ANL/EAD/TM/05-1

Bayesian Approaches for Adaptive Spatial Sampling: An Example Application

Environmental Assessment Division
Argonne National Laboratory

ARGONNE IS OPERATED BY THE UNIVERSITY OF CHICAGO FOR THE U.S. DEPARTMENT OF ENERGY OFFICE OF SCIENCE

Notice

This technical memorandum is an information product of Argonne's Environmental Assessment Division (EAD). It presents results of ongoing work or work that is more limited in scope and depth than that described in formal reports issued by EAD. This memorandum has undergone internal technical review and has been edited according to EAD's quality assurance requirements. In contrast to a formal technical report, this memorandum has not been externally peer reviewed.

For more information on the division's scientific and engineering activities, contact:

Director, Environmental Assessment Division
Argonne National Laboratory
Argonne, Illinois 60439
Phone: 630-252-3107
e-mail: ead@anl.gov

Publishing support services were provided by Argonne's Technical Services Division.

About Argonne National Laboratory

Argonne is operated by The University of Chicago for the U.S. Department of Energy Office of Science, under contract W-31-109-Eng-38. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

Availability of This Report

This report is available, at no cost, at http://www.osti.gov/bridge. It is also available on paper to U.S. Department of Energy and its contractors, for a processing fee, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone (865) 576-8401
fax (865) 576-5728
reports@adonis.osti.gov

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor The University of Chicago, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, Argonne National Laboratory, or The University of Chicago.

ANL/EAD/TM/05-1

Bayesian Approaches for Adaptive Spatial Sampling: An Example Application

Prepared by

R. Johnson, D. LePoire, A. Huttenga, and J.Quinn
Environmental Assessment Division
Argonne National Laboratory

April 2005

Work sponsored by U.S. Department of Energy, National Nuclear Security Administration

This report is printed on recycled paper.

CONTENTS

NOTATION v

ABSTRACT 1

1 INTRODUCTION 1

2 MEASURING PERFORMANCE 7

3 EXAMPLE DESCRIPTION 9

4 TRADITIONAL PROGRAM DESIGN PERFORMANCE 11

5 ADAPTIVE PROGRAM PERFORMANCE 15
- 5.1 Noninformative Prior 15
- 5.2 “Good” Prior 16
- 5.3 “Bad” Prior 24
- 5.4 Spatial Autocorrelation Range Assumptions 28
- 5.5 Modifying Acceptable Uncertainty Levels 28
- 5.6 Inside-Out versus Outside-In Strategies 31

6 DISCUSSION AND CONCLUSIONS 33
- 6.1 Adaptive Sampling Program Performance 33
- 6.2 Stopping Criteria 35
- 6.3 Spatial Autocorrelation Range Assumptions 36
- 6.4 Certainty Level Specifications 37
- 6.5 Initial Conceptual Site Model 37
- 6.6 Approach Limitations 38

7 REFERENCES 39

TABLES

1 Contaminated Areas Estimated with Gridded Programs 11

2 Comparison of Sampling Program Performance When Determining Contamination Footprints 13

3 Sampling Program Details 18

FIGURES

1 Different Forms of Contamination Clustering 5

2 Example Site 9

3 Decision Point Grid for Example Site 10

4 Traditional Gridded Sampling Programs 12

5 Contamination Footprints Based on Gridded Programs 14

6 Adaptive Program Progression with Noninformative Prior 17

7 Initial Conceptual Site Model Using Prior Information 23

8 Adaptive Sampling Progression Based on “Good” CSM 25

9 Initial Conceptual Site Model Based on Misleading Information 26

10 Adaptive Sampling Progression Based on “Bad” CSM 27

11 Adaptive Sampling Progression with Correct Range and “Good” CSM 29

12 Adaptive Sampling Progression with Relaxed Confidence Level Requirements 30

13 Inside-Out Approach to Contamination Delineation 32

14 Expected and Observed Error Rates for Different Sampling Scenarios 34

15 Site Categorization as a Function of Sample Numbers 35

16 Clean Area Gain as a Function of Sample Numbers 36

NOTATION

The following is a list of the acronyms, initialisms, and abbreviations (including units of measure) used in this document.

ACRONYMS, INITIALISMS, AND ABBREVIATIONS

BAASS	Bayesian Approaches for Adaptive Spatial Sampling
CSM	Conceptual Site Model
ft	foot (feet)
ft^2	square foot (feet)
pdf	probability density function

BAYESIAN APPROACHES FOR ADAPTIVE SPATIAL SAMPLING: AN EXAMPLE APPLICATION

by

R. Johnson, D. LePoire, A. Huttenga, and J. Quinn

ABSTRACT

BAASS (Bayesian Approaches for Adaptive Spatial Sampling) is a set of computational routines developed to support the design and deployment of spatial sampling programs for delineating contamination footprints, such as those that might result from the accidental or intentional environmental release of radionuclides. BAASS presumes the existence of real-time measurement technologies that provide information quickly enough to affect the progress of data collection. This technical memorandum describes the application of BAASS to a simple example, compares the performance of a BAASS-based program with that of a traditional gridded program, and explores the significance of several of the underlying assumptions required by BAASS. These assumptions include the range of spatial autocorrelation present, the value of prior information, the confidence level required for decision making, and "inside-out" versus "outside-in" sampling strategies. In the context of the example, adaptive sampling combined with prior information significantly reduced the number of samples required to delineate the contamination footprint.

1 INTRODUCTION

BAASS (Bayesian Approaches for Adaptive Spatial Sampling) is a set of computational routines designed to support the design and deployment of adaptive sampling and analysis programs (ASAPs). In particular, BAASS is intended to support spatial sampling programs that have the following characteristics:

- The decisions to be made are binary (i.e., contamination either is above some predefined threshold or it is not).

- The purpose of sampling is to delineate the spatial extent of contamination.

- Spatial autocorrelation is present. Spatial autocorrelation is present when the results of samples that are spatially close exhibit correlation. The spatial autocorrelation range is the sample separation distance beyond which no correlation in sample results is observable.

- "Soft" conceptual information is available that indicates, at the outset, where contamination is likely to be found and where it is not likely to be found. Soft information refers to information that is indicative of the presence or absence of contamination at a particular location, but not definitive. In contrast, a sample result would be considered definitive.

- A "real-time" (i.e., able to return measurement results quickly enough to affect sampling program decisions) measurement technology is available that can reliably discern whether contamination is above or below the predefined threshold for the location being measured.

As the example illustrates, in this setting BAASS can produce sampling programs that achieve much better delineation results with far fewer samples than a traditional gridded program can accomplish.

As presently formulated, BAASS is not applicable to situations where:

- The primary goal is to estimate the average level of contamination over an area.

- Contamination is highly "spotty," that is, it displays very little spatial autocorrelation or exists in spots too small to be detected by on-site sampling or scanning.

- Sampling programs where the entire program needs to be designed and implemented without the benefit of "real-time" results.

