LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-231455

Material-based Stratification

D. H. MacQueen

June 4, 2007

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Material-based Stratification

Don MacQueen
Lawrence Livermore National Laboratory
June 28, 2007

Abstract

A simple probability model was applied to detection sampling in a room or space in which different surface materials are present. The model assesses the overall detection capability when the sampling and analytical methods have different performance properties for the different materials. The results suggest that some common sampling strategies may not be ideal. In particular: (1) In a single room or area that includes different surface types with different detection properties, do not use a single sampling grid with a common spacing throughout. (2) If it is known or strongly suspected that one material has better detection properties than the other, place *all* samples on that material. (3) When it is completely unknown which material has the better detection properties, allocate the samples equally between them.

Introduction

Consider a situation in which a large indoor space is to be sampled in order to find out if a contaminant (for example, viable *Bacillus anthracis* (*B.a.*)) is present, and the entire space is being considered for sampling. Such a circumstance could arise, for example, during clearance sampling, when a sampling grid covering the flat surfaces (floor, wall, ceiling) of the whole space is used as a back up to focused sampling. Or, it could occur during characterization, in portions of a facility where there is essentially no information about which areas are more likely to be contaminated, and therefore no basis for judgmental selection of sampling locations. It could also be the case that judgmental samples are being collected in parts of the space where contamination is considered more likely (e.g., near doors coming from more contaminated parts of the facility or beneath air supply registers), and a grid or simple random sampling design is being used in the remainder of the space.

Suppose, further, that in this large space about 30% of the flat surface area is carpeted, and the remaining 70% consists of a hard, smooth material such as terrazzo tile. The sampling team may consider using different sampling methods on the different surfaces, such as wipes on the terrazzo and vacuum samples on the carpet (this is not meant to imply they *should* use different methods, only that they *may* use different methods).

If the sampling team chooses to use different sampling methods, then the sampling strategy could be described as using “material-based” stratification. The question then arises: what proportion of the samples should be collected from each of the surface types? If a regular grid covering the whole space, including both materials, were to be used to determine sampling locations, then about 30% of the samples would be on the carpet and 70% on the terrazzo. The samples would be allocated in proportion to the sizes of the strata.

Stratified Sampling

In the statistical literature, "stratified sampling" is a general term that refers to a sampling strategy that "… makes use of prior information to divide the target population into subgroups that are internally homogeneous" (Gilbert, 1987). Here, the surface material defines the subgroups. The statistical approach to stratification uses some aspect of the subgroups to optimize the number of samples within each stratum (given a fixed total number of samples) so as to do the best possible job of meeting the sampling goal. A typical example in the statistical literature would be for estimating the average value of an attribute (such as contaminant concentration), in which case the samples would be allocated in proportion to the variability of the attribute in the subgroups. If the attribute varies more within one of the subgroups, more of the samples should be collected from that subgroup, in order to improve the estimate of the average in the subgroup where it is more difficult to estimate.

Detection sampling

The focus of this analysis is on detection sampling. That is, if a contaminant is present in a specified area, but we don't know that it is, we want our sampling to discover that fact. The specified area could be a single room, a boarding concourse in an airport, the 6th floor of an office building; a city park…any well defined area.

When there is sufficient information, or when contaminant levels are high enough, and/or the contaminant has spread throughout, discovery sampling is easy. For example, if a release is large, and the release location is known, then we have good reason to expect high levels of contamination near the release location, and we expect samples near the release location to be very likely to have positive detections of the contaminant. If it is known that someone walked through a release area shortly after the release, and then went into the next room, there is a very good chance that samples in the next room near the door will have positive results. If the contaminant is both widespread in a room, and at high levels, it does not matter much where in the room a sample is collected – it is highly likely to have a positive result. In cases like these, it should not take very many samples to discover the presence of the contaminant, and it may not matter where samples are collected, or what the surface material is.

This analysis focuses on situations where discovery is difficult. For example, when assessing the extent of contamination in a large transportation facility, there will be rooms near the furthest extent of contamination in which contamination, if present at all, will be at low levels, and probably sparse. After a decontamination fumigation that almost, but not quite, succeeded, the density of viable spores on surfaces may be low (e.g., barely detectable), or the locations of viable spores may be patchy and sparse (e.g., the fumigant was not distributed uniformly). In cases like these, the probability of detection for a randomly placed sample will be low, or equivalently the false negative rate of the sampling and analysis process will be high. Then, differences between surface materials might become critical, and questions of how many samples to place where are more important than in areas where discovery is easy. For this reason, the analysis of stratification presented below is illustrated with examples where the probability of detection is low.

