Title

Characterizing buildings for airflow models: What should we measure?

Permalink

https://escholarship.org/uc/item/1jm9f04g

Authors

Price, P.N.
Chang, S.C.
Sohn, M.D.

Publication Date

2004-06-01

LBNL - 55321

Characterizing Buildings for Airflow Models: What Should We Measure?

P.N. Price, S.C. Chang, M.D. Sohn

Environmental Energy Technologies Division
Indoor Environment Department
Lawrence Berkeley National Laboratory
Berkeley CA 94720, USA

June 2004

This work was funded by the Defense Threat Reduction Agency of the Department of Defense, through the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Characterizing Buildings for Airflow Models: What Should We Measure?

P.N. Price, S.C. Chang, M.D. Sohn

Indoor Environment Department,
Lawrence Berkeley National Laboratory, Berkeley CA 94720, USA
email: pnprice@lbl.gov http:// eetd.lbl.gov/ied/APT/APT.html

Summary: *Airflow models of buildings require dozens to hundreds of parameter values, depending on the complexity of the building and the level of fidelity desired for the model. Values for many of the parameters are usually subject to very large uncertainties (possibly an order of magnitude). Experiments can be used to calibrate or "tune" the model: input parameters can be adjusted until predicted quantities match observations. However, experimental time and equipment are always limited and some parameters are hard to measure, so it is generally impractical to perform an exhaustive set of measurements. Consequently, large uncertainties in some parameters typically remain even after tuning the model. We propose a method to help determine which measurements will maximally reduce the uncertainties in those input parameters that have the greatest influence on behavior of interest to researchers. Implications for experimental design are discussed.*

Keywords: *airflows, modelling, sensor placement, COMIS, CONTAM*
Category: *Sensors and measurement techniques*

1 Introduction

Experiments can measure airflow and pollutant transport in buildings, but these experiments are difficult and time-consuming. In practice, there are never enough resources to monitor pressures and concentrations in every area of interest, so experimenters must ask: Where should the limited number of available sensors be deployed to best estimate the air flows, pollutant transport, or other quantities that are most important? Implicit in this question is another question: what do we mean by "most important"? The flow rate through a closet may be unimportant in most contexts, but may be very important if toxic chemicals are stored there.

Often experiments are designed to collect data that can be used to create or improve a building model, which can then be used to investigate questions of interest and to understand the answers to those questions.

For the work described here, a specific, quantitative question that is of primary interest must be selected; parameters are important inasmuch as they affect the answer to that question. For illustration, we'll consider a specific six-story building, and the question of primary interest is: what is the time-average concentration on the fourth floor, over the first twenty minutes after onset of a continuous first-floor release of a tracer gas? This sort of question is of great interest to building operators or security personnel who are concerned about a possible chemical or biological attack.

Our approach to determining the best measurements to make begins with a computer model of the building. Computer programs such as Contam [1] and COMIS [2] can be used to predict airflow and pollutant transport. In these programs, a building is modeled as a collection of "zones" that are linked together. A building model specifies each zone's volume, its connections to other zones, and driving forces such as temperature differences, exterior winds, and ventilation fans. Given the building geometry and the assumptions built into the modeling approach, the input parameters (including the assumed wind pressures and other driving forces), collectively determine the differential pressures across every pair of zones, as well as the airflows between zones and thus the pollutant transport in the building.

Parameters describing the connection between zones can be hard to measure. For instance, the "mass flow coefficient" (MFC), one of the parameters that controls the mass flow rate as a function of the pressure drop between the zones[1], is closely related to the total area of all of the cracks, ducts, and other openings that connect the zones. A zone, such as a set of offices, may be connected to a ceiling plenum by an enormous number of cracks and penetrations: cracks around thousands of ceiling tiles, penetrations by pipes and conduits, ventilation grilles, and so on. Even if visual inspection is possible, the total effective cross-sectional area of these connections can be estimated only with large uncertainty.

