Approaches to Modeling the Concentration Field for Adaptive Sampling of Contaminants during Site Decontamination*

William K. Sieber, James S. Bennett, Abera Wouhib, Joe Fred Gonzalez, Jr., Myron J. Katzoff, and Stanley A. Shulman

1 National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, WSieber@cdc.gov
2 National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, JBennett@cdc.gov
3 National Center for Health Statistics, Centers for Disease Control and Prevention, AWouhib@cdc.gov
4 National Center for Health Statistics, Centers for Disease Control and Prevention, JGonzalez@cdc.gov
5 National Center for Health Statistics, Centers for Disease Control and Prevention, MKatzoff@cdc.gov
6 National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, SShulman@cdc.gov

1 Introduction

Following the anthrax outbreaks of October–December 2001, there has been an increasing awareness of the potential for biological and chemical contaminant releases in a closed environment such as an office building. Characterizing the spread and distribution of such contaminants so as to minimize worker exposure to them, and to maximize the effectiveness of field procedures used in removing contaminants, has been a priority. During the anthrax outbreaks, for example, extensive environmental sampling was done by the Centers for Disease Control and Prevention (CDC)/National Institute for Occupational Safety and Health (NIOSH) with over 100 individuals involved in taking approximately 10,000 environmental samples. Between 4% and 50% of environmental samples taken from a given site were positive, depending on location sampled [SMW03]. Thus there was a high degree of variability in sample results depending on sampling location. Further, many sites continue to undergo remediation to remove all traces of Bacillus anthracis. If optimal sampling locations could be determined where bacteria were most likely present, and if

* The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health or the National Center for Health Statistics.
environmental sampling were done at these locations, it is possible that the cost of complete building remediation could be significantly reduced. Decontamination efforts, including those continuing ones in postal facilities, are not only associated with the direct and indirect economic impacts of current sampling procedures, but also with the psychological costs such as:

- Loss of faith in management claims about safety of the work environment and subsequent strained relations between labor and management.
- Postal system processing delays and imbalances in the use of institutional resources.
- Loss of work time, wasteful use of medical services, and lawsuits.

Various approaches may be used to characterize the spread and distribution of contaminants in a building. Two such approaches are computational fluid dynamics (CFD) and multizone (MZ) modeling. In CFD simulations, contaminant release information (such as from a mail sorting machine or digital bar code sorting machine) is used to form a set of boundary conditions. The equations for conservation of mass, momentum, and energy for dispersion of the contaminant (gas, liquid droplets, or solid particles) throughout the environment are solved iteratively on a high-resolution mesh until the imbalances in the equations (the residuals) decrease to an acceptably small number (iterative convergence). MZ modeling is a coarser rendering of the conservation laws and is more practical for buildings because it is less computationally burdensome. It is a network model dividing the space into zones that may also have architectural relevance such as rooms. The zones are connected by an airflow path that is often the heating, ventilation, and air conditioning (HVAC) system.

Real transport of contaminants in occupied spaces is a complicated phenomenon, made up of knowable and unknowable features, particularly when trying to reconstruct a past event. Dispersion may occur via mechanisms that can be modeled deterministically: airflow in a ventilation system, air movement due to pressure differences between areas, temperature gradients, and via activities that are difficult or impossible to characterize, such as office mail delivery and sorting, or foot traffic.

Another approach, initially developed for identification of metal ore deposits, has been through adaptation of geostatistical models for identification of concentrations of environmental contaminants. CFD modeling of the concentration field of a contaminant is a deterministic approach while the geostatistical approach is a probabilistic one.

It has been suggested that, once a concentration field for the contaminant is known, an algorithm could be developed so as to optimize the taking of further environmental samples by intensively sampling locations where there would likely be a high probability of extreme concentrations of contaminant present. Such an algorithm could implement taking additional environmental samples in locations adjacent to those where extreme levels of contaminant were present as determined in previously sampled locations. Such a process
using information from previous sampling locations to determine the next sampling location is called adaptive sampling [KWG03a, KWG03b].

In this paper, we consider approaches to modeling of the concentration field in a building following release of a contaminant and subsequent environmental sampling procedures for decontamination of that site. We first consider dispersion characteristics and concentration levels of contaminants that might be released in the workplace. We then consider models of the concentration field using probabilistic approaches such as geostatistical or spatial techniques, and deterministic ones such as CFD. A procedure for site characterization and decontamination incorporating adaptive sampling is introduced. Finally, investigation of contaminant dispersion in a room is proposed using mathematical modeling of tracer gas concentrations in a ventilation chamber, and limitations and advantages of the modeling process are discussed.

