We present a probabilistic approach to designing an indoor sampler network for detecting an accidental or intentional chemical or biological release, and demonstrate it for a real building. In an earlier article, Sohn and Lorenzetti developed a proof of concept algorithm that assumed samplers could return measurements only slowly (on the order of hours). This led to optimal “detect to treat” architectures that maximize the probability of detecting a release. This article develops a more general approach and applies it to samplers that can return measurements relatively quickly (in minutes). This leads to optimal “detect to warn” architectures that minimize the expected time to detection. Using a model of a real, large, commercial building, we demonstrate the approach by optimizing networks against uncertain release locations, source terms, and sampler characteristics. Finally, we speculate on rules of thumb for general sampler placement.

KEY WORDS: CONTAM; indoor airflow; optimization; sampler networks

1. INTRODUCTION

Many private and public agencies are developing hardware to detect the presence of airborne chemical or biological agents in or near buildings. Detecting a contaminant would allow acting to minimize adverse health effects, for example, by evacuating the building, manipulating air supplies, and mobilizing medical response. However, this range of possible responses—plus practical constraints imposed by the hardware, and uncertainty about the operating conditions under which it must function—complicate the design and operation of a monitoring network that balances risk appropriately.

The designer must decide, for example, on the number of samplers to deploy, their operating characteristics (e.g., sampling frequency and detection limit), where to place them, and what release scenarios to try to detect. To address these questions, Sohn and Lorenzetti(1) proposed a probabilistic approach to network design and demonstrated its application in a synthetic building. Many other approaches have been taken(2–4) that advance the state of this research, but none account for the relative likelihoods of uncertain conditions. This article extends the work in Ref. 1 by: (1) developing a more complete analysis framework, (2) adding a new metric for evaluating network performance, and (3) applying the resulting algorithms to a real building.

The approach taken here, while developed to protect against airborne plumes of chemical or biological material (see also Refs. 5 and 6), is relevant to wider problems of monitoring indoor air quality,(7) building energy,(8) occupancy,(9) thermal comfort,(10) lighting,(11) and more.(12,13) Whenever samplers are too expensive to deploy widely throughout a building, a probabilistic optimization approach may help balance the competing design constraints and goals of the sampler network. Furthermore, whenever the
network must operate under uncertain or variable conditions, a probabilistic approach, such as the one described here, may be needed.

2. PROBABILISTIC ALGORITHM FOR SAMPLER DEPLOYMENT

Consider designing an air-monitoring network in order to maximize some measure, \( \phi \), of the network quality. Whatever the metric, uncertainty and variability in the operating conditions mean that the network quality cannot be defined deterministically.

Stochastic effects arise in the source (for example, the release location, rate, and time, and the material degradation and deposition rates); in sampler characteristics (probability of detecting a given concentration, or the time needed to process samples); in environmental conditions (outside temperature and wind direction); and in the building operation (status of the ventilation system, condition of filters, position of doors and windows, and leakiness of the ductwork). Uncertainty also arises from the models used to assess the contaminant dispersion (for example, due to simplifications in the model physics, and the extent to which model parameters have been tuned to match the actual building operation).

2.1. Expected Performance

In the face of such probabilistic effects, the quality metric should reflect the statistically expected performance of the network. The probabilistic approach to sampler siting (PASS)\(^{(1)}\) finds the expected network performance by aggregating the outcomes of many deterministic model runs, each drawing its input parameters from distributions of likely values.

In this approach, the key sources of uncertainty and variability that might affect the performance of the sampler network—the source and sampler characteristics, environmental conditions, building operation, model structure, and so on—are first identified and characterized. Assigning probability distributions to these uncertain conditions can be done using past measurements or engineering judgment. While specifying these distributions is not trivial, a key feature of probabilistic algorithms is that they allow testing for the effect the distributions have on sampler placement and network performance. Sampling from these distributions yields a suite of scenarios, or test cases, against which to evaluate candidate networks. A pollutant fate and transport model is then used to simulate each scenario. Finally, PASS finds the expected performance of each candidate sampler network, taking into account the relative likelihood of each scenario.

Let \( \phi_i \) give the value of some quality metric, as applied to scenario \( i \). Because each scenario is defined deterministically, \( \phi_i \) also is a deterministic measure of how well a particular sampler network performs given a specific scenario. Combining across all \( I \) scenarios in the suite yields the expected performance, as

\[
E[\phi] = \sum_{i \in I} \phi_i \cdot P[i].
\]

where \( P[i] \) gives the relative likelihood of scenario \( i \).

