parameter that is estimated using the BMC algorithm. The library is augmented to represent realizations with release times that begin 1 s, 1 min, 2 min and 3 min prior to when the first datum is received.

The door fusion algorithm follows the description in Figure 6.5. Combinations of door positions were generated as described by Table 6.2. These door positions were treated as hypothetical door sensor information available for real-time fusion. In this section, the door sensor algorithm makes the following assumptions: the door sensors are assumed to be perfect, thus a realization must have a perfect match to the door sensor positions ( $\delta=1$ ) in order for its prior probability to be increased; and the model is

assumed to treat the effect of the doors perfectly ( $\gamma=1$ ). These assumptions ensure that the maximum potential benefit (or penalty) of door sensors is evaluated.

#### **6.4.3.2 Results using experimental data**

In the investigation to follow, each door is equipped with a door-position sensor and the number of contaminant sensors is varied from one to eleven. A full range of possible contaminant sensors and sensor placements was investigated to understand the tradeoffs among door sensors, and number and placement of chemical concentration sensors. For each contaminant sensor network configuration, two networks were evaluated: one without any door sensors and one with a full set of door sensors.

The figures that follow compare the ability of networks, with and without door position sensors, based on Experiment 1, to identify the true release location and characterize the release mass. The results obtained with door position sensors are referenced to the y-axis and the base results are referenced to the x-axis. Each frame in the figures that depict the identification of the release location and the uncertainty reduction of the released mass lists the mean and standard deviation of the performance difference between the networks with door sensors and without door sensors. The difference,  $\Delta$ , is defined as the performance with door sensors minus the performance without door sensors. A positive performance difference indicates a higher probability at which the release location is identified, in the case of the location figures, and a greater reduction in the uncertainty of the released mass, in the case of the mass figures.

To summarize the results for Experiment 1, door position sensors are helpful towards the identification of the release location for many sparsely populated networks (i.e., 2 or 6 contaminant sensors). The median mass results are not significantly impacted by the

fusion of door-position information. However, networks with two contaminant sensors and door-position sensors are able to estimate the mass to a narrower confidence interval.

Figure 6.8 shows the probability of correctly locating the contaminant source for networks based on Experiment 1, with 2, 6 and 10 contaminant sensors, with and without door position sensors, at 5 and 10 min after the release. Figure 6.9 shows the identification of either the return intake or the room containing the intake (Room 1.2a). This latter metric is less strict than isolating the release location to the return intake, but still useful.

Door-position sensors are helpful towards finding the release origin for many networks with 2 contaminant sensors, as evidenced by Figure 6.8 and Figure 6.9. For a fewer number of networks, especially for 10-minute analysis periods, multisensor fusion has either a negative or minimal impact. Using the less strict source identification metric (Figure 6.9), there are fewer networks that are negatively impacted in a significant way. Based on the physics, door sensors were not expected to yield significant benefit in interpreting Experiment 1 data, because the release occurred at the return duct intake and the tracer would be dispersed via the AHU ducts, rather than through doorways.

It is possible that the impacts, both positive and negative, from using door sensors is a consequence of the post-fusion reduced library. The post-fusion library is smaller and thus expedites the identification of the release location. In cases where sensor placement is poor, the door sensors may hinder source characterization.

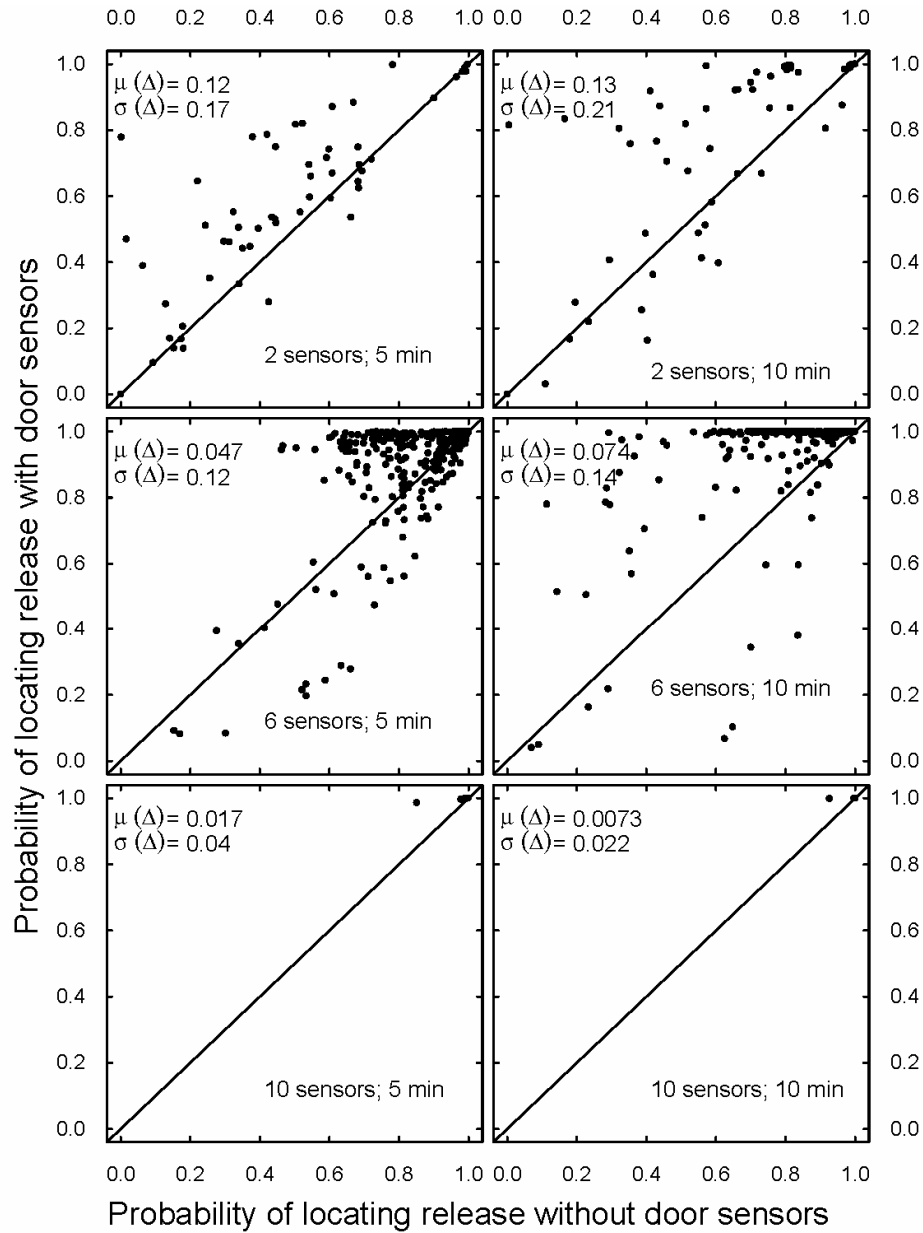


Figure 6.8. Experiment 1: probability of correctly locating release with and without door sensors. Results are shown for all networks containing 2, 6 and 10 contaminant sensors. The mean and standard deviation of the difference between the performance with door position sensors and with door position sensors are listed on each frame (i.e.,  $\Delta \equiv$  probability with door – probability without door).

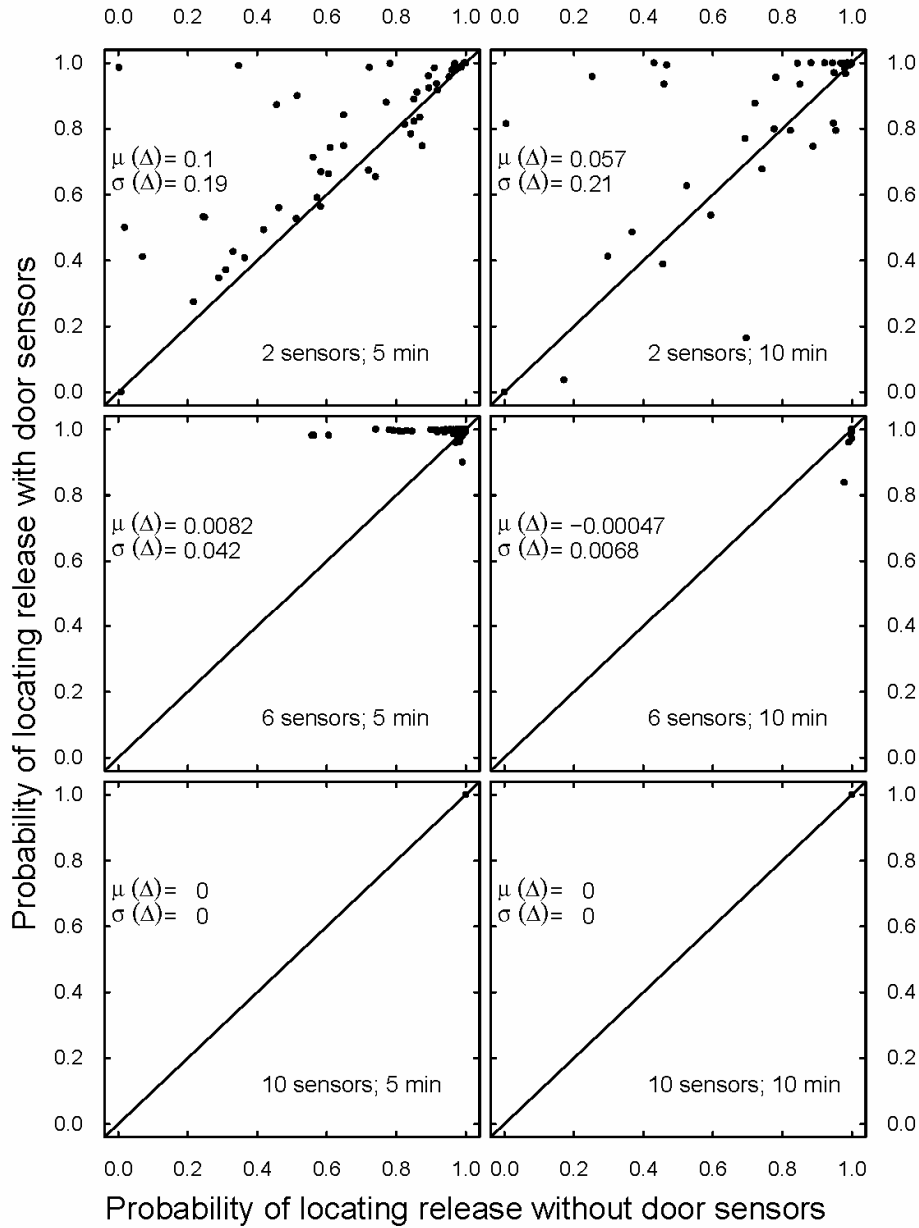


Figure 6.9. Experiment 1: probability of correctly locating either return intake or Room 1.2a as the release location, with and without door sensors. Results are shown for all networks containing 2, 6 and 10 contaminant sensors.

Further analysis suggests that model-to-measurement error in some zones may explain the negative results. When door-position information is fused, the algorithm must select from a fewer number of realizations that match the measured concentrations.

The few anomalous cases where door-position sensors reduce the ability of the system to identify the release location, even using the less strict metric, may represent network placements for locations with greater model-to-measurement error.

Networks containing 6 chemical sensors show similar results to networks with 2 chemical sensors: door sensors facilitate source identification for a majority of networks and impair a smaller subset of networks, particularly when applying the stricter metric (Figure 6.8). Here, also, the positive effect may arise from a reduced library set and negative impact from model-to-measurement error or poor sensor placements. When door-position information is fused, the algorithm favors Room 1.2 as the true release location, possibly because the post-fusion library contains fewer realizations that match the measured concentrations for a release originating in the return duct. In the case of 10 sensors, there is little-to-no impact from using door sensors. The results for the ten-sensor networks suggest that, as expected, for networks that otherwise easily and quickly locate the contaminant origin, door-position sensors are unnecessary.

Figure 6.10 and Figure 6.11 show the estimation of the median mass and the uncertainty reduction for the same networks. The uncertainty reduction is defined as the relative reduction of the 2-sided 80% confidence interval of the estimated mass:

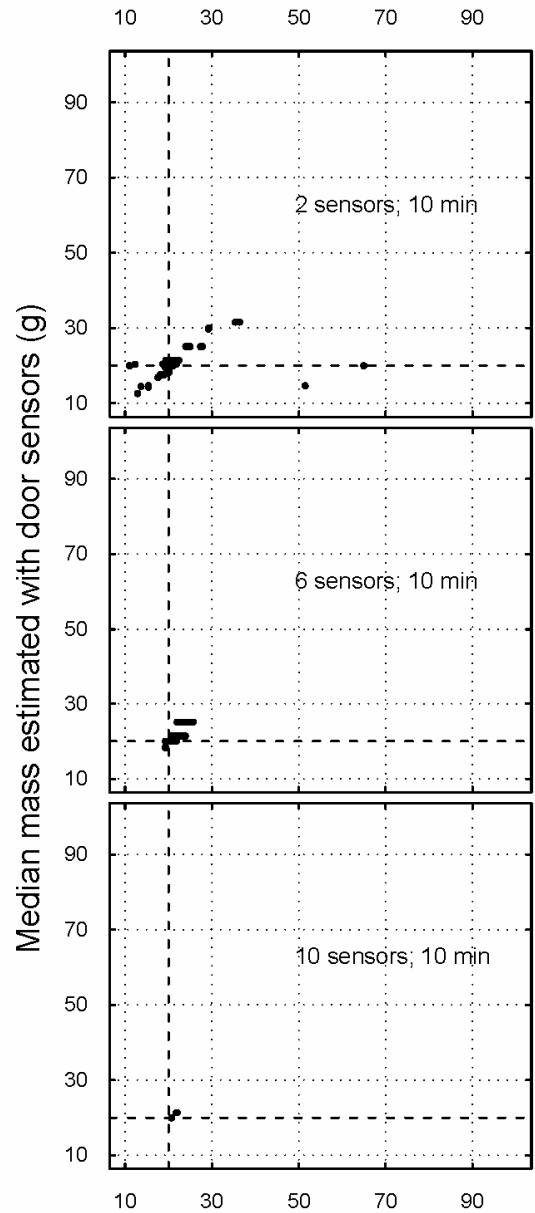
$$UR(t) = \left( 1 - \frac{q_{Mass,90\%}(t) - q_{Mass,10\%}(t)}{q_{Mass,90\%,prior} - q_{Mass,10\%,prior}} \right) \times 100\% \quad (6.2)$$

For networks containing 6 and 10 sensors, there is little effect of adding door-position sensors. Both the median mass and the uncertainty reduction are nearly unimpacted after 10 minutes of updating. This result is consistent with the notion that door positions have minimal impact for the release conditions of the experiment. Furthermore, based on the mass-balance principle, the task of estimating the mass

released is less demanding than identifying the source location. Thus, in meeting this objective, there may be little benefit from the reduced library that is achieved with fusing information from door-position sensors.

For most networks containing 2 sensors, there is little difference in the ability to estimate the median mass with and without door sensors. However, networks with door sensors are able to estimate the mass to a more narrow confidence interval (by  $\sim 7\%$ ). With only 2 sensors, the task of characterizing any source parameter becomes challenging. It is logical that adding door-position sensors can improve source characterization in this case. It does so through narrowing the post-fusion library from 5000 simulations to approximately 200.





Median mass estimated without door sensors (g)

Figure 6.10. Experiment 1: median mass for networks with and without door-position sensors. Results shown for all networks containing 2, 6 and 10 contaminant sensors after 10 min of updating.

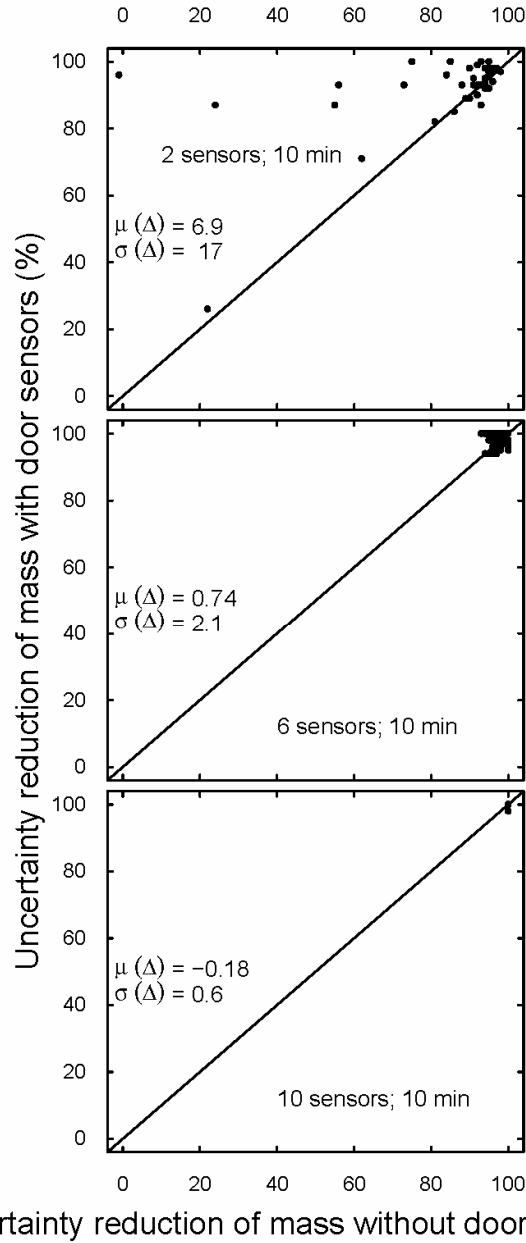


Figure 6.11. Exp 1: Effect of door-position sensors on uncertainty reduction of mass estimate. Results shown for all networks containing 2, 6 and 10 contaminant sensors after 10 min of updating. The mean and standard deviation of the difference between the performance with door position sensors and with door position sensors are listed on each frame (i.e.,  $\Delta \equiv$  uncertainty reduction with door – uncertainty reduction without door).

The following figures compare the performance of networks based on Experiment 4 data. The results, here, are quite different compared to the results for Experiment 1. The

fusion of door sensor positions prevents many networks with two or six sensors from correctly identifying the release location to a high probability. The median mass results are impacted as well by the presence of door sensors for some network configurations. Networks with two or six contaminant sensors and door sensors estimate the mass to a more narrow confidence interval. However, the median mass predictions suggest that these narrow confidence intervals do not bound the actual release mass.

Figure 6.12 shows the probability of correctly identifying the release room in five or ten minutes for networks containing 2, 6 and 10 contaminant sensors. The results for Experiment 4 differ from those describing Experiment 1. Here, the effects of door-position sensors vary from having no impact to a negative impact for networks containing 2 and 6 contaminant sensors. Networks containing 10 sensors are insensitive to door-position sensors.

The networks that show a “negative” impact from door sensors warrant some explanation. Analyses revealed that these negative results are attributable to algorithm and model complications. The multizone model errors are significant for this experiment, particularly in two specific zones. As discussed in Chapter 5, further effort is required to develop a likelihood function that can treat more accurately the model specification error for releases with longer intra- and interzonal mixing times. Because the library already contains few if any realizations with the true parameter set that closely match the data, the smaller library that results from door sensor fusion contains even fewer accurate matches.

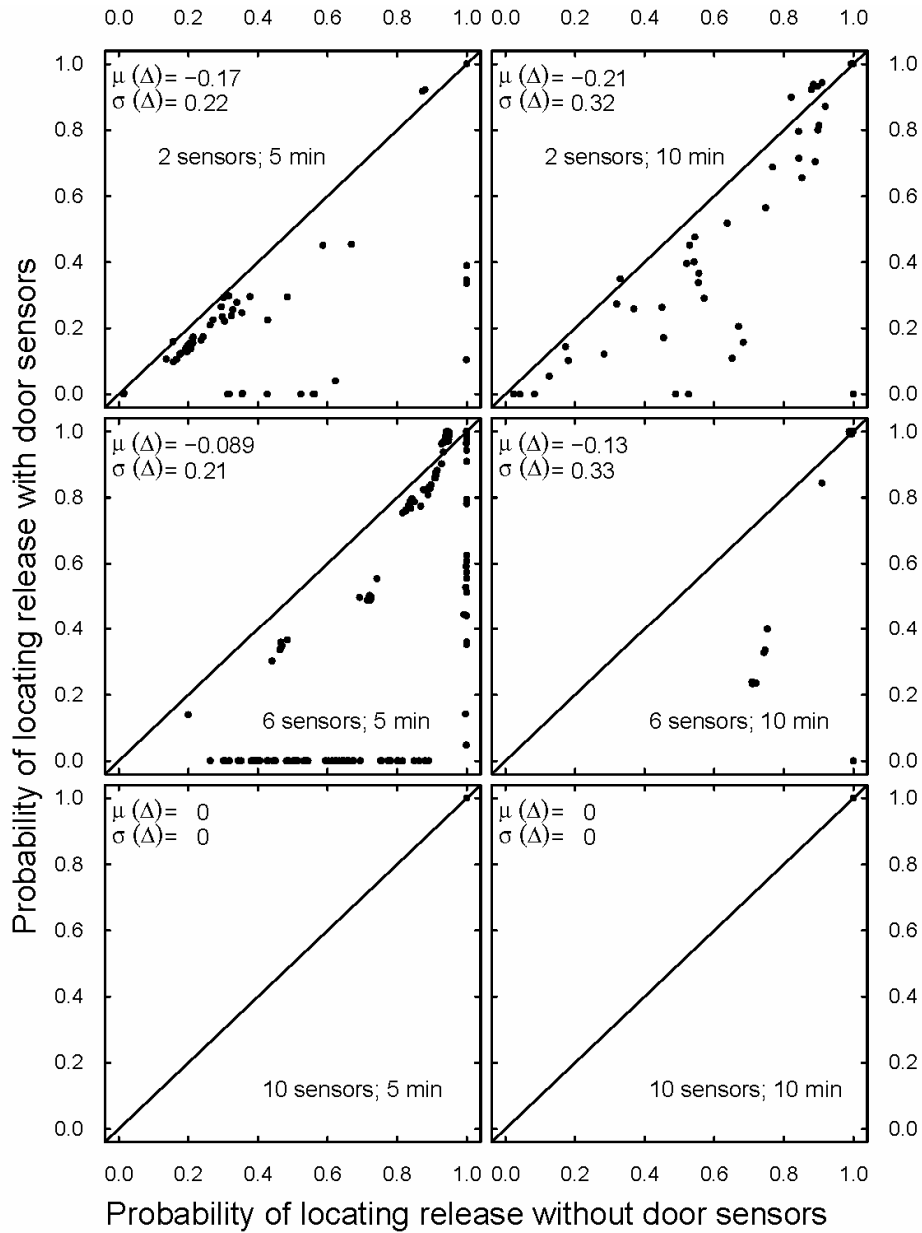
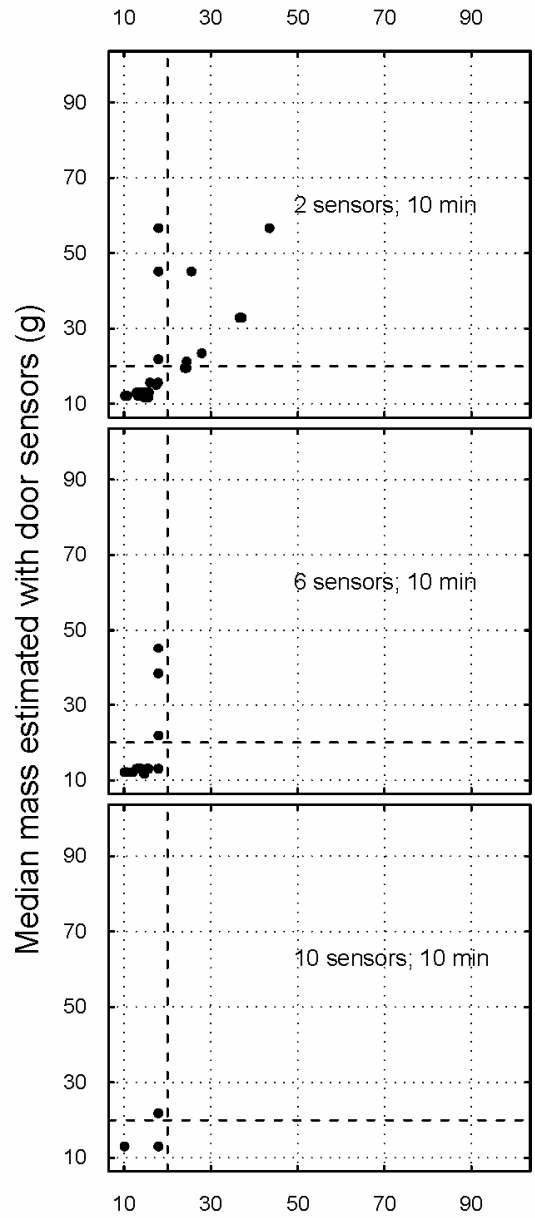


Figure 6.12. Experiment 4: probability of locating release with and without door sensors. Results are shown for all networks containing 2, 6 and 10 contaminant sensors.

Figure 6.13 and Figure 6.14 show the median mass and uncertainty reduction of the mass estimate for the same networks. For networks with 10 contaminant sensors there is little difference between the median mass estimates for networks with and without door sensors. The results for networks with 2 and 6 sensors are slightly different. Here, a few

networks estimate a higher median mass (true mass is 20 g) for networks with door sensors. This result is not surprising, given the algorithm and model complications cited earlier. That is, the negative consequences of model specification error are amplified for networks with door sensors because of the reduced-size library that results.