2 Dispersion of Contaminants in the Workplace

2.1 Example — Dispersion of Anthrax Spores

Investigators have determined that *Bacillus anthracis* spores contained in a letter arriving at a postal facility can be aerosolized during the operation and maintenance of high-speed, mail-sorting machines, potentially exposing workers and possibly entering HVAC systems [CDC02c]. Spores could be transported to other locations in the facility through the ventilation system, by airflow differentials between work areas, or during everyday work activities as shown in Fig. 1. In a building such as a postal processing and distribution facility, factors such as the use of compressed air to clean work surfaces, movement of personnel and carts throughout the building, and a large-volume building with multiple doors being opened and closed may also alter ventilation patterns and change contaminant transport patterns. Indeed, among environmental samples taken at postal processing and distribution facilities by investigators from the NIOSH, positive samples for *Bacillus anthracis* were obtained at diverse locations including furniture and office walls. Some of these locations were some distance away from the mail-sorting machine (Table 1) [SMW03]. Because of the potentially ubiquitous distribution of contaminants, recommendations for protecting building environments from airborne chemical, biological, or radiological (CBR) attack have been developed [NIO02]. These include maintenance and control of the building ventilation and filtration systems, isolation of areas where CBR agents might enter the building (lobbies, mailrooms, loading docks, and storage areas), and control of pressure/temperature relationships governing airflow throughout the building.

Environmental Sampling for Anthrax

Recognized techniques for collecting environmental samples to detect the presence of *Bacillus anthracis* include wipe, swab, and vacuum surface sampling
techniques [CDC02b]. Environmental samples are cultured and then tested for the presence of the bacillus [CDC02a]. Often only binary outcomes (presence/absence of anthrax spores) are reported following culture, although continuous measures such as concentration (spores per area sampled) may sometimes be reported [SHT02].

![Diagram](image)

**Fig. 1.** Possible contaminant transport mechanisms.

| Sampling Location                        | # of Samples | # and Percent Positive |
|------------------------------------------|--------------|------------------------|
| Mail-sorting devices                     | 435          | 151 (34.7%)            |
| Other postal machines/equipment          | 288          | 59 (20.5%)             |
| Office furniture                         | 49           | 20 (40.8%)             |
| Office equipment                         | 32           | 8 (25.0%)              |
| Ventilation system                       | 26           | 10 (38.5%)             |
| Windows                                  | 24           | 11 (45.8%)             |
| Mailbag/pouch/box                        | 16           | 13 (81.3%)             |
| Wall/wall boxes                          | 14           | 2 (14.3%)              |
| Floor                                    | 5            | 2 (40.0%)              |

*Surface samples taken by NIOSH investigators.
2.2 Contaminant Distributions

The behavior of contaminants in occupational settings has been studied extensively [LBL77, Rap91]. Concentrations of contaminants such as aerosols or dusts in occupational and environmental samples are commonly considered to follow a lognormal distribution with probability density function:

\[
p(c) = \frac{e^{-(\log((c-\theta)/m))^2/(2\sigma^2)}}{(c-\theta)\sigma \sqrt{2\pi}} \quad c \geq \theta; m, \sigma > 0,
\]

where \( c \) is the contaminant concentration, \( \sigma \) is the shape parameter, \( \theta \) is the location parameter, and \( m \) is the scale parameter. Typically we consider \( \theta = 0 \) and \( m = 1 \) for the standard lognormal distribution:

\[
f(c) = \frac{e^{-(\log(c))^2/(2\sigma^2)}}{(c\sigma \sqrt{2\pi})} \quad c \geq \theta; \sigma > 0.
\]

The lognormal distribution is completely determined by the median or geometric mean (GM) and geometric standard deviation (GSD). Conditions appropriate for the occurrence of lognormal distributions in environmental and occupational data are [LBL77]:

- The concentrations cover a wide range of values, often several orders of magnitude.
- The concentrations lie close to a physical limit (zero concentration).
- The standard deviation of the measured concentration is proportional to the measured concentration.
- A positive probability exists of very large values (or data “spikes”) occurring.

It should be noted that the distribution of contaminants in a building environment following an airborne CBR attack might not necessarily follow a standard lognormal distribution considered above, however, due to outside factors affecting dispersion. Such factors could include airflow in the HVAC system, airflow in the occupied space driven by the ventilation system, air movement due to pressure differences between room areas and temperature gradients, and general office activities such as mail delivery, sorting, or general foot traffic (see Fig. 1).

A hypothetical lognormal distribution of anthrax spores in environmental samples collected using a high-efficiency particulate air (HEPA) vacuum in work areas of a postal facility is shown in Fig. 2. The curve in Fig. 2 was developed using measured values of \( c = 8.93 \) (GM of spore concentrations) and \( \sigma = 1.87 \) (GSD of spore concentrations).
3 Remediation: How Clean Is Clean?

The desired concentration levels of contaminants may determine methods used during remediation, and thus the cost of remediation of any facility. Thus the extent of the cleanup effort may well be determined by contaminant levels deemed to be acceptable by governments, workers, and employers. For some agents any presence is unacceptable. This is the case for anthrax spores, for which there are currently no occupational or environmental exposure standards, resulting in the massive remediation efforts, closings of certain postal facilities, and large numbers of samples taken to test for the presence or removal of anthrax following the anthrax investigations of October–December, 2001. It should be remembered that even with large numbers of samples, uncertainty still may exist since the limits of detection of present sampling and analytical methods for anthrax spores are unknown and because of limitations on the number of locations tested. In addition, there are currently no validated sampling and analytical methods specifically for *Bacillus anthracis* in environmental samples.