[1] The relationship between flow rate and pressure is nonlinear (and in fact, the exponent of pressure, s, is itself a parameter of the model), so the MFC has complicated units that would be distracting to quote each time they are needed. We will discuss numeric values of the MFC throughout this paper, the units are: Kg m^3 Pa^{-s}.

Most experiments cannot directly measure the parameters that are used as inputs to the model. Instead, an experiment might release tracer gas in a zone and measure concentrations in other zones, along with zone pressures; the data must then be interpreted to estimate the parameters that describe the links.

Suppose we have a small number of sensors, each of which can measure the pressure difference, or "differential pressure," between two adjacent zones. (Differential pressure is not normally an input parameter). Where should we place these sensors to allow us to optimally estimate the parameter values that control the time-average concentration on the fourth floor?

A potential approach to this problem would be to first determine which input parameter is most highly correlated with average concentration and then to try to identify which pressure measurement would give the most information about that single parameter. This approach ignores the possibility that some other pressure measurement could give information about several parameters simultaneously, and thus could provide more overall reduction in the uncertainty in average concentration. This is far from a purely theoretical concern, as we will show.

A better approach is to directly investigate the relationship between the pressure measurements and the time-average concentration. Attacking the problem in this way automatically takes into account the complicated dependence between model inputs and outputs.

2 Methods

Our computational approach is as follows.

1. Define a main question of interest that is to be addressed, and that could be determined from the building model if the input parameters were known; for illustration here, the value of interest is average concentration on the fourth floor during the first 20 minutes of the first-floor release of one unit per second of tracer gas.
2. Create a building model, and define parameter distributions for those input parameters that are uncertain.
3. Perform Monte Carlo sampling from the distributions of input parameters, and use these parameters as inputs to COMIS to generate predictions of both the main output of interest and of quantities that could be measured in an experiment (differential pressures, in this case).
4. Select the differential pressure measurement that has the greatest correlation with time-average fourth-floor concentration. If multiple measurements can be made, then use a standard procedure known as CART (described below) to select the set of measurements that jointly predict the time-average fourth-floor concentration.

We created a COMIS model of a real-world six-story commercial building [3]; a small part of the model is illustrated in Fig. 1. The model is a substantial simplification of the building; for example, each entire floor of the building is modeled as a single well-mixed zone. Even with these simplifications, the model has about 100 adjustable parameters. We estimated each of these parameters for the real building, based on observation and engineering judgment, and created statistical distributions that summarize the uncertainties in the parameters. For example, the MFC for the link between the first floor and the ceiling plenum was assigned a lognormal distribution with a geometric standard deviation (GSD) of 3.76: the actual value of the coefficient is probably within about a factor of four of our estimate, but an error of a factor of fifteen (or more) is not out of the question.

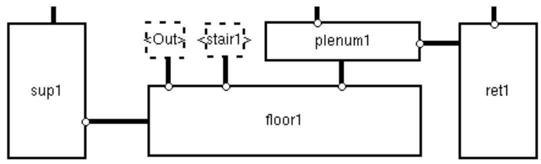

Figure 1. A small part of the building model used for this work. Each zone (box) requires several parameters, as does each link (line) between zones. Even this simplified model the model for a six-story building requires about 100 input parameters, some of them highly uncertain.

Monte Carlo sampling from the uncertainty distributions produced 500 sets of input parameters for the COMIS computer program, which simulated the airflow and pollutant transport in the case of a unit-per-second tracer-gas release on the first floor under normal operating conditions for the ventilation system. Each simulation is a single "realization" of the building model.

Figure 2 shows the predicted 20-minute-average fourth-floor concentration versus time for fifty of the realizations, and also shows a histogram of the time-average concentrations for all 500 simulations. The large variability in predicted concentrations and time averages reflects the large uncertainties in many of the input variables.

