Title

Influence of indoor transport and mixing time scales on the performance of sensor systems for characterizing contaminant releases

Authors

Sreedharan, Piya
Sohn, Michael D.
Nazaroff, William W.
et al.

Publication Date

2007-04-22

Influence of indoor transport and mixing time scales on the performance of sensor systems for characterizing contaminant releases

Priya Sreedharan [a,b,*], Michael D. Sohn [a], William W Nazaroff [a,c], and Ashok J. Gadgil [a]

[a] *Indoor Environment Department, Lawrence Berkeley National Laboratory*

[b] *Department of Mechanical Engineering, University of California, Berkeley*

[c] *Department of Civil and Environmental Engineering, University of California, Berkeley*

[*] Corresponding author. Indoor Environment Department, Lawrence Berkeley National Laboratory, One Cyclotron Road, Mail Stop 90R3058, Berkeley, CA 94720, USA. Tel.: + 1 510 486 4093; fax: + 1 510 486 6658. *E-mail address*: psreedharan@lbl.gov (P. Sreedharan)

Abstract

Optimizing real-time sensor systems to detect and identify relevant characteristics of an indoor contaminant event is a challenging task. The interpretation of incoming sensor data is confounded by uncertainties in building operation, in the forces driving contaminant transport, and even in the physical parameters governing the transport. In addition, any simulation tools used by the sensor interpretation algorithm will introduce modeling uncertainties. This paper explores how the time scales inherent in contaminant transport influence the information that can be extracted from real-time sensor data. In particular, we identify three time scales (within room mixing, room-to-room transport, and removal from the building) and study how they affect the ability of a Bayesian Monte Carlo (BMC) sensor interpretation algorithm to identify the release location and release mass from a set of experimental data, recorded in a multi-floor building. The research shows that some limitations in the BMC approach do not depend on details of the models or the algorithmic implementation, but rather on the physics of contaminant transport.

This in turn has implications for the design of sensor systems.

Keywords: Sensor system; multizone model; contaminant transport; inverse problem; Bayes Monte Carlo; buildings; time scales

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 sensor system. Ideally, a sensor system should not only detect an event, but should also provide information, such as location and strength of the source, and prediction of the future concentrations, for guiding an effective response in real-time. An integrated system design approach that collectively considers the contaminant transport physics, sensor characteristics, and sensor interpretation algorithms can help achieve these goals.

Our research group has been working on the development and evaluation of indoor sensor systems for characterizing high-risk contaminant releases. The approach is based on a two-stage Bayes Monte Carlo (BMC) algorithm. The first stage occurs before monitoring. A library of simulations with time-dependent chemical concentrations is generated using a multizone indoor air pollutant transport and fate model. Each scenario in the library is generated by sampling model inputs (such as release location and amount, weather, and mechanical ventilation conditions) from ranges of likely values. Thus each scenario represents a possible chemical release in the building. 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.

This approach to monitoring and interpreting pollutant releases was first described by Sohn et al. (2002). In that paper, the two-stage approach was demonstrated using synthetic concentration data from a hypothetical 5-room building. Subsequently, Sreedharan et al. (2006) showed that this approach could also work for interpreting data from alarm-type sensors in a real building. That effort underscored the importance of a systems perspective for selecting the optimal mix of sensor performance characteristics, such as response time and accuracy.

While a systems approach using the BMC algorithm has shown promise towards identifying optimal sensor system characteristics, several questions remain. How does the contaminant fate and transport influence the ability of a sensor system to characterize the release? How do 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? Ultimately, how can an understanding of these issues inform an optimal (or near-optimal) sensor system design?

This paper provides partial answers to these important questions. We 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. This discussion is used to explore how limitations of a multizone transport model may affect the estimation of release parameters. We 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, we develop a likelihood function, using empirical data, that aims to characterize model specification error.

The scope of this paper is restricted to instantaneous, single-point release events that occur indoors. Additional research would be required to address sensor systems designed to diagnose releases from multiple locations, of extended duration, or that may originate outside.

2. Background

2.1. Sensor system based on a Bayes Monte Carlo algorithm

The sensor system operates as follows. A contaminant is released over a short duration somewhere in, or near a building. A network of sensors measuring concentrations in near real-time operates to detect the contaminant. When any sensor measurement exceeds the detection limit, it is inferred that a release may have occurred. The sensor data are interpreted to characterize the release.

An effective sensor system design must address several issues. One issue concerns how individual sensor attributes, such as response time, sensitivity, precision, accuracy, dynamic range, and interferences, may affect system performance. A second issue concerns the network architecture, such as the number of devices and their placement. For the BMC algorithm, a third issue concerns the model used to generate the simulation library, because the model results drive the statistical processing of the sensor data and the ensuing parameter estimates. A related, fourth issue is the accuracy of the likelihood function that the BMC algorithm uses to describe the relationship between the modeled and measured values, and what impact that has on system performance.

2.2. Multizone model

In this research, we use a multizone model to generate the simulation library. A multizone model describes a building as a network of well-mixed chambers, termed *zones*, connected by flow paths. The outdoor environment is represented as an additional zone with unbounded volume. Mechanical air handling units (AHUs) may be represented as zones where outdoor air mixes with return air (i.e., with air recirculated from the zones) before being supplied back to the 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 connected them. Feustel (1999) describes the physical foundations of COMIS, the multizone model used in this research.

Multizone models have limitations. The physical assumptions in the model may result in inaccurate airflow estimates (Lorenzetti, 2002). An important concern for the present application is the assumption that a contaminant instantaneously mixes within a particular zone. A localized indoor release of a contaminant generates initially nonuniform concentrations within a zone (Baughman, et al., 1994; Drescher et al., 1995).

While computational fluid dynamics (CFD) provides higher resolution, whole-building CFD remains prohibitively expensive.