BAASS incorporates a combination of Bayesian techniques and ordinary indicator geostatistics into its routines. Bayesian analysis provides a quantitative way to update one's understanding of a probabilistic phenomena as more information becomes available. For any particular location, whether contamination exists above some predefined threshold or not is not known definitively unless a sample is collected at that point. We can infer the probability, however, on the basis of other information available for that location and/or sampling results from neighboring locations. Bayesian statistics captures the likelihood of contamination at a location in a "prior" probability density function (pdf). As new nearby sample results become available, Bayesian analysis provides a mechanism for updating the prior pdf, converting it to a posterior pdf. The posterior pdf reflects the probability of contamination existing at a location given all the information available up to that point in time.

Geostatistics is a set of tools for the analysis of spatially correlated data. The two key components of a geostatistical approach are the variogram and kriging. The variogram is a measure of the spatial autocovariance of data sets as a function of separation distance. Data are correlated up to a separation distance termed the range. Beyond the range, data are uncorrelated. Kriging is an interpolation methodology that is based on sample data and the variogram model; it provides an unbiased minimum-variance estimate of values throughout the spatial domain. Ordinary kriging is a form of kriging in which the local mean data values do not match the global

mean of the data. Indicator geostatistics works with data that have been transformed into 0 and 1 values, representing the presence or absence of some characteristic of interest. Geostatistics allows us to infer the probability of contamination existing at a location where no samples have been collected or measurements made, on the basis of the neighboring sample or measurement locations.

BAASS requires the following:

- A set of decision points, usually in the form of a systematic rectangular grid, that is superimposed over the area of concern. The selection of the appropriate decision point density is driven by two conflicting requirements: spatial resolution needs and computational tractability. In general, the decision point grid needs to be dense enough so that grid point separation is less than the presumed spatial autocorrelation range for the contaminant(s) of concern, but not so dense that computation time becomes impracticable. BAASS provides estimates of the likelihood of contamination being present above some predefined threshold at each decision point, and identifies new sampling locations that will improve likelihood estimates.

- Two-parameter beta pdfs assigned to each decision point. The “prior” beta pdf for any given decision point reflects the probability of finding contamination at that point above the predefined threshold on the basis of all initial information and the certainty associated with that assessment. In a hazardous waste site setting, the set of prior pdfs represents the initial Conceptual Site Model (CSM) for the site. For hazardous waste sites, initial information would likely have come from a site reconnaissance, interviews, and a historical records review. In an environmental emergency response setting where an atmospheric release of contaminants has taken place, an example of initial information would be the results from National Atmospheric Release Advisory Center (NARAC) fate and transport computer modeling. BAASS provides a mechanism for accommodating situations where no prior information exists, making use of “noninformative” priors. As sampling results are generated, BAASS provides the routines for updating prior beta pdfs into “posterior” pdfs that reflect both initial beliefs and more recently acquired measurement results. The role that prior pdfs have on the initial development of a sampling program and its ultimate efficacy are illustrated by the example application that follows.

- Some assumptions must be made about the spatial autocorrelation present for the contaminant(s) of concern. This spatial autocorrelation is associated with the indicator status of contamination (i.e., above or below the threshold), not with raw data values. Variograms provide one method for assessing the presence and nature of spatial autocorrelation present in sample or measurement results. Practical experience has shown that indicator variograms are superior to parametric variograms of raw data, which typically do not show clear structure. Of primary importance is specifying the shape of

the presumed covariance function and its range. The initial selection of a covariance function range can be based on available information (i.e., for situations where fate and transport modeling provides some sense of likely contaminant spatial distribution, the covariance function could be derived from the predictive modeling results), or it can be based on what is the minimum size of contamination footprints one wishes to find. In the latter case, a rule of thumb is to set the indicator spatial autocorrelation range to the average "width" of contiguous contamination footprints one wishes to identify. Alternatively, if a suitable number of sample or measurement results are available at the onset, then the range may be determined through a variogram analysis using geostatistical packages external to BAASS.

BAASS methodology details can be found in Johnson (1996). The advantages of an adaptive sampling program for contaminant footprint delineation are greatest when contamination footprints are contiguous or blocky and "well-defined." Suppose a site is broken into a set of regularly gridded cells, some of which are contaminated, on average, above a threshold, and some of which are not. A measure of the blockiness of a contamination footprint is the ratio of the number of "clean" cells that share at least one side with a contaminated cell to the number of contaminated cells (Thompson and Seber 1996).

Figure 1 illustrates the concept of blockiness. Figure 1 shows three different contamination events, each of equal area. The more "scattered" contamination is (e.g., Scenario C), the higher the ratio. This is an important concept because as this ratio increases, the benefits of using BAASS decrease and area characterization moves from delineating areas of contamination to simply being able to correctly identify isolated contamination occurrences. For the three cases illustrated by Figure 1, this ratio ranges from 1.3 for Case A, to 2.0 for Case B, to 2.6 for Case C. In theory, at the one extreme this ratio approaches 0 for cases where contamination footprints are very large and consolidated. At the other extreme, the ratio will have a value of 4 when contamination is isolated to individual cells that are separated from each other by at least two clean cells. In the case of the environmental release of radionuclides and subsequent deposition on surfaces, one would expect in general that contamination footprints would be relatively well-defined and contiguous, giving rise to relatively small blockiness ratios.