In addition to predicting the concentrations shown in Fig. 2, the model also predicts zone pressures. If a differential pressure measurement is made in the actual building, for most of the realizations it will be substantially different from the model prediction. The sets of parameter values that are inconsistent with the measurement can be discarded, leaving only those sets of parameter values that are consistent with the measurement. Thus, experimental data can be used to reduce the uncertainties in those parameters that affect the pressure (see [4] for an example in a slightly different building-related context). This is one way to "tune" the model to data.

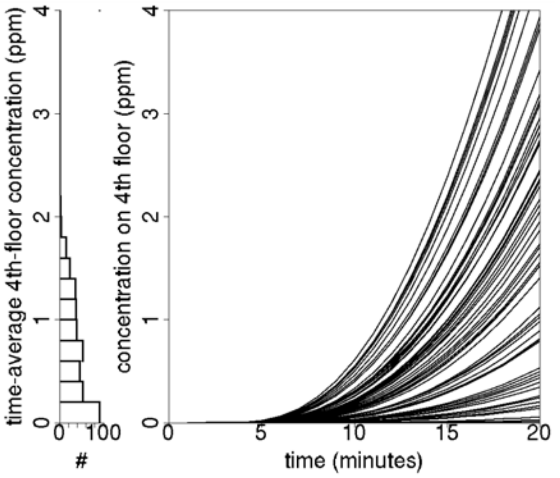

Figure 2. Left: Histogram of predicted 20-minute-average fourth-floor concentration, showing the number of realizations in each bin of average concentration. Right: Time series of predicted fourth-floor concentration for 50 of the 500 realizations of the model.

Suppose some input parameters that strongly affect a given pressure measurement also affect the time-average concentration on the fourth floor. In this case, realizations that are consistent with the pressure measurement will have a restricted range of time-average concentrations. To put it another way, if one of the differential pressures is highly correlated with the fourth-floor time-average concentration, measuring that pressure will give us information about the parameters that affect the concentration. Finding a differential pressure that is highly correlated with the time-average concentration will guarantee that the pressure measurement will help tune the important parameter values. The appeal of this approach is that it can be automated: there is no need to attain a detailed understanding of how each input parameter affects the pressures and the concentrations. In a sense, the model itself contains all of the "understanding" that is required.

The approach described above will identify the single pressure measurement that is optimal, but what if we have more than one pressure sensor (as is usually the case)? Selecting all of the pressure measurements that have relatively high correlations with the time-average fourth-floor concentration will probably not be optimal, because many of these pressure measurements will be highly correlated with each other and will thus convey information about the same input parameters. If the pressure drop from the first floor to its ceiling plenum is likely to be very close to the pressure drop from the second floor to its ceiling plenum, then there is little point in measuring both, and the second sensor would be better used elsewhere.

The best method for choosing multiple sensor locations is a subject of current research. One possibility is known as "Classification and Regression Trees," or CART [5]. CART, a standard statistical procedure, creates a "tree" for predicting a quantity of interest, such as the time-average fourth-floor concentration. At each branching point in the tree, following the left branch leads towards lower values of the concentration, while following the right branch leads to higher values.

Suppose the first branch of the tree is based on a pressure measurement in location 1: differential pressure lower than 12 Pa. leads to the left branch, higher leads to the right. Following the left branch, the distribution of time-average concentrations has variance V1; following the right branch, the distribution has variance V2. Given the model-predicted concentration measurements and differential pressures, CART analysis chooses the measurement (location 1, in this example) and the cut-point (12 Pa.) so as to minimize Max(V1,V2). Each branch is treated independently, e.g. once the first branching is decided, the next question is "considering the realizations for which the pressure measurement in location 1 is less than 12 Pa., what additional pressure measurement and cut-point will minimize the concentration variances?" The result can be a tree like the (hypothetical) one shown in Figure 3, in which different branches can rely on different cut-points or even on different measurements. In this case, three pressure measurements are needed in order to traverse the tree.