Stratification for detection sampling

It turns out, for our carpet/terrazzo example, that in order to have the best chance of detecting the contaminant when it is present, the samples should not be allocated in proportion to the sizes of the areas. Quite the opposite; all of the samples should be collected from the material with the best detection probability. That is, all of the samples should be collected from one of the two surface types, and the other should be completely ignored.

The following probability calculation demonstrates this result.

Probability equations

Refer to the two strata as S1 and S2 respectively. Define:

N	The total number of samples
n_1	The number of samples collected in S1; n_1 can range from 0 to N
n_2	The number of samples collected in S2; $n_2 = N - n_1$
p_1	Probability that a single sample collected in S1 will detect the agent (also referred to as "detection capability" herein)
p_2	Probability that a single sample collected in S2 will detect the agent (also referred to as "detection capability" herein)
$q_1 = 1 - p_1$	Probability that a single sample collected in S1 will fail to detect the agent (the false negative rate in S1)
$q_2 = 1 - p_2$	Probability that a single sample collected in S2 will fail to detect the agent (the false negative rate in S2)

The overall probability of detecting the agent is

$$\begin{aligned}
\Pr(\text{detect}) &= 1 - \Pr(\text{fail to detect}) \\
&= 1 - \Pr(\text{fail to detect in S1 } \textit{and} \text{ fail to detect in S2}) \\
&= 1 - \Pr(\text{fail to detect in S1})\Pr(\text{fail to detect in S2}) \\
&= 1 - q_1^{n_1} q_2^{n_2} \\
&= 1 - q_1^{n_1} q_2^{N-n_1} \\
&= 1 - q_1^{n_1} \frac{q_2^N}{q_2^{n_1}} \\
&= 1 - \left(\frac{q_1}{q_2}\right)^{n_1} q_2^N
\end{aligned}$$

In order for Pr(detect) to be as large as possible, the ratio $\left(\frac{q_1}{q_2}\right)^{n_1}$ should be as small as possible, since q_2^N is constant.

If $q_1 < q_2$, then the ratio is smallest when n_1 is largest, i.e., $n_1 = N$, or all samples are collected in S1. Note that if

$q_1 < q_2$, then

$1 - p_1 < 1 - p_2$, and

$p_1 > p_2$.

That is, all the samples should go in S1 when the detection capability in S1 is better than in S2.

When all of the samples are in S1,

$$\Pr(\text{detect}) = 1 - \left(\frac{q_1}{q_2}\right)^N q_2^N = 1 - q_1^N \ .$$

The converse is true as well: if the detection probability in S2 is better than in S1, then all the samples should go in S2 (making n_1 small will make the ratio $\left(\frac{q_1}{q_2}\right)^{n_1}$ as small as possible, because $q_1 / q_2 > 1$). In that case,

$$\Pr(\text{detect}) = 1 - \left(\frac{q_1}{q_2}\right)^0 q_2^N = 1 - q_2^N \ .$$

This result is illustrated by example in Figure 1. We imagine that deposition is light, so that surface contamination levels are near the limit of detection. At low levels, detection is difficult, so we set the probability of detection in the second stratum, S2, rather low, at 0.1 ($p_1 = 0.1$, $q_1 = 0.9$). In S1, the sampling and analytical methods are somewhat more effective, so that the detection rate in S1 is larger. Two values for the detection rate in S1 are illustrated.

Since the example detection rates in S1 are greater than in S2, the ratios $\left(\frac{p_1}{p_2}\right)$ are greater than 1, and the ratios $\left(\frac{q_1}{q_2}\right)$ are less than one. Figure 1 shows two values for $\left(\frac{q_1}{q_2}\right)$, 0.6 and 0.8, corresponding to detection rates in S1 of 0.46 and 0.28 ($0.46 = 1 - 0.6 \times 0.9$).

The total number of available samples is $N = 10$, and these can be distributed with all samples in S2 ($n_1 = 0$), or all samples in S1 ($n_1 = 10$), or anywhere in between. The two curves in Figure 1 show the overall detection probability as a function of n_1 ranging from 0 to 10.