2.3. Likelihood function

As noted, the BMC monitoring scheme involves two stages. Here, we discuss the second stage, in which Bayes' rule is applied to compare the library of simulations to the observations. In this process, the probabilities of individual simulations are computed as data arrive sequentially in time. These probabilities are used to estimate values of the unknown input parameters, such as the release location and amount.

Given an array of sensor measurements, O, the probability that these correspond to simulation k, with its corresponding concentration values Y_k, is denoted as $p(Y_k|O)$. From Bayes' rule:

$$p(Y_k \mid O) = \frac{p(O \mid Y_k)\, p(Y_k)}{\sum_{i=1}^{N} p(O \mid Y_i)\, p(Y_i)} \tag{1}$$

where $p(O|Y_k)$ is the likelihood function, $p(Y_k)$ is the prior probability of Y_k (i.e., an a priori estimate), and N is the number of simulations in the library. The vector O denotes a collection of

measurements that differ by space (owing to the placement of the sensor) and time.

The likelihood function is the conditional probability of observing the data given a model prediction. Model-to-measurement mismatches due to either model or measurement error must be quantified by the likelihood function. Small and Fischbeck (1999) and Sohn et al. (2000) discuss methods for reconciling mismatches due to systematic errors in the observations.

2.4. Case study: Sudden tracer releases in a three-story building

Our study uses data obtained from tracer gas experiments conducted in one unit of a building at the Dugway Proving Grounds, Utah (Sextro et al., 1999). The interior volume is 660 m^3, with approximately 280 m^2 of total floor area on three levels. An AHU supplied 100% recirculated air to the lower two floors. Sextro et al. (1999) conducted fan pressurization tests to determine leakage rates. They also performed 12 tracer-gas experiments over a period of five days (Table 1). In each experiment, approximately 20 g of propylene was instantaneously released either in the return duct, or in an interior room. Concentrations in each room and at each staircase level were recorded at 20 s intervals for the first 20 min, and at 60 s intervals thereafter. In eleven of the twelve experiments, the AHU was operating.

Sextro et al. (1999) measured air leakage and AHU characteristics, and developed a multizone airflow and gas transport model using COMIS (Feustel, 1999). We used this COMIS 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. We 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 2. The resulting library comprised 5000 possible release scenarios, each with a simulated set of time-dependent concentrations for each sensor location.

3. Transport time scales in buildings

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

In the discussion that follows, we assume that a building is equipped with a conventional commercial building ventilation system that discharges conditioned air from overhead registers to promote rapid mixing in a room. We define time scales associated with three physical processes. The first, denoted τ_1, represents the time required for mixing in a room. Imagine that a contaminant is suddenly released at time $t = 0$ at a point within a room. Then τ_1 is defined such that for times $t << \tau_1$ 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 $\tau_1 \sim 18\, \rho^{1/3} V^{5/9} P^{-1/3}$, where ρ 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 $V = 100$ m^3, $v = 2$ m s^{-1}, and $A = 0.13$ m^2 gives $P \sim 0.7$ W and $\tau_1 \sim$ 5 min. The Drescher et al. experiments were performed in a room without any air exchange outside of the room, with mixing induced by internal blowers. To our knowledge, there are no analogous correlations for 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

τ_2, can be defined as follows. We consider a situation where air exchange from outside the building is suppressed. A contaminant is suddenly released into one room at time $t = 0$. For times $t << \tau_2$, 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 τ_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. If we apply a rough AHU sizing rule (18 m^3/h per square meter of floor area, or 1 ft^3/ft^2 min^{-1}) to a building with room heights of 3.7 m, and neglect non-mechanically induced airflows, τ_2 for a group of rooms served by a single AHU is ~12 min. From this line of reasoning, τ_2 may be longer, but is expected to be within an order of magnitude of τ_1.

The third time scale, τ_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 25th and 75th percentiles of 0.5 and 2.6 h.

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 is to assign less relevance to the measurements taken during this earliest time period. Therefore, it may be challenging to extract useful information from the measurements within this time period due to these model-to-measurement errors.

For times $\tau_1 < t < \tau_2$, conditions are likely to be ideal for determining the release location.

There are concentration differences among the rooms, and in addition, we expect greater model-to-measurement agreement than for times $t < \tau_1$. Consequently, there are no special statistical challenges that must be overcome in comparing model with measured concentrations.

For times $\tau_2 < t < \tau_3$, information about the release location is most likely 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 is lost, since for $t >> \tau_3$ 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, τ_2, will lengthen, and may exceed the ventilation time scale, τ_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 τ_1 may be relevant.

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

systems.

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. We discuss here a method for reconciling model specification errors for the illustrative case study.

4.1. Likelihood function development

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

For the seven development experiments, we 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 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 due to the multizone model's assumption of well-mixed conditions. At later times, errors arise in part from the model misspecification of interzonal airflow rates.

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

standard deviations:

$$\sigma(t) = Ae^{-Ct} + B \tag{2}$$

where A = 1.74, B = 0.46, C = 0.37 min^{-1}. Figure 1 shows the modeled $\sigma(t)$. Since equation 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.

Equation 2 was developed using data from instantaneous releases. To populate a realistic simulation library, we require a likelihood function that includes the effect of release duration. Although we do not have experimental data, we synthesized a likelihood function for longer duration releases. In the interest of space, and because this investigation is preliminary, we only summarize the procedure. We generated hypothetical data for non-pulse releases, but within the limit of the durations considered in the generation of the library (i.e., 5 min), using linear superposition. We assumed that the absolute error (i.e., units of g/m^3) at time t for a particular release can be calculated by scaling equation 2 to the released mass at time t, and dividing by the interior volume. This procedure resulted in a reasonable fit to the simulated release errors, and we extended equation 2 to equations 3.

$$\sigma(t) = \left(Ae^{-Ct} + B\right) \times \frac{t}{\tau_{release}}, t \le \tau_{release} \qquad \sigma(t) = Ae^{-Ct} + B, t > \tau_{release} \tag{3}$$

where $\tau_{release}$ is the release duration in minutes. Equation 3 represents a preliminary attempt to incorporate longer release durations into the likelihood function. Experimental data are necessary to thoroughly investigate the effects of duration on the sensor system performance.