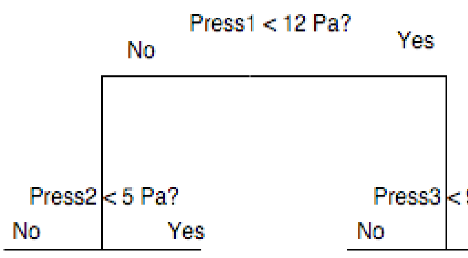

Figure 3. Tree showing the outcome of a hypothetical CART tree for predicting the time-average fourth-floor concentration from several pressure measurements. At each step, CART analysis chooses a variable (pressure1, pressure2, or pressure3) and a cutpoint so that following the tree will minimize the variance in the parameter of interest. The left-most node of the tree has the lowest average concentrations, right-most has the highest, and other nodes are intermediate.

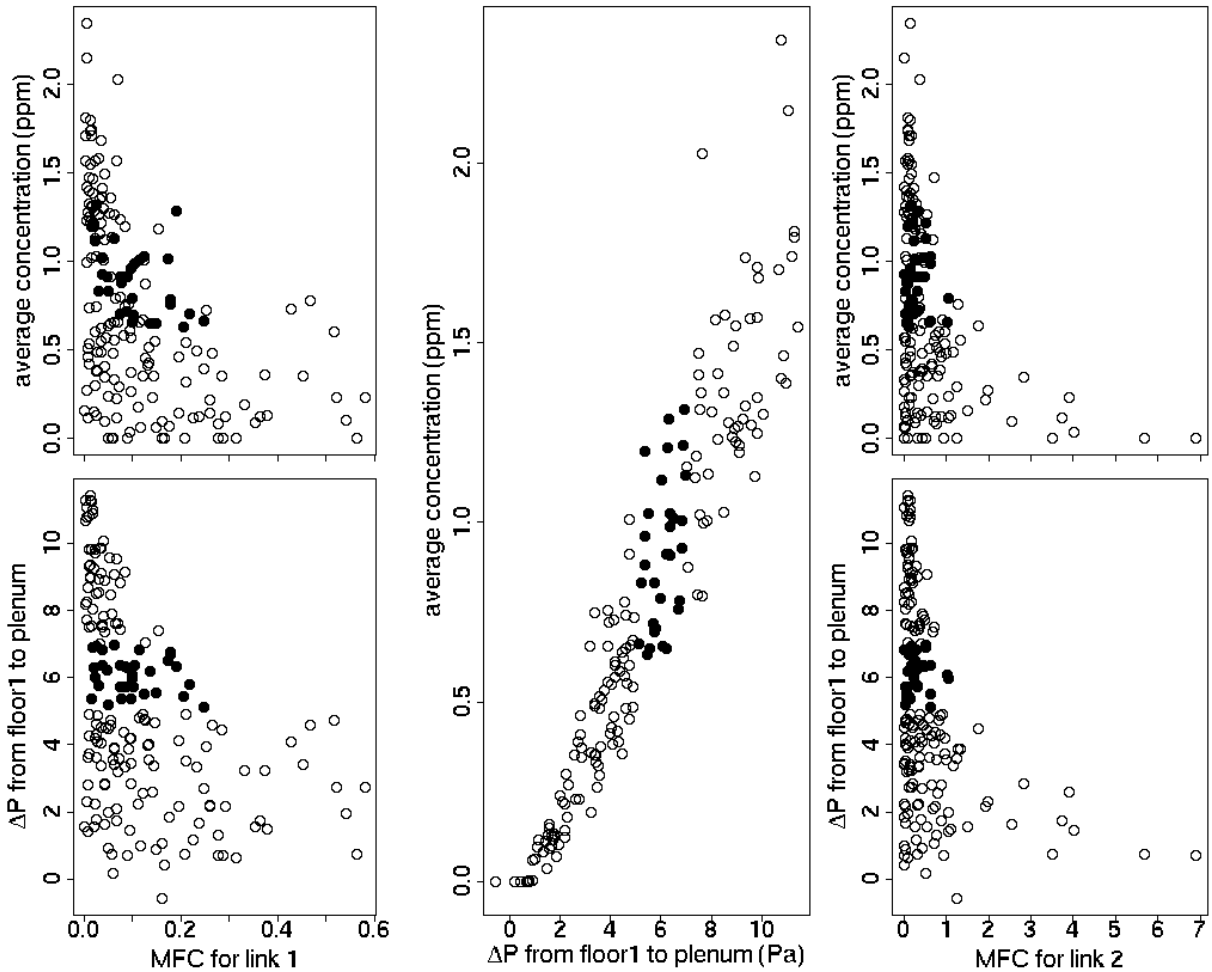

Figure 4. Left column: 20-minute-average fourth-floor concentration (top) and pressure drop from first floor to ceiling plenum (bottom), plotted against mass flow coefficient for link #1. The bottom axis label applies to both plots. Right column: 20-minute-average fourth-floor concentration (top) and pressure drop from first floor to ceiling plenum (bottom), plotted against mass flow coefficent for link #2. Center: 20-minute-average fourth-floor concentration plotted against pressure drop from first floor to ceiling plenum. In all of the plots, dark circles show realizations with a predicted pressure drop from first floor to ceiling plenum of about 6 Pa. The left and right columns plot output predictions against input parameter values; the central column plots output predictions against one another.

Results

Figure 4 shows the interplay between a few of the input parameters and output predictions for our exemplary six-story building with its large parameter uncertainties; the significance of the two different plotting symbols is explained below. Links are numbered for convenience; their physical locations are summarized in Table 1. Each point in Figure 4 represents a single model realization (a random sample of the 500 realizations is shown).

Each realization should be conceptualized as a different possible six-story building, not a different condition in a single building. For instance, considering the central plot, it is tempting to think that the model predicts that increasing the pressure difference between the first floor and the ceiling plenum will also increase the time-average concentration on the fourth floor. In fact, there is no way to "increase the pressure difference" while holding all else constant: the pressure difference is a model output, and depends in a complicated way on many input parameters. A more correct interpretation is "if a six-story building like this one is has a relatively large pressure difference between the first floor and the ceiling plenum, it will probably also have a relatively high time-average concentration on the fourth floor."

The MFC values are random draws from lognormal distributions that are based on our assessment of the likely values of these parameters, which is why there are many realizations that assume low values for these parameters, and only a few that use higher values.

The central panel plots two model predictions against each other: the 20-minute-average fourth-floor concentration, in the case of a unit-per-second release on the first floor, and the differential pressure between the first floor and the ceiling plenum. (In this building, first floor is pressurized relative to the plenum because the plenum is part of the ventilation return system). There is a very strong correlation between this pressure difference and the average fourth-floor concentration. For our relatively simple six-story building model it is possible to trace through the model and see why this is so, but it is not necessary to do this in order to conclude that measuring this pressure drop will provide information about the parameters that affect the time-averaged fourth-floor concentration.

To see how this works, suppose a measurement of the differential pressure from the first floor to the ceiling plenum finds and that it is about 6 Pa. (The realizations for which the differential pressure is around 6 Pa. are indicated by black dots in all of the plots in Figure 4.) As can be seen in the lower left and lower right plots of Fig. 4, a differential pressure near 6 Pa. only occurs for certain ranges of the MFC values: between 0–0.2 or so for link #1, and between 0–1 for link #2 (see footnote 1 for units). According to the model, a low value of the MFC for link #1 is not *sufficient* to guarantee a differential pressure near 6 Pa.—many other parameters are involved too—but it is *necessary* to generate a differential pressure near 6 Pa. This is the principle that allows the pressure measurements to be used to tune the input parameters.