Some evidence on lethal contaminant levels comes from event outbreaks, epidemiological case–control studies, and animal studies. For the anthrax example, information about the quantity of spores needed for health effects comes from an accidental release of anthrax spores in Sverdlovsk, Russia, in 1979 [Gui99]. From epidemiological analysis, the number of spores calculated to cause infection in half the exposed population (*LD*$_{50}$) was found to be between 8,000 and 40,000 spores, and the typical incubation period was 2–6
days with an illness duration of 3–5 days [BB02, BB03]. Extrapolation from animal data suggests that the $LD_{50}$ is 2,500 to 55,000 inhaled spores [IOH02]. Inhalation anthrax is often fatal, whereas cutaneous anthrax usually is not. Such information has been used to model the extent of an anthrax epidemic, and to determine the role public health measures such as vaccination may play in minimizing an epidemic [BB03, JE01, MGH94].

Alternatives may exist for deciding that a building is “clean” other than requiring complete removal of a contaminant. Statistically, a building might be declared “clean” following remediation efforts: (a) if the upper 99% confidence limit for the average concentration (spore count) is less than a threshold determined by experiment, or (b) if the probability of obtaining a determination greater than the threshold is small enough, say less than $\alpha$. In setting appropriate limits, it should also be remembered that the probabilities in (a) and (b) will be affected by environmental sampling and analytical limitations in spore detection. Therefore high inferential limits and low threshold values are desirable. In the anthrax example where any presence is unacceptable, the threshold value would be 0 spores.

It should be mentioned that in certain cases, such as during a first response effort where the presence of contaminant is known and the sources must be quickly identified, it may be that an entirely statistically based sampling strategy may be undesirable or impractical. This case is in contrast to ones in which characterization and estimation of contaminant levels are needed, for which the modeling approaches discussed here may be used. In first response cases, the rationale for clearance sampling might be based on an empirical approach (observation, good practices). Since sample size and surface sampling areas cannot be calculated based on known risk, sampling in such situations might better be determined based on practical experience, observation, and good industrial hygiene practices, guided by statistical considerations.

4 Approaches to Modeling the Concentration Field

4.1 Probabilistic Methods

Geostatistical models originally developed in the mining industry for identification of high concentrations of materials have been adapted for identification of high concentrations of environmental contaminants [WE92]. Examples of models that may be of interest for these data include kriging models that make use of variogram estimates. Suppose that the measurement of a contaminant at location $s_i$ is $Z(s_i)$. For a shift value of $h$, a model of the form

$$\text{Var}(Z(s_i + h) - Z(s_i)) = 2\gamma(h)$$

has been found useful. The quantity $2\gamma(h)$ on the right-hand side is called the variogram. For the prediction problem, measurements are taken at \(\{s_i, i = 1, 2, \ldots, n\}\). The measurements are assumed to follow the model $Z(s_i) =$
The predictor is $P(Z) = \sum_{i=1,2,\ldots,n} \lambda_i Z(s_i)$, where $\sum_{i=1,2,\ldots,n} \lambda_i = 1$ [Cre93]. The optimal predictor in ordinary kriging minimizes the mean square prediction error over the class of linear predictors by using the variogram of the data and assuming that $\mu(s_i)$ is constant. There are many variants of this procedure, including use of the logarithms of the measurements $Z$, application of robust procedures, kriging on ranks, and replacement of the mean $\mu(s_i)$ by a linear combination of known functions.

In the applications of interest here, geostatistical methods could be used as follows. After an initial phase of sampling in a contaminated building, these methods could be used to predict locations of high levels of contaminant, and these locations could then be used in a second sampling phase. An application of this sort has been described by Englund and Heravi [EH94]. In the present work, differences from that study are the use of these methods in an indoor environment, with biological contaminants, which are distributed in three dimensions. The inclusion of knowledge of ventilation systems, building geometry, and effect of human occupation makes this new application a considerable challenge.

### 4.2 Computational Fluid Dynamics

CFD is a powerful deterministic model of contaminant transport that requires solving conservation laws in scenarios such as rooms. One approach, the control volume method, involves division of the physical space of the room into discrete control volumes called cells. The partial differential equations (Navier–Stokes equations) that govern fluid motion are integrated over each control volume to form simplified algebraic equations. To illustrate using $\phi$ for a general scalar variable such as mass, a momentum component, or energy, the continuum form of the general steady-state conservation equation is [BFK03, BCS03, Flu98]

$$\int \varrho \phi \mathbf{v} \cdot d\mathbf{A} = \oint \Gamma_\phi \nabla \phi \cdot d\mathbf{A} + \int S_\phi dV,$$

where

- \( \varrho \) = fluid density
- \( \mathbf{v} \) = velocity vector
- \( \phi \) = scalar variable
- \( \mathbf{A} \) = surface area vector
- \( S_\phi \) = source of \( \phi \) per unit volume
- \( \Gamma_\phi \) = diffusivity for \( \phi \)
- \( V \) = cell volume