2.2. Performance Metrics

The algorithm reported in Ref. 1 maximized the expected probability of detecting a release. This goal implicitly acknowledged the fact that first-generation samplers required many hours to collect and analyze samples before returning results. The resulting networks were optimal \textit{detect-to-treat} architectures, which sought mainly to identify the fact that a release took place.

A new generation of samplers, able to provide data on the order of minutes, offers the promise of \textit{detect-to-warn} architectures. Such systems, by focusing mainly on fast detection, will enable actions intended to minimize exposures, for example, by evacuating the building, or manipulating fresh air supplies. However, this greater capability further complicates the network design: while higher sampling rates may let the network detect a release earlier, they can also lead to noisier data, to lower detection probabilities (since shorter sampling windows present the sampler with less airborne mass to detect), or to more false positives.

Suppose a sampler returns a new result at intervals of length \( \tau \). Each such interval constitutes a \textit{sampling window}. Let \( P[S_{i,z,w}] \) give the probability, for release scenario \( i \), that sampler \( z \) will alarm during a particular window, \( w \). Then a network comprising a set \( Z \) of samplers will alarm during window \( w \) if one or more of those samplers alarms:

\[
P[N_{i,w}] = 1 - \prod_{z \in Z} (1 - P[S_{i,z,w}]).
\]

Note that Equation (2) uses \( P[\text{at least one occurs}] = 1 - P[\text{none occurs}] \). If the network concept of operation demands multiple samplers to alarm, for
example, in order to guard against false positives, a more complicated expression results.

Looking across all sampling windows, the cumulative probability that a network will alarm for scenario $i$ after sampling across $W$ windows is

$$P[N_i] = 1 - \prod_{w=1}^{W} (1 - P[N_{i,w}]).$$  \hfill(3)

Combining Equations (2) and (3),

$$P[N_i] = 1 - \prod_{w=1}^{W} \prod_{z \in Z} (1 - P[S_{i,z,w}]).$$  \hfill(4)

Since the order of multiplication is immaterial, one may precompute the products $1 - P[S_{i,z,w}]$ across the windows $W$ of interest, then combine them according to the sampler selections in $Z$.

For detect-to-treat architectures, we define the performance metric in scenario $i$ as the probability that the network in question will detect the release:

$$\phi_i = P[N_i].$$  \hfill(5)

with $W$ chosen sufficiently large. Note that the optimal network will maximize the expected value of this performance metric over all scenarios.

We now turn to the goal of fast detection. As described above, many of the distributions that define the scenarios affect the probability of detecting a release. Therefore, detection is itself a stochastic phenomenon. Accordingly, we let the network designer specify a desired level of confidence, $\beta$, that the network will alarm. Then

$$T_i = \tau \cdot \min\{W : P[N_i] > \beta\}$$  \hfill(6)

gives the time at which a particular network can detect scenario $i$ with at least $\beta$ probability.

If a network does not detect the release in scenario $i$, Equation (6) leaves $T_i$ undefined. In this case, the designer must specify some appropriate value, for example, by estimating the time it would take to detect the release by some other means (e.g., when a large number of occupants experience health effects).

For detect-to-warn architectures, we define the performance metric in scenario $i$ as the time required to detect the release:

$$\phi_i = T_i,$$  \hfill(7)

and note that the optimal network will minimize the expected value of this performance metric over all scenarios.

In this article, we consider only two performance metrics, probability of detection and time to detection. However, the algorithm presented here can use any performance metric, including occupant exposure\(^{5,6}\), health consequences, or total cost of operation.

3. APPLICATION TO A CONVENTION CENTER

To demonstrate the sampler network design approach, we apply the algorithm to a realistic model of a large building. Fig. 1 shows a modified schematic of a real convention center in which Lawrence Berkeley National Laboratory (LBNL) performed tracer gas experiments\(^{14}\). In addition to the main convention space floor, the building has two floors of offices. The building is served by 67 heating, ventilation, and air conditioning (HVAC) units.