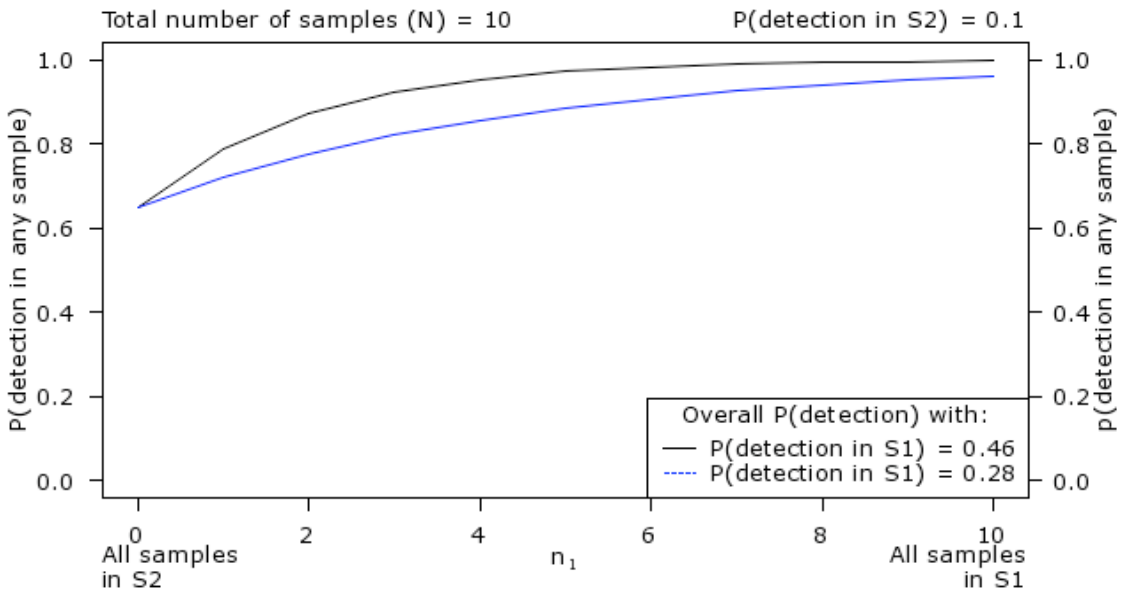

Figure 1. Probability of detecting the presence of a contaminant when distributing N = 10 samples between two strata having different probabilities of detection.

As can be seen by the two curves, the overall probability of detection is largest when all samples are in S1, the area with the better detection capability, and lowest when all samples are in S2, the area with the poorer detection capability.

It should be noted that this elementary probability calculation is similar to one of the most common rationales for judgmental sampling: sample where the agent is believed most likely to be present (i.e., "most likely to be present" is similar in concept to "most likely to be detected").

In using judgmental sampling, however, the reasons why the agent is believed most likely to be present are typically be due to event-specific information, such as the release location ("follow the letter trail"), or specific information about how an air-handling system was operating, or specific movements from place to place of people who were present, or results of numerical models that predict the contaminant fate and transport.

In the above scenario, that kind of information is considered to be not available. Instead, the different types of material lead to different detection capabilities, because there may be different collection and extraction efficiencies for different sampling methods, or for the same sampling method on different materials (compare wipes on terrazzo with vacuum samples on carpet, for which both collection and extraction efficiencies may be different; or compare vacuum on carpet with vacuum on terrazzo, for which extraction efficiencies should be the same, but collection efficiencies may be different).

Analysis for more than two material types

The principle extends to multiple strata, but the equation for Pr(detect) does not have a nice solution as above. With *K* strata,

$$\Pr(\text{detect}) = 1 - \prod_{i=1}^{K} q_i^{n_i},$$

subject to the constraint that $\sum n_i = N$. To make the probability of detection as large as possible, make product of the q's as small as possible. To do that, raise the smallest q_i to the largest possible power, which means set $n_i = N$, and all other $n = 0$. That is, put all samples in the stratum with the smallest q, and thus the largest p. If even a single sample is shifted to a stratum with a larger q, then one of the terms in the product is replaced with a larger q, and that increases the product, making Pr(detect) smaller.

Analysis when it is unknown which stratum has the better detection capability

Suppose there are two materials that have different detection capabilities (different false negative rates), but it is completely unknown which one has the better capability. Figure 2 illustrates a case where S2 has the poorer detection capability ($p_2 = 0.04$) and the ratio $q_1/q_2 = 0.5$ [$p_2 = 0.04$; $q_1 = 0.5 \times q_2 = 0.5 \times (1 - 0.04) = 0.48$; $p_1 = 1 - 0.48 = 0.52$].