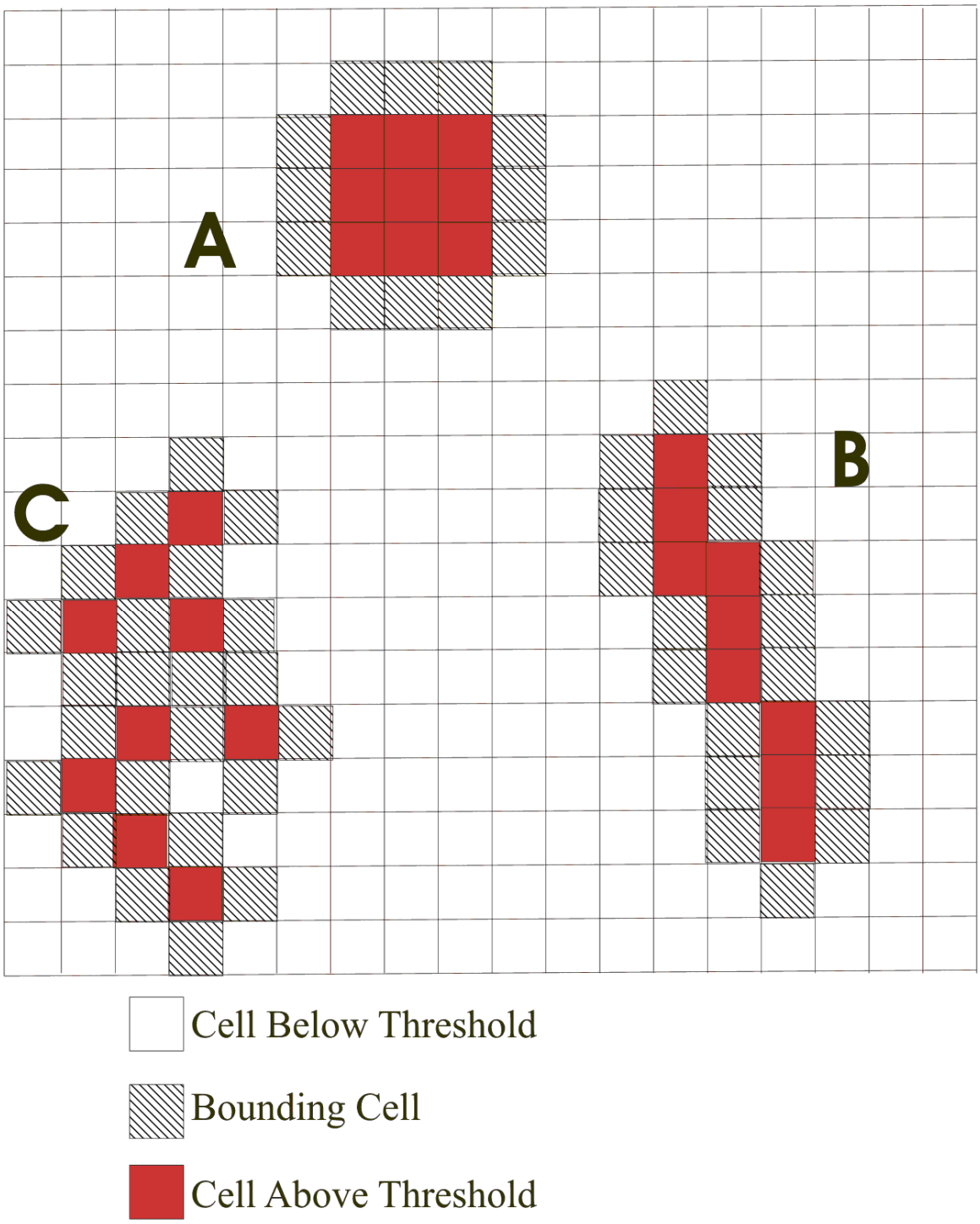

FIGURE 1 Different Forms of Contamination Clustering

2 MEASURING PERFORMANCE

BAASS can address two basic sampling program goals: (1) estimating the area (in two dimensions) or volume (in three dimensions) of material contaminated above some predefined threshold, and (2) determining the precise spatial location of material that exceeds the predefined threshold. In the context of these two sampling goals, the question is how one can measure the performance of any particular sampling program design.

For the first goal, there are two basic metrics for measuring the ability of a sampling program to estimate the area or volume of contamination. The first (assuming that the methodology is inherently unbiased) is the degree of uncertainty associated with the estimate for any given set of samples. The second is the rate at which that uncertainty is reduced with increasing sample numbers.

There are several ways to address the uncertainty associated with a contaminated volume or area estimate (and to generate the estimate itself). One approach assumes that samples are systematically or randomly distributed across the region of concern and that the probability of any sample exceeding a guideline follows a binomial probability distribution. The contaminated volume/area estimate is then simply represented by the ratio of sample numbers exceeding the guideline to total samples collected, multiplied by the area or volume of the region of concern. If the number of samples is sufficiently large, then consistent with the Central Limit Theorem, the distribution of this ratio can be approximated by a normal distribution; confidence intervals can then be calculated for the ratio, and consequently for the area/volume estimate.

A second approach is to lay a grid of points across the region of concern and infer the probability of contamination exceeding the guideline at each of the grid points, on the basis of available sample results. Indicator kriging provides one methodology for inferring grid-node-specific exceedance probabilities. Once grid-node-specific exceedance probabilities have been calculated, an area/volume estimate can be constructed by finding the expected probability of exceedance across the grid node set. Alternatively, an area or volume estimate could be estimated by the fraction of grid nodes with probability exceeding 0.5 multiplied by the area or volume of the region of concern. In the latter case, uncertainty estimates could be formed by looking at the number of grid nodes that exceed a lower probability threshold (e.g., > 0.1) and an upper probability threshold (e.g., > 0.9).

For the second goal, determining the location of contamination, two errors can be made when identifying a particular location as contaminated (contamination concentrations above a predefined threshold) or clean (contamination concentrations below a predefined threshold). These are “false positive” conclusions (i.e., locations are identified as contaminated when in fact they are clean) and “false negative” conclusions (i.e., locations are identified as clean when in fact they are contaminated). A metric that can be used to measure the ability of a sampling program to identify correctly a contamination footprint is the observed false positive and false negative rates.

3 EXAMPLE DESCRIPTION

The site of interest (Figure 2) is 250 ft by 250 ft in size (or slightly larger than 1 acre). It contains a waste lagoon, a storage building, and is bounded on two sides by roads. In addition to the location of the lagoon, there is limited topographic information for the site. The hypothetical contamination event involves a breach of the lagoon's banks and overland flow of contaminated liquid, resulting in surface soil contamination. Figure 2 shows the site and the actual footprint of soil with contamination above the regulatory threshold. The site owner is aware that contamination occurred, but does not know the extent or exact footprint of contamination that will need to be addressed. The purpose of the sampling program in this instance is two-fold: (1) estimate the contaminated surface area, and (2) delineate the contamination footprint. An alternative way of stating the latter is that the goal is to determine which portions of the site are likely clean, at some prespecified probability level.