When this partial differential equation is discretized in the control volume method, it becomes
Modeling the Concentration Field for Site Decontamination

\[
\sum_{f} v_f \phi_f A_f = \sum_{f} \Gamma_{\phi}(\nabla \phi)_n A_f + S_\phi V, \quad (5)
\]

where

- \( f = \) cell face
- \( N_{faces} = \) number of faces enclosing cell
- \( \phi_f = \) value of \( \phi \) convected through face \( f \)
- \( v_f = \) mass flux through face \( f \)
- \( A_f = \) area of face \( f \)
- \( (\nabla \phi)_n = \) magnitude of the gradient of \( \phi \) normal to face \( f \)
- \( V = \) cell volume

The algebraic equations are then solved iteratively, starting at the boundaries of the physical space. The calculations are repeated until the conservation laws are satisfied to an acceptable degree in each cell. The degree of imbalance, called the “residual,” in the conservation law is computed as the difference between the value of a variable in a cell and the value that would be expected based on what is flowing into and out of that cell via adjacent cells. A global measure of the imbalance in the conservation law is formed by summing the residuals for all cells, then dividing by the sum of the variable in all the cells. This quantity is termed a “normalized residual.”

Sensitivity of CFD models to boundary conditions (e.g., building geometry and HVAC parameters) represents both strengths and limitations of the method. While accurate reconstructions of contaminant transport and deposition are practical in simple environments, real buildings and the activities therein are often prohibitively complex and uncertain. A simplified approach for airflow in buildings, known as MZ modeling, treats building zones as nodes and the HVAC system as links in a network model [DW02, pp. 131–154]. Sandia National Laboratories has developed this method intensely, beginning in 1995, in response to the sarin gas release in Tokyo’s subway system. Their technology has been refined now to the point where building blueprints (CAD (computer-aided design) drawings) are used to more rapidly develop flow models that incorporate detailed physics of contaminant behavior. Such information helps optimize detector placement, emergency response strategies, and decontamination tactics [SNL03]. It is also possible to combine a local CFD model with a building-scale MZ model. Sextro et al. [SLS02], for example, modeled the spread of anthrax in buildings using a combination of MZ and CFD modeling.

The results of these numerical models are a predicted concentration field that can be treated as a population, from which sparser samples can be drawn, for the evaluation of statistical approaches to estimating a complete field. The fidelity of the numerically predicted fields must be assessed before they can be used as a meaningful population of concentration values. The authenticity of CFD solutions can be decomposed into two areas, verification and validation. The American Institute of Aeronautics and Astronautics defines verification
as “the process of determining that a model implementation accurately represents the developer’s conceptual description of the model and the solution to the model.” Validation is defined as “the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model” [OT02]. The following verification and validation tests would be performed.

**Spatial Discretization.** The solutions are brought to convergence using the second-order upwind scheme. Also, the results from the original grid are compared to a new grid where the number of cells has increased by a factor of 8 in regions of large gradients. This grid convergence test will be evaluated using Roache’s grid convergence index (GCI):

\[
GCI = F_S \left| \frac{f_2 - f_1}{1 - r^p} \right|,
\]

where \(F_S = 3\) is a safety factor designed to approximate errors in convergence in the coarse-grid and fine-grid studies, \(f_2\) is the coarse-grid solution, \(f_1\) the fine-grid solution, \(r\) the ratio of cell sizes, and \(p\) the order of the discretization [Roa98, pp. 107–136].

**Solution Convergence.** The solution process is iterated until the normalized residuals for each conservation equation are less than \(10^{-3}\). The second-order solution will require adjustment of the under-relaxation parameters in order to converge to this level.

**Experimental Measurements.** A CFD simulation of a space that is also characterized experimentally provides a basis for comparison of variables such as velocity and concentration. Concentration is a useful endpoint variable to determine whether contaminant transport has been simulated accurately. If not, the velocity field agreement may be used to understand how to improve the contaminant transport simulation.

The verifications and validations should be looked at as a set. Experimental measurements by themselves are not a sufficient yardstick of CFD accuracy, given that experiments also have error and that agreement between CFD and experiment can occur by chance.

## 5 Adaptive Sampling for Site Characterization and Decontamination

In general, adaptive sampling refers to sampling designs in which the procedure for selecting sites or units to be included in a second sample may depend on values of interest observed or predicted from previous samples. With this in mind, the primary application of an adaptive sampling procedure is to capture as many of the units of interest as possible in the sample based on a well-established linking mechanism. In the anthrax example, the main purpose of the adaptive sampling procedure is to identify units of a given building with concentration of spores above the threshold for remediation.
As with MZ modeling, the discretization of a building into adaptive sampling units may be done room by room or by another rule with contaminant transport relevance, such as connection via the ventilation system. It is desirable to choose units small enough to be relatively homogeneous with respect to the level of contamination within, while not making the total number of units impractically large.