Door-position sensors have no impact on the mass uncertainty reduction for networks with 10 contaminant sensors. This result is consistent with the estimation of the source strength and origin. Networks with 2 and 6 sensors show slightly different results. Here, the presence of door sensors improves the uncertainty reduction especially for two-sensor networks – that is, networks with door sensors estimate narrower 80% confidence intervals of the mass. Here, also, model specification error may account for these results. In cases where the errors are large, a small subset of realizations are likely to dominate the results. This outcome leads to a greater uncertainty reduction, even if the distribution is not centered around the correct mass (as is the case for several networks).



Median mass estimated without door sensors (g)

Figure 6.13. Experiment 4: median mass for networks with and without door sensors. Results are shown for all networks containing 2, 6 and 10 contaminant sensors after 10 min of updating.

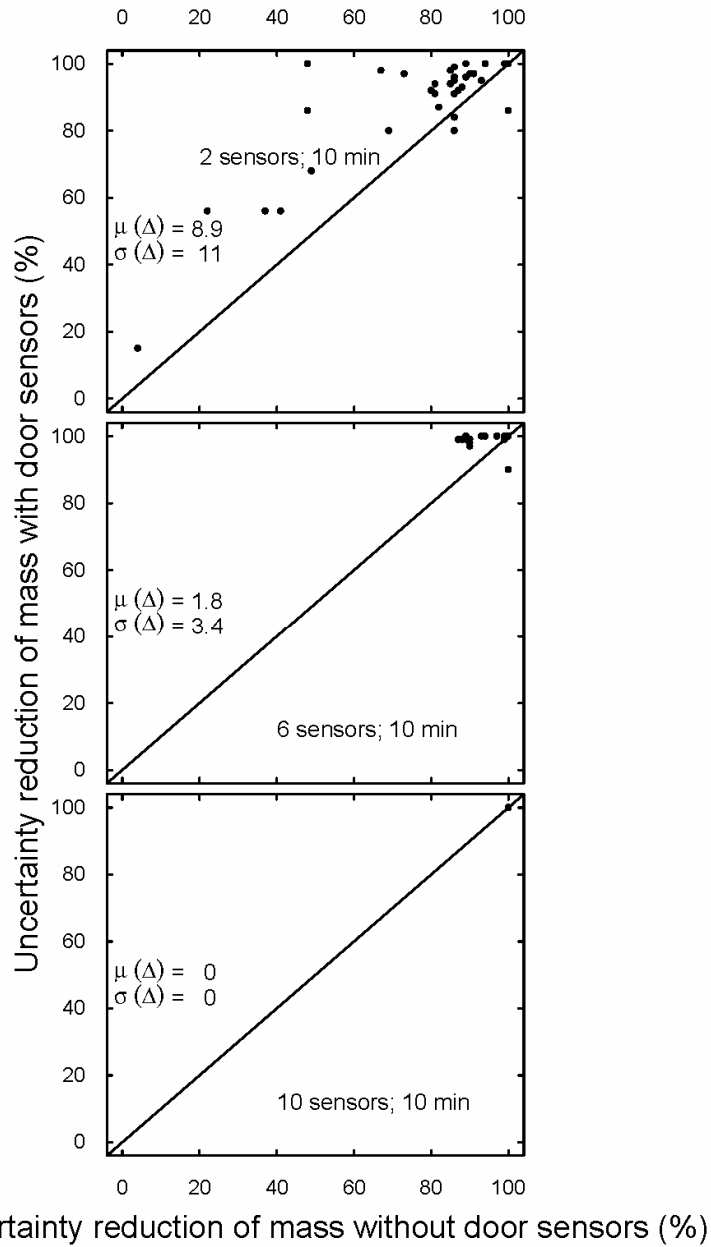


Figure 6.14. Experiment 4: effect of door sensors on uncertainty reduction of mass estimate. Results are shown for all networks containing 2, 6 and 10 contaminant sensors after 10 min of updating.

The results for Experiment 4 illustrate research opportunities for improving the likelihood function to better represent model errors. However, the complications from this particular data set and limitations in the modeling tools are obstacles to thoroughly

exploring multisensor fusion using door-position sensors. Consequently, simulated data are used to continue the investigation in the next section.

The results of this section indicate benefits towards identifying the release location for some networks based on Experiment 1 but not Experiment 4. In fact, the door sensors had a distinctly negative impact towards locating the release for Experiment 4. While the results for Experiment 1 are reasonable (some benefit for many networks, minimal negative impact), the results for Experiment 4 defy physical intuition, as door sensors were expected to be slightly more beneficial for Experiment 4 than for Experiment 1. Modeling errors and data idiosyncrasies seem to obscure the potential benefit from using door-position sensors. Overall, the investigations reported in this section have failed to identify physically meaningful conditions for which door-position sensors may be valuable. The next section strives to understand if there are scenarios, with mechanical airflow, for which door sensors yield a clear improvement for sensor-system performance.

#### **6.4.3.3 Generation of simulated data, likelihood function and updating**

This section further investigates the value of door-position sensors in mechanically driven airflow conditions, using simulated data. Data were generated to replicate (a) the release conditions of Experiment 4 and (b) to simulate a pulse release originating in Stairwell 2.

To generate simulated data, the base COMIS model (used to generate the library) was first exercised to model spatially and temporally varying concentrations, following an instantaneous releases from either Room 1.3 or from Stairwell 2. For the release originating from Room 1.3, the stairwell doors were closed and all others were open



(consistent with the conditions during Experiment 4). For the release originating from Stairwell 2, all doors were kept open.

Next, the modeled concentrations were perturbed to simulate sensor error. The sensor error was assumed to follow a lognormal distribution. The statistical distribution of the measurement,  $X$ , is indicated by equation 6.3

$$f(X) = \frac{1}{\sqrt{2\pi} \ln(GSD)X} \exp\left\{-\frac{1}{2}\left(\frac{\ln X - \ln(GM)}{\ln(GSD)}\right)^2\right\} \quad (6.3)$$

where the geometric mean, GM, and geometric standard deviation, GSD, are set to the modeled concentration, and to  $e^{0.2} = 1.22$ , respectively. For each model prediction, a random variate was generated by sampling from  $f(X)$ . In cases where the modeled concentration was 0, the sensor measurement was also set to 0. The perturbed concentrations were adjusted in two additional ways: a sensor resolution and detection limit comparable to those in the original tracer data were imposed on the final simulated data set. Figure 6.15 and Figure 6.16 show the simulated data and idealized model concentrations for both release conditions.

Because the likelihood function quantifies the model-to-measurement error, equation 6.3 also serves as the appropriate basis for the likelihood function in these investigations. The likelihood function,  $p(O_i|Y_{k,i})$ , given “measurement”  $O_i$  and modeled concentration  $Y_{k,i}$  is calculated using equation 6.4.

$$p(O_i | Y_{k,i}) = \frac{1}{\sqrt{2\pi} \ln(GSD) \times O_i} \exp\left(-\frac{1}{2}\left(\frac{\ln O_i - \ln Y_{k,i}}{\ln(GSD)}\right)^2\right) \quad (6.4)$$

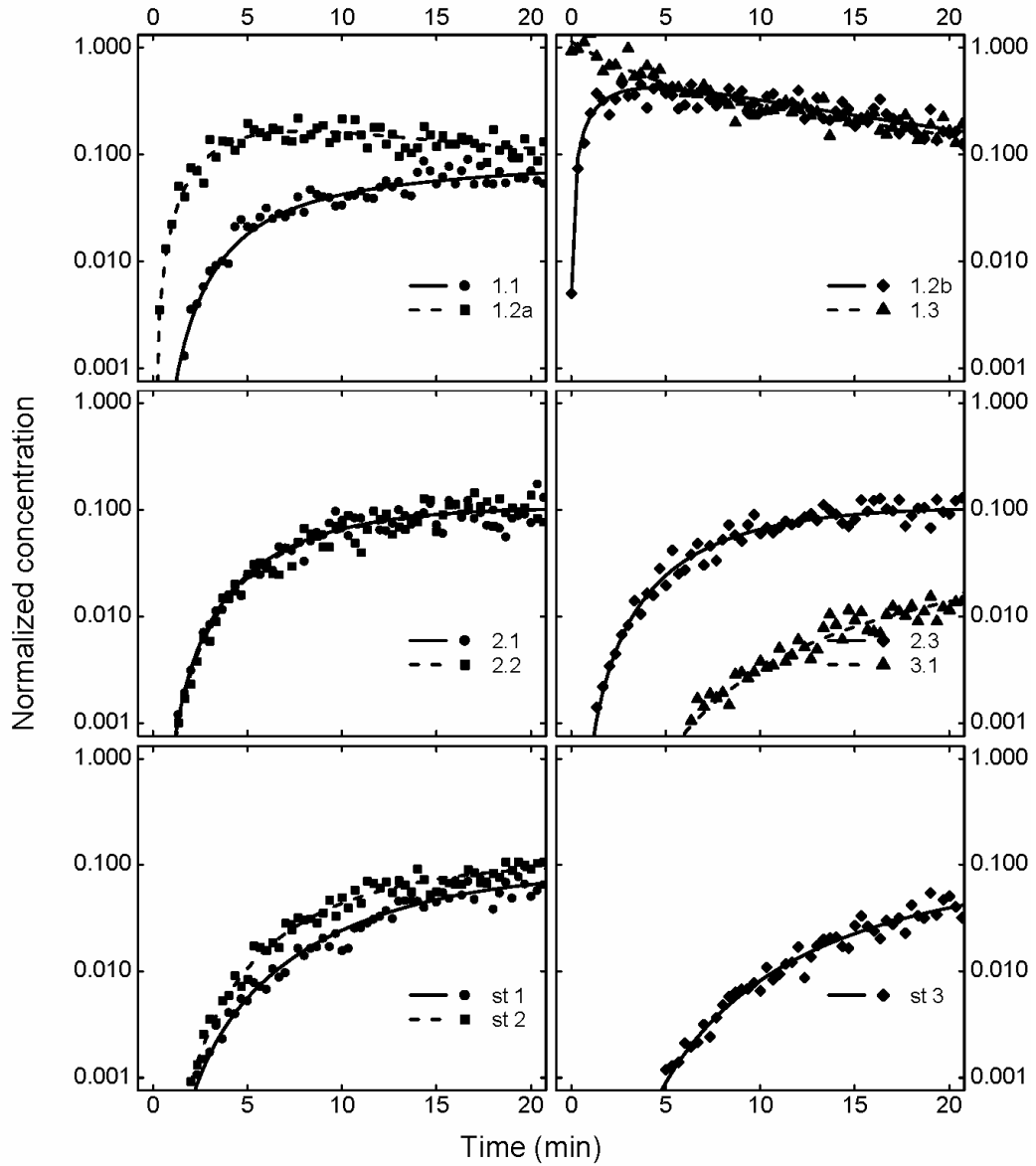


Figure 6.15. Simulated data for Experiment 4 conditions with modeled concentrations. Concentrations are normalized by theoretical peak concentration in the experiment; time is in reference to the time of the release; upper frame: first floor rooms; middle frame: second floor and third floor rooms; lower frame: staircases; 20 g of an inert tracer was released in Room 1.3 at time = 0.

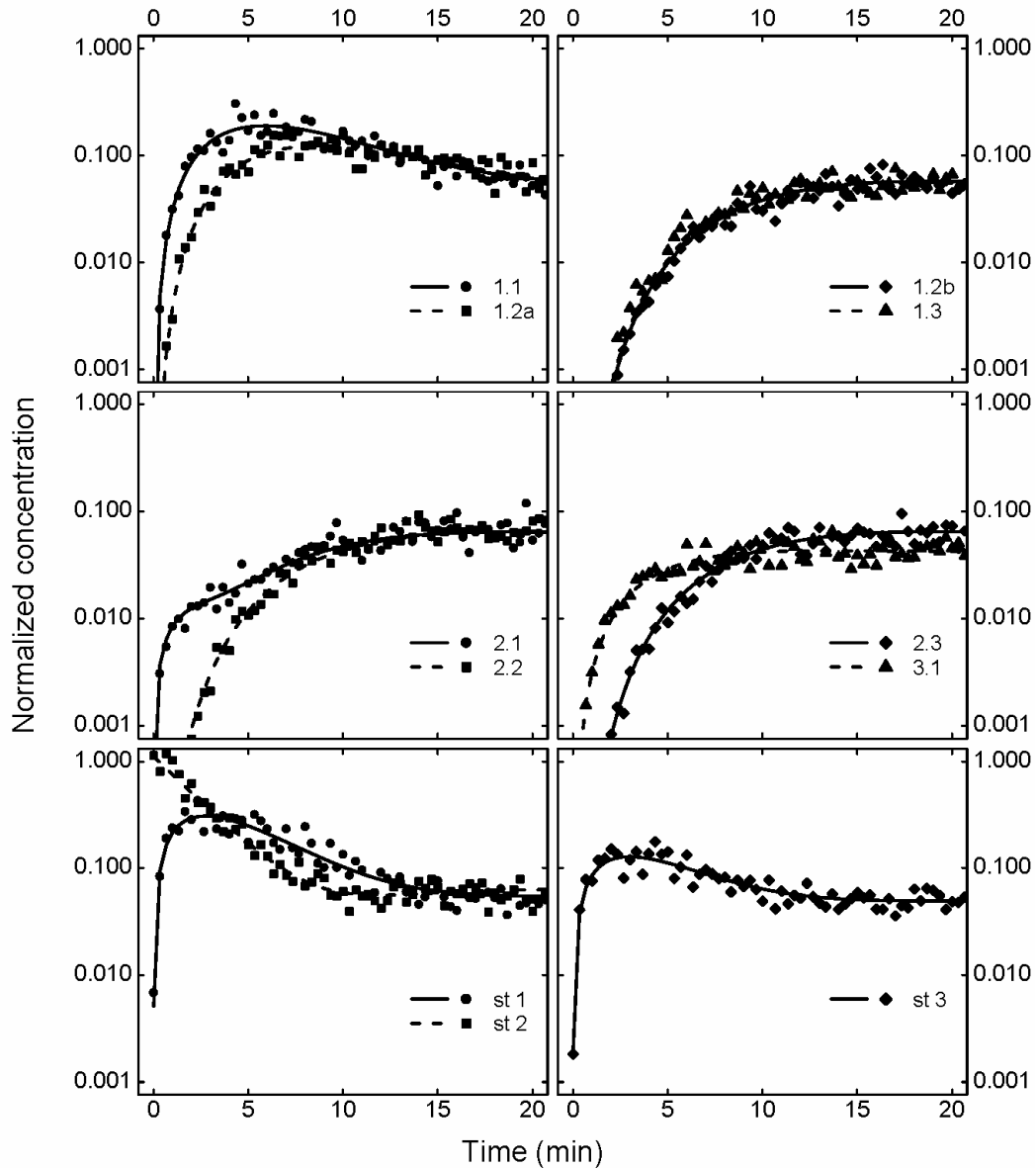


Figure 6.16. Simulated data and modeled concentrations for a 20 g release in Stairwell 2.

The Bayesian updating was adjusted to accommodate cases where either the modeled or sensor concentrations were 0. In both these cases, the values were set to the sensor detection limit. An outcome from this adjustment is that the algorithm has less discriminating power among realizations that predict very small concentrations. This effect happens to be a positive one for the sensor system. One potential problem

encountered with the BMC algorithm is that it can prematurely discriminate among different library realizations. The adjustment employed here serves to equalize the posterior probabilities of realizations that predict small concentrations.

Door-position sensor fusion was implemented as described in Section 6.4.1. In these investigations, the parameters  $\delta$  and  $\gamma$  were both set to 1, which signifies that the door-position sensors were assumed to be perfect and the maximum amount of library prior filtering was performed.

As in the previous analyses, the release time was treated as an unknown parameter.

#### **6.4.3.4 Results using simulated data**

The simulated data were used to evaluate hypothetical sensor networks ranging from 1 to 11 chemical sensors with and without door sensors. Based on the degree of influence concept, door sensors were expected to be nominally beneficial towards characterizing the release from Room 1.3 and more useful for the Stairwell 2 release.

Figure 6.17 compares the performance of networks containing 3, 5, 7 and 9 contaminant sensors for correctly locating the release that originated in Room 1.3. In many cases the presence of door sensors yields a significant penalty as exhibited by the numerous points that fall below the diagonal line. In some cases, there is little difference. The penalties may occur because the library is reduced to a small subset by filtering with the door sensors. If the parameter has a small degree of influence on contaminant transport, as is the case in this example, then removal of the parameter may eliminate realizations that otherwise help to characterize the release more rapidly.

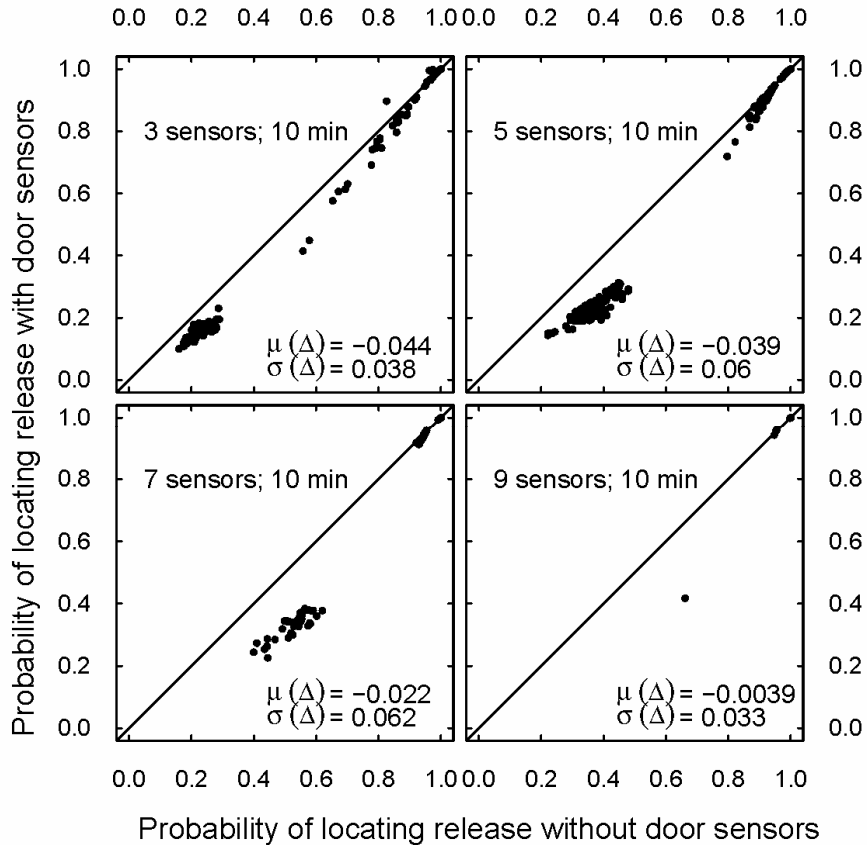


Figure 6.17. Effect of door-position sensors for identifying release location for Exp 4 simulation. Probabilities of correctly identifying release location after 10 min of updating are shown for networks containing 3, 5, 7 and 9 contaminant sensors.

A related explanation for the nominal benefits or penalties, for these release conditions, is that a random sampling peculiarity exists within the library for releases originating in Room 1.3. This possibility was explored by regenerating the library using a newly generated random set of simulation parameter combinations. The evaluations for all networks containing two sensors were repeated (Figure 6.18). Results for two-sensor networks using the original library are included for comparison purposes. The comparison suggests this hypothesis is true.

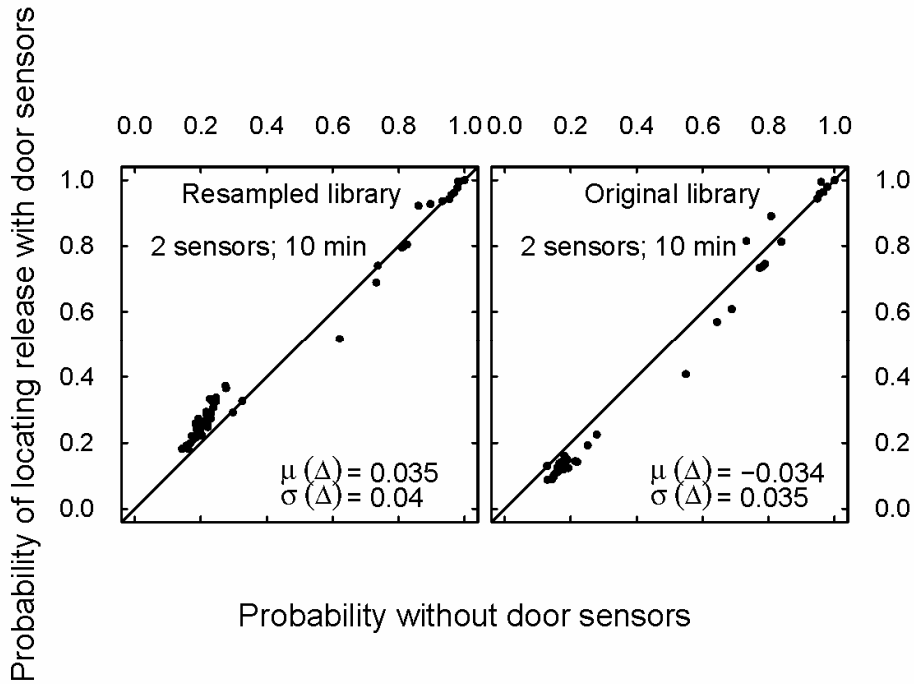


Figure 6.18. Reevaluation of simulated data from Exp 4 using a newly generated library. Probability of finding release location after 10 min of updating for networks containing 2 contaminant sensors. Comparable results using original library are included for reference.

Accurate quantification of the release mass is also an important task of the sensor system. The presence of door sensors has little impact on the median value of the estimated mass, as shown in Figure 6.19. As noted in previous chapters, the mass is often easier to determine than the release location, in part because the information pertaining to the mass is less dependent on sensor placement, unlike locating the source. Also, information needed to estimate the release mass is available up to the removal time scale for the building, whereas the location fingerprint typically fades on a typically faster time scale. Door-position sensors may bring less value to the task of quantifying released mass, as it is a relatively easy task for many networks.

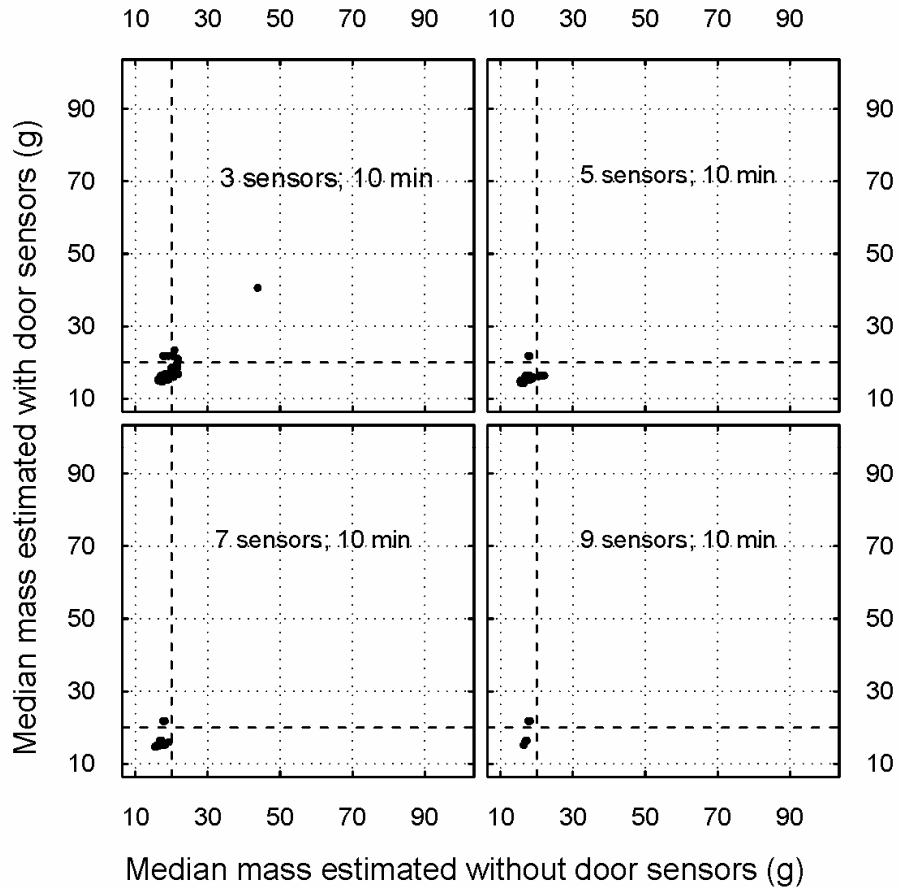


Figure 6.19. Effect of door-position sensors for quantifying the mass released for the Experiment 4 simulation. Shown are results for all networks containing 3, 5, 7 and 9 contaminant sensors after 10 min of updating.

Figure 6.20 shows the reduction in the 80% 2-sided confidence intervals of the released mass. In many cases the networks with door-position sensors estimate the mass to a narrower confidence interval than those networks without door sensors. This effect can be understood through considering the reduced library size that results from fusing information from door-position sensors. A smaller posterior library size means that the realizations that fit the data well are assigned proportionally higher probabilities, relative to the case for the full library.