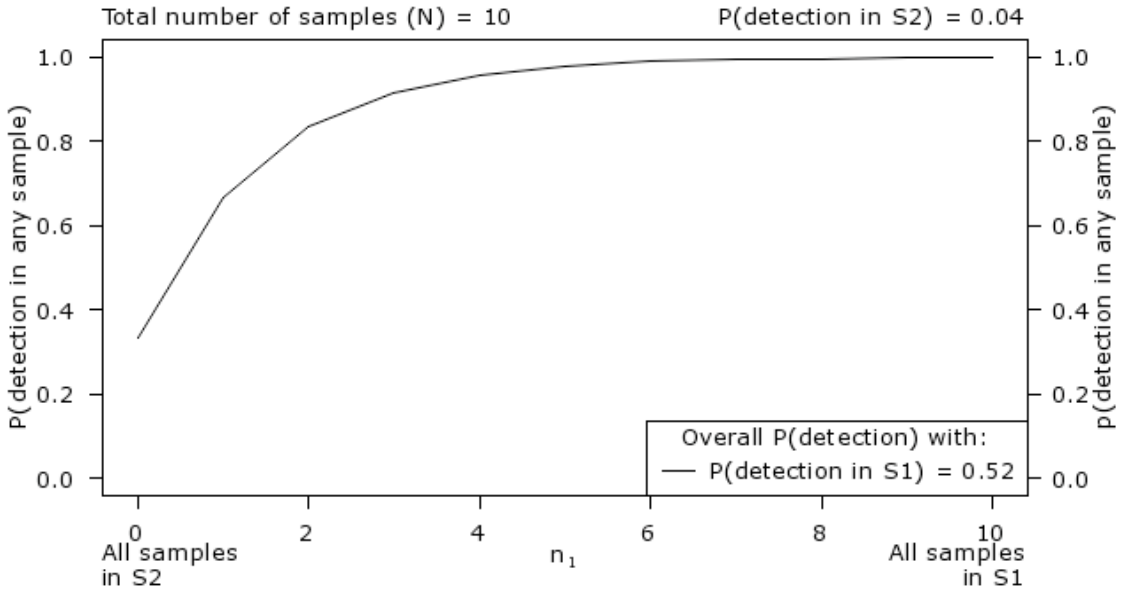

Figure 2. Overall detection probabilities for an example where one material has a poor detection capability and the other is better, to illustrate the case in which it is completely unknown which is which.

If half of the samples are placed each area, the overall probability of detection is 0.98. If the samples are allocated in a different proportion, for example, 2 and 8, then the detection probability is either 0.67 or 0.99, depending on which stratum gets which number of samples. However, since we don't know which stratum is better, we don't know which of those two overall detection probabilities will apply. By changing from 5 in each stratum to 2 in one and 8 in the other, we can either increase our overall detection probability from 0.98 to 0.99, or reduce it from 0.98 to 0.67. We get either a negligible gain or a substantial loss; obviously the change is not worth the risk. Therefore, if it is completely unknown which stratum is better, the best strategy is to place half our samples in each one. Samples are not placed in proportion to the sizes of the areas. This means

that placing a single equally spaced grid over the entire area, including both types of material, is not a good idea.

Analysis when it is uncertain which strata has the better detection capability

Now suppose it is suspected that detection capability is better on one of the materials, but there is some uncertainty about this. For example, one might be willing to say, "Detection capability in S1 is probably better than in S2, but we're not completely sure."

To model this situation, we fix the detection capability in one of the strata, and consider a range of possible detection capabilities in the other. This is illustrated by the curve in Figure 3. Detection capability in S1 is considered to range from about 0.01 up to 0.6, with capabilities at the low end being more likely (the higher the curve in Figure 3 the more likely). Detection capability in S2 is set at 0.1. Thus, the curve in Figure 3 represents the idea that detection capability in S1 is probably better than in S2, but that there is some possibility it is poorer. The expected value of the detection capability in S1 (the weighted average over the range from 0.01 to 0.6) is about 0.25, which is better than the 0.1 in S2.