What if we were to measure a lower value of the differential pressure? As the lower left and lower right panels of Fig. 4 show, according to the model a pressure measurement of, say, 2 Pa. could be generated by any of a very wide range values for the MFC of link 1 and link 2. Does that mean that if we find that the pressure difference between the first floor and the ceiling plenum is low, we haven't learned anything about any of the input parameters? No. The predicted relationship between pressure and average concentration must be caused by some input parameter or parameters. In this example there are two phenomena: (1) some other parameters come into play, and (2) although the MFC for either link #1 or #2 can fall within a wide range and still lead to a low average concentration, low concentrations usually occur only when MFC for link #1 is low and for link #2 is high, or vice versa.

Table 1. Description of links between zones, for four out of the 60 links in the model.

Link	Desciption
1	Connection between HVAC supply penthouse and outdoors.
2	Connection between ceiling plenum and HVAC return duct.
3	Connection between supply duct for the fourth floor and supply duct for the lower floors.
4	Connection between the fifth floor and its ceiling plenum.

Figure 5 shows the predicted 20-minute-average fourth-floor concentration plotted against the MFC for four links in the model; realizations for which the differential pressure between the first floor and the ceiling plenum is around 2 Pa. are plotted with black dots. In contrast to the situation when the differential pressure is 6 Pa., a pressure of 2 Pa. is consistent with a wide range of values for the MFC of link #1 and link #2 (top row of plot). However, the MFC values for some other links (bottom row shows two examples) are more constrained. This is not some happy accident; given the central plot of Fig. 4, .

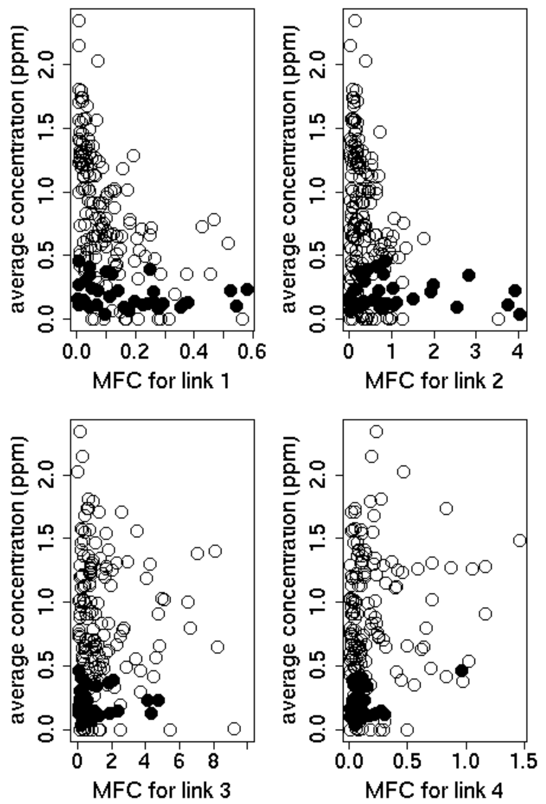

Figure 5. Predicted 20-minute-average fourth-floor concentration plotted against the MFC values for several links. Realizations that lead to a differential pressure near 2 Pa. from the first floor to the ceiling plenum are plotted with black dots.

For our six-story building model, the pressure drop from the first floor to the ceiling plenum is so highly correlated with the fourth-floor 20-minute-average concentration that, if we believe the model, this measurement is by far the most important one. But if we have two differential pressure sensors, where should we place the second one?

Figure 6 shows the CART tree for predicting 20-minute-average fourth-floor concentration from differential pressures, for a unit release per second on the first floor. At each terminal node, the average ppm value is shown for the realizations that comprise the node. For several of the routes through the tree, the result is just an inefficient way of duplicating the relationship, shown in Fig. 4, between the first-floor-to-ceiling-plenum pressure and the fourth floor concentration. But CART finds that if the first-floor-to-plenum pressure drop is high (above 7 Pa.), then the pressure drop from the supply duct to the fourth floor affects the fourth-floor concentrations substantially. If it is possible to make two differential pressure measurements, then these are the two to measure.