4.2. Implementation of likelihood function

If the release mass, time, and duration are known, then the standard deviation of the

absolute error can be estimated using equation 2 or 3, 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 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, we 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 sensors.) The initial mass is set to a large value, and duration to a small value, to place less emphasis on the initial data.

We address the problem of 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 3, 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. Investigating transport and mixing time scales by blinding data

To explore how time scales affect the interpretation of sensor data, we applied the Bayesian algorithm to Experiments 1 and 4, retaining data from all 11 sensors. We assessed the importance of early data by blinding the BMC algorithm to measurements taken at the beginning

of the release, then comparing its performance for different blinding periods.

For each run of the Bayesian estimator, an input data set was constructed by removing, from the experimental measurements, the first 4 to 160 min of data. Because the experimental data were taken at denser intervals in the beginning stages of the release, this means the interpretation algorithm received data at 20 s intervals for the first 20 min of the release, and at 60 s intervals thereafter. In this exploration, we assumed the release time was known in order to explore the time-dependent utility of the data, without introducing the complication of an unknown release time.

The 4- to 160-minute windows of missing data straddle the expected τ_1 - τ_3 time scales for the case study. Consider the within-room mixing time. Using the correlation from Drescher et al. (1995), we estimate τ_1 to be 12 min and 10 min for Experiments 1 and 4, respectively. We note that the correlation may not be appropriate for the release conditions of Experiment 1, since the release occurred adjacent to the return duct intake and presumably entered the ventilation system without significant mixing in the room. The experimental data show that the fluctuations observed by the sensors effectively cease by ~ 5 min, suggesting a lower value of τ_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 to blow partly into an adjacent room. Consequently, the effective volume may be larger than that used to estimate τ_1 ~ 10 min.

For both experiments, we estimate an interzonal mixing time τ_2 ~ 17 min, using the mechanical ventilation flows and the volume served. Finally, we estimate a removal time scale of τ_3 ~ 2 h, based on the rate of air exchange with outdoors given by the multizone model.

Figures 2 (a) and (b) illustrate the ability of the BMC algorithm to identify the release location. We plot the probability that the estimator 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 any confidence.

Clearly the data at the earliest part of the release is more rich in the information needed to identify the source location. We refer to this information as the "location fingerprint," and relate its loss to the τ_1 - τ_3 time scales. Blinding the estimator 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 in the ventilation system (see Table 1), the strong location fingerprint 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 forces, identified with τ_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 has begun to fade.

Experiment 4 shows a similar pattern (Figure 2(b)),; we are able to find the 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, τ_2, as characterizing movement of a contaminant from zone to zone. As with Experiment 1, by the time that τ_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. Figures 2 (c) and (d) depict 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 due to measurement error.

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 τ_2. This observation reflects the fact that interzonal mixing simply redistributes contaminant mass around the building. Because the sensors continue to send this information to the Bayesian estimator, the mass fingerprint does not significantly fade until time scales at least on the order of τ_3 have elapsed, i.e. when removal from the building becomes important.

For Experiment 4, the median value of the estimated release mass is 30% below 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 3 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.

6. Evaluating different sensor systems

We next describe the performance of networks that vary in the number of sensors, their placement, and in sensor response time. We further explore the relevance of the time scales in the context of these system design parameters. 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?

We investigated the performance of all possible sensor networks composed of from 1 to 11 sensors, for Experiments 1 and 4. 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 non-zero concentration.

Figure 3 shows 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 locate the source with 90% probability, using only 10 min of data. If we seek 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 identify the release location to a probability of at least 90%.

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. Note the differences in the patterns between 1 min and 5 min, as compared to 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 min < t < 5 min) than later (5 min < t < 10 min). This 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.

Turning to Experiment 4, Figure 3 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, a larger proportion of networks can 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). These results suggest that the source may be 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. We 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 prior mass, as follows:

$$UR(t) = \left(\frac{q_{Mass,90\%}(t) - q_{Mass,10\%}(t)}{q_{Mass,90\%,\,prior} - q_{Mass,10\%,\,prior}} \right) \times 100\% \quad (4)$$

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 4 shows the estimates of the median release mass and the uncertainty reduction of the 80% two-sided confidence interval for the same networks considered in Figure 3. For Experiment 1, the 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 mass during the 1 min to 5 min monitoring interval than during the 5 min to 10 min period. Comparing Figures 3 and 4, fewer sensors are required to estimate 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 ($t << \tau_3$), which suggests that perhaps even fewer sensors may be able to estimate the mass to a high confidence, if we believe that there is mass-relevant information until $t \sim \tau_3$.

We also explored how time scales relate to the performance of sensor systems with varied sensor response time. Figures 5 and 6 compare the performance of sensor systems with different response times for all networks containing an odd number of sensors. We 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. Figure 5 shows 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 paper, 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. Figure 6 depicts the median release mass, and corresponding uncertainty reduction of the 80% two-sided confidence interval as defined by Equation 4. For Experiment 1, we observe that the median mass is relatively insensitive to sensor response time, but that the final confidence intervals are 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 estimate the mass to a slightly higher confidence than networks with a greater number of slower sensors. In Experiment 4, we see 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, we anticipate that as time progresses, the results produced by slower sensors will approach the results for networks with

faster sensors.

7. Conclusions

Real-time sensor systems have the potential to help reduce building occupant exposure to contaminants. This paper shows how an analysis of the relevant time scales for contaminant transport may help explain the performance of sensor systems that use Bayesian interpretation to estimate release conditions, based on concentration measurements. We show that the sensor system design should consider the physics governing contaminant transport, as well as the limitations that arise from the transport model.