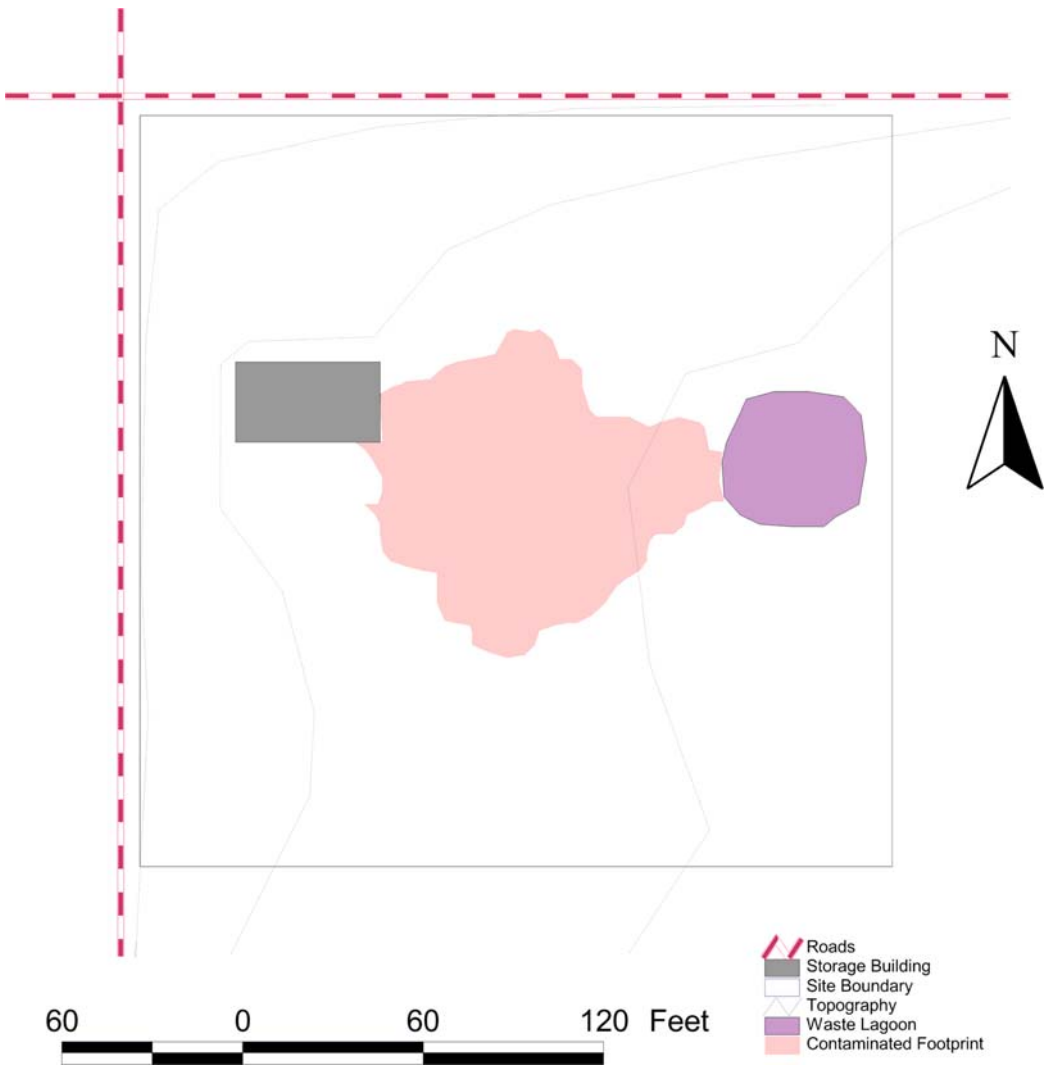

FIGURE 2 Example Site

To facilitate the analysis of the performance of various sampling program designs, a rectangular 25-by-25 node grid is superimposed over the site, resulting in a grid spacing of 10 ft. Each grid node is assumed to represent the conditions of its 100-ft^2 grid cell. Figure 3 shows the locations of these nodes, or decision points. With this set of grid nodes, the minimum number of samples required to completely delineate the boundary would be 86. For the purposes of this example, it is assumed that sampling a decision point definitively determines whether the area represented by that decision point contains contamination above or below the threshold.

For this particular example, the ratio of boundary decision points to the number of contaminated decision points is 0.45, indicating that it is a likely candidate for an application of BAASS.

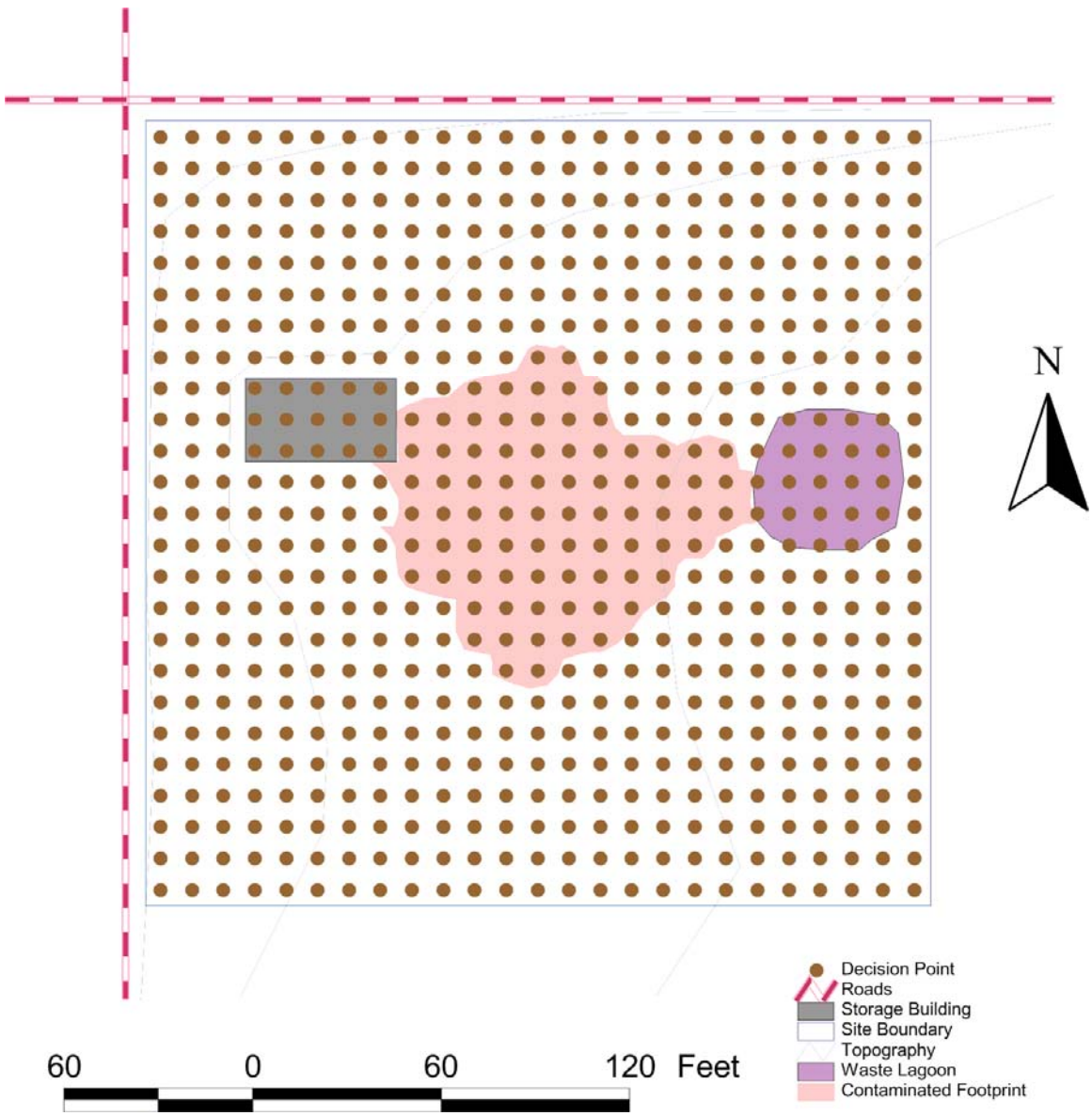

FIGURE 3 Decision Point Grid for Example Site

4 TRADITIONAL PROGRAM DESIGN PERFORMANCE

In a traditional sampling program, samples are typically laid out on a regular grid across the area of interest, sampled, and then all of the samples analyzed by an off-site laboratory. This kind of program assumes that one can estimate the appropriate number of samples required and their locations prior to the fieldwork. Table 1 shows the results for 5 different traditional sampling programs, ranging from 12 samples to 203 samples. In each case, a triangular grid was used with a random start. Figure 4 visually shows the sampling results for each of these programs.