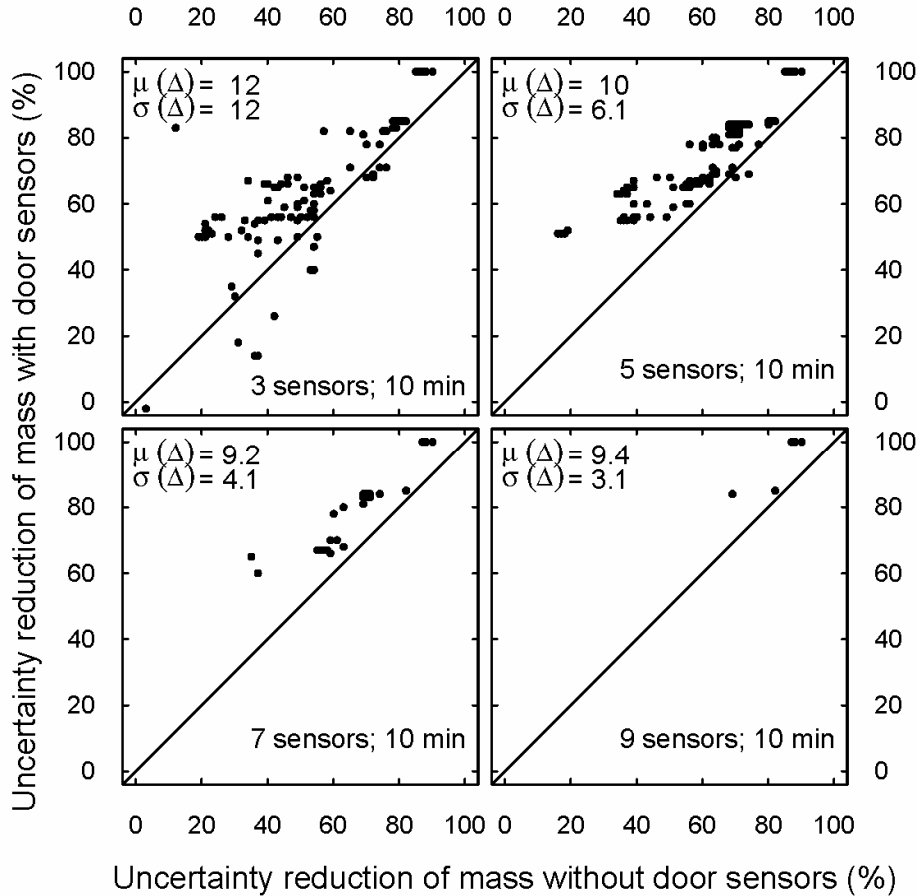


Figure 6.20. Experiment 4 simulation: effect of door-position sensors on uncertainty reduction of mass released. Results are shown for all networks containing 3, 5, 7 and 9 contaminant sensors after 10 min of updating.

In summary, the results for the case of a release originating in Room 1.3 are unsurprising. The relatively low degree of influence of the doors with respect to contaminant transport translates into low information value added by door-position sensors. The cases where there are nominal performance advantages or disadvantages represent more of a statistical fluctuation than a systematic improvement.

To explore a case in which door-position sensors are expected to be more valuable, a second pulse release was simulated, in this case originating in Stairwell 2, with all interior doors open (i.e., stairwell doors and doors between the rooms). Because the



stairwell is not actively fed by the AHU, the contaminant is spread to the remainder of the building by spreading through the vertical stairwell shaft and then passing into the rooms through the stairwell doorways. Thus, the position of the doors, particularly the stairwell doors, is expected to have a significant influence on the rate at which the contaminant disperses.

The simulated data were used to generate several hypothetical sensor networks ranging from 1 sensor to 11 sensors. Each of these networks were investigated with and without door sensors, in which all door sensors were included.

Figure 6.21 compares the ability of the sensor system to identify the release location at 5 min and 10 min following the release, with and without door-position sensors. For many networks, door sensors help the system to identify the release location to a higher probability at each time, as indicated by the general tendency of the points to lie above the diagonal lines in these frames. Thus, in this case, for an equivalent amount of chemical sensor information, systems with door sensors perform slightly better on average. On the other hand, for some networks, door sensors impose a slight penalty. These networks appear to be those with poor sensor placements, as indicated by their relatively low probabilities in the “without door sensors” cases. Thus, door sensors may not be able to compensate for the negative effects of poor placement, such as for those networks with no sensors in close proximity to the release source.

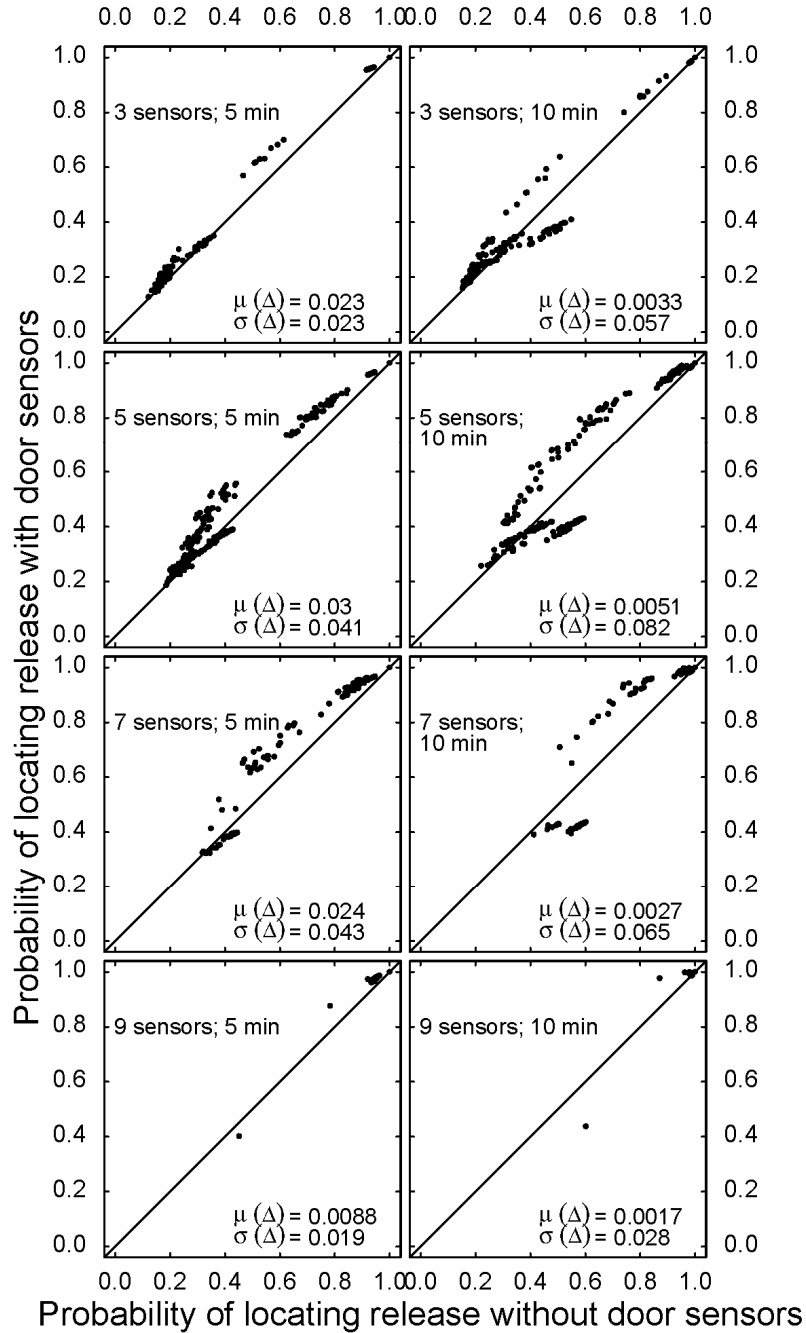


Figure 6.21. Effect of door sensors for correctly identifying Stairwell 2 as the site of the contaminant release, using simulated event data. Probabilities are plotted for 5 and 10 min of updating for networks containing 3, 5, 7 and 9 contaminant sensors.

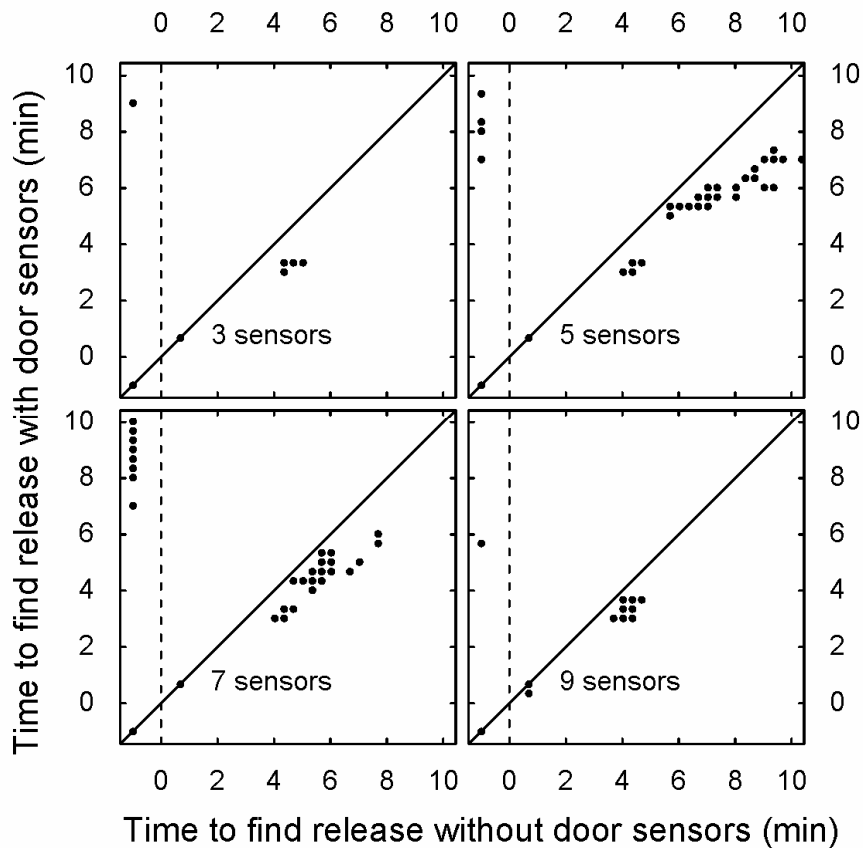


Figure 6.22. Time to locate release to 0.9 probability with and without door-position sensors for a simulated release event in Stairwell 2. Results are shown for networks containing 3, 5, 7 and 9 contaminant sensors. A negative time indicates that a probability of 0.9 was not reached by 11 min.

Identifying the release location quickly to a high degree of certainty is an important task of the sensor system. If door-position sensors can help with this task, then they bring value to the system. Figure 6.22 shows that the networks incorporating door-position sensors are able to locate the release to a probability of 0.9 more rapidly than those networks without door-position sensors, on a consistent basis and sometimes several minutes more quickly. Some sensor systems with door-position sensors are able to identify the release with a probability of 0.9 in 6-10 min while systems without door-position sensors are unable to do so within 11 min. Identifying the release location

in a timely manner is important because information pertaining to the release location erodes as the elapsed time from the onset of an event approaches and then exceeds the interzonal transport time scale.

There is little difference in the median estimates of mass released between the two groups of sensor systems (Figure 6.23). However, systems using door sensors are able to estimate the released mass to a more narrow confidence interval (~5-11% as shown on Figure 6.24), and more quickly. In this case, for equivalent deployment of chemical sensors, systems using door-position sensors are able to characterize the release more quickly and with higher confidence.

The results for the Stairwell 2 release are not surprising because the dispersion of the contaminant, in this example, is influenced by the stairwell door positions. By comparing the results between the release originating in Stairwell 2 and Room 1.3 it is demonstrated that the door position sensors are more useful in conditions for which they exhibit a higher degree of influence on contaminant transport.

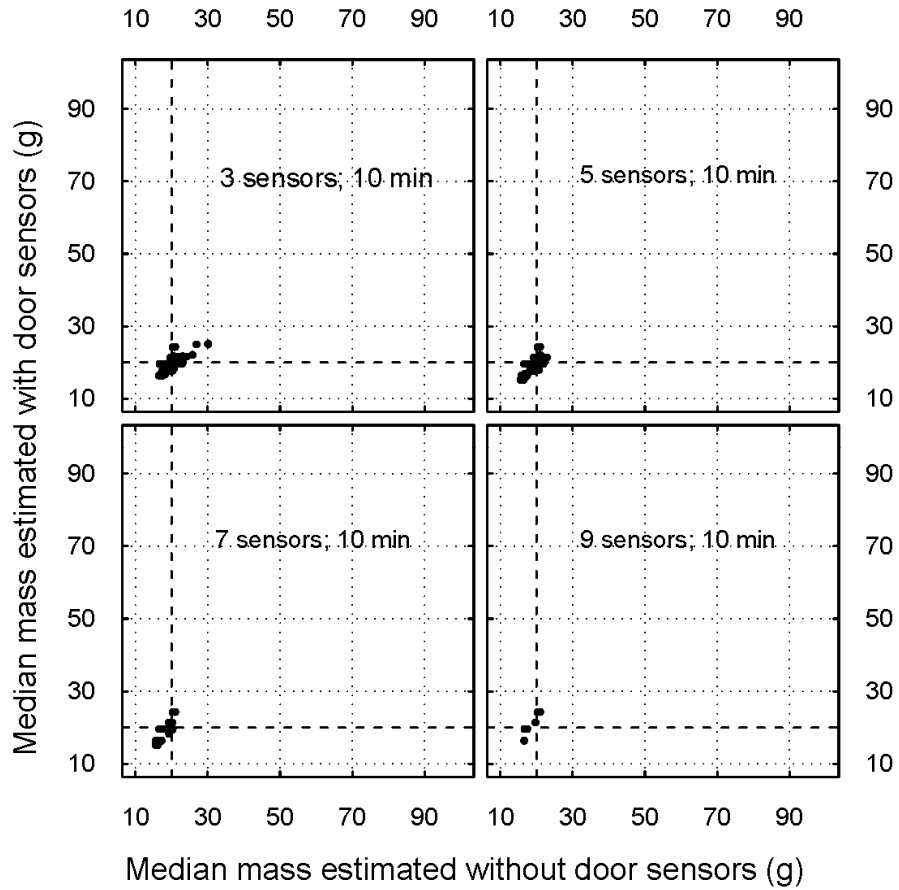


Figure 6.23. Effect of door-position sensors for estimating the mass released for the simulated case of a release into Stairwell 2. Results are shown for all networks containing 3, 5, 7 and 9 contaminant sensors after 10 min of updating.

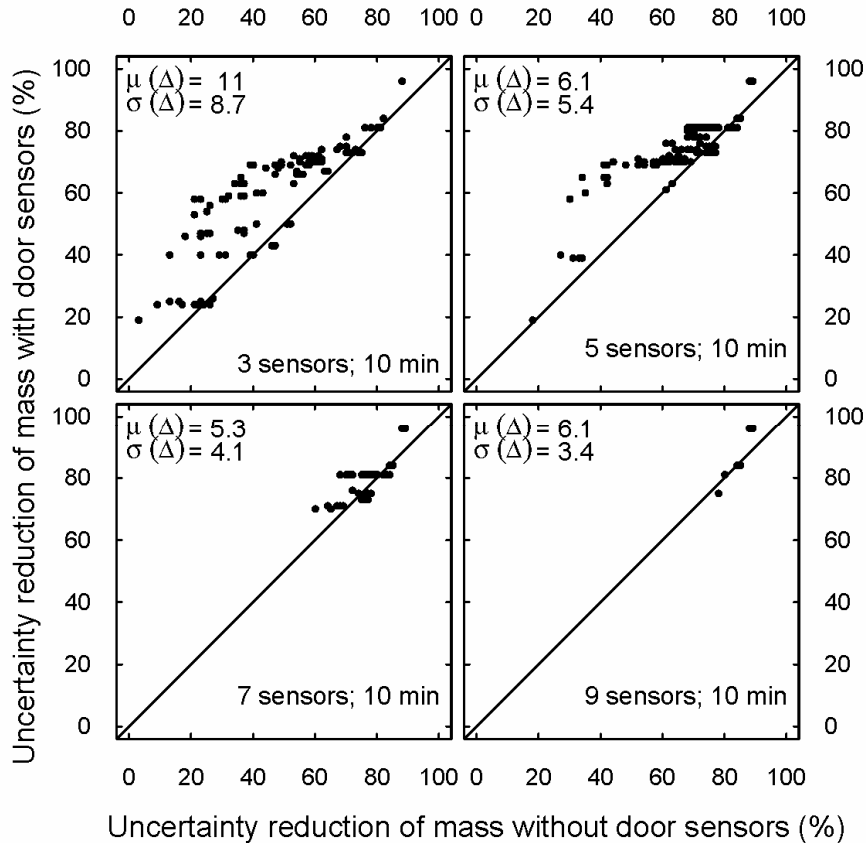


Figure 6.24. Effect of door-position sensors on reducing uncertainty in the estimate of mass released for the simulated case of a release into Stairwell 2. Results are shown for all networks containing 3, 5, 7 and 9 contaminant sensors after 10 min of updating.

The simulations reported in this section indicate that placement of the contaminant sensors is critical and more important than the presence or absence of door sensors. In fact, there are many cases where a network with fewer sensors, both with and without door sensors, can more effectively characterize the release, than a more fully instrumented network. Door-position sensors expedite characterization of the release event, but do not resuscitate an otherwise poorly performing network.

#### **6.4.4 Naturally driven airflow conditions**

This section investigates the utility of door-position sensors for airflow conditions in which there is no mechanically driven airflow. Door-position sensors are expected to be more useful here, because, in the absence of forced flow through ducts, the door positions would have a greater influence on contaminant transport. Owing to the limitations of the experimental data, this investigation uses simulated data. (Only one experiment at the test facility was conducted with the mechanical unit off.) Simulated data were generated to replicate the releases investigated in Section 6.4.3.4 but with the AHU not operating. Additional releases were simulated to further investigate the effect of varying door positions, release location, and duration.

The simulated data sets are used to test the performance of hypothetical sets of networks with different numbers of contaminant sensors and placement, with and without door sensors. Additional analyses were conducted to further explore if the results from fusing door sensors can be explained physically.

##### **6.4.4.1 Library, data generation and likelihood function**

A library was generated to represent possible release conditions with the AHU off. The procedure outlined in Section 6.4.3.1 and Table 6.2 was followed with one difference: the AHU was off in the base COMIS model.

Incidentally, the state of the AHU represents an additional opportunity for sensor fusion. In a real implementation, the library might contain two large subsets of simulations, respectively representing conditions with and without mechanical airflow.

Because the AHU operating status is an input signal that could be accessed through the control system, it could be used in a straightforward manner to sort the library.

Hypothetical release data were generated following the procedure outlined in Section 6.4.3.3. Table 6.3 summarizes the simulated release conditions of this section. Figures illustrating these simulated release events are presented in Appendix 6.A.

In each case, 20 g of contaminant were released. “Tier 1” replicates the release conditions from Section 6.4.3.4 and includes two additional sets of door configurations. In Tier 1, I evaluate many sensor network configurations using a combinatorial approach. “Tier 2” investigates additional release scenarios with different release locations, door positions, and release duration. In Tier 2, a smaller subset of networks are evaluated.

Table 6.3 Description of simulated release conditions with AHU off.

Simulated experiment number	Release location	Door position <sup>a</sup>	Release duration
“Tier 1”			
AHUoff-1	Room 1.3	Stairwell doors closed, other interior doors open	1 s
AHUoff-2	Stairwell 2	All doors open	1 s
AHUoff-3	Room 1.3	All doors closed	1 s
AHUoff-4	Stairwell 2	Stairwell doors closed, other interior doors open	1 s
“Tier 2”			
AHUoff-5	Room 1.1	Stairwell doors closed, other interior doors open	1 s
AHUoff-6	Room 2.2	All doors open	1 s
AHUoff-7	Stairwell 3	All doors open	1 s
AHUoff-8	Room 2.2	All doors open	120 s

<sup>a</sup> Door position information applies to interior doors and stairwell doors. The exterior door on Floor 1 was modeled as “closed” in all cases.

The likelihood function follows the description in Section 6.4.3.1.



#### 6.4.4.2 Tier 1 release scenarios

All possible networks containing three and eight contaminant sensors are evaluated, without and with a full set of door-position sensors. In all simulations, the door fusion algorithm assumed full confidence and maximum filtering. That is, the door fusion algorithm was implemented with parameters  $\delta = \gamma = 1$ .

##### *Results of door sensor fusion*

The value of door-position sensors depends on both the release and airflow characteristics. With the mechanical unit off, door sensors appear to add little value towards characterizing releases that originate in Room 1.3 (AHUoff-1 and AHUoff-3). Figure 6.25 illustrates the ability of the networks to locate Room 1.3 as the source at 10 min after the release occurred. Door-position sensors variably exhibit nominal benefits and nominal penalties towards identifying the release location. The nominal benefits appear to be associated with networks that may otherwise successfully identify the release location.

The nominal penalties appear to be associated with networks characterized by poor placement (as exhibited by the relatively low probabilities) and may be due to two factors. The removal of realizations from door sensor fusion leaves a smaller subset of the library that may not sufficiently describe the release, for the particular network configuration. The second reason for the nominal penalties may be due to a sampling peculiarity of this particular library, as was described in Section 6.4.3.4 with the AHU operating.

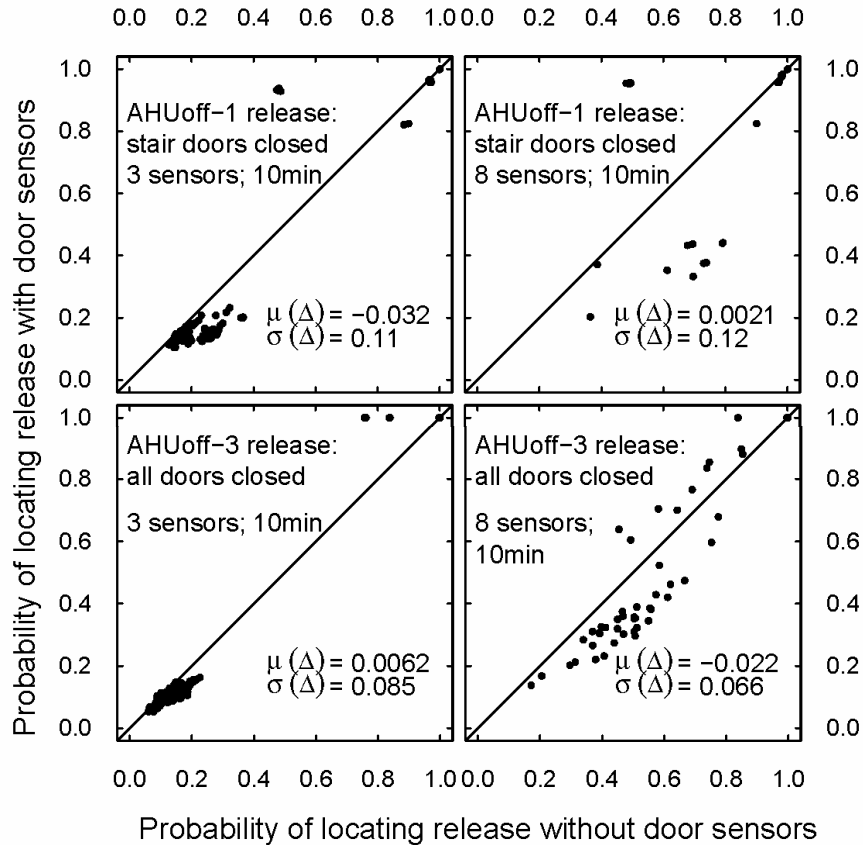


Figure 6.25. Effect of door-position sensors for correctly identifying the locations of releases that occurred in Room 1.3 (Experiments AHUoff-1 and AHUoff-3). Probabilities after 10 min of updating for networks containing 3 and 8 contaminant sensors are shown.

For both releases originating in Room 1.3 there are some cases in which networks locate the source to a probability of 0.9 more quickly with the door-position sensors than without (Figure 6.26). In these cases, the door-position sensors bring a computational and efficiency advantage. It is not entirely clear, however, if this advantage is a purely statistical one (owing to a reduced library), or if it results from an explicit dependence of transport on the door positions. Regardless, even if sensor fusion yields benefits primarily because it can be used to reduce the scale of the library of simulations that must be assessed in the second stage of the BMC algorithm, there can be significant

computational advantages and minimal penalty provided that the library has been generated properly.