We discussed three relevant time scales, and demonstrated how estimating different release parameters requires different kinds of information. 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, we expect that the concepts are more broadly relevant. This work should be expanded to consider 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, we developed a likelihood function that quantifies model uncertainty, which is often a challenge for modeling contaminant dispersion. Our 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, are needed to build upon this initial effort.

Although we 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 us connect the algorithm's performance directly to the transport physics, the conclusions reached here may apply to other, non-Bayesian, sensor interpretation algorithms.

Future research efforts may 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. Non-chemical sensors (such as door-position sensors) are worth exploring to further improve the sensor system performance and to reduce the overall system costs, particularly for buildings that are more complex than the one considered here. Advances along these lines may be useful not just for this application, but possibly in other building and environmental monitoring applications.

Acknowledgements

This work was supported by a fellowship from the National Science Foundation and by the Office of Chemical Biological Countermeasures of the Science and Technology Directorate of the Department of Homeland Security, and performed under U.S. Department of Energy Contract No. DE-AC03-76SF00098. We thank Richard Sextro, Darryl Dickerhoff, Helmut Feustel and Corina Jump for collecting the Dugway data and generating the original COMIS model. We also thank David Lorenzetti and Mark Sippola for providing useful review comments on a draft manuscript.

References

Baughman, A.V., Gadgil, A.J., Nazaroff, W.W., 1994. Mixing of a point source pollutant by natural convection flow within a room. Indoor Air 4, 114-122.

Drescher, A.C., Lobascio, C., Gadgil, A.J., Nazaroff, W.W., 1995. Mixing of a point-source indoor pollutant by forced convection. Indoor Air 5, 204-214.

Feustel, H.E., 1999. COMIS — an international multizone air-flow and contaminant transport model. Energy and Buildings 30, 3-18.

Lorenzetti, D.M., 2002. Assessing multizone airflow simulation software. Report LBNL-49578, Lawrence Berkeley National Laboratory, Berkeley, California. Proceedings of the 9th International Conference on Indoor Air Quality and Climate — Indoor Air 2002, Monterey, California, vol. 1, pp. 267–271.

Persily, A.K., Gorfain, J., Brunner G., 2006. Survey of ventilation rates in office buildings. Building Research & Information 34, 459-466.

Sextro, R.G., Daisey, J.M., Feustel, H.E., Dickerhoff, D.J., Jump, C., 1999. Comparison of modeled and measured tracer gas concentrations in a multizone building. Proceedings of the 8th International Conference on Indoor Air Quality and Climate – Indoor Air 99, Edinburgh, Scotland, vol. 1, pp. 696-701.

Small, M.J., Fischbeck, P.S., 1999. False precision in Bayesian updating with incomplete models. Human and Ecological Risk Assessment 5, 291-304.

Sohn, M.D., Small, M.J., Pantazidou, M. 2000. Reducing uncertainty in site characterization using Bayes Monte Carlo methods. ASCE Journal of Environmental Engineering 126, 893-902.

Sohn, M.D., Reynolds, P, Singh, N., Gadgil, A.J., 2002. Rapidly locating and characterizing pollutant releases in buildings. Journal of the Air and Waste Management Association 52, 1422–1432.

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.

Table 1. Description of twelve tracer gas experiments. [a]

Experiment number [b]	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	Interior room, 1st floor [c]	Stairwell doors closed, other interior doors open	On
13	Interior room, 1st floor [c]	Stairwell doors closed, other interior doors open	Off

[a] See Figure 1 of Sreedharan et al., 2006, for a schematic of the test building.
[b] There was no experiment numbered 7.
[c] This room is numbered 1.3 on the schematic.

Table 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 captions

Fig 1. Normalized model-to-measurement error (small points) for seven tracer-gas experiments (Experiments 5,6 & 8-12). Also shown are calculated standard deviations for errors grouped into two-minute intervals and modeled standard deviation of normalized errors using exponential fit, equation 2.

Fig 2. Probability of correctly identifying release location after four rounds of data for (a) Experiment 1 and (b) Experiment 4. Estimated mass released after updating on four rounds of data for (c) Experiment 1 and (d) Experiment 4. In each case, 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.

Fig 3. For Experiments 1 and 4, probability of correctly identifying release location at 1 min, 5 min, and 10 min after the release; 10^{th} and 90^{th} percentiles (dashed bars) and median (bullets).

Fig 4. Median release mass and uncertainty reduction of the 80% confidence interval at 1 min, 5 min, and 10 min; 10^{th}, 90^{th} percentiles (error bars), and median (bullets) of the median release mass; 10^{th}, 90^{th} percentiles (solid lines), and median (dashed line) of uncertainty reduction.

Fig 5. Probability of correctly identifying release location at 10 min after release event for 3 different sensor response times (20 s, 60 s, and 120 s); 10^{th} and 90^{th} percentiles (error bars) and median (bullets).

Fig 6. Median release mass and uncertainty reduction of the 80% confidence interval at 10 min after release event for 3 sensor response times (20 s, 60 s, and 120 s); 10^{th} and 90^{th} percentiles (error bars) and median (bullets) of the median release mass; 10^{th} and 90^{th} percentiles (solid lines) and median (dashed line) of uncertainty reduction.