In each case, the estimate of the contaminated surface area was the fraction of samples that exceeded the threshold multiplied by the area of the site. Confidence limits can be calculated for these estimates by assuming that the indicator results from the sampling program followed a binomial distribution with a probability of contamination equal to 0.5 (an assumption that yields the largest confidence interval for any given number of samples and their associated results). In this instance, confidence intervals can be calculated using a normal distribution approximation. Table 1 includes the confidence intervals calculated in this manner. The negative lower confidence interval for 12 samples reflects the approximate nature of our normal assumption when sample numbers are small (i.e., under an assumption of normality, negative values are possible; in actuality, the lower confidence interval is bounded by zero). As one would expect, the confidence intervals tighten as sample numbers increase. One would expect the rate of tightening to be proportional to the inverse of the square root of sample numbers, that is, increasing sample numbers by a factor of 4 would shrink the confidence interval by a factor of 2. This rate is confirmed by the values in Table 1. With 203 samples, the 90% upper confidence level on the mean is still 29% greater than the best estimate. This methodology provides an

TABLE 1 Contaminated Areas Estimated with Gridded Programs

	Best Estimates and Uncertainty (ft^2)		
Program	Lower 90% Confidence Limit	Best Estimate	Upper 90% Confidence Limit
Reality	Not applicable	10,000	Not applicable
Gridded 12 samples	-6,353[a]	5,208	16,769
Gridded 20 samples	420	9,375	18,330
Gridded 42 samples	7,213	13,393	19,572
Gridded 105 samples	6,211	10,119	14,027
Gridded 203 samples	7,041	9,852	12,663

[a] The negative area results from using a normal approximation when computing the confidence interval with a small number of contributing samples.

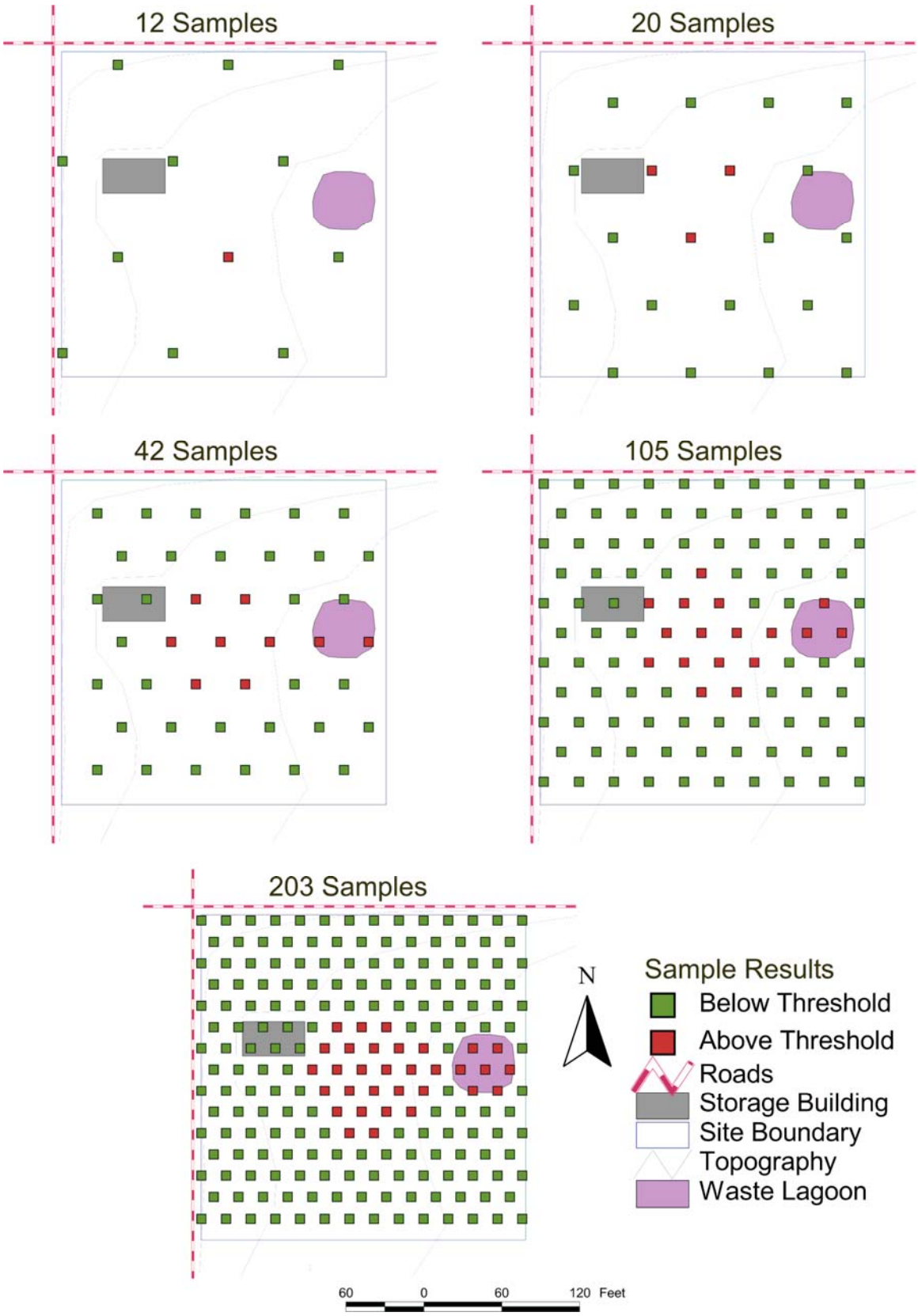

FIGURE 4 Traditional Gridded Sampling Programs

estimate of the area of contamination along with confidence levels; it does not, however, provide any information about exactly where that contamination is spatially located.

For the purposes of contamination delineation, the indicator sample results (0,1) for each traditional sampling program were interpolated to the decision point grid using ordinary kriging, and assuming an exponential covariance function with a range equal to 50 ft. In this particular case, 50 ft was selected as representing the width of the smallest contiguous area that might be contaminated. A search neighborhood of 50 ft was also applied, with a maximum of eight sample results (the closest eight) used to support the interpolation at any particular location. The end result was an estimate at each decision point of the probability of exceeding the threshold.

For the purpose of this discussion, if decision points had a contamination probability less than 0.1, the associated cell was assumed to be clean. In the context of either hazardous waste site decision making, or environmental emergency response, if the probability of contamination is not sufficiently low for a particular location, action would have to be taken (e.g., remediation or evacuation). For this example, decision points with contamination probabilities greater than 0.1 and their associated cells were treated as dirty. Error rates were then determined by comparing the conclusions drawn from interpolating the sample results with the actual contamination footprint for the complete set of decision points. False negative numbers represent decision points with calculated contamination probabilities less than 0.1, but which, in fact, were contaminated. False positive numbers represent decision points with calculated contamination probabilities greater than 0.1, but which, in fact, were clean. Table 2 summarizes the error rates observed for each of the traditional programs. Figure 5 shows the interpolated footprint for each of the programs (i.e., the set of decision points with probability of contamination greater than 0.1), with the actual footprint overlain.

As one would expect, false positive and false negative errors in general decrease as sample numbers increase. False negative errors are reduced to zero for programs that have 42 samples or more. However, even with 203 samples, the resulting interpolated footprint still encompasses 35% more area than the actual footprint contains.