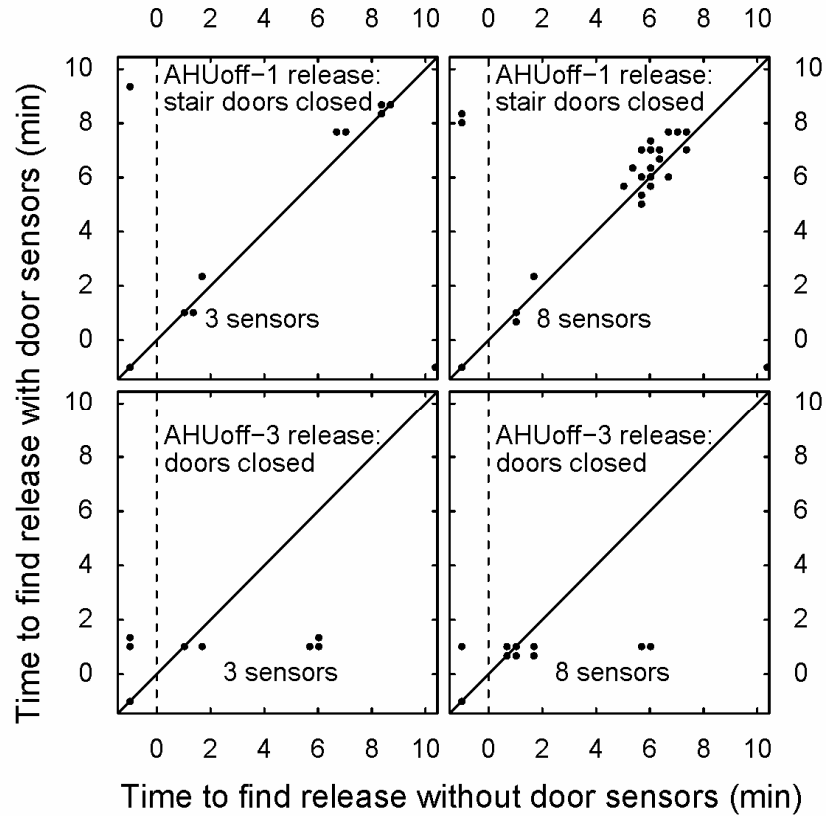


Figure 6.26. Time to locate release to a confidence of 0.9 or better for releases that occur in Room 1.3 (Experiments AHUoff-1 and AHUoff-3), for networks with and without door-position sensors. A negative time indicates that a probability of 0.9 was not reached by 11 min.

Figure 6.27 and Figure 6.28 show the estimation of the released mass for the releases originating in Room 1.3. Some networks augmented with door-position sensors are able to estimate the release mass to lower levels of uncertainty, particularly for the AHUoff-1 experiment (Figure 6.28). However, for many networks, the estimation of the mass at 10 min is characterized by relatively broad confidence intervals. Door-position sensors

might help reduce the overall uncertainty that describes the release by eliminating sets of realizations from the library.

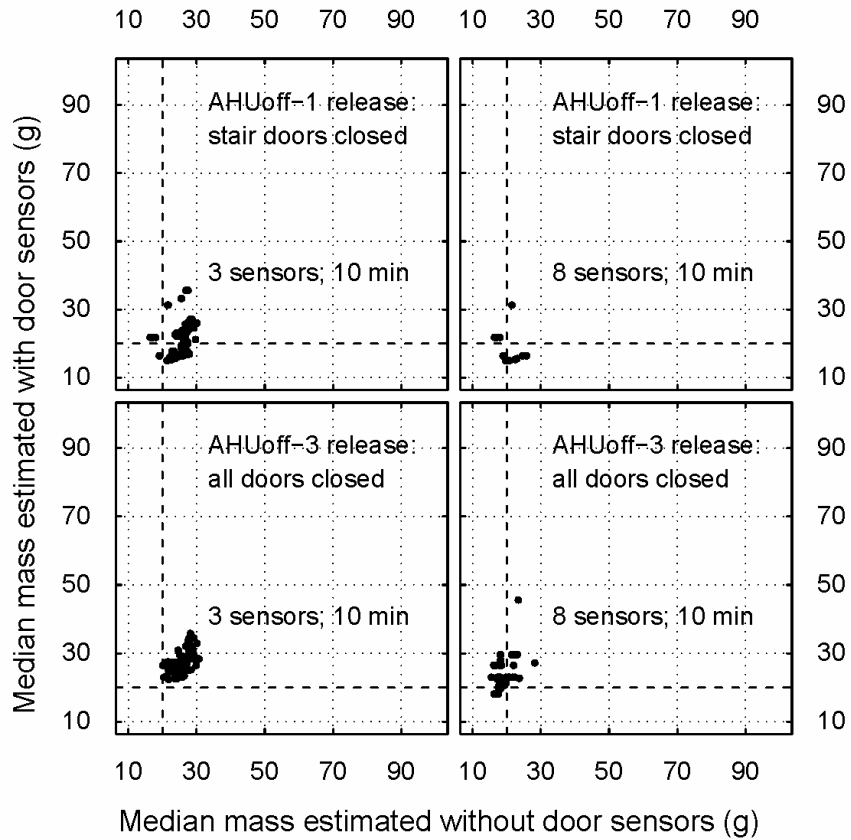


Figure 6.27. Effect of door-position sensors for estimating mass released for experiments AHUoff-1 and AHUoff-3. Results are shown for all networks containing 3 and 8 contaminant sensors after 10 min of updating.

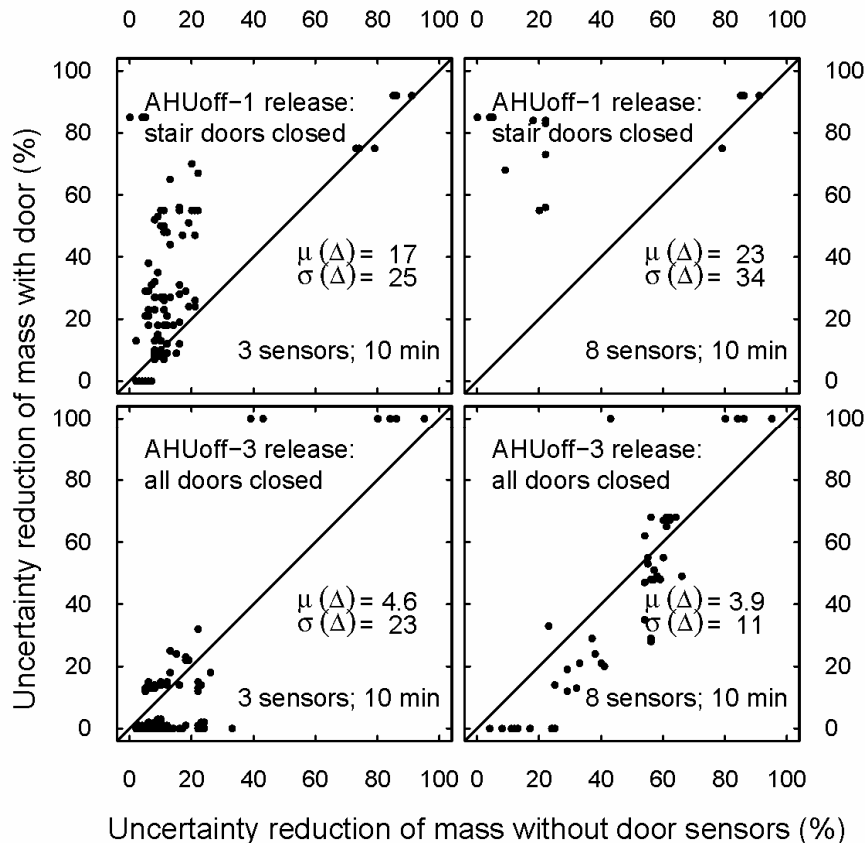


Figure 6.28. Effect of door-position sensors on reducing the uncertainty of the estimated mass release for experiments AHUoff-1 and AHUoff-3.

Door-position sensors may have genuine value in conditions for which the door positions significantly influence the overall contaminant concentrations, as is expected for the experiments AHUoff-2 and AHUoff-4, in which the releases occurred in Stairwell 2. In both cases door-position sensors help the networks to identify the release location more efficiently (i.e., with fewer chemical sensors) and quickly. For a given amount of chemical sensor information, networks with door sensors commonly identify the release location to a higher probability by  $\sim 2\text{-}3\%$  for eight-sensor networks and  $\sim 6\text{-}7\%$  for three-sensor networks, as shown in Figure 6.29. If a sensor-system goal is to locate the release to a probability of at least 0.9, then networks incorporating door-

position sensors have a clear advantage that is on the order of minutes (Figure 6.30). After more than 10 minutes of updating, many three-contaminant-sensor networks without door sensors do not identify the source to a 0.9 probability or better, while networks with door-position sensors consistently do.

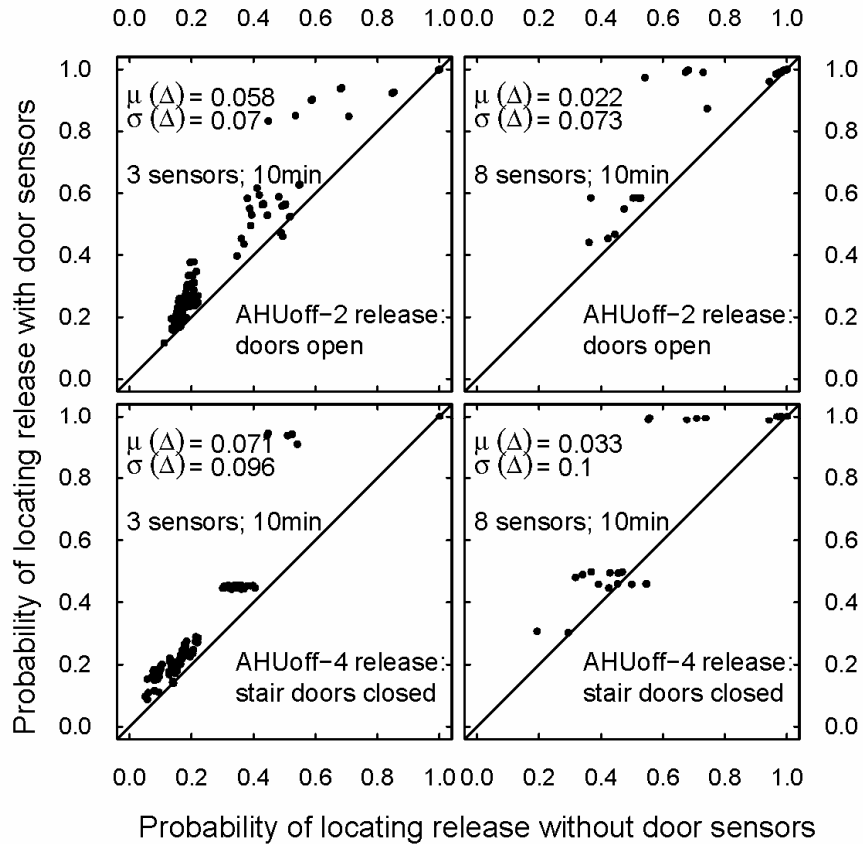


Figure 6.29. Effect of door-position sensors for identifying the location of releases that occurred in Stairwell 2 (Experiments AHUoff-2 and AHUoff-4). Probabilities are shown of being correct after 10 min of updating for networks containing 3 and 8 contaminant sensors.

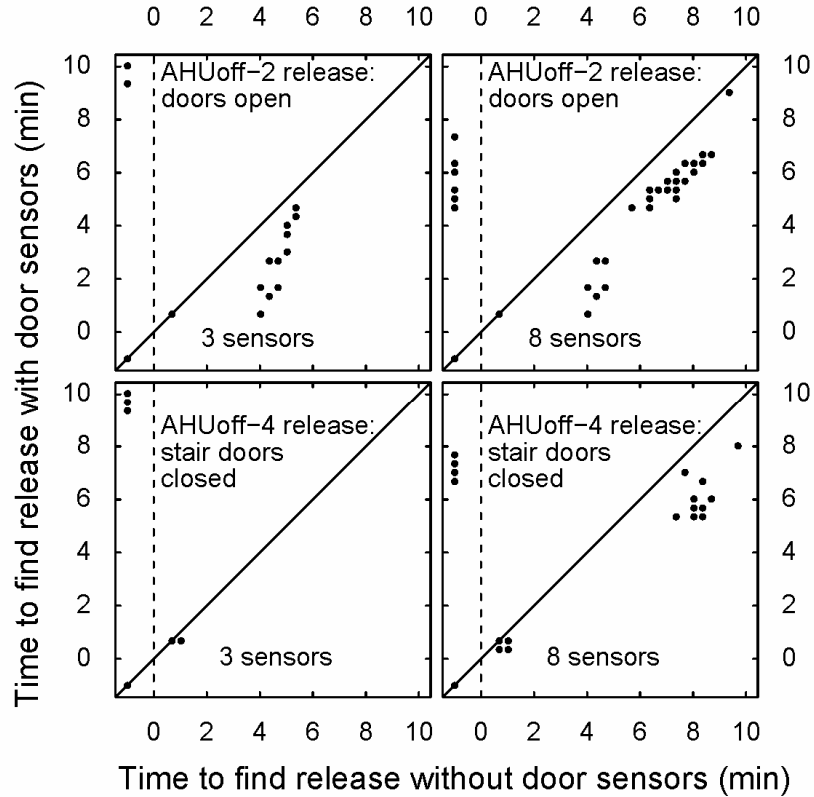


Figure 6.30. Time to locate release to 0.9 probability for releases occurring in Stairwell 2 (Experiments AHUoff-2 and AHUoff-4) for networks with and without door-position sensors. A negative time indicates that a probability of 0.9 was not reached by 11 min.

Door-position sensors also facilitate estimating the released mass to more narrow confidence intervals by  $\sim 6\text{-}17\%$  in both experiments (Figure 6.31 and Figure 6.32). Both three-sensor networks exhibit a greater difference in the median mass that is estimated between networks with and without door sensors. However, this result, if accompanied by a broad confidence interval, is expected. The information from some three-sensor networks is not sufficient to characterize the released mass to a narrow confidence interval with a median that approximates the true mass.

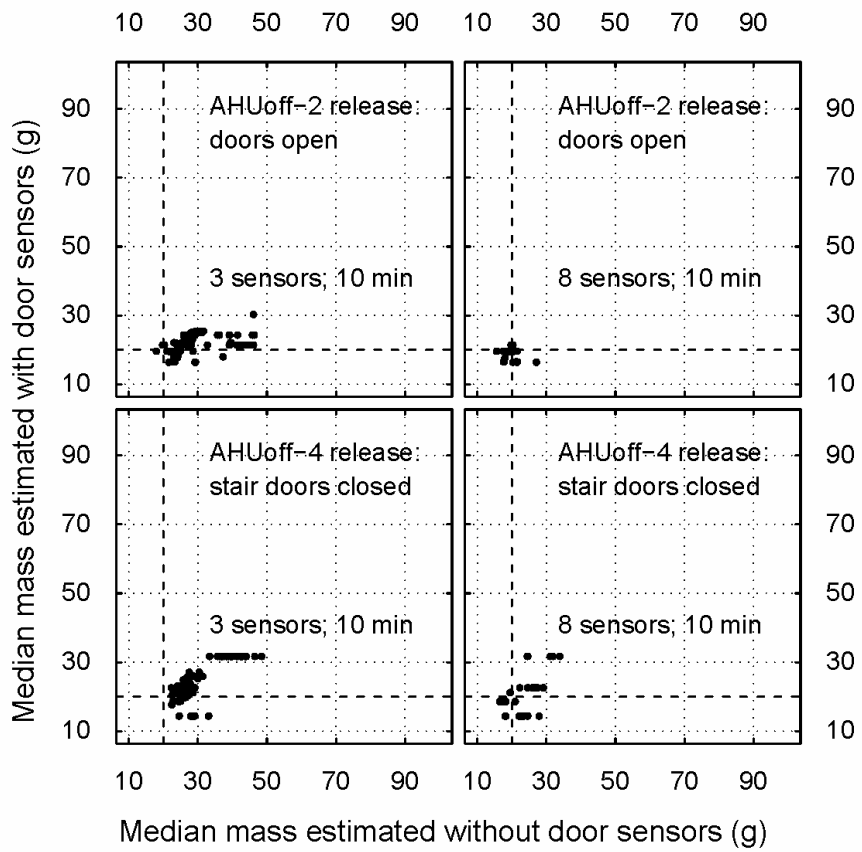


Figure 6.31. Effect of door-position sensors for estimating mass released for Experiments AHUoff-2 and AHUoff-4). Results are shown for all networks containing 3 and 8 contaminant sensors after 10 min of updating.



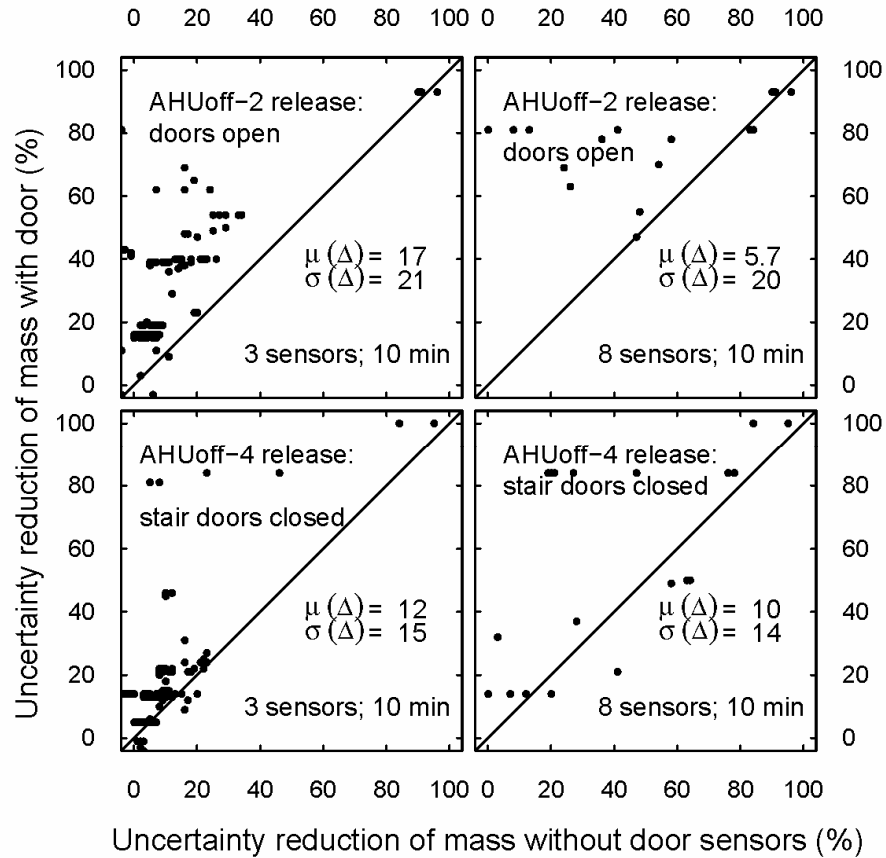


Figure 6.32. Effect of door-position sensors on reducing uncertainty of the estimated mass release for Experiments AHUoff-2 and AHUoff-4). Results are shown for all networks containing 3 and 8 contaminant sensors after 10 min of updating.

#### 6.4.4.3 Tier 2 release scenarios

Tier 2 explores additional release conditions in which door-position sensors may be beneficial. Releases originating from two different interior rooms (Rooms 1.1 and 2.2) and a stairwell (Stairwell 3) are simulated. Also in this exploration, a longer duration release is investigated for one release location.

This section evaluates how a “typical” network might perform, rather than evaluating a full combinatorial set of networks. I defined a typical network as follows: a total of 5 sensors were included; no sensor was placed in the release zone; sensors were distributed

throughout the floors; no two sensors were placed in adjacent zones; if a floor had two sensors (as did Floors 1 and 2), one sensor was located in the stairwell and one in the main interior space. Table 6.4 lists the network configurations for each release scenario.

Table 6.4 Tier 2 network configurations

Simulated experiment number	Release location and duration	Sensor locations
AHUoff-5	Room 1.1 (1 s)	Rooms 1.2b, 2.3, 3.1, Stairwells 1 and 2
AHUoff-6	Room 2.2 (1 s)	Rooms 1.3, 2.3, 3.1, Stairwells 1 and 2
AHUoff-7	Stairwell 3 (1 s)	Rooms 1.3, 2.3, 3.1, Stairwells 1 and 2
AHUoff-8	Room 2.2 (120 s)	Rooms 1.3, 2.3, 3.1, Stairwells 1 and 2

Door sensors influence each network in Tier 2 differently. Figure 6.33 and Figure 6.34 demonstrate the time-dependent capability of the networks to locate the release to a high probability and to estimate the released mass.

For Experiment AHUoff-5, in which the source originated in Room 1.1, the sensor network has difficulty locating the release, whether or not the chemical sensors are augmented by door-position sensors. In contrast, both networks can characterize the mass released to reasonable confidence intervals after 10 minutes of analyses (see upper frames of Figure 6.34). These results likely occur because releases that originate in multiple locations would generate similar concentrations at the sensed locations. Although the sensor system cannot isolate the true release location, the concentration information is sufficient to characterize the released mass. This example illustrates the importance of evaluating multiple network configurations, sensor network densities, and release conditions.

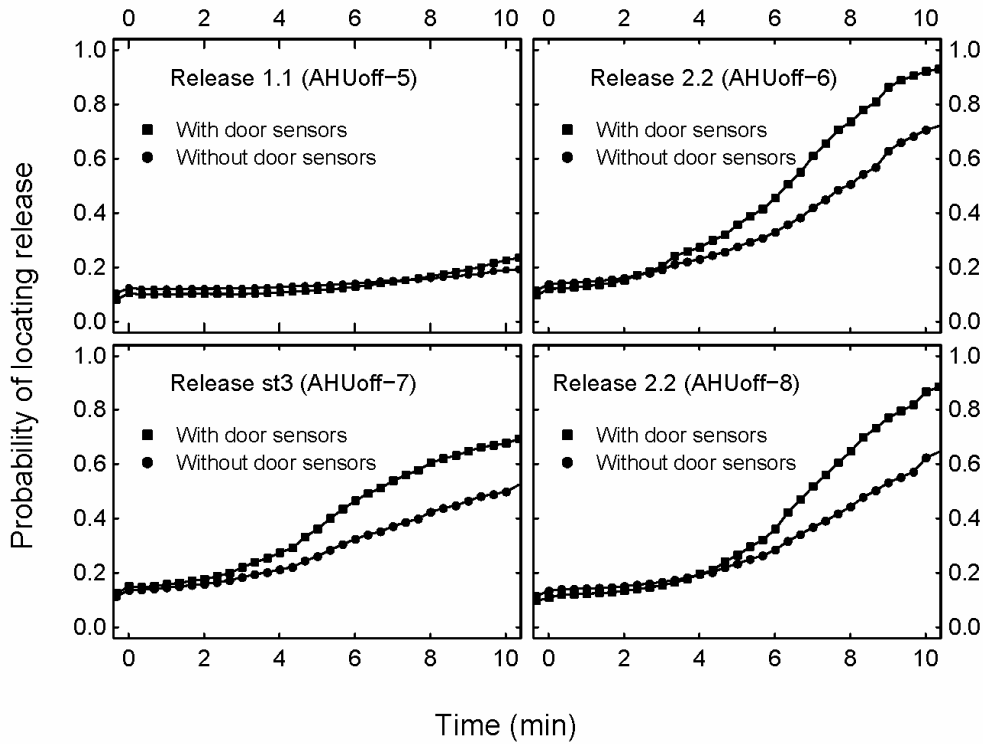


Figure 6.33. Time-dependent probability of locating a release with and without door-position sensors for the Tier 2 simulated release experiments.

For each of the other simulated experiments, the presence of door sensors helps to correctly identify the release location to higher probabilities for a given monitoring time. In the case of the releases originating in Room 2.2, networks with door sensors calculate a probability of  $\sim 0.9$  within the first 10 min of updating (Figure 6.33). Door-position sensors also enable the sensor systems to estimate the release mass to narrower confidence intervals (Figure 6.34). It was anticipated that the benefits of door sensors may be less for releases of longer duration. However, this is not the case here, as seen in comparing the results of simulated experiments AHUoff-6 and AHUoff-8.