The utility of these two pressure measurements makes sense for our building, because the first-floor measurement is related to how quickly the gas will be pulled into the ventilation system, and the supply duct measurement is related to how quickly the gas in the ventilation system, which re-circulates some air, will be supplied to the fourth floor. (From this explanation, it is "obvious" that these are the two measurements to make, but this was not obvious to us before we started this analysis!)

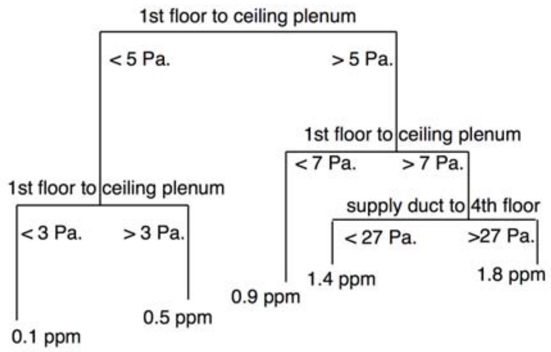

Figure 6. CART tree for predicting 20-minute-average fourth-floor concentration from differential pressures, for a unit release per second on the first floor. At each terminal node, the average ppm value is shown for the realizations that comprise the node.

Discussion

Buildings are complicated. Indeed, even simplified building models are complicated. Consequently, it can be hard to understand how the parameters that describe the building are related to the airflows and pollutant transport in the building. It can be hard to figure out what parameters are the most important to measure, and it can be hard to figure out what measurements will provide the most information about those parameters. In practice, designers of experiments rely on their judgment in determining which parameters need the most attention and in deciding where to place measurement devices. They simply have no other option, since experiments are typically completed before any modeling is begun, even though the goal of experiments is often to allow construction of an acceptably accurate model.

An alternative exists, as discussed in this paper. Create a preliminary model that relies only on the connectivity of the zones (what is connected to what) and on easily observable parameters such as zone volumes. Describe other parameters with uncertainty distributions. Sample from the parameter distributions and exercise the preliminary model, and analyze the results to determine what measurements will most reduce the uncertainties in the parameters that affect a specific question of interest. When the ultimate goal of the experimental program is to collect information that allows creation of a building model, this procedure doesn't require much extra work, but rather a shift of some modeling work from post-experiment to pre-experiment.

Of course, this approach is only a starting point for selecting candidate measurement locations. Even something as simple as the connectivity in the preliminary model could be incorrect. If the building model is substantially wrong, then of course the predictions from the model may not be useful for selecting measurement locations.

The approach described here will at least produce a list of recommended measurement locations, and the means for the person designing the experiment to investigate why they are recommended by making plots such as Figures 4, 5, and 6. She can then investigate and judge whether these recommendations are reasonable.

Acknowledgements

We thank Matthew McLeod and Ashok Gadgil for helpful advice. This work was funded by the Defense Threat Reduction Agency of the Department of Defense, through the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

References

[1] Walton GM. CONTAM96 User Manual. Gaithersburg: US Department of Commerce, National Institute of Standards and Technology, Report NISTR 6056, 1997.

[2] Feustel HE. COMIS—An International Multizone Air-Flow and Contaminant Transport Model. *Energy and Buildings* 30 (1999) 3-18.

[3] Black DR, Thatcher TL, Delp WW, Derby EA, Chang S-C, and Sextro RG. Data Summary Report of Commercial Building Experiments in Salt Lake City, UT from May 17 to June 10, 2002. Lawrence Berkeley National Laboratory Report LBNL-53801, 2003.

[4] Sohn MD, Reynolds P, Singh N, and Gadgil AJ. Rapidly Locating and Characterizing Pollutant Releases in Buildings. *J. Air & Waste Manage. Assoc.* 52 (2002) 1422-1432.

[5] Breiman L, Friedman JH, Olshen RA, and Stone CJ. (1984) Classification and Regression Trees. Wadsworth, New York.