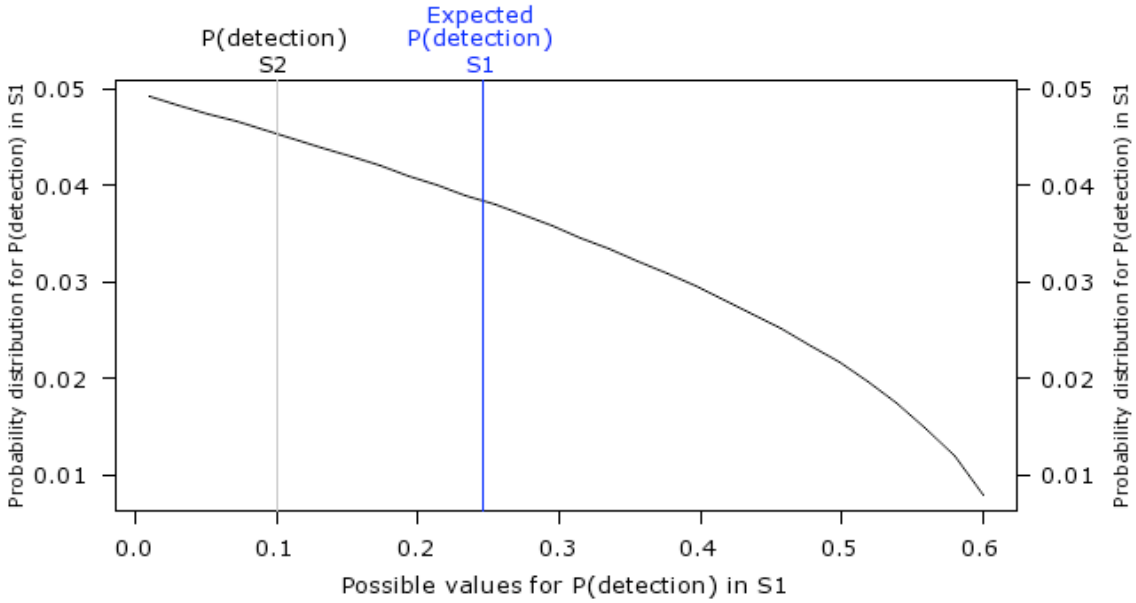

Figure 3. Example probability distribution for the detection capability of the sampling and analytical process in area S1.

To find the best value of n_1 we average over the range of possible values of q_1. We simplify the calculation by using a set of equally spaced discrete values for q_1, that is, $q_1 = \{q_j, j = 1...n_p\}$.

$$E(\text{overall probability of detection, given } n_1) = \sum_{j=1}^{n_p} P(q_1 = q_j)\left[1 - \left(\frac{q_j}{q_2}\right)^{n_1} q_2^N\right]$$

The results of this calculation, for n_1 from 0 to $N = 10$, are shown in Figure 4.

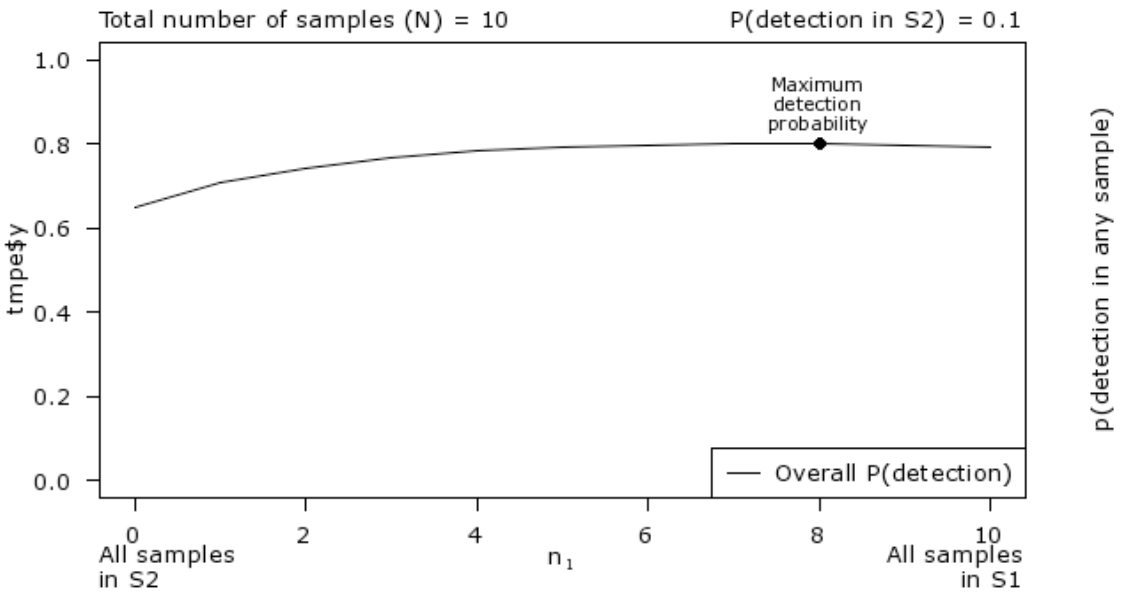

Figure 4. Example overall discovery probability when it is uncertain which stratum has the better detection capability.