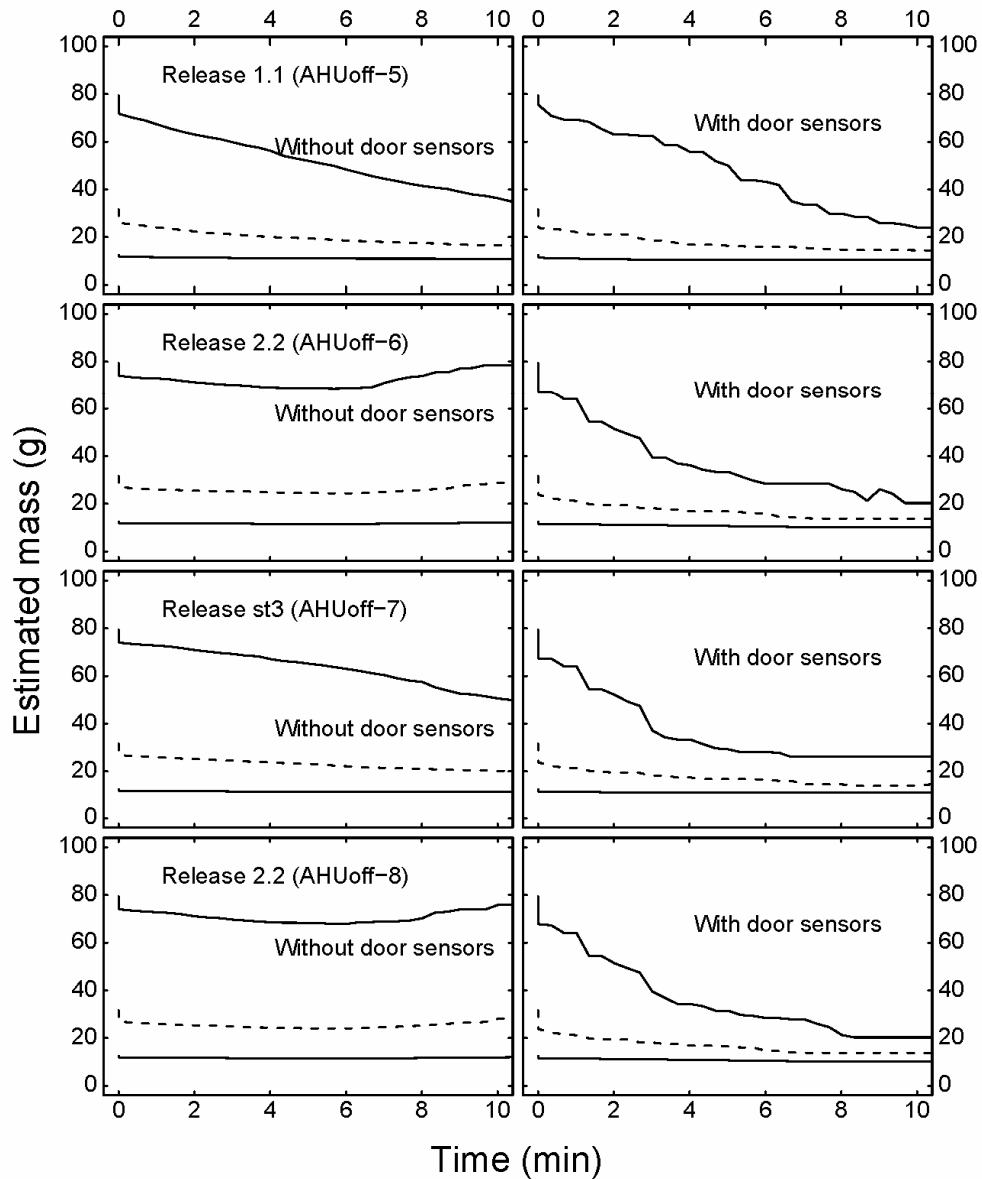


Figure 6.34. Time-dependent estimates of mass released for Tier 2 simulated experiments, with and without door-position sensors; the true released mass is 20 g. Plotted are the median (dashed line), and the 10<sup>th</sup> and 90<sup>th</sup> percentiles (solid lines).

In addition to benefits from characterizing the release parameters, improved knowledge about future concentrations may help plan an effective evacuation strategy. Figure 6.35-Figure 6.37 display the evolution of concentration profiles for the simulated releases originating in Room 2.2 for networks with and without door-position sensors.

Three conditions are shown: no data (Figure 6.35), after 5 min of Bayesian updating (Figure 6.36), and after 10 min of data have been processed (Figure 6.37). Species concentrations in one zone from each floor is plotted, none of which had sensors in the network (Table 6.4). All three figures include the effects of an unknown release time. The algorithm considered equally possible release times occurring at time 0 (the true release time), 60 s, 120 s or 180 s prior to time 0.

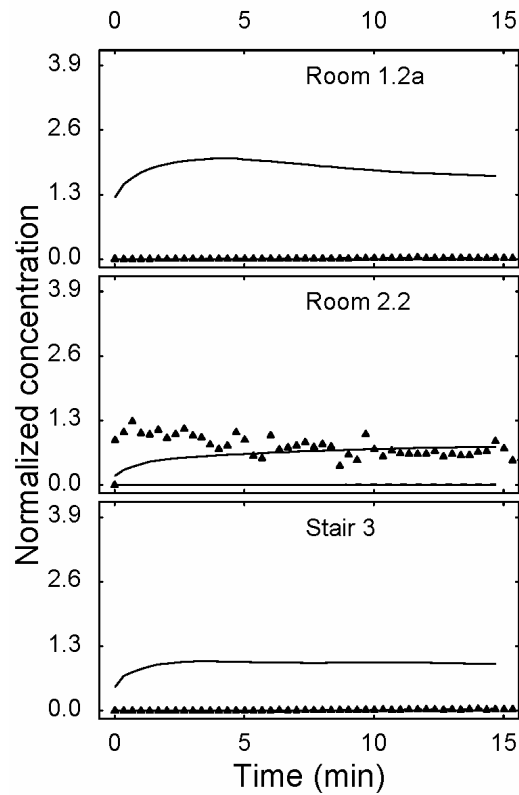


Figure 6.35. Prior confidence intervals of time-dependent concentrations with and without door sensors, displaying median (dashed line), 10th and 90th percentiles (solid lines), and (simulated) concentration data for pulse release in Room 2.2 (points). The release occurred at time 0. The prior uncertainty distributions include the possibility of releases occurring at time 0, 60 s, 120 s, or 180 s prior to time 0.

The door-position sensors influence the prediction of concentration profiles to different degrees in each zone. Figure 6.35 illustrates the confidence intervals of the

concentration profiles in each of the selected zones, based on the prior assumptions embodied in the simulation library. The actual results for the release zone (Room 2.2) are not enclosed within the confidence interval. This outcome is reasonable given that there are eleven other possible release zones which, before the event occurs, are assumed to be equally likely locations for the release. The early confidence intervals do not converge to zero because of the uncertainty of the actual release time.

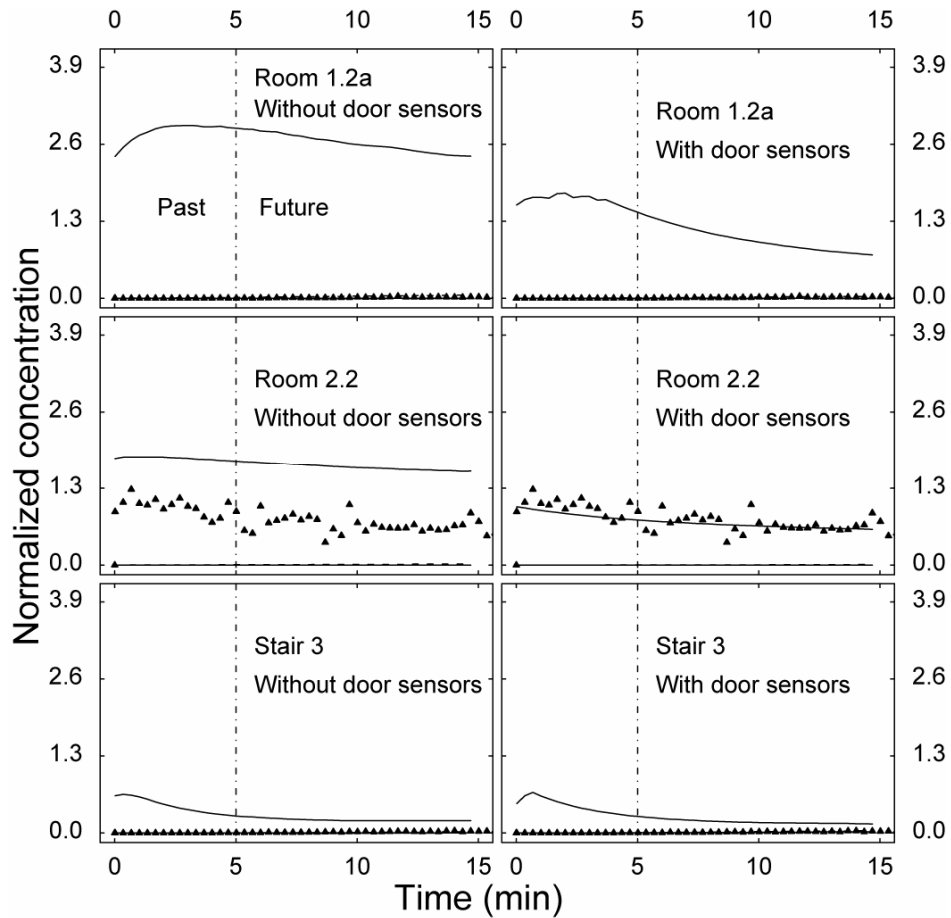


Figure 6.36. Updated confidence intervals of concentration profiles with and without door sensors after performing 5 minutes of Bayesian updating. Plotted are median (dashed line), 10th and 90th percentiles (solid lines), and (simulated) concentration data for pulse release in Room 2.2 (points). Vertical lines denote the updating or “current” time (5 minutes).

After 5 min of updating (Figure 6.36) both networks (with and without door sensors) are able to determine with small uncertainty bounds that the past and future concentrations in Stair 3 are small. In both cases, there is more certainty that Room 2.2 may have been a release location, which accounts for the wider confidence intervals, which now enclose the actual data. Similarly, the wide confidence intervals for Room 1.2a suggest that both sensor networks, with and without door position sensors, consider that the release may have originated in the first floor, after processing 5 minutes of data. However, the networks augmented with door sensors are able to further reduce the uncertainty of the past and future concentrations in both Rooms 1.2a and 2.2. Door position sensors, here, are valuable for characterizing the concentration profiles after only 5 minutes of Bayesian updating.

After the system processes 10 min of data (Figure 6.37), both networks estimate concentration profiles that better reflect the true data. While at 5 min, Room 1.2a was considered a possible release zone or in close communication to the release zone, by 10 min the algorithm considers it to be a less likely source zone, as reflected by the very narrow confidence intervals centered about much lower concentrations. The network with door sensors is able to much more narrowly estimate the past and future concentration uncertainty distributions of Room 2.2., as compared to the corresponding plot for the network without door-position sensors. The wider confidence interval estimated by the base network for Room 2.2 is coincident with the results for estimating the mass released (Figure 6.34), which were also characterized by wider confidence intervals, compared to the networks with door sensors. Thus, heterogeneous sensor

networks can help reduce uncertainty of past and expected future concentrations, even for zones that are not instrumented with sensors.

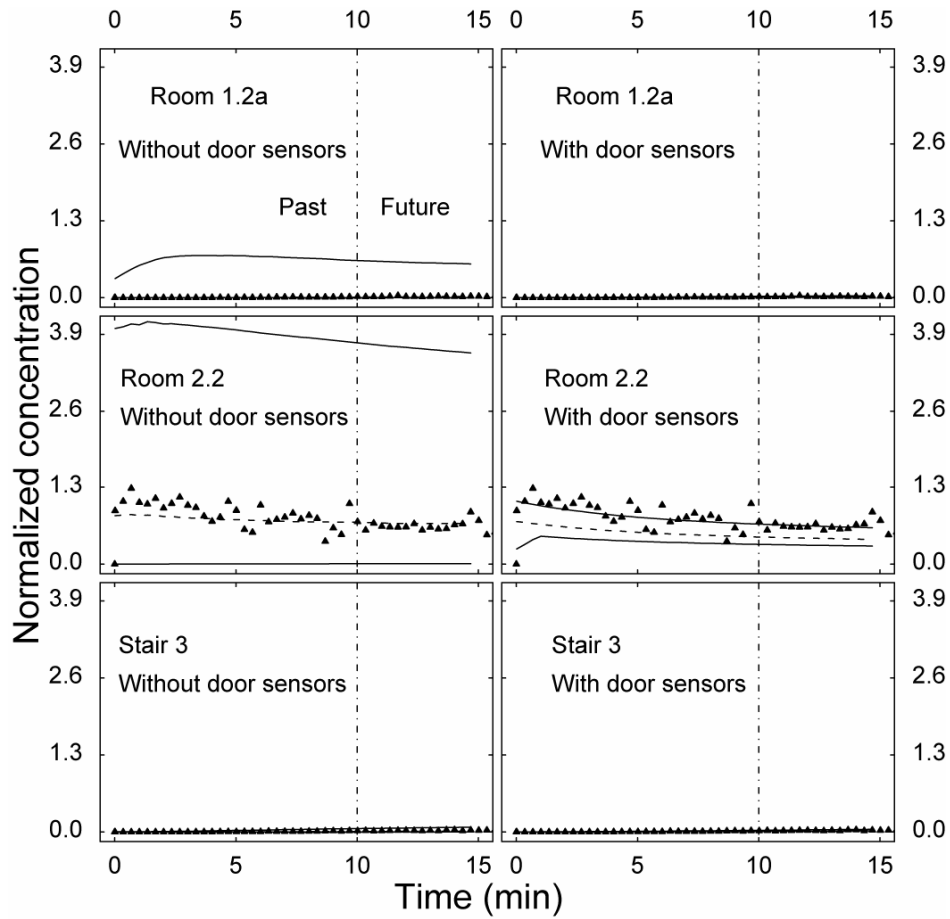


Figure 6.37. Updated confidence intervals of concentration profiles with and without door sensors after performing 10 minutes of Bayesian updating. See caption to Figure 6.44 for additional information.

Overall, the investigation of releases that occur in naturally driven airflow conditions further illustrates conditions in which door position sensors complement chemical sensors in establishing effective sensor networks. Door sensors are likely to be more valuable to a sensor system in conditions where the door position has a greater degree of influence on contaminant transport. The influence of door positions depends on release conditions and more specifically the extent to which the door position state affects concentrations early



after a release. Both investigations, i.e., with and without mechanical airflow, demonstrate that door position sensors are unlikely to displace contaminant measuring sensors or to compensate for poor sensor placement. However, door sensors may be able to help the sensor system characterize the release more quickly, efficiently, and with significantly less computational effort. More generally, the investigation substantiates the physics- and algorithm-based perspectives described early in the chapter.

## **6.5 Advanced fusion methods and statistical issues**

The illustrative example suggests that a heterogeneous sensor system may offer performance advantages compared to a sensor system composed only of chemical sensors. Although in the example shown, door-position sensors could not compensate for poor sensor placement or for an inadequate number of sensors, they enabled sensor systems to characterize a contaminant release to lower levels of uncertainty for an equivalent amount of chemical information.

In addition to offering a performance advantage, a mix of sensors may help prevent the failure of a sensor system owing to unexpected uncertainties. Various sources of uncertainty can cause the system characterization (i.e., the “library” in BMC) to be inaccurate. One potential problem is the use of an inadequate contaminant transport model. Another potential problem arises from unanticipated dynamic operation. The BMC approach models the system prior to a release event. Dynamic building operation can cause the system description to be inaccurate and cause the algorithm to perform poorly. Dynamic building operation also poses challenges for incorporating different types of sensors into a monitoring network.

This section proposes advanced methods for treating sources of uncertainty, such as dynamic building operation. Where applicable, the role that heterogeneous sensors may play in such methods is highlighted.

### **6.5.1 Sensor fusion using benign tracer gas releases**

A key challenge in implementing the BMC method successfully lies in developing an appropriate likelihood function that can characterize the model-to-measurement errors. In Chapter 5 a time-varying likelihood function was developed to represent the nature of errors associated with the multizone model. However, as observed in Chapter 5 and in Section 6.4.3.2, a more sophisticated likelihood function is needed to account for how the errors vary according to different factors such as the release origin and duration. It was also observed that a poor quality likelihood function hindered the ability to effectively use information from door-position sensors.

In a more complex building, it may be even more difficult to develop generalizable tools, such as a likelihood function that can universally account for the effects associated with specific release conditions. Information fusion using data from benign tracer-gas releases may help the sensor system respond to changing operating conditions and avoid problems due to modeling error.

The concept is as follows. Over the lifetime of sensor-system operation, small quantities of a benign tracer gas, such as carbon dioxide ( $\text{CO}_2$ ), would be released episodically into the building. The conditions of the release, such as the release mass, duration, source, would be controlled and therefore known. In addition, all other operating parameters, such as HVAC operating conditions, occupancy, temperatures, etc,

would be recorded to fully characterize the system. Carbon dioxide is an attractive possibility because it is inexpensive to measure (air handling units often come with CO<sub>2</sub> sensors) and is relatively harmless. A network of CO<sub>2</sub> sensors would be dispersed throughout the building. One challenge to using CO<sub>2</sub> as a tracer gas is that occupants emit CO<sub>2</sub>. However, monitoring could be conducted to assess typical background concentrations of CO<sub>2</sub> and the background measurements could be subtracted from real-time measurements taken during the test releases. Alternatively, experiments could be restricted to times when the building is unoccupied.

Data from many releases would be recorded in a database. If the BMC algorithm is implemented, the multiple release scenarios could augment the library; if enough releases are generated, they could populate the library entirely and supplant the modeling effort. When a real event occurs, the true sensor network records concentrations of the real contaminant, as usual. The concentration data could be interpreted as in previous chapters. The likelihood of each realization in the library would be calculated based on the error characteristics of the contaminant sensors. Then, using the posterior probabilities, each release parameter could be estimated. In the case where real-time information of the HVAC operating conditions, and other release conditions are known, the library could be filtered to generate a smaller subset.

The method described is one approach to fusing benign tracer-gas information. In a complementary approach, benign tracer gas releases could be used to diagnose and test the library and algorithm, rather than to build it. Irrespective of the fusion method, such measurements represent the *real* behavior of the system, rather than an a-priori, possibly

outdated description. Thus, they offer an opportunity for the sensor system to adapt to new conditions.

### **6.5.2 Dynamic library generation and advanced filtering**

There are many sensors that can help update system characterization. Several measurements may be available through the Energy Management Control System (EMCS). Other sensors are relatively inexpensive and could be deployed as part of a sensor system. Measurements of airflow, pressure, economizer operation contain essential information towards understanding how a contaminant may be transported within and out of the building. The dynamic state of the measurement has implications for how sensor signals are to be interpreted by an algorithm.

This dissertation has used a BMC framework to characterize release events. I propose here that dynamic library generation and library filtering methods can be developed as approaches for facilitating the fusion of information from heterogeneous sensors within the BMC framework. The idea of dynamic library generation is as follows. In situations where operational parameters change over time scales that are longer than those required to generate the library, measurements of the parameters could be used to generate an updated library. Generally, such measurements would fall under the pseudo-static model input classification discussed in Section 6.3.2.1. For example, consider economizer operation. Economizer operation is typically changed on a time scale not less than an hour, rather than on the order of minutes. If a new set of recirculating and outdoor air flow rates are known, then these could be used to update the base contaminant transport model, which could in turn be used to generate a new library.

Pressure measurements in different zones of the building could be incorporated in the same way.

Advanced filtering methods may be required to accommodate the fusion of pseudo-static model inputs that change more quickly than library generation time. Here, it would be necessary to shift the computational effort to an earlier time. Rather than generating a new transport model and library based on real-time measurements of pseudo-static inputs; sets of libraries could be generated a priori to accommodate a wide range of possible operating conditions. To generate this wider library, first, a discrete set of different operating conditions that feed into the contaminant transport model would be determined. Then, multiple sets of libraries would be generated. Finally, in real-time, the measurements could be used to filter this larger set of realizations into a representative sublibrary that would be available to the BMC algorithm in the event of a detected release.

Both dynamic library generation and advanced filtering are methods that may allow the BMC framework to accommodate changing operating conditions, incorporating information from heterogeneous sensor types. These approaches may help to extend the utility of the BMC framework and in some cases render the method more computationally efficient. Both methods discard superfluous parameter values; by discarding unnecessary parameter values in the library generation step, higher resolution can be realized in sampling input variables. Rather than casting a “wider net”, this higher resolution, more narrowly targeted net may help to characterize a release more accurately and more efficiently.

### 6.5.3 Methods that combine BMC, MCMC and SMC

Chapter 2 discussed different statistical inference methods. The Bayes Monte Carlo (BMC) and sequential Monte Carlo (SMC) methods are easily implemented in real time because they are not iterative procedures. Markov Chain Monte Carlo (MCMC), though considered to be a superior statistical sampling algorithm, is not amendable to real-time parameter estimation. However, it may be possible combine elements from each algorithm, generating a superior hybrid method.

SMC methods are geared towards identifying and re-identifying, in real-time, sets of samples that better approximate the target probability distribution. In the SMC literature, realizations are called particles. Based on the extent to which a particle's updated posterior probability differs from the posterior probabilities of other particles, it can be discarded, or assigned a new probability. The differences in the probabilities (termed "weights" in the SMC literature) of each particle can be used to diagnose how efficiently the algorithm is performing. In BMC, the analog to resetting particle weights in SMC is "Bayesian windowing". If at some time, the probabilities of some realizations are deemed to be insignificant, those realizations can be discarded, and the probabilities of remaining realizations can be reset to equal values. Sohn et al. (2000) used a Bayesian windowing approach in modeling groundwater contaminant transport.

An innovation on Bayesian windowing and SMC particle filtering may be capable of diagnosing in real time whether the library is representative of the release. In cases in which the library is not representative, it may be more efficient to discard a large subset of realizations, reset the probabilities, and if necessary, generate a new set of realizations.

While MCMC is not practical to implement as the sole parameter estimation tool, it may be possible to implement MCMC at some stage of the updating to generate a more refined characterization of the release. One limitation of the BMC method is that the library must necessarily cover a very wide range of possible release scenarios. The resolution of parameter values is limited to keep the library size manageable. This limitation means that more relevant values of model input parameters may not be sampled at sufficient density to accurately characterize the release.

In a hybrid BMC-MCMC approach, BMC could be deployed to target the general parameter space that characterizes the release. BMC is very good at eliminating or nearly eliminating realizations that describe a release poorly. Once a more probable region of the solution space has been determined using BMC, it may be possible to use an MCMC algorithm (such as Gibbs sampling) to refine the parameter estimates. As discussed in Chapter 2, MCMC does not necessarily demand that the modeling be done in real time, simply that the conditional probabilities (i.e., the likelihood functions) be evaluated in real-time, through an iterative sampling process. Perhaps at the MCMC stage, a higher resolution, but more narrowly targeted library could be brought online to improve the characterization of the release event.

#### **6.5.4 Using the sensor system as a general system diagnostic tool**

Building operation can change unexpectedly and, thus, a static algorithm may not be able to accurately characterize a contaminant release event owing to a new state of airflow and contaminant transport conditions. I have suggested that it may be beneficial, if not essential, to monitor the operation of the HVAC system and adjust the contaminant

sensor system's interpretive algorithm to ensure that the system does not fail. While much of the chapter proposes methods for adjusting the sensor system and BMC algorithm to changing, unexpected conditions, this section proposes that a sensor system could also be used to detect unexpected operating conditions, which are often due to undesirable, faulty operation.

There are different ways in which the basic BMC sensor system or advanced systems could be used to detect faulty HVAC operation. Consider a BMC algorithm that is implemented with a library, which considers all possible scenarios of reasonable operation. In the event that combinations of real-time measurements of pseudo-static model inputs suggest new and unmodeled combinations, two possibilities exist. Either (a) the library is inadequate and must be augmented to reflect this new operating state, or (b) the new operating state is outside of the realm of typical operation. In the latter case, the sensor system may offer value beyond chemical detection by possibly identifying faulty operation that could result, for example, in increased energy use.

Consider the implementation of the benign tracer gas method described in Section 6.5.1. Such releases could be used not only to augment the sensor-system library, but also to troubleshoot and diagnose the completeness of the library. If the existing library is relatively well populated (either by modeling or through a combination of modeling and tracer-gas release monitoring) then any benign tracer-gas release should be well characterized by the sensor system. If the sensor system is unable to characterize the release to a reasonable level of uncertainty, then there is a strong possibility that the system is in a new operating state, which is potentially faulty.



By identifying conditions where there is a discrepancy between the library and real behavior, there are two possible benefits. First, the physical system (i.e., HVAC system) can be restored to a desirable operating condition. Second, the sensor system can be adjusted to reflect new operating conditions that may otherwise pose performance problems.

## **6.6 Conclusions**

This chapter explores information fusion from heterogeneous sensors. The objective was to investigate how different types of nonchemical sensors can contribute towards a less expensive, more robust, and more efficient sensor-system design. This matter cannot be addressed without an understanding of how most commercial building systems are operated and controlled. Thus, the background sections of the chapter provide a brief, but essential review of HVAC control and operation, including a discussion of sensors that are likely to be available to a high-risk contaminant sensor system.

Sensor fusion is discussed at a conceptual level with physics- and algorithm-based frameworks that can be used to assess any sensor. The physics-based perspective evaluates the potential benefit from a sensor by considering its degree of influence on overall contaminant transport. Sensors reflecting state parameters that exert a high degree of influence are likely to contribute usefully to sensor-system design. The algorithm perspective proposes two types of classification. One is based on the model input/output role of the measurement. A second classification is based on the dynamical state of the measurement. Both of these classifications have implications for how the information from a sensor can be fused into the monitoring system. These classifications

are then related to the specific BMC algorithm used in this dissertation. Both frameworks are used to describe how information from different sensors can be incorporated into a sensor system.

The sensor fusion concepts are illustrated with an example of door-position sensors. The investigation highlights the connection between the degree of influence (for model inputs) and the value of a sensor's data, demonstrating how the fusion of data from heterogeneous sensors can be assessed through the algorithm frameworks. Information from door-position sensors is fused by adjusting the prior probabilities of the library realizations. An algorithm is proposed that can be exercised to allow for sensor uncertainty and model uncertainty. Many networks are evaluated, using different numbers of contaminant sensors, with and without door-position sensors.

Door-position sensors improved the identification of the release location on average by  $\sim 1$  to 10% for networks based on Experiment 1. The benefit depended strongly on sensor placement and was characterized by standard deviations that ranged from 10 to 20%. While door sensors could facilitate the identification of the source, the success of networks with 2 or 6 contaminant sensors depended more on sensor placement. Door-position sensors had little influence on the median of the released mass, although for networks with few sensors, they enabled the network to estimate narrower confidence intervals of the released mass.