To maximize the number of sampled units with concentration above the threshold, the procedure could be composed of several rounds of adaptive sampling depending on the outcomes during the survey. At the first round, an initial sample of $n$ units is taken, which could be any probability sample, for instance, a simple random sample, a stratified random sample, or a systematic sample. The sampling units could be defined as a volumetric grid of spatial units of uniform dimensions for the building, which are then indexed. It may be convenient to have this grid coincide with an equal or lower frequency subset of the numerical grid, though the numerical grids are often not spatially uniform. Interpolation between grid points then becomes necessary. If the concentration population was formed by a MZ model, the sampling units can coincide with the zones. We then examine each unit in the initial sample to determine if it contains “above the threshold level of spores.” When it does, we add units that are connected by some linking mechanism, such as a ventilation system, ordinary foot traffic, interoffice mail distribution route, or adjacency to other units that may contain the spores above the threshold. In other words, if adjacency is the linking mechanism and there are grid units in the initial sample that contain spore concentrations in excess of a prescribed threshold, add grid units adjacent to each contaminated sample grid unit according to a fixed “neighboring units” pattern. Continue to add units adjacent to these grid units in excess of the predefined threshold value until no other units need be added. Note that the threshold could be chosen to be 0, but that the resulting sample sizes required might be large. Alternatively, it is possible that the modeling proposed in Sect. 4 above can provide information concerning which units to add to the screening sample. We continue applying the procedure until all the units that are connected by the linking mechanism are exhausted.

The collection of the sampled units (created by an adaptive sampling procedure) with spore concentrations greater than the threshold value is called a network. We could have multiple networks of various sizes in the building. If a unit with concentration of spores above the threshold is selected in the initial sample and is not connected to any other unit with concentration above the threshold by the established linking mechanism, then this unit would be considered as a network by itself. We can have $k$ networks where $k$ does not exceed the initial sample size $n$.

All grid units in the $k$ networks are units with above the threshold concentration of spores and should be decontaminated. After the initial remediation based on the findings of the first round of adaptive sampling procedure, another sample of units could be selected randomly from the area that was not covered in the first round of networking. The adaptive sampling procedures
would be applied in a similar fashion as before. If units with concentration of spores above threshold are found in this round, they are decontaminated and further sampling would be done of the area not covered in the previous two rounds. If no unit above the threshold is found in the second round, we stop the sampling. In other words, if no unit in the initial sample contains a concentration of spores above the threshold in the first round, one might consider declaring that there is no evidence that the site is not clear and stop further environmental sampling. Alternatively, a second sample could be chosen for further verification of the initial findings.

After decontaminating those units with levels above the threshold, as found in the successive adaptive sampling procedures, a new sample from the whole building should be taken. At this point a nonparametric statistical approach or asymptotic distribution theory could be used to test the hypothesis whether spore concentrations in the decontaminated areas are elevated relative to the whole building sample. In this case it should be noted that the remaining contaminants no longer have the distribution they had before remediation. From the concentration field previously generated, draw a “screening” or initial sample of grid units of sufficient size to estimate the proportion of spatial grid units that contain lethal spore concentrations with acceptable precision and reliability. For example, to determine the number of grid units to be initially considered, one might use the criterion that we want to be 98% confident that estimated contaminant concentrations will have an absolute error of 1% or less. If the sampling results are binary (indicating only whether spores are present or absent), then an appropriate criterion might be that the fraction of grid units (or the total number of grid units) with spores present should be estimated with 98% confidence. It should be noted that the method used in assigning selection probabilities to the grid units is an important part of this approach and that confidence levels will be greater for subsequent samples.

Katzoff et al. [KWG03a] have used computer simulation procedures to study adaptive sampling procedures, assuming a critical spore count per unit was required. They found that the final sample of units with lethal spore counts was orders of magnitude greater for adaptive procedures than for the corresponding nonadaptive procedures. Their results also showed that consideration of the sources of contamination was important, as was the importance of an understanding of the airflow in and between the various subunits.

A sampling scheme to characterize the extent of site decontamination need not be adaptive, but could be based on expert judgment or probabilistic models as discussed in Sects. 3 and 4 above. By subtracting expected values of concentration of the contaminant (from CFD and MZ models) from observed values measured using monitors, residuals may be obtained for examination. These could be modeled to obtain information about model goodness of fit and covariance structure of the observations. If the number of samples is sufficiently large, geostatistical techniques such as variogram estimation may also be used to model the information in these residuals. The aim of this process is the approximate validation of the deterministic approaches, followed
by the improvement of these approximations by statistical means. This will ensure that the distribution of contaminants in the building environment is understood.

Environmental sampling thus has important uses:

- Identification of areas of high concentrations of contaminant.
- Comparison of results from adaptive sampling with results from CFD and MZ models.

Successful modeling of precleanup data may also help identify the location of the sources of the contaminant, if that is unknown. This can be useful both for the cleanup operation and for law enforcement.

6 Modeling Contaminant Dispersion: Proposed Investigations

Experimental data may be used to validate and improve CFD models of contaminant concentrations as described in Sect. 5. Contaminant transport experiments in buildings can validate and improve MZ modeling. Then, the numerical models can be used to validate and enhance the statistical approaches such as adaptive sampling and kriging. Adaptive sampling is one of a family of techniques, including kriging, that estimate a field at a spatial resolution higher than what is initially known. CFD may be useful to investigate the accuracy of such estimates, through its high-resolution rendering of fields. These may be viewed as populations, from which samples may be drawn for the adaptive sampling process whose estimates can then be compared to the population. Also, CFD shares with MZ the prediction of transport. Whereas an adaptive sampling routine might look neutrally in all directions from a local maximum to find other high values, the numerical airflow model would predict that higher values occur downwind of a source. In the case of an MZ model of a multiroom system, adjacent rooms fed by different air handlers may have very different concentrations, whereas a neutral adaptive routine would assume a spatial correlation.