TABLE 2 Comparison of Sampling Program Performance When Determining Contamination Footprints

	Contaminated Area Estimate (ft^2)			Errors	
Program	Probability > 0.1	Best Estimate	Probability > 0.9	False Negative (ft^2)	False Positive (ft^2)
Reality	Not applicable	10,000	Not applicable	Not applicable	Not applicable
Gridded 12 samples	8,100	6,100	4,800	5,600	3,700
Gridded 20 samples	17,700	8,200	6,800	2,300	10,000
Gridded 42 samples	19,100	10,700	4,400	0	9,100
Gridded 105 samples	16,400	10,100	5,200	0	6,400
Gridded 203 samples	13,500	9,900	6,600	0	3,500

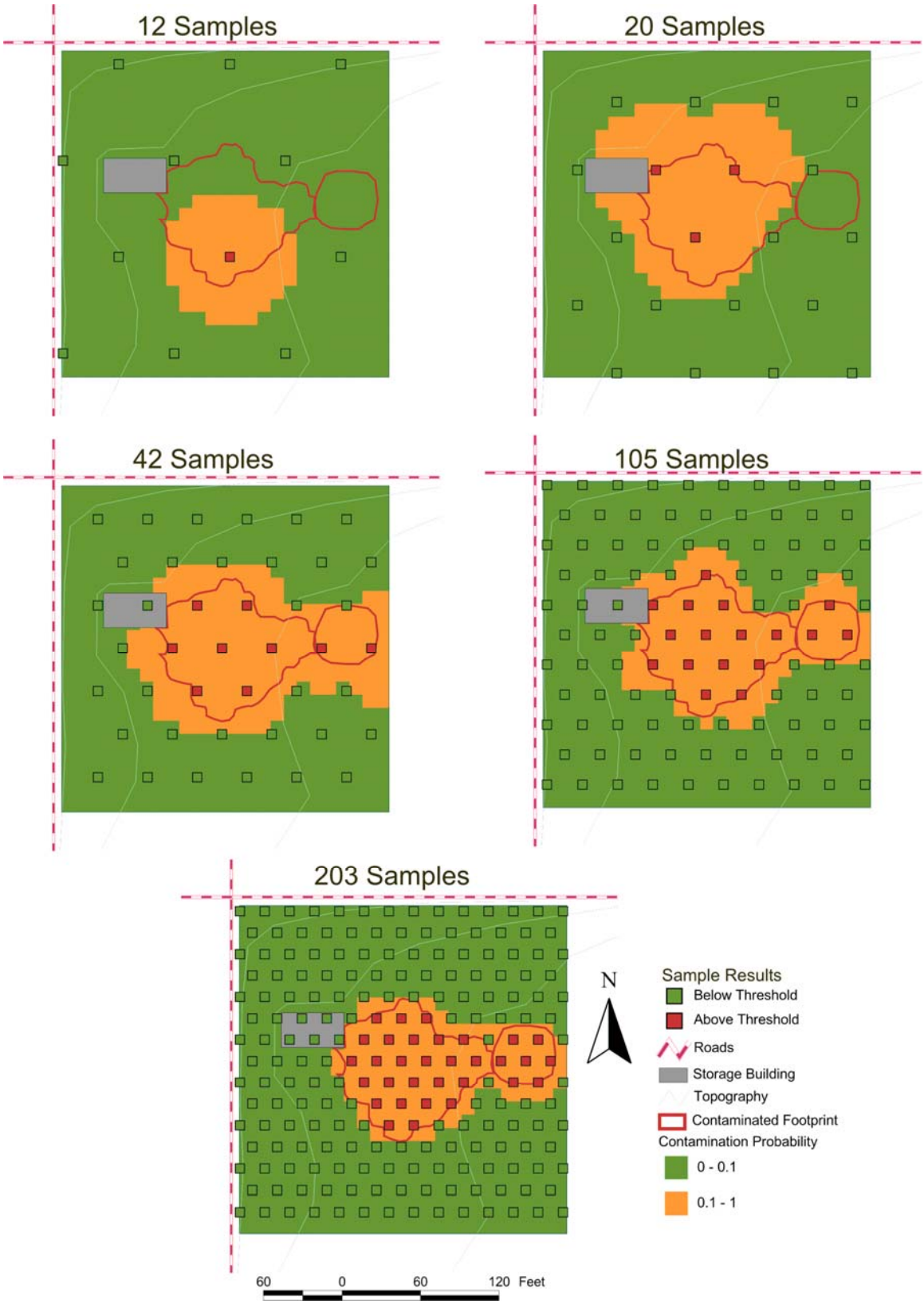

FIGURE 5 Contamination Footprints Based on Gridded Programs

5 ADAPTIVE PROGRAM PERFORMANCE

The BAASS methodology can improve sampling program performance in two ways. The first is by bringing soft information into the analysis in a manner that influences both the estimate of contamination area or volume and the design of the sampling program to improve that estimate. The second is by adapting or adjusting the sampling program in response to real-time measurement results. This section reviews the performance of BAASS using this example for a number of different scenarios and evaluates the sensitivity of BAASS's performance to soft information, spatial autocorrelation assumptions, required decision-making confidence levels, and "inside-out" versus "outside-in" sampling strategies.

For the purposes of this analysis, the same 625-node grid of decision points described previously is used. The presumed spatial autocorrelation is captured by an exponential covariance function with range equal to 50 ft. A search neighborhood is applied to the indicator interpolation process with radius equal to the assumed spatial autocorrelation range and a maximum number of sampled locations contributing to any particular interpolation limited to the nearest eight. Capping the number of sampling locations contributing to the interpolation enhances computational performance. Implementing a search neighborhood both improves computational performance and addresses the fact that contamination distributions across a site are nonstationary.

Contamination status at decision points is determined by the probability at each point that contamination exceeds the requirement. Decision points with a contamination probability less than 0.1 are presumed clean. All other decision points are treated as contaminated. The primary goal of the majority of the adaptive programs that will be discussed is the same: to maximize the number of decision points that can be shown to be clean (i.e., have a contamination probability < 0.1). As will become clear from the examples, this goal translates into delineating contamination footprints, working from the outside inwards (i.e., from areas known to be clean toward areas known to be contaminated).

In each case (except where noted), the implementation of the adaptive program is the same: a location is selected that maximizes the clean area expected, a measurement is taken, the beta pdfs for every decision point are updated on the basis of the sampling result, the next location is selected, and the process is repeated until the goals of the sampling program have been achieved.

5.1 NONINFORMATIVE PRIOR

Soft information regarding the contamination status of an area is captured by the prior beta pdfs that are applied to each of the decision points. These pdfs reflect the initial belief about contamination distribution for the site (i.e., the likelihood at any particular location that contamination is present at levels exceeding guidelines). In the unlikely absence of prior information, a noninformative prior can be used. A noninformative prior assigns the likelihood of contamination being present to 0.5 for each grid node by using small (<1) values for the beta

distribution parameters, alpha and beta. For the beta distribution, setting alpha equal to beta results in an expected probability of 0.5 (alpha and beta must be nonnegative). Small alpha and beta values reflect the fact that there is little information contributing to this expected probability. Small alpha and beta parameter values ensure that as samples are collected and beta pdfs updated, that the value of information provided by sample results will be greater than that associated with the initial noninformative prior. For the purposes of this example, alpha and beta are set to 0.3 for each decision point.