The results for Experiment 4 exhibited a negative impact from the fusion of door-position information. These results arise because of the model-specification error for this experiment. If the likelihood function does not adequately capture the error structure, the consequences due to model-specification error are exacerbated when

door-position information is fused. (Suggestions for constructing more sophisticated likelihood functions are discussed separately.)

Networks constructed from simulated data were investigated to compensate for the limited experimental data and to further explore conditions under which door-position sensors could improve the sensor-system performance. Door-position sensors were not useful for the simulated release in Room 1.3 (which replicates Experiment 4). This result was unsurprising because the AHU was on and the door positions exhibited little influence on the contaminant dispersion.

Door-position sensors were useful for characterizing the release that originated in Stairwell 2. Door sensors improved the identification of the release location marginally (on average by 1–3%), however, for some networks the inclusion of door-position sensors allowed the network to identify the release location to a 90% probability within 11 min. The uncertainty of the released mass was further reduced (on average by an additional 6 to 11%). Although the AHU was on, the door positions influence the contaminant transport because the AHU does not directly supply air to the stairwell.

Several simulated releases were explored with the AHU off. With few anomalies, door-position sensors were found to be consistently useful for releases originated in Stairwell 2 for all network configurations. The release location was identified to a higher probability, on average by 3 to 7%, with increased benefit for networks with fewer sensors. Door-position sensors enabled some networks to identify the release location to a 90% probability more quickly. Door-position sensors enabled the estimation of the released mass to lower uncertainty levels, on average  $\sim$  5 to 20%. Sensor placement, however, had a greater influence on the overall system performance.

Additional releases were explored with the AHU off and with a “realistic” sensor network. The performance benefit from door-position sensors ranged from 5% to 20% improvement in source identification and ~ 40 to 85% reduction in uncertainty of the released mass. For these releases, also, contaminant sensor placement influenced the benefit from fusing door-position information because door positions do not always influence the contaminant concentrations significantly at all sensed locations.

Overall, it was found that, while door-position sensors cannot compensate for poorly placed contaminant sensors, or for an inadequate number of contaminant sensors, they can improve the efficiency with which the sensor-system characterizes the release. In many cases, door-position sensors contribute a nominal benefit towards identifying the site of the release. In some cases, there is a nominal penalty, which may be a statistical effect, particularly, if the library has not been adequately sampled. It was found that when the door position exerts little influence on contaminant transport, fusing information from door-position sensors can hinder the performance of the sensor system, since it acts to eliminate otherwise useful realizations.

Generally, the extent to which door-position sensors are beneficial depends on the degree of influence that the door positions exert on contaminant transport. The degree of influence depends on specific release characteristics, such as the release location and mass, which are unknown to a monitoring system. Thus, the *realized* degree of influence cannot be known in advance of a release. The concept of “realized influence” could be applied to evaluate conditions for which disparate sensors are less or more likely to be useful, which in turn could be used towards making system design decisions.

While the example of integrating door-position sensors into a sensor system is a simple one, the results suggest that there are indeed benefits from applying the heterogeneous sensor fusion concepts for improving the task of characterizing high-risk pollutant releases. Several other types of sensors, such as pressure sensors, airflow sensors, and control-system input signals should be evaluated within the proposed sensor fusion framework. The fusion of information from some sensors may require more sophisticated implementations of the BMC algorithm and possibly the implementation of other algorithms. More sophisticated fusion algorithms such as benign tracer-gas sensor fusion, dynamic library generation, and advanced filtering techniques may facilitate the fusing of information from many different types of sensors. Advances in these areas may enable the design of sensor systems that are more robust against faulty performance owing to changing operating conditions and may also help the system to perform more efficiently. Ultimately, the fusion of disparate sensors into a monitoring system could be extended to use the sensor system as a diagnostic tool that can detect faulty HVAC operation that results in increased energy use.

## **6.A. Appendix: Simulated data**

This appendix contains figures for the releases that were simulated with the air handling unit off. In all, eight releases were simulated, which are described in Table 6.4. Figures 6.A.1-6.A.8 show the simulated data, which includes randomly generated error, and the modeled concentrations. In all figures, concentrations are normalized by the theoretical peak concentration in the experiment; time is in reference to the time of the release; and the upper frame plots data for the first floor rooms, the middle frame plots data for the second floor and third floor rooms, and the lower frame plots data for the staircases.

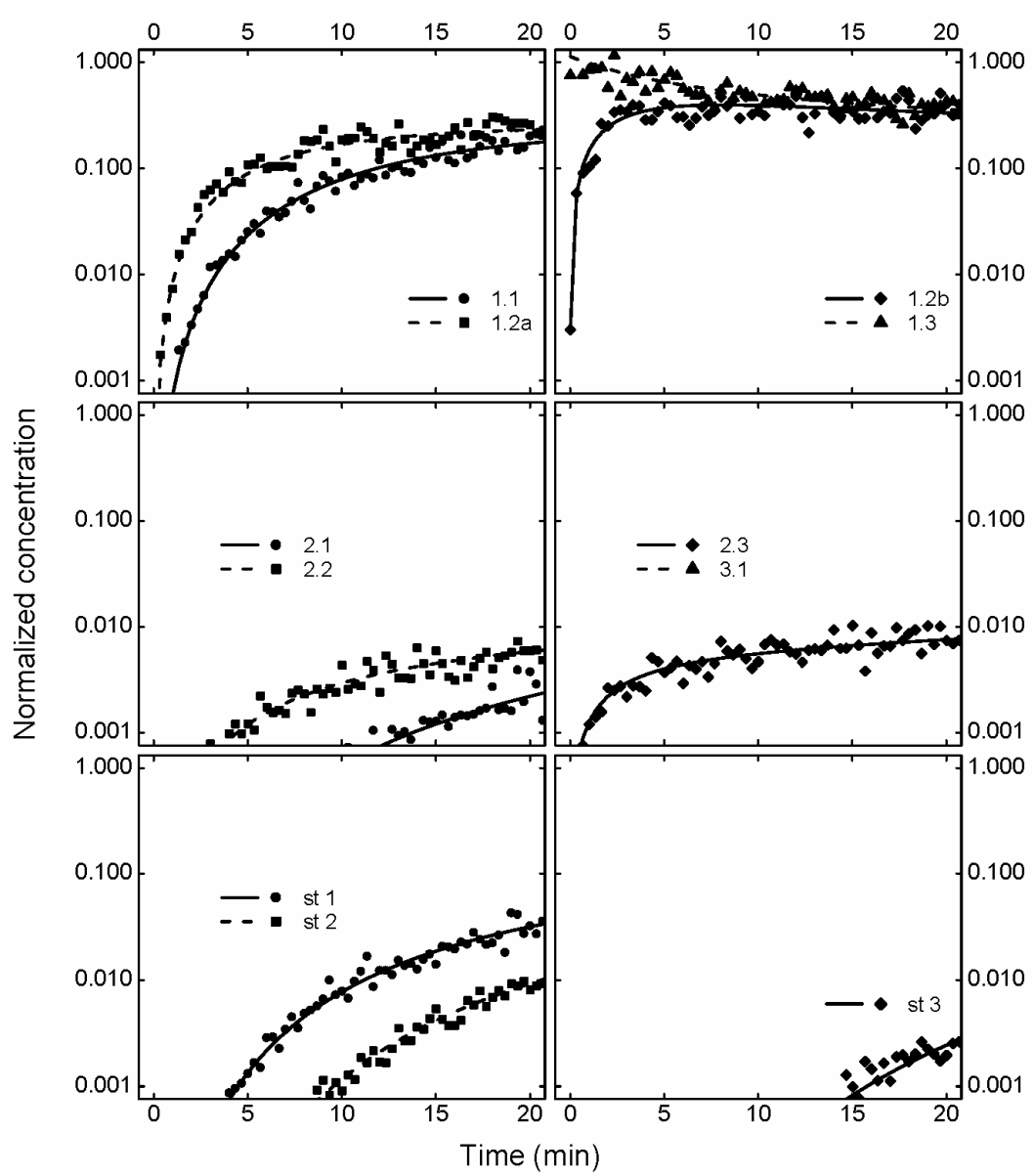


Figure 6.A.1 Simulated and modeled concentrations for Room 1.3 release, Experiment AHUoff-1. .

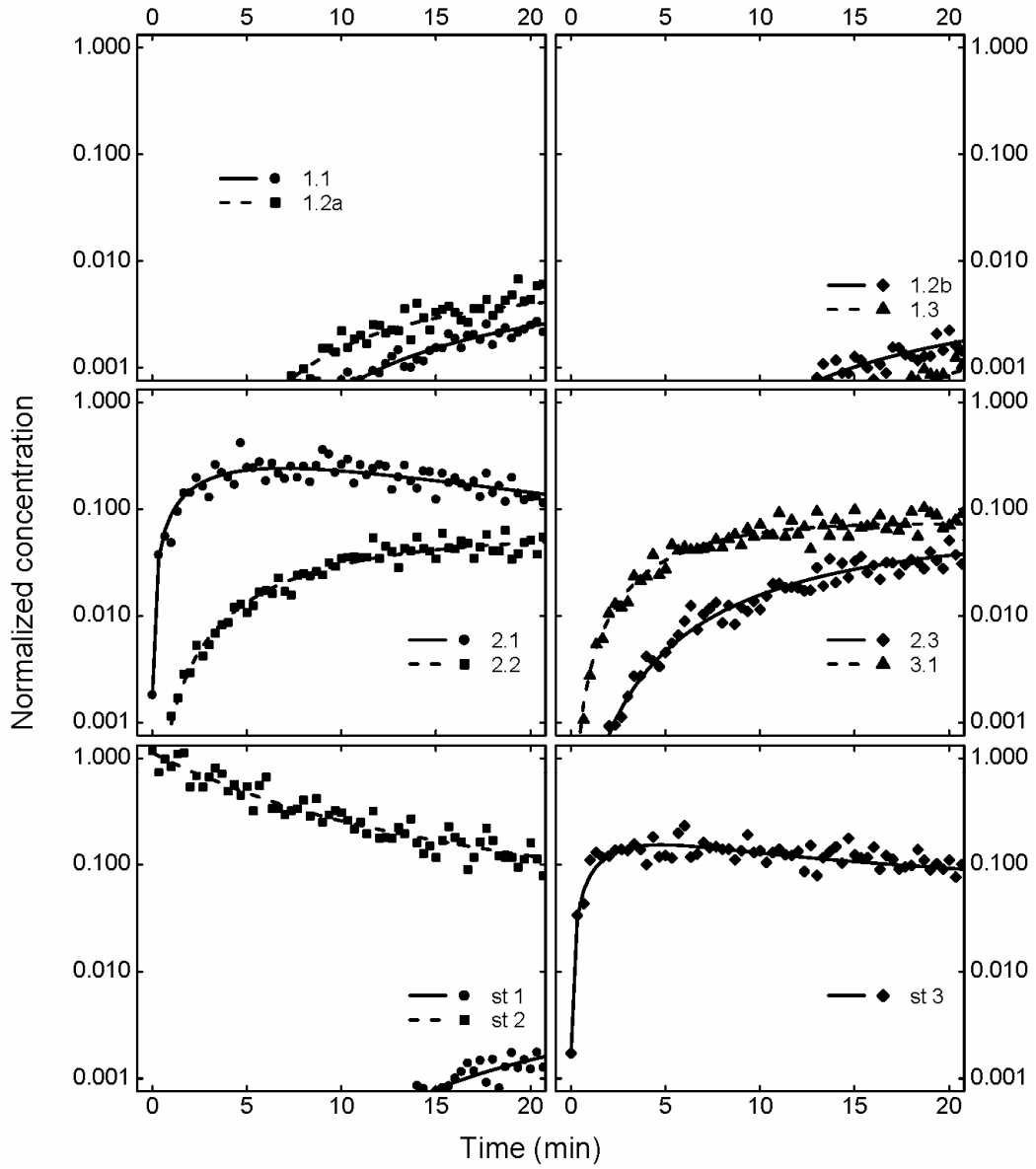


Figure 6.A.2 Simulated and modeled concentrations for Stairwell 2 release, Experiment AHUoff-2.



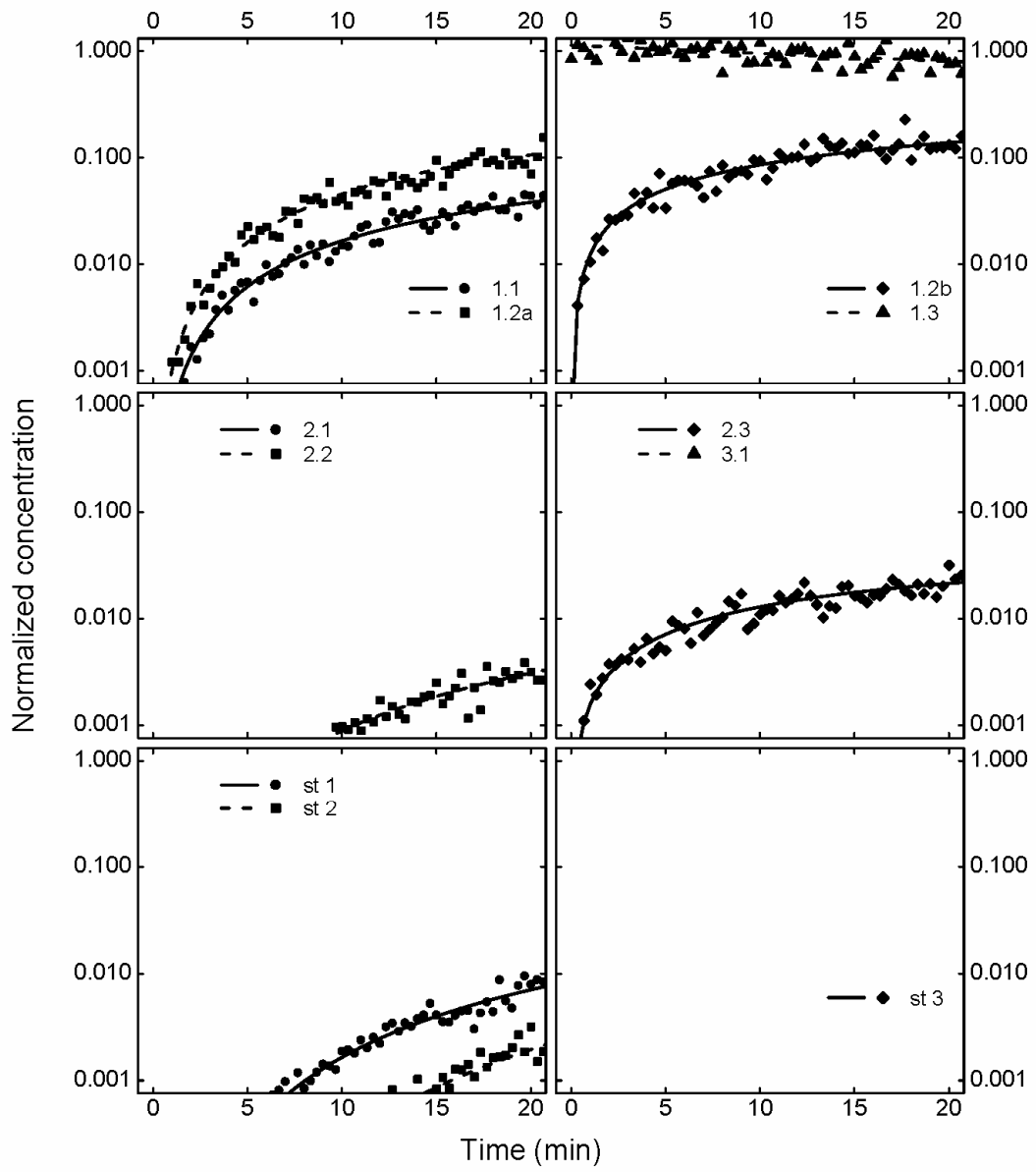


Figure 6.A.3 Simulated and modeled concentrations for Room 1.3 release, Experiment AHUoff-3.

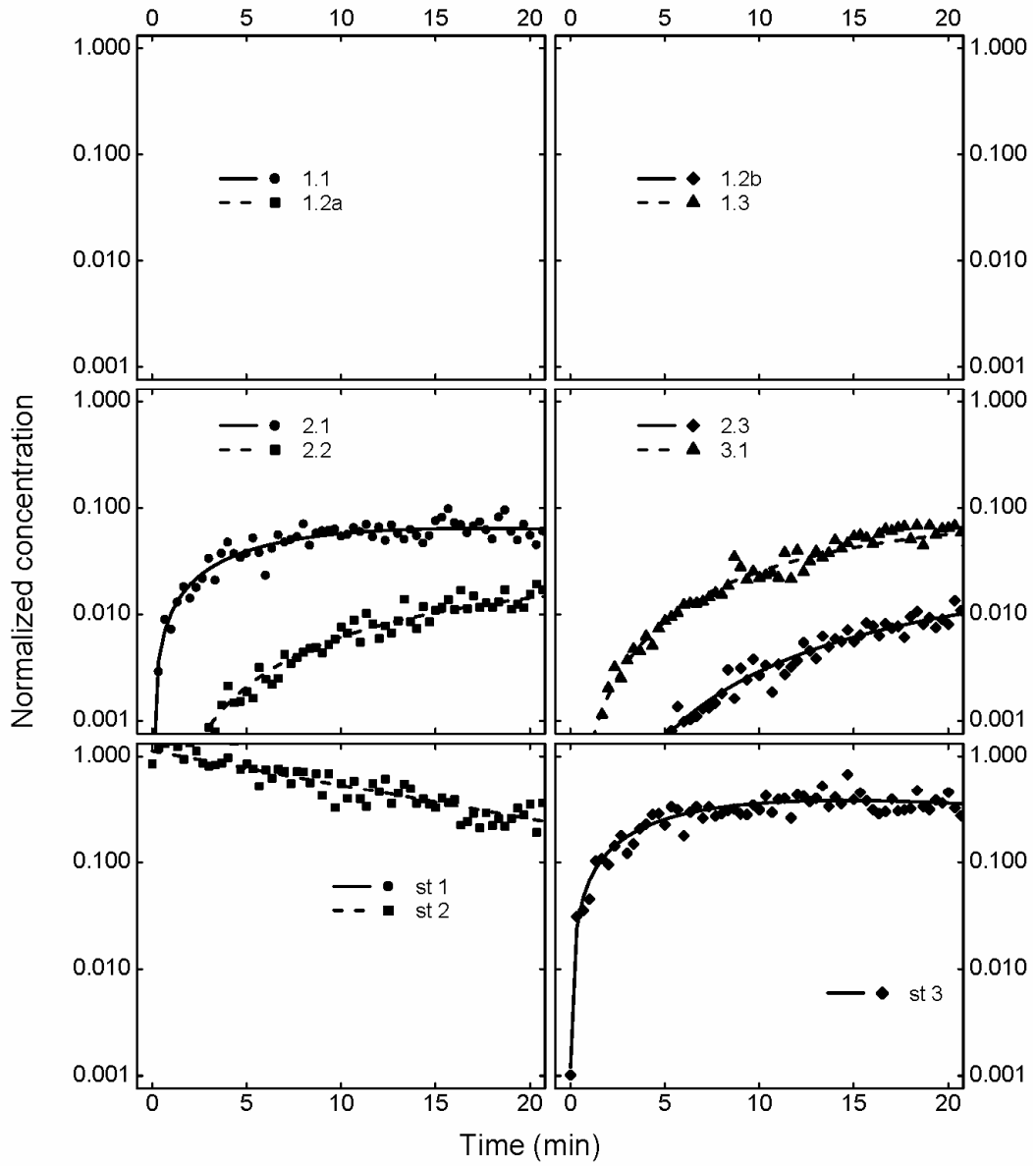


Figure 6.A.4 Simulated and modeled concentrations for Stairwell 2 release, Experiment AHUoff-4.

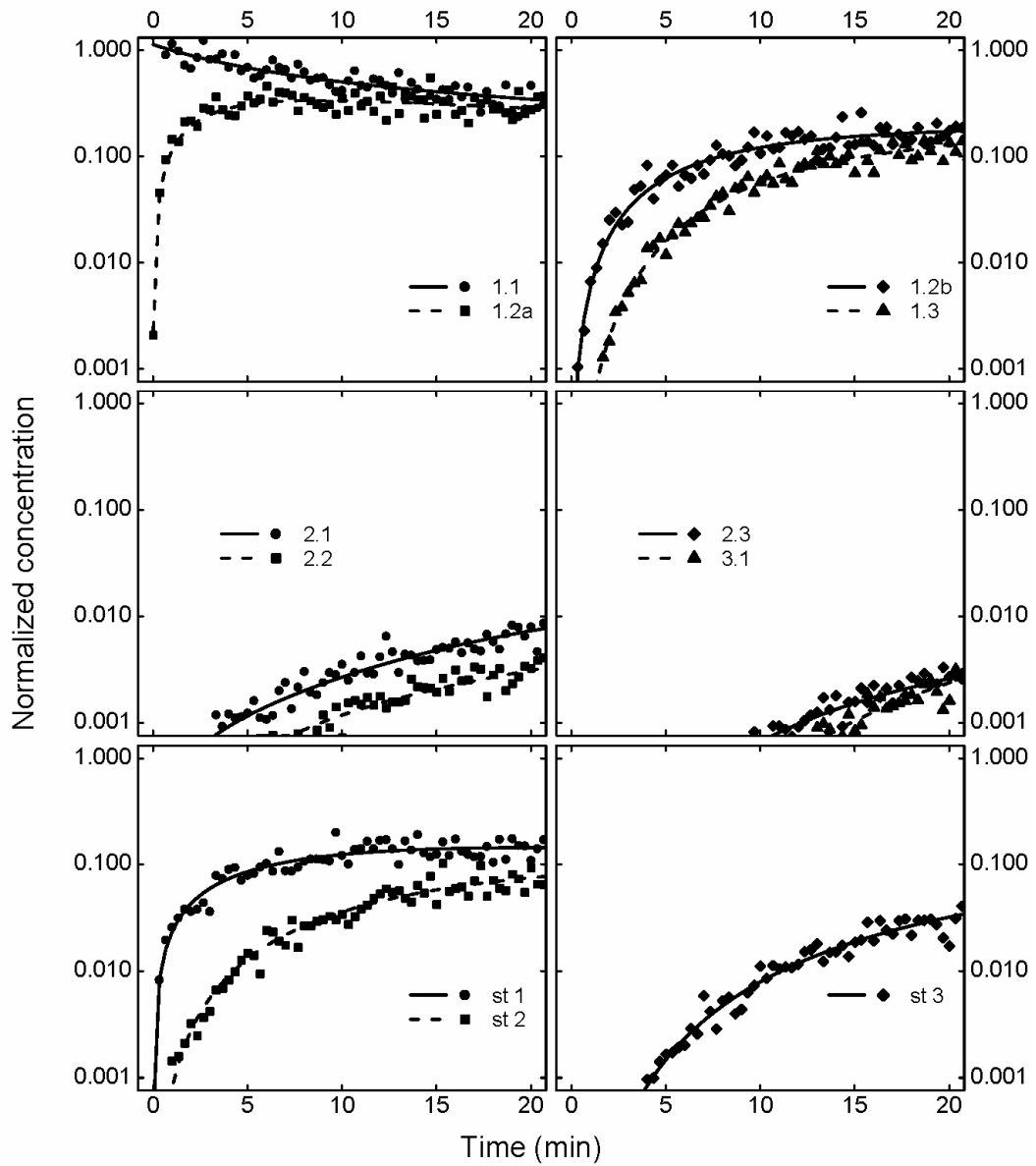


Figure 6.A.5 Simulated and modeled concentrations for Room 1.1 release, Experiment AHUoff-5.