Figure 1

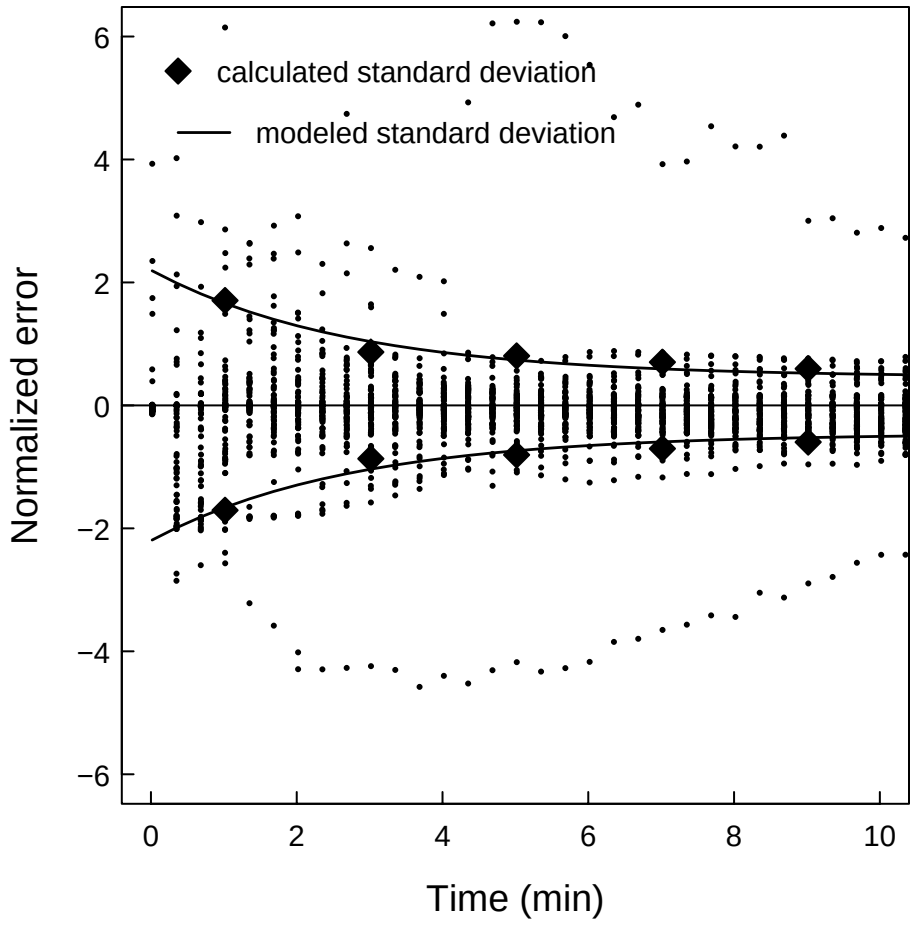

calculated standard deviation
modeled standard deviation
Normalized error
Time (min)
6
4
2
0
−2
−4
−6
0
2
4
6
8
10

Figure 2

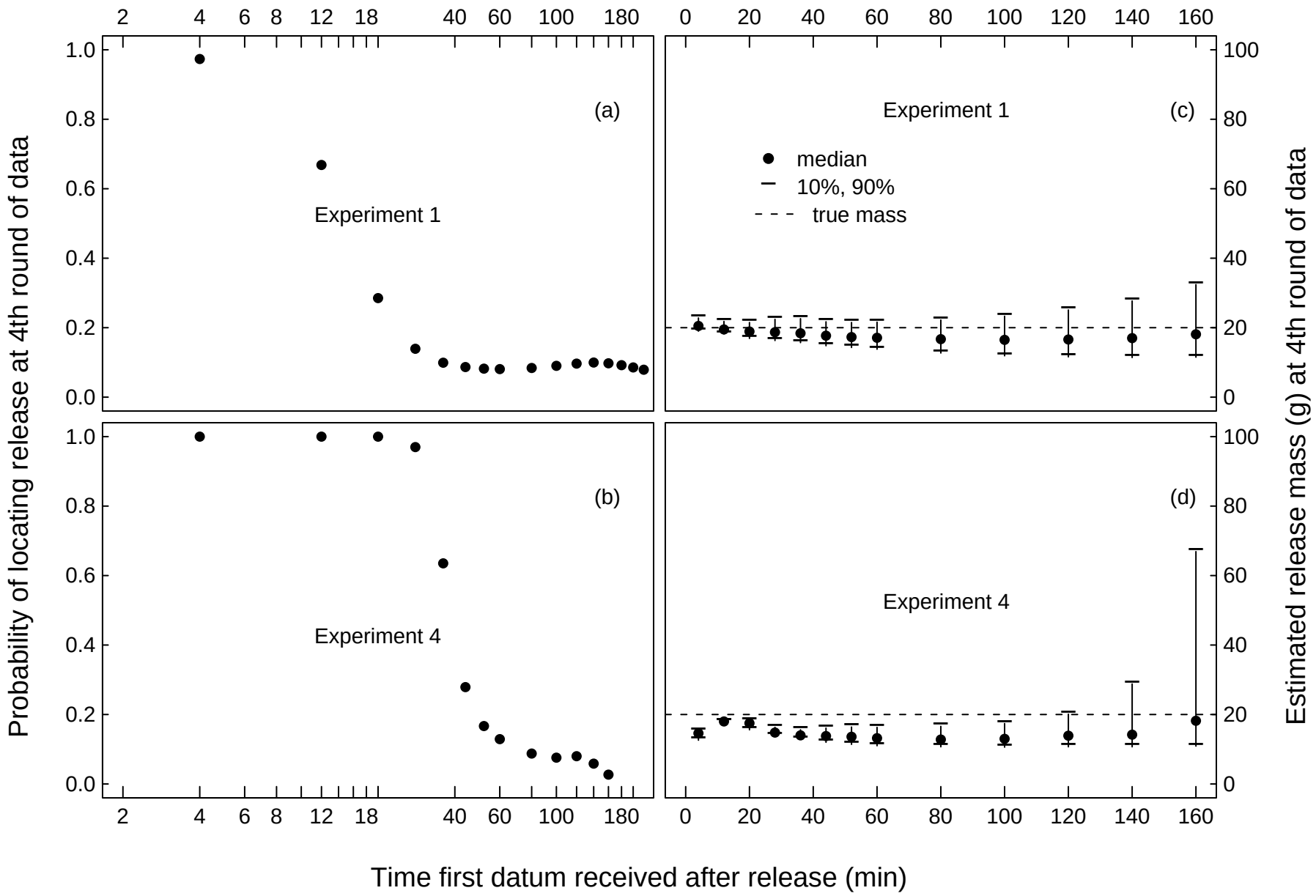

Probability of locating release at 4th round of data
Estimated release mass (g) at 4th round of data
Time first datum received after release (min)
(a)
(b)
(c)
(d)
Experiment 1
Experiment 4
median
10%, 90%
true mass
2
4
6
8
12
18
40
60
100
180
0
20
40
60
80
100
120
140
160
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