The following is a more detailed look at preliminary experiments in a single room in the ventilation laboratory, performed for the purpose of generating a concentration field to compare with CFD. An outline of the proposed experiment is shown in Fig. 3.

The experimental layout and geometry of the ventilation chamber are shown in Fig. 4. As indicated, concentrations of tracer gas are measured in real time at both locations A and B. Distances between measuring points and velocities at airflow boundaries are also shown in Fig. 4.

Results of measurements taken in the ventilation chamber are shown in Table 2 and Fig. 5.

For this example, measurements taken at locations A and B were fit to a negative exponential model of the form
Generate concentration fields using tracer gas in a ventilation chamber

Use probabilistic approach to model concentration field (spatial modeling)

Use deterministic approach to model concentration field (CFD)

Take a sample of results from each experiment/model above for a given location

Use statistical adaptive sampling algorithm to determine locations of high levels of contaminant for further environmental sampling

**Fig. 3.** Outline of proposed experiment.

**Fig. 4.** Experimental layout and CFD geometry viewed from top of room.
\[ C(t) = \beta_0 (1 - \exp(-\beta_1 t)), \]  

(7)

where \(C(t)\) is concentration at time \(t\), to generate Fig. 5. Notice that the range of measured concentrations is larger at location A than at location B, and that concentration at stationarity is greater at location A (29.2 mg/m\(^3\) in the fitted model, versus 28.1 mg/m\(^3\) for location B), and that it took longer to reach this value at location A (118 seconds versus 66 seconds at location B). Since location A is farther from the source, a long-term diffusion action may be responsible for the longer time to stationarity. The main and more rapid transport mechanism in this room under mixing ventilation is convection. A range of concentrations was measured at both locations.

![Concentration vs Time Graph](image)

Fig. 5. Predicted concentration* versus time at two monitoring locations. Concentrations predicted under a negative exponential model (7).

| Concentration (mg/m\(^3\)) | Time (seconds) |
|-----------------------------|---------------|
| Location A                  | Location B    |
| 0                           | 0             |
| 10                          | 100           |
| 20                          | 200           |
| 30                          | 300           |

Table 2. Results of fitting observed data to negative exponential growth curve

| Location B Location A | Observations (mg/m\(^3\)) | Fitted model |
|-----------------------|-----------------------------|--------------|
| Parameter \(\beta_0\) | 29.2                        | 28.1         |
| Parameter \(\beta_1\) | 0.11                        | 0.20         |
| Time to reach \(\beta_0\) (seconds) | 118 | 66 |
Experimental measurements taken in the ventilation chamber may be modeled using an equation of the form of (7), but one in which error and autocorrelation of measurements would also be taken into account. Predictions from such a model could be compared to concentrations from CFD predictions. Once the CFD predictions are validated, the entire field produced by CFD can be used as a source for adaptive sampling. The predictions obtained from adaptive sampling from the CFD grid of values will be compared to CFD mean values. If the use of adaptive sampling can be demonstrated successfully, CFD will be used to generate grids for a variety of contaminant situations, on which further adaptive sampling will be carried out and compared to CFD predictions. CFD calculations will be performed using Fluent 6 software. Fluent 6 is a commercial code that has been widely used in academia, government, and industry for many years [Flu98].

6.1 Modeling Dispersion of Tracer Gas in a Room — Limitations and Further Research

As mentioned earlier, the tracer gas experiment has limitations that include the following issues:

- **Location**: The point source(s) in the tracer gas experiment are known and fixed, whereas source location may be an unknown in a real attack.
- **Agent**: The tracer is a gas measured in air, whereas the contaminant of interest may be an aerosol in air or deposited on surfaces.
- **Generality**: Can agreement of this particular experiment with CFD modeling say anything about expected agreement in other rooms, building floors, and whole buildings?

The question of whether experimental validation of the CFD techniques in one situation says anything about accuracy in another situation deserves some attention. It is not proposed to use a CFD solution for the test room as information for the wider field of contaminants and potentially contaminated environments. Rather, the specifics of the process of arriving at the CFD solution, when validated, provide information about how that same process would perform for airflow in another occupied space. To require laboratory validation for every new situation is an unnecessary burden. The aim is to first validate CFD in a situation that is feasibly measurable, then to also apply CFD in the situation of interest not amenable to direct validation. CFD is capable of tracking particles of any reasonable aerodynamic diameter from their source to their fate on a surface. Its ability to accurately predict a tracer gas field in a test room says much about the fidelity of computed particle paths in a building. On the other hand, CFD/MZ methods may provide imperfect predictions in a real-world workplace. Perhaps by combining data and CFD/MZ values, predictions of concentrations of contaminants in unsampled units may be obtained that are better than those produced by either geostatistical methods or CFD/MZ methods by themselves. Since there is also
no guarantee that the assumptions required for second-order stationarity and ergodicity (which ensure convergence of the average of the sample values to the mean for large samples) will be met, it may be that alternative modeling approaches including transformation of data or considerations of lognormality may be needed. Furthermore, even if residuals in the modeling process appear to be appropriate, systematic components may still not be completely taken into account.