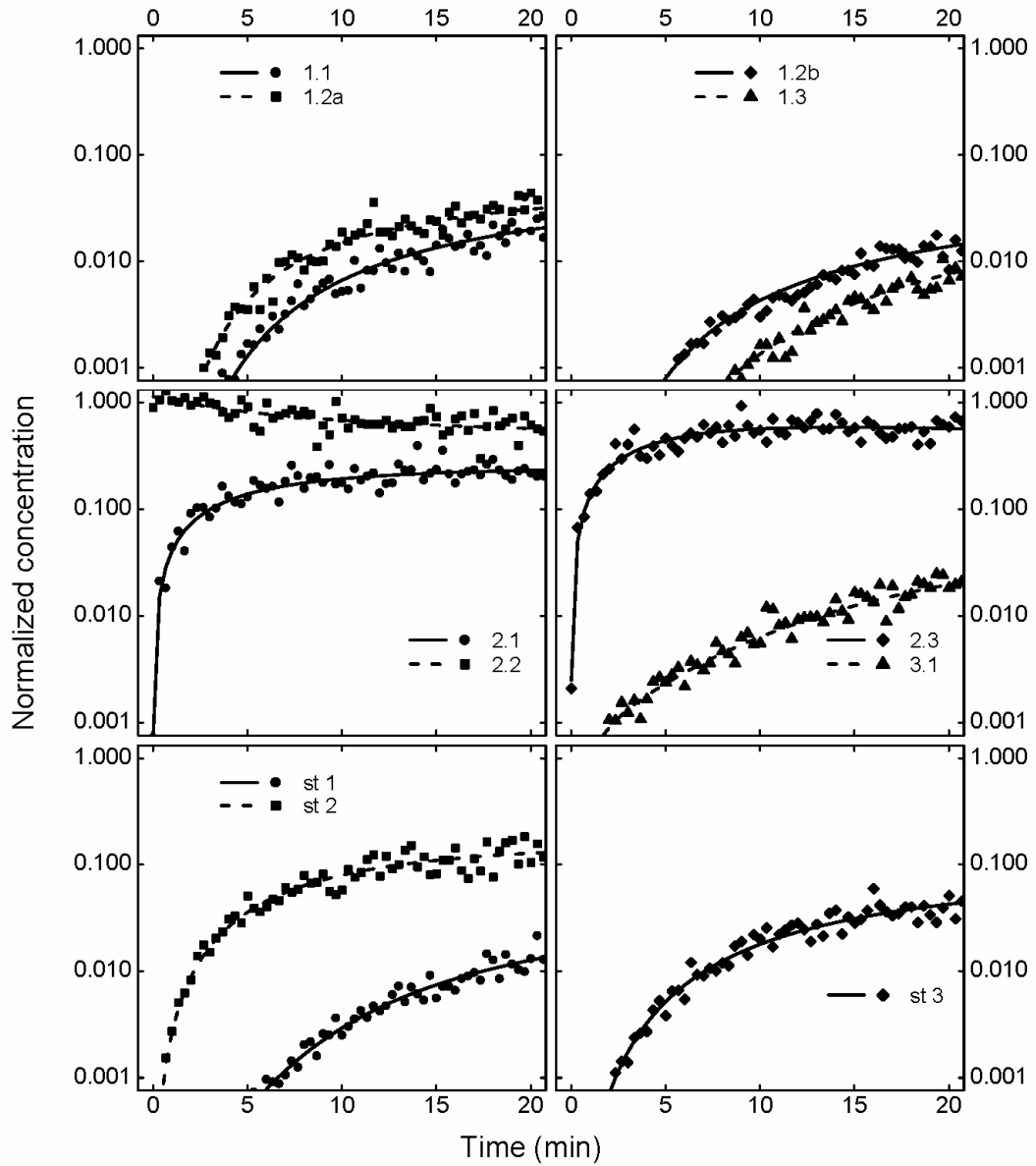


Figure 6.A.6 Simulated and modeled concentrations for Room 2.2 release, Experiment AHUoff-6.

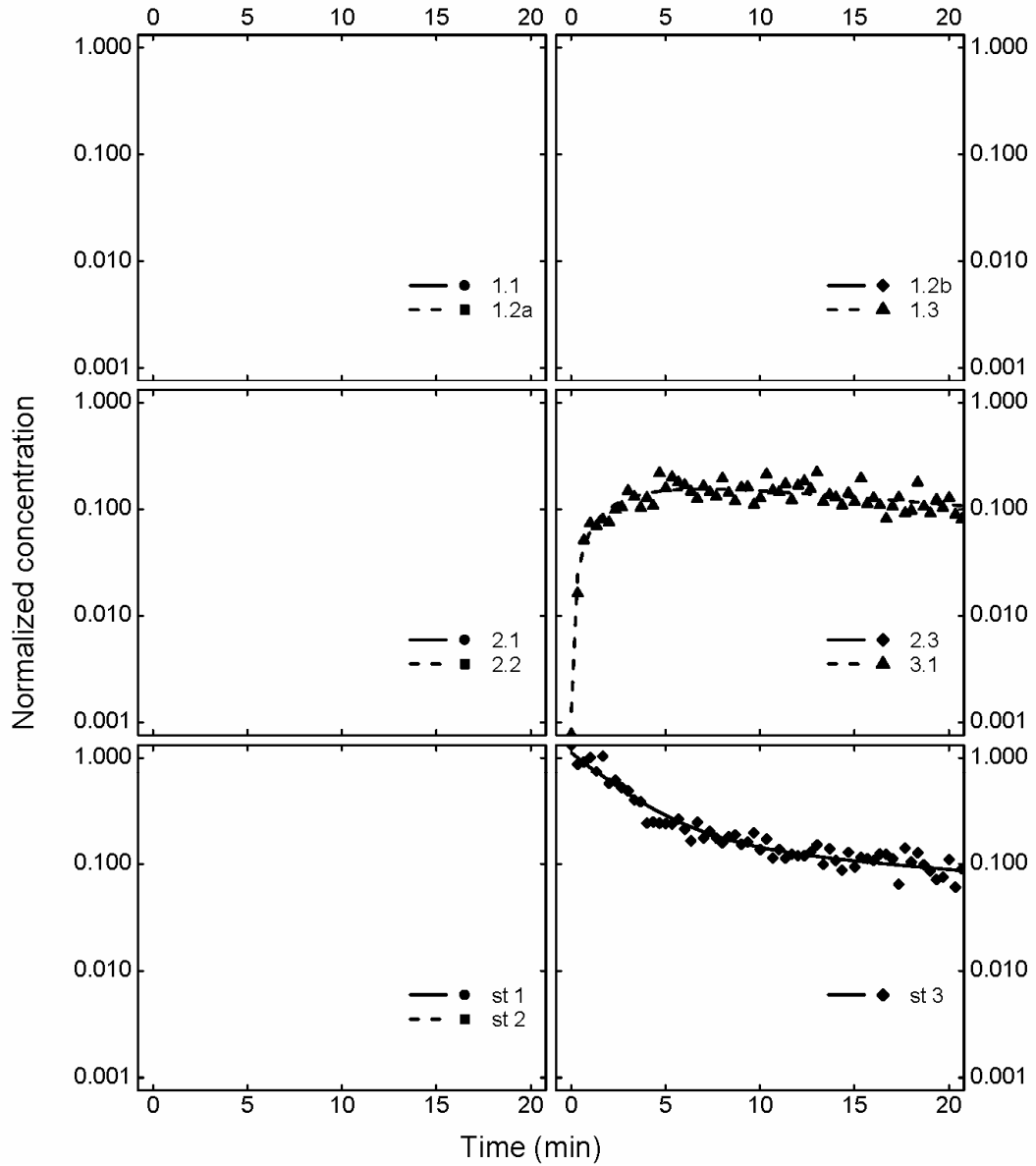


Figure 6.A.7 Simulated and modeled concentrations for Stairwell 3 release, Experiment AHUoff-7.

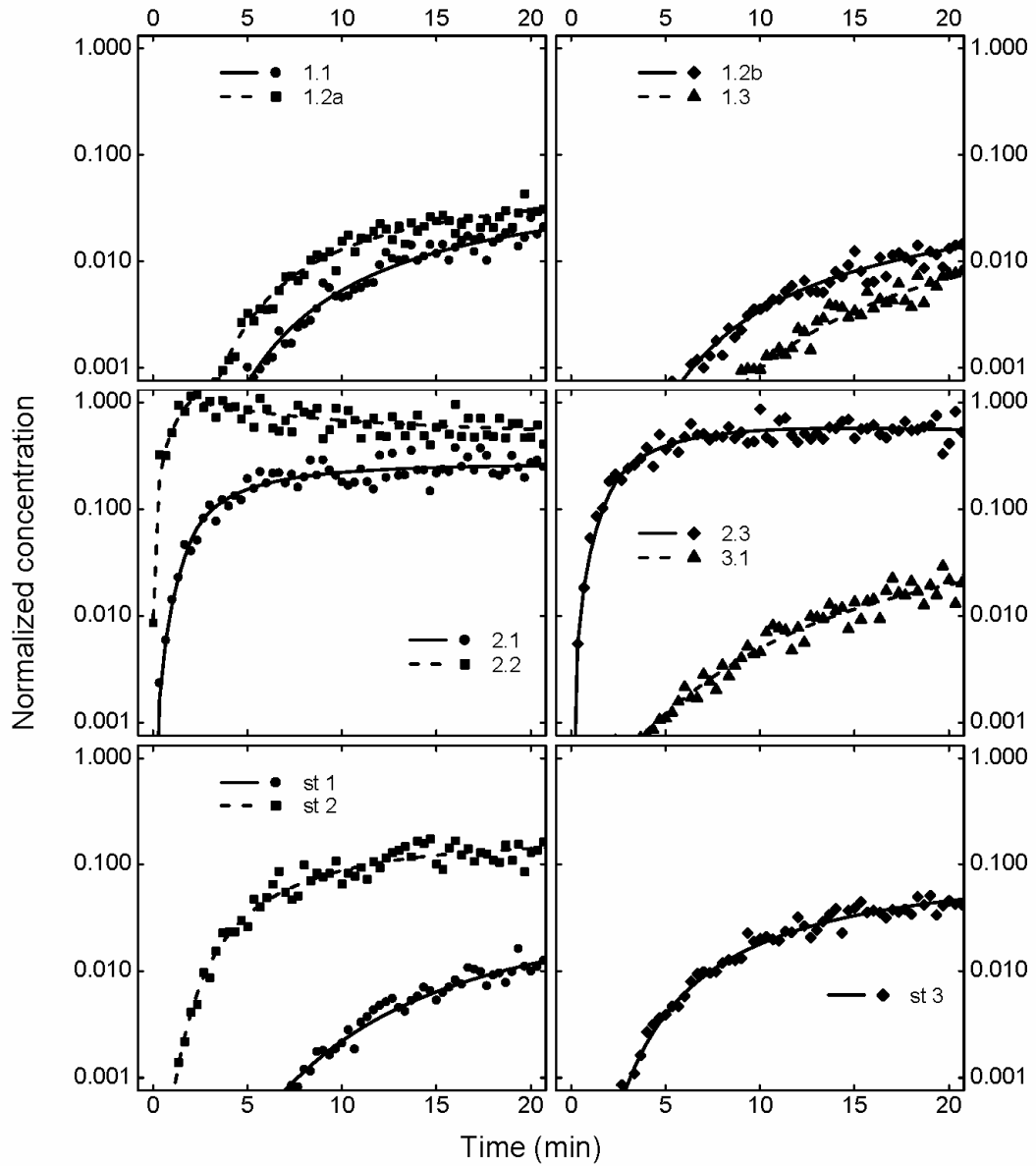


Figure 6.A.8 Simulated and modeled concentrations for Room 2.2 release, Experiment AHUoff-8.

# Chapter 7

## 7 Conclusions and future work

Chapter 7 is organized into three sections. I first summarize the methods and findings from the core chapters (4-6). Next, specific future research questions are discussed. They include methodological and application oriented questions. The third section discusses broader research questions for the design of sensor systems and for problems that extend beyond the application addressed in this dissertation.

### 7.1 Summary

Chapter 1 sets the tone. This dissertation seeks to develop tools and concepts to facilitate the design of sensor systems for characterizing high-risk airborne contaminant releases. The basic goal of the sensor system is framed as an inverse problem, which is solved using a Bayes Monte Carlo (BMC) approach. The objective of the dissertation is to explore how design decisions regarding sensor measurements and the contaminant transport model influence the overall system performance and how these relationships are linked to generalizable characteristics of the physical system.

Chapters 2 and 3 provide essential background. Chapter 2 discusses environmental sensor systems. The Bayes Monte Carlo algorithm is presented in detail along with related algorithms and alternative statistical inference techniques. The physics of indoor

contaminant transport and modeling methods are summarized. Chapter 3 describes the case study that provides the experimental data for the dissertation.

Chapter 4 is the first core chapter. It investigates how sensor characteristics, specifically of threshold- or alarm-type sensors, are related to the overall performance of a sensor system. There are two primary objectives. The first is to understand whether a Bayesian based system can characterize a release, given the large uncertainties of threshold data. The second goal is to demonstrate how improved design decisions can be made by analyzing the sensor characteristics using a system-level analysis. The concentration data from the case-study tracer-gas experiments were converted into threshold sensor data by simulating the effects of varied sensor response time, threshold level, and accuracy.

Threshold signals were found to be successful at characterizing a contaminant release even with sensors exhibiting relatively high error rates. In many cases, the release location was identified to a 90% probability within 2 min and the release mass was estimated to a narrow confidence interval within 4 to 10 min. Relationships among each of the sensor characteristics and the pollutant profile parameters were identified. For example, higher threshold levels were found to reduce the uncertainties of all the parameters in general, compared to lower threshold levels. Different release parameters were better informed by different types of sensor information: identifying the source location depends significantly on the signaling-on information, whereas the mass estimation benefits from signaling-off information. The performance trade-offs between different characteristics were mapped. For example, for the conditions studied, slower, more-accurate sensors performed better than faster, less-accurate sensors. Higher



threshold signals behave as more selective filters and reduce the overall uncertainty to a greater level as compared to lower threshold signals.

Chapter 5 introduces the concept of using contaminant mixing and transport time scales to diagnose, interpret, and improve the performance of sensor systems. The time scales are used to help explain the modeling errors that arise when using a multizone modeling approach and the implications of the transport and removal time scales for extracting useful information about a release from concentration measurements. The chapter uses the time-scale concepts to contrast the opportunities and challenges in determining the release location versus determining the mass released.

An empirical time-varying likelihood function is developed to represent the modeling errors. Two sets of investigations are conducted. The first study blinds the algorithm to portions of the data immediately following the release. This exercise evaluates the value of the early data and explores how its removal affects the estimation of the release parameters. The second study investigates the relationships among the time scales and sensor network density, placement and sensor response time.

The examples illustrate the time-scale concepts and offer insight into how sensor characteristics and network architecture can be exercised such that the sensor system can characterize the release location and contaminant mass. Data acquired shortly after a release event are important for identifying the release location; ideally, these data should be acquired within the interzonal time scale. Although the early data are error-prone because the multizone model does not account for finite intrazonal mixing times, the algorithm could still extract relevant information from these early data and identify the release location. In contrast to the release location, information pertaining to the released

mass remained available to a monitoring system up to several multiples of the characteristic removal time scale. Thus, the latter data still contain information that are relevant to determining the release mass.

As expected, sensor placement became less relevant for more densely populated networks and was more influential for sparser networks and for systems with shorter transport time scales. Nevertheless, a majority of networks with as few as 3 sensors (for Experiment 4) and 4 sensors (for Experiment 1) could identify the release location to a probability exceeding 90%. Sensor response time appears to have less impact on overall performance for release conditions with longer mixing time scales. Comparing the estimation of release mass and the identification of release location, sensor placement is less critical and fewer sensors may be required to characterize the mass to a reasonable level of uncertainty. For both experiments, networks with as few as 3 sensors could estimate the released mass with a 90% uncertainty reduction. The relationships identified can serve to guide the selection of useful sensor characteristics and also to facilitate robust placements of sensors in the network.

Chapter 6 considers the fusion of information from a heterogeneous mix of sensors. The chapter has three distinct parts. A background section includes a discussion regarding existing systems that control thermal conditioning and air distribution systems in buildings. These systems offer potential information that can be used for heterogeneous sensor fusion. The second part is a conceptual treatment of fusing information from heterogeneous sensors. Physics-oriented and algorithm-oriented frameworks are described. The physics framework suggests that the degree of influence of the sensed parameter on the contaminant transport should be considered in selecting

sensors. The algorithm framework categorizes sensors based on two classifications: their role in the model (whether it is an input or output, or both as in the case of multiply connected models), and their dynamical state. The fusion of information from several sensors is described at a conceptual level, using both frameworks and including implementation issues that would arise within the BMC algorithm.

The third section of Chapter 6 demonstrates the concepts of sensor fusion, using door-position sensors in the case study as an example. A door-position fusion algorithm is developed that adjusts the prior weights of the library realizations based on the degree of match between the door sensors and library parameter sets. The algorithm includes parameters that can be tuned to address sensor and library uncertainty.

Heterogeneous networks combining contaminant with door-position sensors are constructed and evaluated using the case-study data. The case study experiments reveal nominal effects from incorporating door-position sensors in the sensor network. The weak influence is a result of the experimental conditions: door positions have little impact on the overall contaminant transport when an air-handling unit is operating. Simulated data were generated to explore additional releases in which door-position sensors were anticipated to have a greater degree of influence on contaminant transport. These scenarios included different release locations, duration, and airflow conditions with the air handling unit both on and off.

Door sensors were found to be consistently beneficial in cases where the door positions exert a substantial degree of influence on transport and could improve the identification of the source by as much as 20%. In fact, with the addition of door-position sensors some networks could identify the release location to a 90%

probability within 11 min when networks without door-position sensors were unable to. The median released mass was less impacted but door-position sensors generated 80% confidence intervals of the released mass that were as much as 85% narrower. While door-position sensors could not compensate for poor sensor placement, they did enable the sensor system to characterize the release to lower levels of uncertainty and with less computational effort. In some cases, door-position sensors are nominally, but unexpectedly helpful; this was found to be a statistical effect that results from a post-fusion reduced library. However, the performance advantage is larger when there is a real influence of door position on contaminant transport.

## **7.2 Specific research suggestions**

### **7.2.1 Chapter specific**

This dissertation has answered several questions that are relevant to the design of real-time indoor contaminant sensor systems. As this line of research is relatively new and as the systems being investigated are complex along several dimensions, many questions remain unanswered. Chapter-specific questions and suggested research are highlighted in this section.

Chapter 4 demonstrated an important point: sensors should be selected within the context of system-level goals and performance. However, Chapter 4 evaluated only one type of sensor: a single-level threshold sensor. For some release scenarios, multilevel sensors may be required. In fact, the findings showed benefits and trade-offs between low-level and high-level threshold levels: a multilevel sensor may offer system-level

performance advantages of both types of sensor. Such sensors should be evaluated in a future study.

There are many straightforward extensions to the studies in Chapter 4 that would be beneficial. One limitation of the work reported is that a fully instrumented network was investigated. The trade-off between the number of sensors in a network and characteristics such as sensor error-rate would be informative. Networks that mix coarse-threshold sensors with more refined concentration sensors would also be useful for system design. One additional limitation of Chapter 4 is that the investigations assumed a known release time. An unknown release time should be assumed in future investigations of threshold-level sensors, as was done in Chapters 5 and 6.

There are many methodological improvements associated with the design of the likelihood function that could help improve the overall performance of a Bayes Monte Carlo sensor system that uses threshold style sensors. One question is this: whether a concentration-based likelihood function is the best way to extract information from threshold sensors. Would a time-based likelihood function be able to better represent the uncertainty of when a signal crosses a threshold level? If so, transforming the algorithm to one based on the time of threshold crossing could yield additional computational benefits, since the library could store the relevant information in a more consolidated form.

The current updating procedure does not treat oscillations about the threshold level effectively. Oscillations are currently a liability since neither the likelihood function nor the multizone model account for the information embedded in this behavior. An oscillation indicates that the true concentration is most likely near the threshold level,

rather than far below or above it. For sensor systems based on threshold sensors, this information adds a potentially important clue for characterizing release events. A more sophisticated likelihood function Bayesian updating scheme may use these signals more effectively than the current approach.

Chapter 5 discusses relevant mixing and transport time scales and how they can be related to the overall system performance. The specific theories are based on the behavior of overhead air-distribution systems. The time-scale concepts are relevant to other building airflow systems, such as displacement ventilation. The ideas introduced in this chapter should be further developed and applied for exploring sensor networks in buildings with alternative ventilation configurations.

An interesting follow-up study would be to apply the lessons of Chapter 5 to explain the behavior of the networks assessed Chapter 4, for example, by performing similar blinding experiments. What if, for example, a threshold sensor's response time exceeds the interzonal mixing times? What can the sensor system learn? Can the time scales be used to define optimal characteristics for all types of sensors, including threshold-level sensors?

Certain results from Chapter 5 suggest the need for more sophisticated likelihood functions. The time-varying likelihood function should be adapted to consider the effect of release location. Clearly, the magnitude of measurement errors in the source zone are likely to exceed the measurement errors in all other zones. However, the current empirical likelihood function treats all zones equally. While the likelihood function worked well for Experiment 1, in which the release occurred at the return duct intake, it was not ideal for treating Experiment 4, where the release occurred in the center of a

room. Devising a source-based likelihood function may not be an easy task, as the source characteristics are unknown a priori. The effect of source duration should also be considered. New experimental data in which release duration varied would be particularly useful for this purpose.

The need for a source-based likelihood function arises because the existing likelihood function does not treat all model-specification errors optimally. Ideally, a likelihood function would self-diagnose its behavior based on estimated probabilities. If the posterior probabilities of certain realizations change too rapidly, the likelihood function may be assigning too much confidence to the data. Similarly, if the posterior probabilities change very slowly, the likelihood function may be assigning too little confidence to the data. A feedback-based likelihood function may alleviate this problem. Such a likelihood function could adjust the level of confidence assigned to the data based on some metric of sensor-system performance, such as the rate of change of the posterior probabilities. Such an approach might lead to a more nearly optimal algorithm that can adapt to changing conditions.

Chapter 6 addresses heterogeneous sensor systems on a conceptual and empirical level. While physics and algorithm concepts are discussed generally and within the Bayes Monte Carlo context, only one relatively simple fusion example is developed. The fusion of other sensors would be worth exploring. In particular, a future study that explores the use of sensors available through an EMCS, such as pressure, control signals, and airflow measurements, is suggested. Information from such sensors may help make the system more robust and avoid failure due to a poor a-priori characterization of airflow and contaminant transport. Overall, many sensors in addition to those for monitoring

door-position and contaminant level could help the system account for unpredicted uncertainties, such as changing operating conditions. The benefits for monitoring systems of incorporating alternative heterogeneous sensor types should be explored more thoroughly.

Methodological improvements of the Bayes Monte Carlo method are also suggested. Dynamic library generation and advanced fusion methods may be necessary to account for measurements that are classified as pseudo-static model inputs. These variables change, so they cannot be treated as fixed inputs, but the changes are slow enough that their state can be determined and incorporated into the BMC algorithm in advance of a release. Advanced filtering methods may also help extend the utility of BMC for a wider range of conditions. In this concept, several sets of libraries would be generated before monitoring, and advanced filtering methods would fuse real-time measurements of variables in the pseudo-static or dynamic model input categories.

Benign tracer-gas measurements present an additional opportunity for sensor fusion that merits further development. In this concept, a benign tracer gas would be deliberately released at intervals over a range of building operating conditions and the measured concentrations resulting from those releases would help populate or refine the library used in the real-time assessment. This strategy would help to compensate for modeling errors, changing operating conditions and, in general, other unanticipated uncertainties. In addition to augmenting or generating a more realistic library of releases, benign tracer gas releases could be used to periodically to test the sensor system to verify proper operation.



### 7.2.2 Common to all chapters

The three core chapters of this dissertation have certain common limitations that should be considered in future research efforts. One unknown is how a sensor system would perform if a release is characterized by a set of parameters that lie outside the bounds considered in the generation of the library. For example, how well will the algorithm perform if the mass released is less than (or greater than) the minimum (or maximum) in the library? How will the algorithm perform if the parameter resolution is too coarse, or if an inadequate number of realizations populate the library? Indirectly, the research in this dissertation supports the hypothesis that a well designed library is essential for achieving robust system performance. Precisely what constitutes a well-designed library is a topic that should be studied in future research.

Another library issue that has not been considered is the effect of different prior distributions assumed in the sampling of the release parameters. The prior distributions determine the regions of the input parameters that are coarsely or densely sampled and control the accuracy at which releases are characterized when the release conditions are within the bounds considered in the library. The prior assumptions in the library also influence the efficiency at which the release is characterized.

One important limitation of this dissertation is that the concepts have been evaluated using a case-study approach that is populated with limited data from one relatively small building. Additional field studies in larger, more complex buildings are warranted to establish a test-bed for further development and refinement of the sensor-system concepts introduced here.

An additional limitation is that a multizone model has been used to link releases to concentrations. Clearly, this kind of model is sometimes inappropriate. Given its foundational assumptions, the multizone model will likely perform best for buildings that are designed to use mechanical ventilation in a way that promotes rapid intrazonal mixing. The performance of a multizone model is likely to not be as good in buildings dominated by natural ventilation (or when the mechanical unit is turned off). Other modeling approaches for linking release events to contaminant concentrations should be evaluated. These may include combined computational fluid dynamics and multizone models, network type models (described in Chapter 2), and physics-free black box models, such as neural networks.

### **7.3 General future research suggestions**

#### **7.3.1 The next generation of high-risk indoor sensor systems**

Many other questions remain that are beyond the scope of the topics addressed in the dissertation. The next generation of high-risk indoor sensor systems may require a more sophisticated algorithm. This dissertation uses a simple implementation of Bayes Monte Carlo for solving the inverse problem in real time. More advanced implementation strategies for BMC were proposed in the context of Chapter 6; hybrid methods that combine elements from Markov Chain Monte Carlo or Sequential Monte Carlo methods could be beneficial. A hybrid MCMC-BMC method could produce an approach that captures the best characteristics of each method.

Fault-tolerant methods are required for successful operation of a sensor system, as illustrated by the results in this dissertation. Airflow and operating conditions could change unexpectedly. Sensor fusion methods may help to prevent some unanticipated uncertainties from causing a monitoring system to fail. However methods dedicated specifically to fault prevention may also increase the robustness of the sensor system. Any number of sensor failures, such as data losses, drift, or saturation, could result in a signal that could cause the algorithm to fail. Such errors may not conform to a standard probability distribution.

Several additional design questions should be addressed in future work. Specific guidelines for selecting sensors, models, and algorithms based on the particular conditions of a building are needed to facilitate the sensor-system design process. With further development, the concepts from this dissertation could be used as a foundation to develop these guidelines. A set of performance metrics for sensor systems should be developed. This dissertation applies a very narrow set of metrics. An expected value of information (EVI) analysis could be conducted using these metrics and the results of the EVI analysis could be used to improve the selection of sensors and network architecture. Ultimately, sensor selection should be informed by a tangible and measurable goal, such as the number of lives that could be saved in the event of a release.