Figure 3

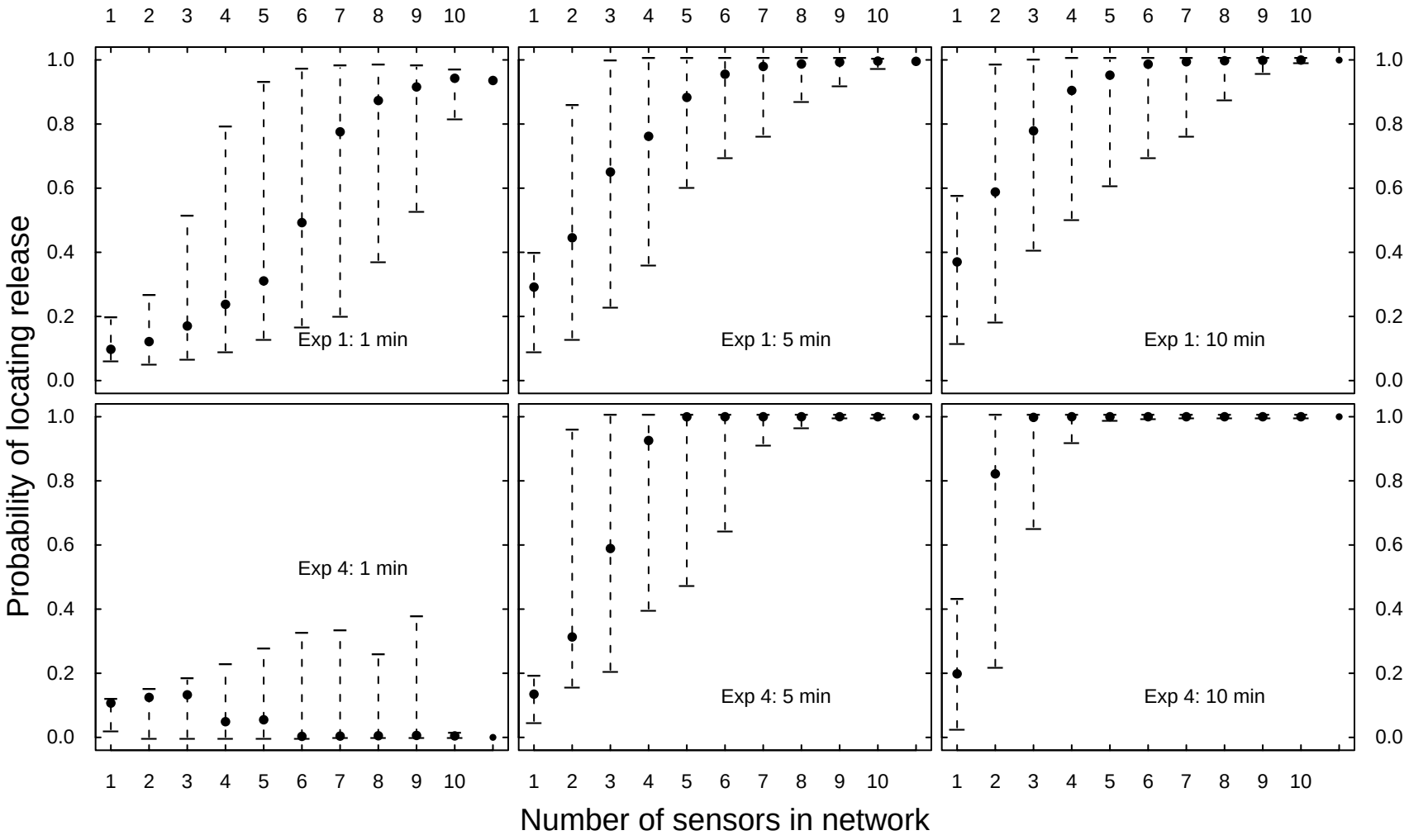

Probability of locating release
Number of sensors in network
Exp 1: 1 min
Exp 1: 5 min
Exp 1: 10 min
Exp 4: 1 min
Exp 4: 5 min
Exp 4: 10 min
0.0
0.2
0.4
0.6
0.8
1.0
1 2 3 4 5 6 7 8 9 10