3.1. Model

In each of six experiments, one or more inert tracer gases were released, and concentrations measured every 1–30 minutes, in approximately 40 locations. The data were used to calibrate a multizone airflow and pollutant transport model of the building using CONTAM\(^{15}\). The model consists of 337 well-mixed zones. Fig. 2 shows typical postcalibration model-to-data comparisons. For this particular building, high ventilation rates mean that the well-mixed assumption is valid, but our algorithm allows any type of pollutant transport model (e.g., CFD) to be used.

Because the model does not perfectly represent the building, it introduces uncertainty into the network design process. A key feature of the approach described here is that the network designer can hedge against this uncertainty by using multiple pollutant transport models of the building. For example, one model could be tuned to match the integrated concentration in each zone (which might be most appropriate when maximizing the probability of detection), while another model could be tuned to match the estimated timing of the peaks (which might be most appropriate when minimizing the time to detection). Using multiple models would mean adding scenarios to the analysis, with the relative confidence in each model reflected in the scenario likelihoods, $P[i]$. In this study, we used only the convention center CONTAM model described in Ref. 14, since it is available for others to use in comparative sampler network design studies.
3.2. Release Scenarios

The most important sources of uncertainty in designing a sampler network are the variables that affect the transport and dispersion of the chemical or biological agent: the source characteristics, environmental conditions and building operation. As demonstrated here, the designer can enumerate the scenarios of interest and assign each a relative likelihood of occurrence. Alternately, the designer can define continuous distributions of parameters such as the release mass and wind speed, then sample from those distributions in order to generate probability-weighted scenarios. For clarity, in this study, we consider only 60 scenarios, each consisting of one of 20 possible release locations (Table I) and one of three possible release rates (Table II). In this study, we vary only release locations and release rates, but our algorithm allows specification of scenarios that vary any uncertain parameters, such as those mentioned in Section 2.

3.3. Sampler Performance

The real performance of a sampler may have a probabilistic component. Given a large amount of contaminant in the air, there is a higher chance that the sampler will detect the agent. However, due to miscalibration, fouling, noise, imperfect mixing, and so on, the presence of an agent in the room air does not guarantee detection—even above the sampler’s nominal detection threshold.

Fig. 3 shows the assumed sampler performance for this study, based on simplified performance curves of actual hardware (mass units are withheld for security reasons). The probability of detection during any given sampling window depends on both the agent mass that passed through the sampler during that time, and the sensitivity of the detection equipment. We assumed that ambient air is pumped through a sampler at 100 L/min. For this building, with large rooms and high ventilation rates, we assumed that the presence of a sampler will not affect the airflow between rooms, and will not significantly change the well-mixed assumption within rooms.

In the analysis that follows, all samplers in a given network have the same operating curve and sampling window. However, a network could include samplers with different detection characteristics—for example, incorporating fast, sensitive samplers to detect a release quickly, along with slower, but less error-prone samplers to confirm a release. Similarly,
a network consisting of samplers with different window lengths, or with windows staggered in relation to one another, might yield a more robust network. The PASS approach can be applied to any of these options, at the cost of having to evaluate more networks in order to find the optimal one.

3.4. Detection Confidence

The sampler network design includes decision points for signaling an alarm, which is a part of the network concept of operation (ConOps). As suggested above, this may include the number and type of samplers that must alarm before taking action. The decision criterion used in this analysis is for a network to alarm as soon as at least one sampler alarms. Alternately, a network could be chosen to alarm when at least two samplers alarm, or when at least two samplers alarm within a given time period. In cases where false alarms can be exceedingly expensive (e.g., whole-building evacuation, deployment of emergency personnel, etc.), a very high confidence criteria might be chosen, but this, in turn, may result in delayed detection.

It is important to distinguish between the ConOps (which determine the operation of the network after deployment) and the calculations of
### Table I. Release Location Probabilities

<table>
<thead>
<tr>
<th>Location</th>
<th>Count</th>
<th>Probability (each)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Floor</td>
<td>16</td>
<td>0.0575</td>
</tr>
<tr>
<td>Second Floor</td>
<td>2</td>
<td>0.02</td>
</tr>
<tr>
<td>Third Floor</td>
<td>2</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*Note:* Release locations were selected to encompass a variety of zone types and ventilation rates. Note that releases on the main floor are assumed more likely than ones on the upper floors.

### Table II. Release Rate Probabilities

<table>
<thead>
<tr>
<th>Rate (g/min)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>0.3</td>
</tr>
</tbody>
</table>

*Note:* All releases are assumed to last for 10 minutes.