For instance, if our initial sample can be viewed as a grid, then we could perhaps combine median polish kriging (which adjusts the mean for rows and columns in the data) with estimates from the CFD/MZ data in one statistical model to provide smoothed estimates of mean concentrations at the sampled locations. To the extent that the modeled CFD/MZ data explain all variation, the additional contribution from the rows and columns might be unnecessary. Another plus of this approach would be that it does not seem to require stationary data. If the explanatory components from the model adjust for all changes in the mean, then the residuals will be stationary [Cre93]. Variograms can be fit to the residuals and these can be combined with the mean structure (from the model for the fitted mean) to obtain the full model for prediction of unsampled locations, in order to determine sampling locations for the second phase of sampling. Another advantage of using the median polish method is the use of medians protects against outliers. A somewhat different approach using medians has been discussed by Kafadar and Morris [KM02].

7 Understanding Contaminant Fields: The Need for New Multidisciplinary Approaches

Understanding the distribution and dispersion characteristics of contaminants that might be released during a terrorist attack poses great challenges. Not only are the types of contaminants varied, including chemical, biological, as well as radiological agents, but standard techniques may not be applicable and appropriate ones must be developed.

An important issue is how and when to carry out probability sampling for further characterization of results obtained following initial judgmental sampling (first response), and for determination of the extent of site decontamination. It may be that at some facilities, where anthrax (or some other toxic substance) will very likely be found, expert judgment alone can be used to choose a sampling location that yields a positive result, such as on the surface of a return air grille. However, for other locations, where the presence of the substance is less likely, the need for probability sampling is greater. The question is how to sample and how many samples to take, to accurately assess the risk to workers. Epidemiological information such as where an index case spent time can be helpful, but limiting sampling to a specific area is often hard to justify, due to the variety of pathways for contaminant transport. The same is true for the influence of the ventilation system. The field of inquiry quickly
expands then to the entire space where contamination is possible. Therefore, response teams initially rely on necessarily sparse sampling of large spaces.

Although initial environmental samples might be considered to be taken at random locations in a given area, other kinds of probability samples can be taken later to characterize the concentration field — for instance, systematic samples. One approach might be to set up a uniform grid in the area under question and designate sample locations on the grid. Modeling of the contaminant concentration field using a deterministic approach such as CFD or a probabilistic spatial analysis approach could help to determine appropriate sampling locations on the grid. Some researchers [Tho02] have demonstrated the utility of systematically sampling on a uniform grid to determine contaminant concentrations, using results to determine locations for further sampling in an adaptive procedure. In actual practice there might be a rule such as, take one sample from every room in a building.

The approach presented here for modeling the concentration field to direct further environmental sampling represents the marriage of two complementary disciplines, fluid dynamics and statistical science. Such an analytic approach not only adds validity and worth to each, but may also serve as a model for analyses of other aerosols and/or agents. It is also an example of the need for multidisciplinary approaches in understanding contaminant distributions and subsequent environmental sampling and decontamination efforts. CFD has been used in combination with other epidemiological findings to model airborne transmission of the Severe Acute Respiratory Syndrome virus (SARS), for example Yu et al. [YLW04].

While CFD can be further developed to account for patterns of airflow, deposition, and resuspension of particles around furniture, office or postal sorting machines, it is not clear that it is the most efficient method, since deterministic models get very complicated (expensive) as the details reach ever smaller resolved scales. Furthermore, accounting for the contaminant transport induced by normal human activity may be unknowable in a deterministic sense. While MZ is more practical for large buildings, it is a less complete deterministic model. In view of the limitation of both numerical techniques, statistical methods of generating the concentration field for a contaminant may be helpful in understanding contaminant distribution for site decontamination. The use of the lognormal distribution to characterize contaminant concentrations was introduced in Sect. 3. Kriging and geostatistical analyses represent alternative approaches to modeling the concentration field of a contaminant. Another approach for dynamic modeling of transport of airflows (air contaminants) in buildings is the method of Markov chain models or use of stochastic differential equations. Markov chains can be used to model turbulent diffusion and advection of indoor contaminants [Nic01]. Ideally the numerical/deterministic and the statistical/empirical methodologies (such as geostatistical displays) can complement each other in setting appropriate locations to apply adaptive sampling techniques.
Conclusions concerning concentrations of contaminants require sampling. Adaptive sampling provides a novel means of using spatial correlation to identify locations of high concentrations of lethal substances. The use of CFD/MZ and geostatistical methods to identify potential locations at which to sample can enhance adaptive sampling through identification of metaspatial or non-proximal correlations based on airflow and containment patterns.

Such multidisciplinary approaches will be useful in understanding the concentration field of contaminants and in determining methods most appropriate to site characterization and remediation. They will also be of value in other potentially hazardous situations such as characterizing the quality of air in offices or the extent of lead dust or respirable silica in construction/demolition activities.