Figure 4

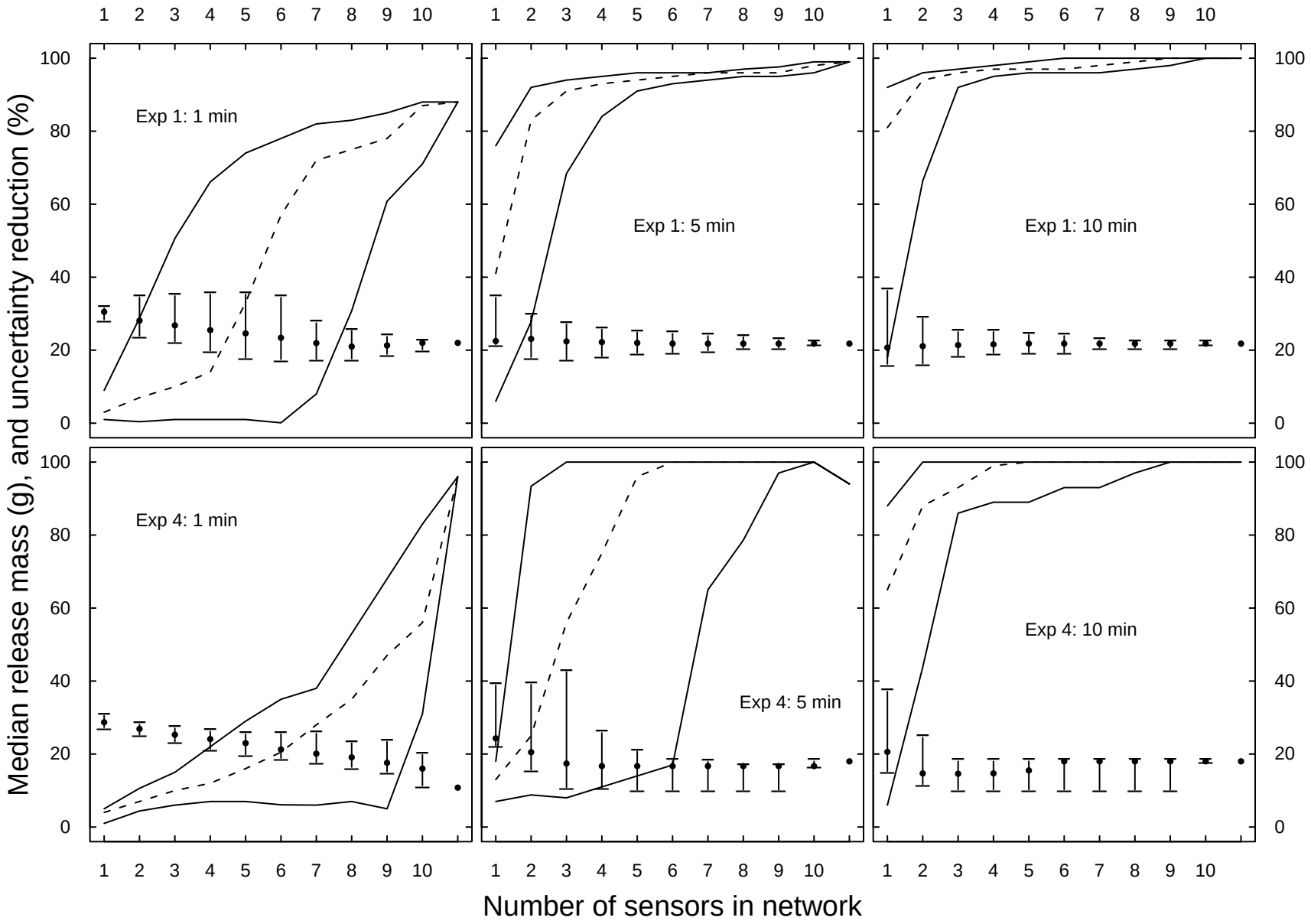

Exp 1: 1 min
Exp 1: 5 min
Exp 1: 10 min
Exp 4: 1 min
Exp 4: 5 min
Exp 4: 10 min
Median release mass (g), and uncertainty reduction (%)
Number of sensors in network
0
20
40
60
80
100
1 2 3 4 5 6 7 8 9 10