---

### 3.5. Candidate Locations

We allow PASS to choose from among 35 possible sampler locations in the convention center, including several types of occupied zones and ventilation return ducts. In principle, any zone of the multizone model defines a possible sampler location. However, in practice, the number of possible sampler networks increases sharply with the number of sampler locations the algorithm is allowed to consider. When PASS optimizes $n$ samplers among possible locations, it must evaluate

$$\frac{(r + n - 1)!}{(r - 1)! n!}$$

networks. For example, placing $n = 5$ samplers among the 35 locations we allow defines 575,757 networks. Doubling the number of possible locations would increase the number of networks by a factor of almost 28. In the examples that follow, computing an optimal two-sampler network with five-minute sampling windows takes less than one minute on a 2.4 GHz processor with 2 GB of RAM, running Mac OS X 10.5. Finding an optimal four-sampler network takes approximately 18 minutes.

### 3.6. Calculations

Simulating the contaminant transport for any given scenario gives the mass that a hypothetical sampler would accumulate in each candidate location, during each sampling window. The mass is then used to determine the detection probabilities, $P[S_i, z, w]$, using the curves in Fig. 3. Aggregating across samplers and sampling windows, Equation (3) gives the probability a network will detect the scenario, while Equation (6) gives the time to detect at the specified confidence level (for convenience in calculating these performance metrics, the first sampling window, $w = 1$, is taken as the window in effect at the time the release begins). Finally, Equation (1) gives the network’s expected performance across all 60 scenarios. All possible networks are compared, in order to find the one with the best expected performance.

### 4. RESULTS

Figs. 4–7 summarize the performance of the optimal networks that PASS identifies. For example, Fig. 4 shows the maximum expected probability of detection, across all the networks tested, as a function of the number of samplers in the network.

---

![Graph](image-url)
Fig. 4. Expected probability of detection of best network, for varying network size, sampler sensitivity, and sampling window.

Fig. 5. Expected probability of detection of best network, for varying network size and sampling window. Results are for low sensitivity samplers. Each curve is labeled with the number of samplers in the network.

Similarly, Fig. 5 shows the probability of detection for the best network as a function of the sampling window duration, and Figs. 6 and 7 show similar plots for the networks with the fastest time to detection.

4.1. Optimal Locations

Comparing the optimal networks, we saw no consistently-favored sampler locations. This contradicts Ref. 1, in which maximizing the probability of detection, across networks of different sizes, tended to place more sensitive samplers in bathrooms (to take advantage of exhaust airflows), and less sensitive samplers in ventilation system return ducts (which effectively sample air from throughout the building).

We attribute the lack of favored sampler locations in the convention center to the large airflows between zones. With no partitions between many zones, and relatively high recirculation rates, the convention center mixes quickly compared to the office-dominated building from the original study. The high mixing rate also explains why, in the convention center, many of the best networks have nearly the same expected performance: if no particular zone has a unique concentration profile, then no particular zone is critical to the sampler network’s performance.

Because many networks have similar quality, the optimal sampler locations are often nonintuitive. For example, the best two-sampler network will not necessarily place a sampler in the same zone as the best one-sampler network. Thus, a “greedy” optimization approach, in which samplers are added one by one to the previous best network, is not ideal for the sampler placement problem.

In a real design exercise, we would treat the relatively small variation in the performance among many networks as an invitation to expand the scope of the investigation. Improving the scenarios considered—for example, by including new building operating conditions, better characterizing the distributions of uncertain parameters, or adding new release locations and amounts—might allow PASS
to better discriminate between the expected performance of the networks. Admitting more possible sampler locations might improve the final network quality. Tightening the confidence limit for estimating $T_i$ might reveal some networks to be more robust than others. Finally, if none of these changes affected the results appreciably, then we would accept that many networks are near-optimal and pick the final sampler locations based on other operational criteria (such as ease of service, or aesthetics).

4.2. Network Size

In Figs. 4–7, the expected network performance improves with network size. The marginal improvement in network performance when adding a sampler is largest for small networks, and, in this application, is virtually negligible for networks with five or more samplers. Intuitively, allowing PASS to place more samplers improves its ability to cover all parts of the building; however, mixing by the ventilation system means that effective coverage does not demand placing a sampler in every zone.