Acknowledgments

The authors would like to acknowledge the thoughtful review and suggestions by Wayne T. Sanderson, Max Kiefer, Marty Petersen, and one anonymous reviewer. Stanley Shulman expresses his gratitude to his family, Jane, Meryl, and Richard, as well as to his sisters, Sherri and Lisa, for their support over the years.

References

[BCS03] Bennett, J. S., K. G. Crouch, and S. A. Shulman. 2003. “Control of wake-induced exposure using an interrupted oscillating jet.” AIHA Journal 64:24–29.

[BFK03] Bennett, J. S., C. E. Feigley, J. Khan, and M. H. Hosni. 2003. “Comparison of emission models with computational fluid dynamic simulation and a proposed improved model.” AIHA Journal 64:739–754.

[BB02] Brookmeyer, R., and N. Blades. 2002. “Prevention of inhalational anthrax in the U. S. outbreak.” Science 295:1861.

[BB03] Brookmeyer, R., and N. Blades. 2003. “Statistical models and bioterrorism: Application to the U. S. anthrax outbreak.” Journal of the American Statistical Association 98 (464): 781–788.

[CDC02a] Centers for Disease Control and Prevention. 2002. Basic diagnostic testing protocols for level A laboratories for the presumptive identification of bacillus anthracis. http://www.bt.cdc.gov/agent/anthrax/lab-testing/.

[CDC02b] Centers for Disease Control and Prevention. Revised April 2002. Comprehensive procedures for collecting environmental samples for culturing bacillus anthracis. http://www.bt.cdc.gov/agent/anthrax/environmental-sampling-apr2002.pdf.

[CDC02c] Centers for Disease Control and Prevention. 2002. “Notice to readers: Protecting building environments from airborne chemical, biological, or radiological attacks.” Morbidity and Mortality Weekly Report 51 (35): 789.
Cressie, N. A. C. 1993. *Statistics for spatial data*, rev. ed. New York: Wiley.

Dols, W. S., and G. N. Walton. 2002. “CONTAMW 2.0 user manual; Multizone airflow and contaminant transport software.” NISTIR 6921, U. S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, Building and Fire Research Laboratory.

Englund, E. J., and N. Heravi. 1994. “Phased sampling for soil remediation.” *Environmental and Ecological Statistics* 1 (3): 247–263.

Fluent 5 user’s guide. 1998. Lebanon, NH.

Guillemin, J. 1999. *Anthrax: The investigation of a lethal outbreak*. Berkeley: University of California Press.

Inglesby, T. V., T. O’Toole, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, J. Friedlander, J. Gerberding, J. Hauer, J. Hughes, J. McDade, M. T. Osterholm, G. Parker, T. M. Perl, and K. Tonat. 2002. “Anthrax as a biological weapon, 2002: Updated recommendations for management.” *Journal of the American Medical Association* 287 (17): 2236–2250.

Johnson-Winegar, A., and D.W. Evans. 2001, August. “Anthrax and other vaccines: Use in the U. S. military.” *Proceedings of the Joint Statistical Meetings*. American Statistical Association.

Kafadar, K., and M. D. Morris. 2002. “Nonlinear smoothers in two dimensions for environmental data.” *Chemometrics and Intelligent Laboratory Systems* 60:113–125.

Katzoff, M. J., A. Wouhib, and J. F. Gonzalez. 2003. “Refinements of adaptive sampling in site decontamination.” *Proceedings of the Joint Statistical Meetings*. American Statistical Association, 2092–2096.

Katzoff, M. J., A. Wouhib, and J. F. Gonzalez. 2003. “Using design-based adaptive sampling procedures in site decontamination.” 35th Symposium on the Interface: Computing Science and Statistics, Security, and Infrastructure Protection, Salt Lake City, UT.

Leidel, N. A., K. A. Busch, and J. R. Lynch. 1977. “Occupational exposure sampling strategy manual.” Publication 77-173, U. S. Department of Health, Education, and Welfare, National Institute for Occupational Safety and Health.

Meselson, M., J. Guillemin, M. Hugh-Jones, A. Langmuir, I. Popova, A. Shelokov, and O. Yampolskaya. 1994. “The Sverdlovsk anthrax outbreak of 1979.” *Science* 266:1202–1208.

National Institute for Occupational Safety and Health. 2003. “Guidance for protecting building environments from airborne chemical, biological, or radiological attacks.” Technical Report DHHS-NIOSH, No. 2002-139, U. S. Department of Health and Human Services, CDC, NIOSH.

Nicas, M. 2001. “Modeling turbulent diffusion and advection of indoor air contaminants by Markov chains.” *American Industrial Hygiene Journal* 62 (2): 149–158.

Oberkampf, W. L., and T. G. Trucano. 2002. “Verification and validation in computational fluid dynamics.” Sandia report SAND2002-0529, Sandia National Laboratories, Albuquerque, NM.

Rappaport, S. M. 1991. “Assessment of long-term exposures to toxic substances in air.” *Annals of Occupational Hygiene* 35 (1): 61–121.