Figure 5

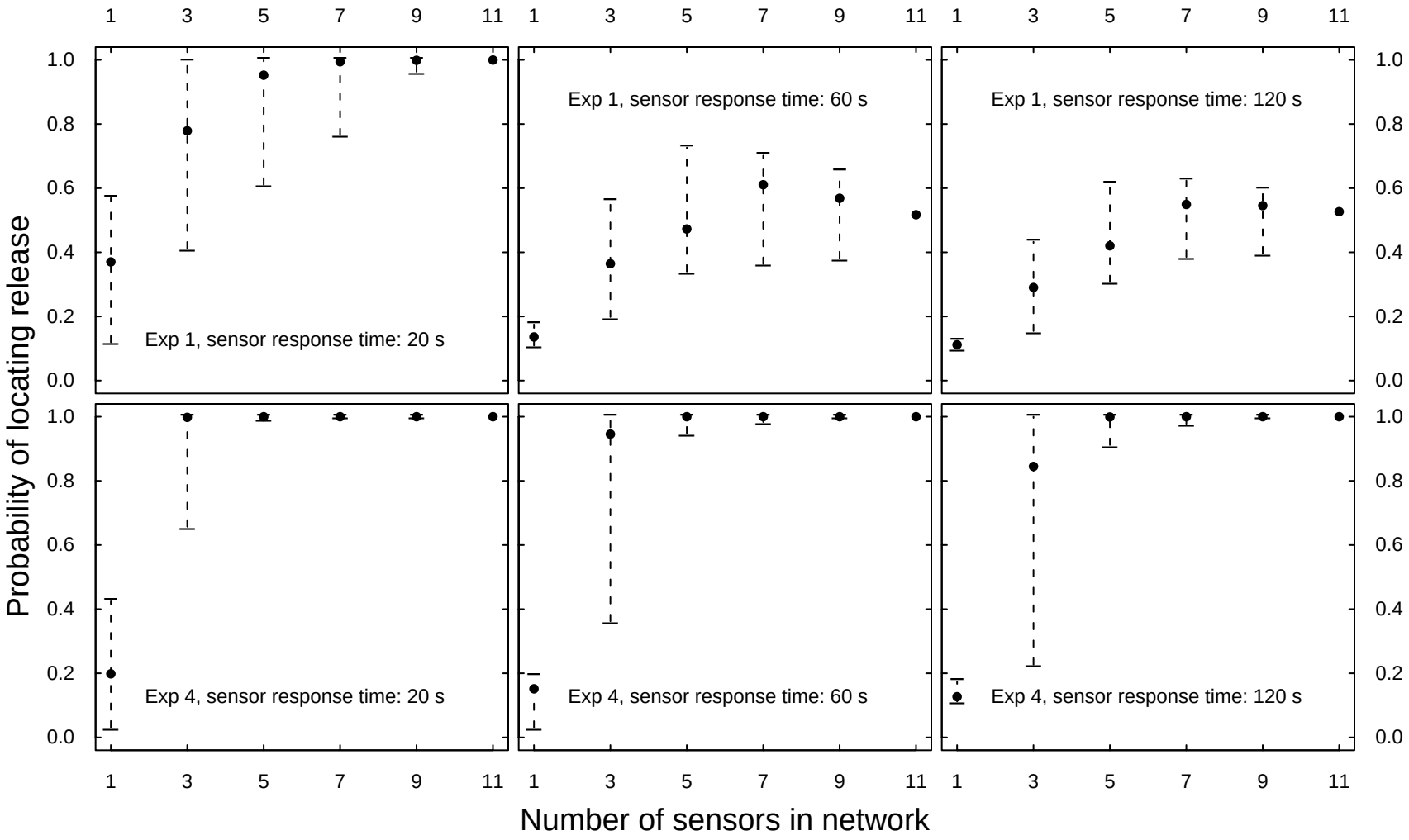
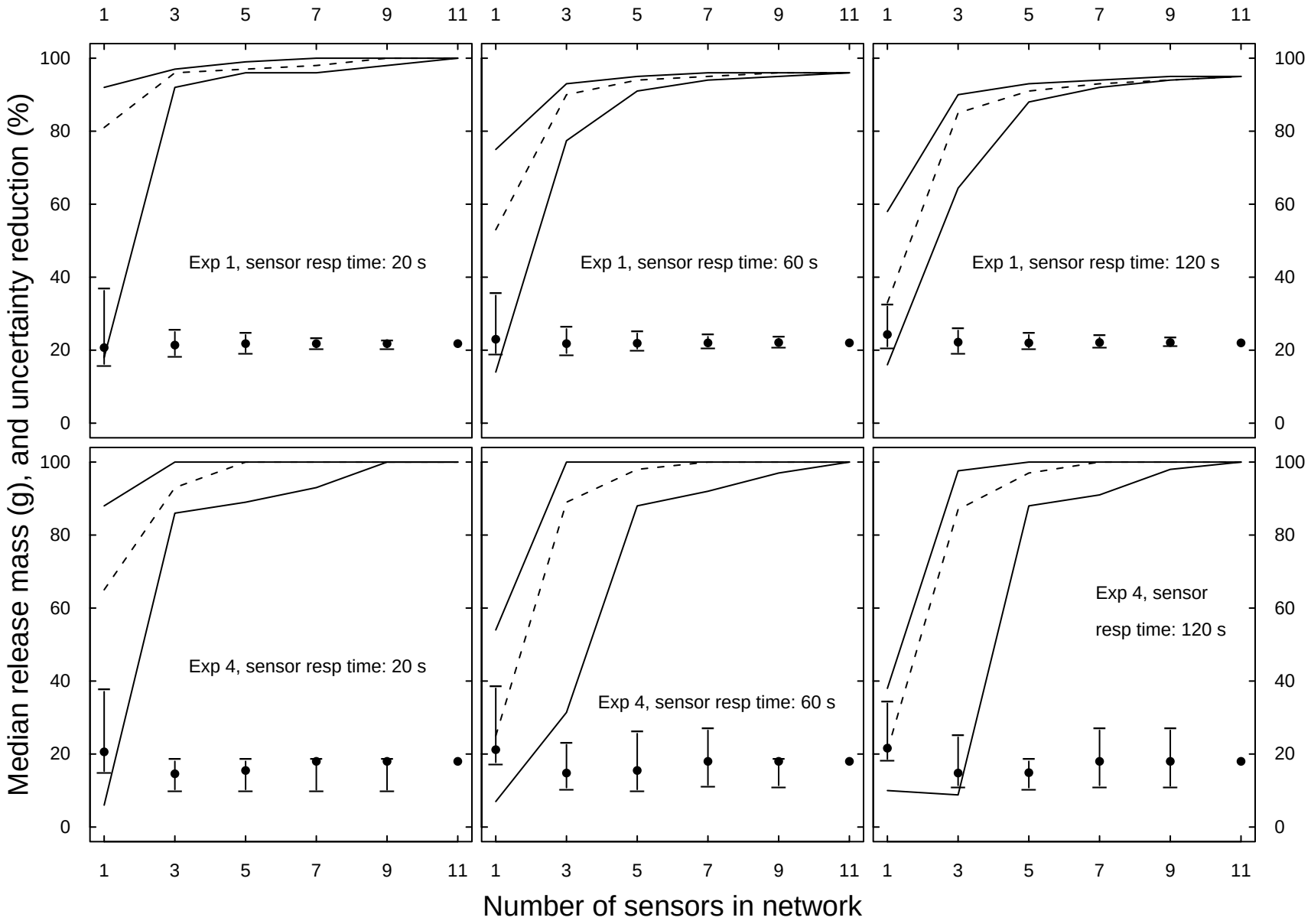

Probability of locating release
Number of sensors in network
Exp 1, sensor response time: 20 s
Exp 1, sensor response time: 60 s
Exp 1, sensor response time: 120 s
Exp 4, sensor response time: 20 s
Exp 4, sensor response time: 60 s
Exp 4, sensor response time: 120 s
0.0
0.2
0.4
0.6
0.8
1.0
1
3
5
7
9
11

Figure 6


Median release mass (g), and uncertainty reduction (%)
Number of sensors in network
Exp 1, sensor resp time: 20 s
Exp 1, sensor resp time: 60 s
Exp 1, sensor resp time: 120 s
Exp 4, sensor resp time: 20 s
Exp 4, sensor resp time: 60 s
Exp 4, sensor
resp time: 120 s
0
20
40
60
80
100
1
3
5
7
9
11