By definition of the noninformative prior, the whole site is presumed to be contaminated at the start (i.e., the expected contamination probability at each decision point is 0.5, which is greater than the 0.1 requirement for any particular decision point to be classified as clean). This means that if the site were remediated on the basis of the noninformative prior alone, all of the contaminated area would have been remediated. The resulting remediated surface area would have been more than six times larger than necessary to address the actual contamination footprint.

Figure 6 shows the progression of the adaptive sampling program for this scenario at several key steps. Samples results are color-coded as either green (results below threshold) or red (results above threshold). In each map, green areas capture decision points with expected probability less than 0.1. Orange areas have a probability of contamination greater than 0.1. The details resulting from this program are contained in Table 3. Table 3 shows the error rates observed at each step of the sampling program. As a point of comparison, the error rates for the traditional programs are included as well.

As Table 3 illustrates, even with a noninformative prior, an adaptive sampling program outperforms a traditional gridded program from an error rate perspective for this example. The initial performance is better for the traditional gridded program because the small number of samples is systematically distributed across the area, whereas the BAASS sampling program evolves from one end of the site working toward the other, with sample spacing initially driven by the assumed spatial autocorrelation range. After approximately 20 samples, BAASS has covered the site and achieved much lower false negative rates than its gridded program counterpart, although false positive rates are still higher. By 42 samples, BAASS is outperforming the gridded program from an error rate perspective. After 110 samples, the BAASS methodology could no longer identify additional sampling locations that were both likely to yield a clean result and increase the number of decision points classified as clean. False negative rates were reduced to zero after 27 samples. After 110 samples, the area incorrectly identified as potentially contaminated was 500 ft^2, significantly lower than the 3,500-ft^2 false positive area that a gridded program with almost twice as many samples produces.

5.2 “GOOD” PRIOR

In this second scenario, additional information is brought to the design and implementation of the adaptive sampling program. In this case, we know that the contamination event involved the movement of contaminated liquid over the ground’s surface, and that the local