The theoretical and practical interpretations of this result differ. Theoretically, the best allocation of samples is $n_1 = 8$ in S1 and $n_2 = 2$ in S2, with an overall detection probability of 0.80. More samples are in S1 because detection capability is probably better in S1 than in S2, but some are placed in S2 to allow for the possibility that detection capability may be better in S2. Practically speaking, however, if all samples are placed in S1, the detection probability is 0.79, and the difference between 0.80 and 0.79 is of course completely negligible. If all samples are placed in S2, the overall detection probability is reduced to 0.65, and that difference, though not necessarily large, is meaningful. Practically speaking, one might as well place all samples in S1.

Dispersion

The discussions above were focused on detection sampling. Detection capability (p_1 or p_2) was defined as the probability of detection given that the contaminant is present at the exact location sampled. In that context, differences in detection capability were attributed to differences in surface type and sampling and analytical processes (e.g., collection and extraction efficiencies).

The discussions were also set in the context of a space or room where there is no basis for judgmental selection of sampling locations, i.e., no reason for expecting higher surface concentrations in some areas. If surface concentrations are uniform throughout then it is differences in collection and extraction efficiencies that drive the detection capability.

In reality, detection probability is a function of both the probability that the contaminant is present (a function of dispersion), and the probability that the contaminant is detected if a sample is placed in a contaminated location (a function of the sampling and analytical process). This is written mathematically as

$$\Pr(\text{detection in a single sample})$$

$$= p(\text{present at the sampled location } \textit{and} \text{ detected})$$

$$= p(\text{detected given present at the sampled location})\,p(\text{present at the sampled location})$$

However, the probability equations in the stratified sampling analysis are ignorant of any real world reasons for differences in detection capability. The probability equations require only that there be a detection probability for a sample placed at random in each stratum.

If dispersion is sufficiently uneven, then a randomly placed sample may or may not land on a contaminated location, and the degree of unevenness affects the detection probability. That is, p_1 and p_2 depend on both the dispersion pattern and on the interaction between surface type and the sampling and analytical methods. If one of the strata is very small and dispersion is highly uneven, then the detection probability in that stratum might be low, even if the sampling and analytical methods are better in that stratum. Therefore, if dispersion is highly heterogeneous and one of the strata is very small, the recommendations described above, to sample in the stratum with the better sampling and analytical methodology, do not apply.

Applicability

In should also be noted that the examples above used a relatively small number of samples ($N = 10$). Suppose that detection is difficult, such as $p_1 = 0.05$ for each sample. The overall detection probability can still be made large by collecting enough samples. Even with as p_1 at 0.05, the overall detection probability can be raised to 0.95 by collecting 60 samples. Thus, if resources (time, personnel, fund, laboratory throughput) are sufficient to collect enough samples, the issue of allocating samples to strata may might not be critical.

Review

This analysis started out as an attempt to optimize the allocation of samples between different surface materials within a single space. After the analysis was complete it became clear that the results are intuitively sensible even without the probability analysis, as follows.

The overall chance of detection in a sampling program depends on the chance of detection for each sample. The larger the chance of detection for each sample, the larger the overall chance of detection. So put as many samples as possible in places where their individual chance of detection is greatest. If the surface material affects the chance of detection (and there is no other information to indicate the chance of detection), place all samples on the surface with the best chance of detection. An ordinary regular grid does not do this.

If a sample is moved from a given location to another location where the chance of detection is smaller, the overall chance of detection decreases. Therefore, if the affect of surface material is completely unknown, a 50-50 allocation is best, because given a 50-50 split, moving even a single sample to the other material may be a move in the wrong

direction. An ordinary regular grid does not provide a 50-50 split (unless, of course, the materials cover equal areas).

Summary

This simple probability model suggests that for detection sampling:

When it is completely unknown which material has the better detection properties, split the samples equally between them.

If it is known or suspected that one material has better detection properties than the other, place all samples on that material.

Do not use a single sampling grid (with the same spacing throughout) that covers different materials, if those materials have different detection properties. That is, in a single room that has different surface types, do not use a single sampling grid.

References

Gilbert, R.O. (1987), *Statistical Methods for Environmental Pollution Monitoring*, New York, NY, Van Nostrand Reinhold.