4.3. Sampler Sensitivity

Figs. 4 and 6 show that improving the sampler sensitivity improves the expected network performance (giving higher detection probability, or faster detection). This effect is more pronounced for smaller networks. However, among the smaller networks, adding a single sampler generally improves the network quality more than does increasing the sampler sensitivity by a factor of 10. This suggests using PASS to explore an interesting practical trade-off, between cost and sensitivity, in real sampler design.

4.4. Sampling Frequency

Figs. 5 and 7 show that network performance is highest for very short sampling windows and improves as the sampling windows get very long. They also show lower overall network performance for intermediate sampling window sizes. For example,
in Fig. 5, the detection probability for the optimal network that samples with 15-minute windows is lower than that of networks having one-minute or 30-minute windows. In Fig. 7, note that this effect is relative to maximum performance—overall, shorter sampling windows yield lower absolute time to detection. This effect is more pronounced for smaller networks.

One explanation for this result may be competing attributes of an optimal network. With very short sampling windows, many samples, each of which individually may have low probability of detection, can result in a high cumulative probability of detection (see Equation (3)). Conversely, with a long-duration window, such as two hours, a large amount of mass is collected and that single sample results in a high probability of detection (see Fig. 3). With intermediate sampling windows, the network benefits from neither many samples nor long samples, and the performance of the networks decreases.

Other possible reasons for the dip include the duration of the release, the residence time of the contaminant in the building, or when samples are taken relative to the beginning of the release. To explore these possibilities, we conducted numerical experiments in which we varied these parameters. While none of these factors individually explained
the shape of the curves in Fig. 5, each had some contribution. Network designers would benefit from a rule of thumb on selecting the ideal sampling window, but further research is needed on this topic.

4.5. Air Exchange Rates

We also explored methods for choosing optimal networks in cases where a contaminant transport model of the building is not available, and would be prohibitively expensive or time-consuming to produce. Building operators may wish to estimate network performance using easily identifiable building characteristics, either in lieu of building a model, or as a feasibility test to determine whether it is worthwhile to construct a building model.

Airborne transport is the most important mechanism for mixing chemical and biological agents through a building, and is essential to the operation of the types of samplers considered here. Furthermore, all else being equal, increasing the amount of airflow through a zone increases the chances it will receive air from a zone that contains the agent release. Therefore, a natural choice for a performance predictor is the air exchange rate in each zone, information which may be easily estimated by building operators (for example, using as-built drawings of the ventilation system).

In Figs. 8 and 9, the performance of a one-sampler network improves somewhat with higher air exchange rates in the zone of interest. However, there is a great deal of variability, particularly for zones with low air exchange rates. Clearly, the air exchange rate for a zone is not a good proxy, at least in this building, for overall mixing of air from other parts of the building through that particular zone.

While we acknowledge that the relationship between network performance and air exchange rates is tenuous, and the accuracy of performance predictions depends on the accuracy of air exchange rate estimates, we believe that there is merit in using easily evaluated building characteristics, and plan to explore this further in future work.

5. CONCLUSION

We presented a probabilistic approach to design an indoor sampler network for the purpose of detecting a chemical or biological agent. The design of such a network is complicated by uncertainty and variability in all aspects of the problem, including building operation modes, agent release conditions, meteorology, contaminant transport modeling, and sampler hardware behavior. These probabilistic effects motivated a statistical approach that optimizes the network’s expected performance, according to the likelihood of a range of possible scenarios.

Past work on this approach maximized the probability of detecting a release. However, advances in sampler hardware have made results available more rapidly. Therefore, in this work, we also minimize the time to detect a release, at a prescribed level of confidence. We demonstrated our approach by designing sampler networks for a large commercial building, using a pollutant dispersion model that was tuned to experimental data from a real building.

Our approach allows comparisons between competing network design parameters and network performance. Therefore, network designers can minimize the hardware, deployment, and maintenance cost of fielding a network with a given level of performance (for example, by trading one high-sensitivity sampler for several lower cost samplers of lesser sensitivity). Similarly, the PASS methodology could also be used by sampler hardware manufacturers, to guide their designs (for example, in deciding whether to build faster or more sensitive samplers).

ACKNOWLEDGMENTS

This research was funded in part by the Defense Threat Reduction Agency and was performed under U.S. Department of Energy Contract no. DE-AC03-76SF00098.

REFERENCES