Moving beyond sensor-system performance per se, how should the information provided by a sensor system be used by first-responders and building facilities managers? How can the EMCS be used to facilitate the implementation of a response strategy? How can the system be ensured to operate robustly over the lifetime of the building? What continued commissioning strategies are needed to guarantee robust performance over

time? The next generation of sensor systems may have improved performance, reduced costs, and save more lives by addressing such questions.

### **7.3.2 Beyond high-risk indoor sensor systems**

The ideas introduced in this dissertation may be extended and applied to solve other environmental and engineering problems. These include, but are not limited to, environmental and energy monitoring, and system design. I discuss some examples.

Buildings have operational goals in addition to high-risk contaminant detection. These goals include maintaining occupant thermal comfort, indoor air quality and achieving energy efficiency. The concepts and lessons learned may be relevant to achieving these goals, such as improving the performance of the building's energy systems. A fault detection system, for example, based on the principles used in this dissertation, may reduce the energy consumption of the building's mechanical systems. More generally, the methods could be used to develop statistically-based fault detection and diagnostic systems for many other engineering processes.

Environmental monitoring applications, in addition to the one addressed in this dissertation, may also benefit from the research developed here. The methods and concepts could be used to develop outdoor high-risk real-time monitoring systems. An immediate research need is to develop monitoring methods for tracking greenhouse gas emissions (GHGs), and in particular, fugitive GHG emissions from agricultural sources. While the emission characteristics and contaminant dispersion are quite different from those addressed in this dissertation, the ideas in this dissertation may provide a basis for reducing the uncertainty of fugitive agricultural GHG emissions.

This dissertation uses a systems-based design approach. A systems-approach may be relevant to a variety of engineering design problems that require a robust treatment of uncertainty. Sustainable building design is one such example. For example, the respective benefits to energy performance from improvements in architectural and engineering systems may not be apparent without analyzing the overall building system. This analysis is complicated by the uncertainties that characterize the processes in the building; the approach in this dissertation can facilitate this analysis. In fact, sustainable building design extends beyond the energy performance of a building and includes the resource intensity of a building over the entire life cycle of the building. Understanding the true sustainability of a building design requires analyzing an even larger number of processes that can be facilitated by a systems approach.

A system-level analyses can be useful for making non-engineering decisions, such as environmental or energy policy decisions. The success of any one policy decision will be based on many factors, each of which are characterized by uncertainty. The methods used in this dissertation offers contributions towards the design and evaluation of environmental policy decisions.

At its most fundamental level, this dissertation models a complex environmental system and evaluates design decisions under conditions of uncertainty. The design problems and scientific questions that fit this fundamental definition are virtually limitless. This dissertation, while generating more questions than answers, nevertheless provides insights, analyses, models and tools that are relevant to a broad array of environmental and systems-based design problems.

# References

- Allocca C, Chen Q, Glicksman L, 2003. Design analysis of single-sided natural ventilation. *Energy and Buildings* 35: 785-795.
- Arulampalam MS, Maskell S, Gordon N, Clapp T, 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50: 174-188.
- Arvelo J, Brandt A, Roger RP, Saksena A, 2002. An enhanced multizone model and its application to optimum placement of CBW sensors. *ASHRAE Transactions* 108 (2): 818-826.
- ASHRAE, 1993. *ASHRAE Handbook Fundamentals*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.
- Aslam J, Butler Z, Constantin AF, Crespi AV, Cybenko AG, Rus AD, 2003. Tracking a moving object with a binary sensor network. In: *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*. Los Angeles, California, USA; pp. 150-161.
- ASTM, 2003. *Standard Test Method for Determining Air Leakage Rate by Fan Pressurization*. West Conshohocken: American Society for Testing and Materials.
- Axley JW, Emmerich SJ, 2002. A method to assess the suitability of a climate for natural ventilation of commercial buildings. In: *Proceedings of the 9th International Conference on Indoor Air Quality and Climate - Indoor Air 2002*. Monterey, California; pp 854–859.

- Axley JW, Emmerich SJ, Walton GN, 2002. Modeling the performance of a naturally ventilated commercial building with a multizone coupled thermal/airflow simulation tool. *ASHRAE Transactions* 108:1260-1275.
- Barrettino D, Graf M, Zimmermann M, Hagleitner C, Hierlemann A, Baltes H, 2003. A microhotplate-based monolithic CMOS gas sensor array. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*; Vol 4, pp. 852-855.
- Baughman AV, Gadgil AJ, Nazaroff WW, 1994. Mixing of a point-source pollutant by natural-convection flow within a room. *Indoor Air* 4: 114-122.
- Bedoya G, Jutten C, Bermejo S, Cabestany J, 2004. Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation. In: *Sensors for Industry Conference, 2004. Proceedings the ISA/IEEE*; pp. 149-154.
- Bennett DH, McKone TE, Evans JS, Nazaroff WW, Margni MD, Jolliet O, Smith KR, 2002. Defining intake fraction. *Environmental Science & Technology* 36: 206A-211A.
- Bergin MS, Milford JB, 2000. Application of Bayesian Monte Carlo analysis to a Lagrangian photochemical air quality model. *Atmospheric Environment* 34: 781-792.
- Bourgeois W, Romain AC, Nicolas J, Stuetz RM, 2003. The use of sensor arrays for environmental monitoring: Interests and limitations. *Journal of Environmental Monitoring* 5: 852-860.
- Brager G, Paliaga G, De Dear R, 2004. Operable windows, personal control, and occupant comfort. *ASHRAE Transactions* 110: 17-35.

- Brand KP, Small MJ, 1995. Updating uncertainty in an integrated risk assessment: Conceptual framework and methods. *Risk Analysis* 15: 719-731.
- Cano-Ruiz JA, Kong D, Balas RB, Nazaroff WW, 1993. Removal of reactive gases at indoor surfaces: Combining mass transport and surface kinetics. *Atmospheric Environment* 27A: 2039-2050.
- Carey P, Etheridge D, 1999. Direct wind tunnel modelling of natural ventilation for design purposes. *Building Services Engineering Research and Technology* 20: 131-140.
- Casella G, George EI, 1992. Explaining the Gibbs sampler. *American Statistician* 46:167-174.
- Chan WR, Nazaroff WW, Price PN, Sohn MD, Gadgil AJ, 2005. Analyzing a database of residential air leakage in the United States. *Atmospheric Environment* 39: 3445-3455.
- Chib S, Greenberg E, 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49: 327-335.
- Coward CW, 1990. System effect - a design fundamental. *ASHRAE Journal* 32 (5): 44-49.
- Csanady G, 1973. *Turbulent Diffusion in the Environment*. Dordrecht, Holland: Springer.
- Cullen AC, Frey HC, 1999. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. New York: Plenum Press.



- Dakins ME, Toll JE, Small MJ, Brand KP, 1996. Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis* 16:67-79.
- Dascalaki E, Santamouris M, Argiriou A, Helmis C, Asimakopoulos D, Papadopoulos K, Soilemes A, 1996. On the combination of air velocity and flow measurements in single sided natural ventilation configurations. *Energy and Buildings* 24:155-165.
- Dhara VR, Dhara R, Acquilla SD, Cullinan P, 2002. Personal exposure and long-term health effects in survivors of the Union Carbide disaster at Bhopal. *Environmental Health Perspectives* 110:487-500.
- Dilks DW, Canale RP, Meier PG, 1992. Development of Bayesian Monte-Carlo techniques for water-quality model uncertainty. *Ecological Modelling* 62:149-162.
- Doucet A, Godsill S, Andrieu C, 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 10:197-208.
- Doucet A, de Freitas N, Gordon N, 2001. Sequential Monte Carlo methods in practice. In: *Statistics for Engineering and Information Science* (Jordan M, Lauritzen SL, Lawless JF, Nair V, eds). New York: Springer-Verlag.
- Dowd M, Meyer R, 2003. A Bayesian approach to the ecosystem inverse problem. *Ecological Modelling* 168:39-55.
- Drescher AC, Lobascio C, Gadgil AJ, Nazaroff WW, 1995. Mixing of a point-source indoor pollutant by forced-convection. *Indoor Air* 5:204-214.
- Fan Y, 1995. CFD modelling of the air and contaminant distribution in rooms. *Energy and Buildings* 23:33-39.

- Ferziger JH, Peric M, 2002. *Computational Methods for Fluid Dynamics*, 3rd ed. Berlin: Springer-Verlag.
- Feustel HE, Rayner-Hooson A, 1990. *COMIS Fundamentals*. Report LBNL-28560, Lawrence Berkeley National Laboratory, Berkeley, CA.
- Feustel HE, 1999. COMIS - An international multizone air-flow and contaminant transport model. *Energy and Buildings* 30:3-18.
- Finlayson EU, Gadgil AJ, Thatcher TL, Sextro RG, 2004. Pollutant dispersion in a large indoor space. Part 2: Computational fluid dynamics predictions and comparison with a scale model experiment for isothermal flow. *Indoor Air* 14:272-283.
- Fisk WJ, De Almeida AT, 1998. Sensor-based demand-controlled ventilation: A review. *Energy and Buildings* 29:35-45.
- Freeze RA, Massmann J, Smith L, Sperling T, James B, 1990. Hydrogeological decision-analysis 1. A framework. *Ground Water* 28:738-766.
- Gadgil AJ, Finlayson EU, Fischer ML, Price PN, Thatcher TL, Craig MJ, Hong KH, Housman J, Schwalbe CA, Wilson D, Wood JE, Sextro RG, 2000. Pollutant transport and dispersion in large indoor spaces: A status report for the large space effort of the interiors project. Report LBNL-44791, Lawrence Berkeley National Laboratory, Berkeley, CA.
- Gadgil AJ, Lobscheid C, Abadie MO, Finlayson EU, 2003. Indoor pollutant mixing time in an isothermal closed room: An investigation using CFD. *Atmospheric Environment* 37:5577-5586.
- Gan G, Riffat S, 2004. CFD modelling of air flow and thermal performance of an atrium integrated with photovoltaics. *Building and Environment* 39:735-748.

- Gelfand AE, 2000. Gibbs sampling. *Journal of the American Statistical Association* 95:1300-1304.
- Gelman A, Carlin JB, Stern HS, Rubin DB, 2004. *Bayesian Data Analysis*, 2nd ed. Boca Raton: Chapman & Hall / CRC.
- Gilks WR, Berzuini C, 2001. Following a moving target - Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 63:127-146.
- Godsill S, Clapp T, 2001. Improvement strategies for Monte Carlo particle filters. In: *Sequential Monte Carlo Methods in Practice* (Doucet A, de Freitas N, Gordon N, eds): New York: Springer-Verlag, pp. 139–158.
- Granderson J, Agogino A, Wen Y, Goebel K, 2004. Towards demand-responsive intelligent lighting with wireless sensing and actuation. In: *Proceedings of the IESNA 2004 Annual Conference*. Illuminating Engineering Society of North America, Tampa, FL.
- Grate J, Rose-Pehrsson S, Venezky D, Klusty M, Wohltjen H, 1993. Smart sensor system for trace organophosphorus and organosulfur vapor detection employing a temperature-controlled array of surface acoustic wave sensors, automated sample preconcentration, and pattern recognition. *Analytical Chemistry* 65:1868-1881.
- Hall DL, Llinas J, 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE* 85:6-23.
- Harmon R, Challenor P, 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecological Modelling* 101:41-59.

- Heinonen J, Kosonen R, 2000. Hybrid ventilation concepts in commercial buildings - Indoor air quality and energy economy perspective. In: Proceedings of the Healthy Buildings. Espoo, Finland; Vol 4, pp 517-522.
- Huang H, Haghghat F, Blondeau P, 2006. Volatile organic compound (VOC) adsorption on material: influence of gas phase concentration, relative humidity and VOC type. *Indoor Air* 16: 236-247.
- Hunter GW, Neudeck PG, Liu CC, Ward B, Wu QH, Dutta P, Frank M, Trimbol J, Fulkerson M, Patton B, Makel D, Thomas V, 2002. Development of chemical sensor arrays for harsh environments and aerospace applications. In: *Sensors, 2002. Proceedings of IEEE*; Vol 2, pp. 1126-1133.
- Iman R, Davenport J, Zeigler D, 1980. Latin Hypercube Sampling (Program Users's Guide): Report 79-1473, Sandia National Laboratory, Albuquerque, NM.
- Ishida H, Nakamoto T, Moriizumi T, Kikas T, Janata J, 2001. Plume-tracking robots: A new application of chemical sensors. *The Biological Bulletin* 200:222-226.
- Jablonski A, Klempous R, Licznarski B, 2003. Diversified approach to methodology and technology in distributed intelligent building systems. In: *Computer Aided Systems Theory - Eurocast 2003*. Berlin: Springer; pp. 174-184.
- Jayaraman B, Finlayson EU, Sohn MD, Thatcher TL, Price PN, Wood EE, Sextro RG, Gadgil AJ, 2006. Tracer gas transport under mixed convection conditions in an experimental atrium: Comparison between experiments and CFD predictions. *Atmospheric Environment* 40:5236-5250.
- Jiang Y, Chen Q, 2003. Buoyancy-driven single-sided natural ventilation in buildings with large openings. *International Journal of Heat and Mass Transfer* 46:973-988.

- Jiang Y, Alexander D, Jenkins H, Arthur R, Chen Q, 2003. Natural ventilation in buildings: measurement in a wind tunnel and numerical simulation with large-eddy simulation. *Journal of Wind Engineering & Industrial Aerodynamics* 91:331-353.
- Kass RE, Carlin BP, Gelman A, Neal RM, 1998. Markov chain Monte Carlo in practice: A roundtable discussion. *American Statistician* 52:93-100.
- Katipamula S, Brambley MR, 2005a. Methods for fault detection, diagnostics, and prognostics for building systems - A review, part I. *HVAC&R Research* 11:3-25.
- Katipamula S, Brambley MR, 2005b. Methods for fault detection, diagnostics, and prognostics for building systems - A review, part II. *HVAC&R Research* 11:169-187.
- Kintner-Meyer M, Conant R, 2004. Opportunities of wireless sensors and controls for building operation. In: 2004 ACEEE Summer Study on Energy Efficiency in Buildings. American Council for an Energy-Efficient Economy, Washington, DC; pp. 3.139-3.152.
- Klepeis NE, Nazaroff WW, 2006a. Mitigating residential exposure to secondhand tobacco smoke. *Atmospheric Environment* 40: 4408-4422.
- Klepeis NE, Nazaroff WW, 2006b. Modeling residential exposure to secondhand tobacco smoke. *Atmospheric Environment* 40: 4393-4407.
- Koracin D, Panorska A, Isakov V, Touma JS, Swall J, 2007. A statistical approach for estimating uncertainty in dispersion modeling: An example of application in southwestern USA. *Atmospheric Environment* 41:617-628.
- Lee PM, 2004. *Bayesian Statistics - An Introduction*, 3rd ed. London: Arnold.

- Linden P, 1999. The fluid mechanics of natural ventilation. *Annual Review of Fluid Mechanics* 31: 201-238.
- Liu DL, Nazaroff WW, 2001. Modeling pollutant penetration across building envelopes. *Atmospheric Environment* 35:4451-4462.
- Liu JS, 2001. *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag.
- Liu Z, Kim AK, 2003. Review of recent developments in fire detection technologies. *Journal of Fire Protection Engineering* 13:129-151.
- Lo SC, Ma HW, Lo SL, 2005. Quantifying and reducing uncertainty in life cycle assessment using the Bayesian Monte Carlo method. *Science of the Total Environment* 340:23-33.
- Lorenzetti DM, 2001. Assessing multizone airflow software. Report LBNL-47653, Lawrence Berkeley National Laboratory, Berkeley, CA.
- Lorenzetti DM, 2002a. Assessing multizone airflow simulation software. Report LBNL-49578, Lawrence Berkeley National Laboratory, Berkeley, CA.
- Lorenzetti DM, 2002b. Computational aspects of nodal multizone airflow systems. *Building and Environment* 37:1083-1090.
- Malve O, Laine M, Haario H, 2005. Estimation of winter respiration rates and prediction of oxygen regime in a lake using Bayesian inference. *Ecological Modelling* 182:183-197.
- McDowell T, Emmerich S, Thornton J, Walton G, 2003. Integration of airflow and energy simulation using CONTAM and TRNSYS. *ASHRAE Transactions* 109: 757-770.

- McWilliams JA, 2002. Review of airflow measurement techniques. Report LBNL-49747, Lawrence Berkeley National Laboratory, Berkeley, CA.
- Miller SL, Leiserson K, Nazaroff WW, 1997. Nonlinear least-squares minimization applied to tracer gas decay for determining airflow rates in a two-zone building. *Indoor Air* 7:64-75.
- Miller SL, Nazaroff WW, 2001. Environmental tobacco smoke particles in multizone indoor environments. *Atmospheric Environment* 35: 2053-2067.
- Mills A, 1995. Heat and mass transfer. Chicago: Irwin.
- Mora L, Gadgil AJ, Wurtz E, 2003. Comparing zonal and CFD model predictions of isothermal indoor airflows to experimental data. *Indoor Air* 13:77-85.
- Murray DM, Burmaster DE, 1995. Residential air exchange rates in the United States: empirical and estimated parametric distributions by season and climatic region. *Risk Analysis* 15: 459-465.
- Nazaroff WW, Gadgil AJ, Weschler CJ, 1993. Critique of the use of deposition velocity in modeling indoor air quality, *Modeling Indoor Air Quality and Exposure*, ASTM STP 1205, Nagda NL (Ed.), American Society for Testing and Materials, Philadelphia, pp. 81-104.
- Nazaroff WW, Weschler CJ, 2001. Indoor air and the public good. *Indoor Air* 11:143-144.
- Nazaroff WW, 2004. Indoor particle dynamics. *Indoor Air* 14 (Supplement 7): 175-183.
- Nicas M, 1996. Estimating exposure intensity in an imperfectly mixed room. *AIHA* 57:542-550.

- Nielsen PV, 2004. Computational fluid dynamics and room air movement. *Indoor Air* 14:134-143.
- Ott WR, Klepeis NE, Switzer P, 2003. Analytical solutions to compartmental indoor air quality models with application to environmental tobacco smoke concentrations measured in a house. *Journal of the Air & Waste Management Association* 53: 918-936.
- Persily AK, Gorfain J, Brunner G, 2006. Survey of ventilation rates in office buildings. *Building Research and Information* 34:459-466.
- Piette MA, Kinney SK, Haves P, 2001. Analysis of an information monitoring and diagnostic system to improve building operations. *Energy and Buildings* 33:783-791.
- Politis K, Robertson L, 2004. Bayesian updating of atmospheric dispersion after a nuclear accident. *Journal of the Royal Statistical Society Series C-Applied Statistics* 53:583-600.
- Price PN, A S, Chan WR, 2006. Indoor outdoor air leakage of apartments and commercial buildings. California Energy Commission, PIER Energy-Related Environmental Research Program. CEC-500-2006-111.
- Qian SS, Stow CA, Borsuk ME, 2003. On Monte Carlo methods for Bayesian inference. *Ecological Modelling* 159:269-277.
- Reichert P, Schervish M, Small MJ, 2002. An efficient sampling technique for Bayesian inference with computationally demanding models. *Technometrics* 44:318-327.



- Richmond-Bryant J, Eisner AD, Brixey LA, Wiener RW, 2006. Short-term dispersion of indoor aerosols: Can it be assumed the room is well mixed? *Building and Environment* 41:156-163.
- Robert CP, Casella G, 2004. *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer.
- Ross SM, 2003. *Introduction to Probability Models*, 8th ed. San Diego: Academic Press.
- Rubin D, 1987. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. *Journal of the American Statistical Association* 82:543–546.
- Russell RA, 2001. Tracking chemical plumes in constrained environments. *Robotica* 19:451-458.
- Sakr W, Weschler C, Fanger P, 2006. The impact of sorption on perceived indoor air quality. *Indoor Air* 16: 98-110.
- Sanderson AC, Hombal V, Fries DP, Broadbent HA, Wilson JA, Bhanushali PI, Ivanov SZ, Luther M, Meyers S, 2006. Distributed Environmental Sensor Network: Design and Experiments. In: *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*; pp. 79-84.
- Seifert J, Li Y, Axley J, Rosler M, 2006. Calculation of wind-driven cross ventilation in buildings with large openings. *Journal of Wind Engineering and Industrial Aerodynamics* 94: 925-947.
- Seinfeld JH, Pandis SN, 1998. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. New York: Wiley.

- Sextro RG, Daisey JM, Feustel HE, Dickerhoff DJ, Jump C, 1999. Comparison of modeled and measured tracer gas concentrations in a multizone building. In: Proceedings of the 8th International Conference on Indoor Air Quality and Climate - Indoor Air 99, Indoor Air 99. Edinburgh, Scotland; Vol 1, pp. 696-701.
- Singer BC, Revzan KL, Hotchi T, Hodgson AT, Brown NJ, 2004. Sorption of organic gases in a furnished room. *Atmospheric Environment* 38:2483-2494.
- Singer BC, Hodgson AT, Destailats H, Hotchi T, Revzan KL, Sextro RG, 2005. Indoor sorption of surrogates for sarin and related nerve agents. *Environmental Science & Technology* 39:3203-3214.
- Singer BC, Hodgson AT, Hotchi T, Ming KY, Sextro RG, Wood EE, Brown NJ, 2007. Sorption of organic gases in residential rooms. *Atmospheric Environment* 41:3251-3265.
- Small MJ, Fischbeck PS, 1999. False precision in Bayesian updating with incomplete models. *Human and Ecological Risk Assessment* 5:291-304.
- Smith J, French S, 1993. Bayesian updating of atmospheric dispersion models for use after an accidental release of radioactivity. *Statistician* 42:501-511.
- Sohn MD, Small MJ, 1999. Parameter estimation of unknown air exchange rates and effective mixing volumes from tracer gas measurements for complex multi-zone indoor air models. *Building and Environment* 34:293-303.
- Sohn MD, Small MJ, Pantazidou M, 2000. Reducing uncertainty in site characterization using Bayes Monte Carlo methods. *Journal of Environmental Engineering-ASCE* 126:893-902.

- Sohn MD, Reynolds P, Singh N, Gadgil AJ, 2002a. Rapidly locating and characterizing pollutant releases in buildings. *Journal of the Air & Waste Management Association* 52:1422-1432.
- Sohn MD, Reynolds P, Gadgil AJ, Sextro RG, 2002b. Rapidly locating sources and predicting contaminant dispersion in buildings. In: *Proceedings of the 9th International Conference on Indoor Air Quality and Climate - Indoor Air 2002*. Monterey, California; Vol 4, pp 211-216.
- Sohn MD, Apte MG, Sextro RG, Lai ACK, 2007. Predicting size-resolved particle behavior in multizone buildings. *Atmospheric Environment* 41: 1473-1482.
- Stewart J, Ren ZG, 2006. COwZ - A subzonal indoor airflow, temperature and contaminant dispersion model. *Building and Environment* 41:1631-1648.
- Swall LJ, Davis JM, 2006. A Bayesian statistical approach for the evaluation of CMAQ. *Atmospheric Environment* 40:4883-4893.
- Tan G, Glicksman LR, 2005. Application of integrating multi-zone model with CFD simulation to natural ventilation prediction. *Energy and Buildings* 37:1049-1057.
- Taylor S, 2000. Comparing economizer relief systems. *ASHRAE Journal* 42 (9): 33-42.
- Thomson LC, Hirst B, Gibson G, Gillespie S, Jonathan P, Skeldon KD, Padgett MJ, 2007. An improved algorithm for locating a gas source using inverse methods. *Atmospheric Environment* 41:1128-1134.
- Wang L, Chen Q, 2007. Theoretical and numerical studies of coupling multizone and CFD models for building air distribution simulations. *Indoor Air* 17: 348-361.
- Weschler CJ, 2004. Chemical reactions among indoor pollutants: what we've learned in the new millennium. *Indoor Air* 14: 184-194.

- Whicker JJ, Rodgers JC, Moxley JS, 2003. A quantitative method for optimized placement of continuous air monitors. *Health Physics* 85:599-609.
- Wide P, Winquist F, Driankov D, 1997. Air-quality sensor system with fuzzy classification. *Measurement Science and Technology* 8:138-146.
- Wilson D, Kiel D, 1990. Gravity driven counterflow through an open door in a sealed room. *Building and Environment* 25:379-88.
- Won D, Corsi R, Rynes M, 2001. Sorptive Interactions between VOCs and Indoor Materials. *Indoor Air* 11: 246-256.
- Wurtz E, Nataf JM, Winkelmann F, 1999. Two- and three-dimensional natural and mixed convection simulation using modular zonal models in buildings. *International Journal of Heat and Mass Transfer* 42:923-940.
- Yanagisawa N, Morita H, Nakajima T, 2006. Sarin experiences in Japan: Acute toxicity and long-term effects. *Journal of the Neurological Sciences* 249:76-85.
- Yang C, Demokritou P, Chen Q, Spengler J, 2001. Experimental validation of a computational fluid dynamics model for IAQ applications in ice rink arenas. *Indoor Air* 11:120-126.
- Zhao F, Guibas L, 2004. *Wireless Sensor Networks: An Information Processing Approach*. San Francisco: Morgan Kaufman Publications.