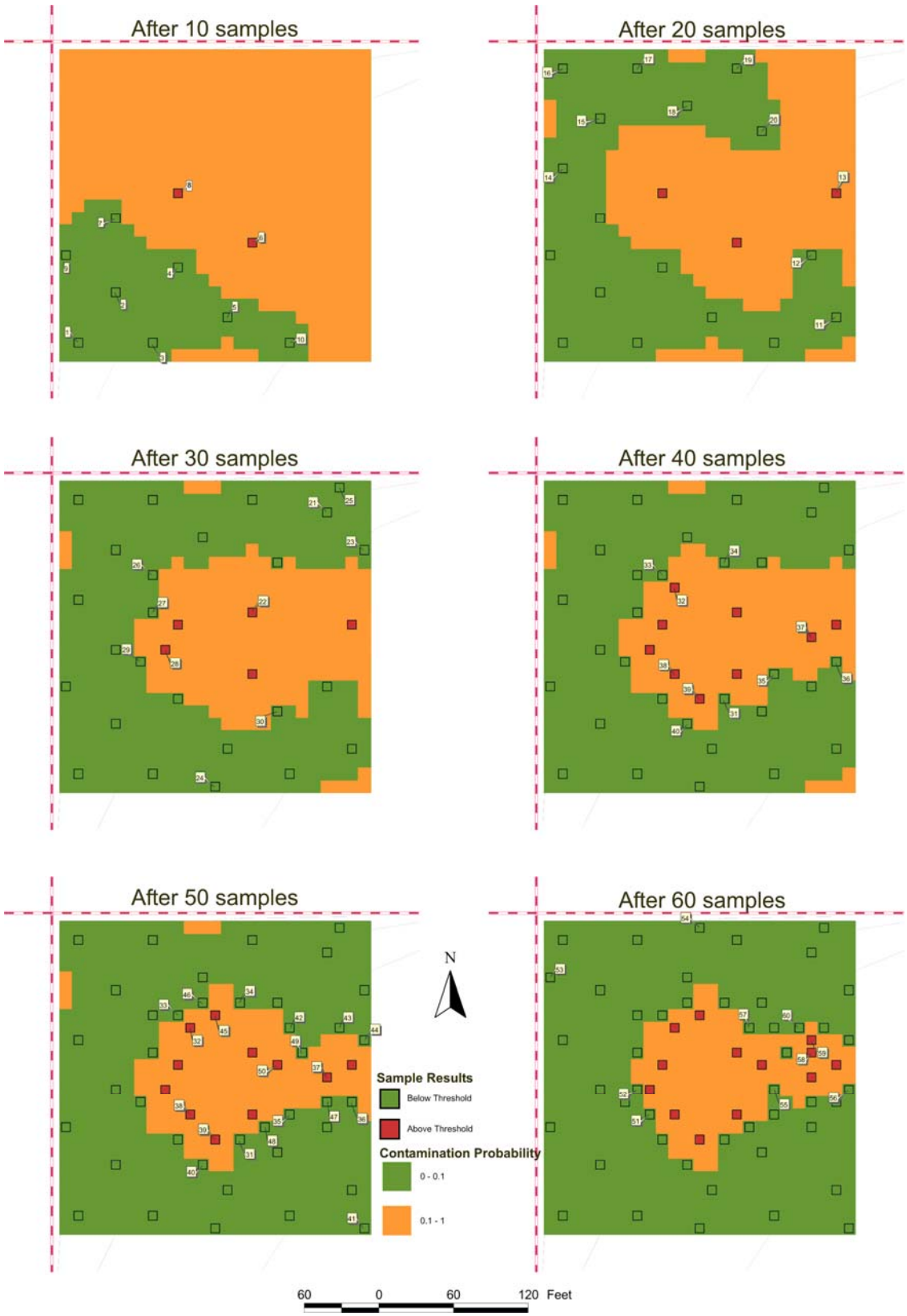

FIGURE 6 Adaptive Program Progression with Noninformative Prior

TABLE 3 Sampling Program Details

	Traditional Grid		Noninformative Prior		Good Prior		Bad Prior		Correct Range		Lower Confidence Level (40%)		Inside-Out	
Sample	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)
1			0	51,600	1,100	5,800	4,600	14,900	0	26,000	4,800	1,400	300	12,500
2			0	49,500	300	7,000	5,200	8,200	0	25,400	2,800	2,200	300	12,500
3			0	47,400	1,300	3,900	5,200	5,300	100	21,300	3,000	700	600	7,800
4			100	45,400	2,600	3,400	5,200	3,200	0	30,400	1,100	2,600	200	9,600
5			200	42,900	2,600	2,200	2,900	5,100	100	25,000	700	6,400	200	8,800
6			100	43,100	2,600	1,800	2,900	4,600	200	21,700	700	4,800	200	8,700
7			100	40,800	2,500	2,900	2,900	3,700	200	18,800	900	3,600	500	8,500
8			100	41,400	2,500	2,000	2,200	4,400	200	11,300	1,000	2,700	300	10,200
9			100	39,300	800	3,500	1,800	6,300	200	8,200	1,000	1,900	300	9,800
10			100	37,300	800	5,100	1,800	5,400	100	12,500	1,200	900	300	9,300
11			100	35,300	800	3,600	1,800	5,000	100	8,100	1,700	600	300	9,700
12	5,600	3,700	100	33,200	0	5,200	1,800	4,400	100	5,800	1,000	900	300	8,300
13			100	33,400	0	4,500	1,800	4,000	100	4,200	900	1,400	400	7,200
14			100	31,400	0	6,000	1,800	4,400	100	6,300	400	2,600	0	8,200
15			100	29,300	0	4,900	1,800	4,000	100	4,100	600	1,500	0	9,500
16			100	27,200	0	7,500	1,800	3,600	0	5,800	500	2,300	0	8,800
17			100	25,100	0	6,600	1,000	5,700	0	3,900	500	1,500	0	10,300
18			100	23,000	0	5,600	1,000	4,900	0	4,300	500	1,100	0	11,600
19			100	20,900	0	4,600	1,000	4,000	0	3,800	500	1,300	0	12,400
20	2,300	10,000	100	18,800	0	4,100	0	4,900	0	3,200	500	900	0	11,700
21			100	16,600	0	3,700	0	7,800	0	3,600	500	500	0	12,200
22			100	17,400	0	2,800	0	7,300	0	3,000	200	800	0	10,900
23			100	15,900	0	2,400	0	6,500	0	3,500	200	600	0	11,000
24			100	15,000	0	1,900	0	5,500	0	3,000	300	500	0	8,700
25			100	14,100	0	1,500	0	4,600	0	2,600	200	700	0	7,600

TABLE 3 (Cont.)

	Traditional Grid		Noninformative Prior		Good Prior		Bad Prior		Correct Range		Lower Confidence Level (40%)		Inside-Out	
Sample	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)
26			100	13,300	0	1,200	0	6,200	0	2,300	200	700	0	8,200
27			100	12,400	0	1,400	0	5,200	0	2,000	200	600	0	8,000
28			0	13,900	0	1,100	0	4,600	0	1,600	200	500	0	8,100
29			0	12,600	0	900	0	4,700	0	1,600	300	500	0	8,800
30			0	11,800	0	1,600	0	4,300	0	1,400	300	500	0	8,100
31			0	10,700	0	900	0	3,900	0	1,800	300	500	0	8,400
32			0	12,500	0	1,100	0	3,600	0	1,400	300	500	0	8,400
33			0	11,300	0	900	0	3,300	0	1,200	300	400	0	8,400
34			0	10,400	0	1,000	0	3,000	0	1,200	300	300	0	9,100
35			0	9,600	0	800	0	2,800	0	1,400	200	300	0	8,400
36			0	8,900	0	800	0	2,500	0	1,200	200	200	0	7,400
37			0	9,200	0	600	0	2,500	0	1,000	100	200	0	7,400
38			0	10,000	0	400	0	2,300	0	1,100	100	200	0	6,600
39			0	11,600	0	200	0	2,100	0	3,400	200	200	0	6,100
40			0	9,900	0	300	0	1,900	0	1,100	200	100	0	6,000
41			0	9,400	0	500	0	1,700	0	1,100			0	6,000
42	0	9,100	0	8,900	0	300	0	1,800	0	1,100			0	6,400
43			0	8,400	0	200	0	1,600	0	1,200			0	5,100
44			0	7,900	0	300	0	1,400	0	1,100			0	5,100
45			0	8,300	0	200	0	1,200	0	1,000			0	5,100
46			0	7,700	0	400	0	1,400	0	800			0	5,100
47			0	7,200	0	300	0	1,100	0	700			0	5,400
48			0	6,700	0	700	0	1,200	0	800			0	5,200
49			0	6,300	0	300	0	1,100	0	700			0	4,300
50			100	6,300	0	500	0	1,000	0	600			0	4,700

TABLE 3 (Cont.)

	Traditional Grid		Noninformative Prior		Good Prior		Bad Prior		Correct Range		Lower Confidence Level (40%)		Inside-Out	
Sample	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)
51			100	6,000	0	300	0	1,000	0	500			0	5,200
52			100	5,400	0	400	0	1,000	0	400			0	5,300
53			100	5,100	0	300	0	900	0	300			0	4,300
54			100	4,800	0	500	0	1,000	0	200			0	4,100
55			100	4,500	0	300	0	900	0	200			0	4,100
56			100	4,200	0	700	0	900					0	3,700
57			0	3,800	0	300	0	1,100					0	3,700
58			0	4,100	0	200	0	900					0	3,900
59			0	5,100	0	100	0	800					0	4,000
60			0	4,000			0	700					0	4,500
61			0	4,000			0	600					0	4,700
62			0	3,700			0	500					0	4,100
63			0	3,400			0	400					0	4,100
64			0	3,100			0	300					0	4,100
65			0	3,600			0	300					0	4,100
66			0	3,200										
67			0	3,000										
68			0	2,800										
69			0	3,000										
70			0	2,800										
71			0	2,600										
72			0	2,500										
73			0	2,300										
74			0	2,100										
75			0	2,100										

TABLE 3 (Cont.)

	Traditional Grid		Noninformative Prior		Good Prior		Bad Prior		Correct Range		Lower Confidence Level (40%)		Inside-Out	
Sample	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)
76			0	2,100										
77			0	2,000										
78			0	2,100										
79			0	2,000										
80			0	2,100										
81			0	2,100										
82			0	2,000										
83			0	2,200										
84			0	2,000										
85			0	2,200										
86			0	2,300										
87			0	2,100										
88			0	2,200										
89			0	2,100										
90			0	2,100										
91			0	2,000										
92			0	1,900										
93			0	2,000										
94			0	1,900										
95			0	1,800										
96			0	1,700										
97			0	1,600										
98			0	1,500										
99			0	1,400										
100			0	1,300										

TABLE 3 (Cont.)

	Traditional Grid		Noninformative Prior		Good Prior		Bad Prior		Correct Range		Lower Confidence Level (40%)		Inside-Out	
Sample	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)	False Neg. (ft^2)	False Pos. (ft^2)
101			0	1,200										
102			0	1,100										
103			0	1,100										
104			0	1,000										
105	0	6,400	0	900										
106			0	800										
107			0	700										
108			0	600										
109			0	600										
110			0	500										
203	0	3,500	0											

topography would have controlled which way the contaminated liquid traveled. By using the topographical information for the site, a "prior" or initial CSM was constructed, as shown in Figure 7. This prior was constructed by setting beta distribution parameters at each decision point to reflect the initial information available. As simple examples, we know with certainty that the lagoon footprint is contaminated. Decision points falling within the lagoon footprint have their alpha and beta values set accordingly (i.e., beta set to zero, and alpha to a large number). Conversely, we know with certainty that the building footprint downgradient from the lagoon is clean. Decision points falling within the building footprint have beta set to a large number, and alpha set to zero. For all other areas of the site, we select alpha and beta to reflect our belief about contamination probabilities, with both parameters kept around one or less, to reflect the fact that these beliefs are not at all certain.

The resulting prior, color-coded in Figure 7, might be referred to as a "good" prior. It captures the majority of the actual contamination, while carving away portions of the site that are

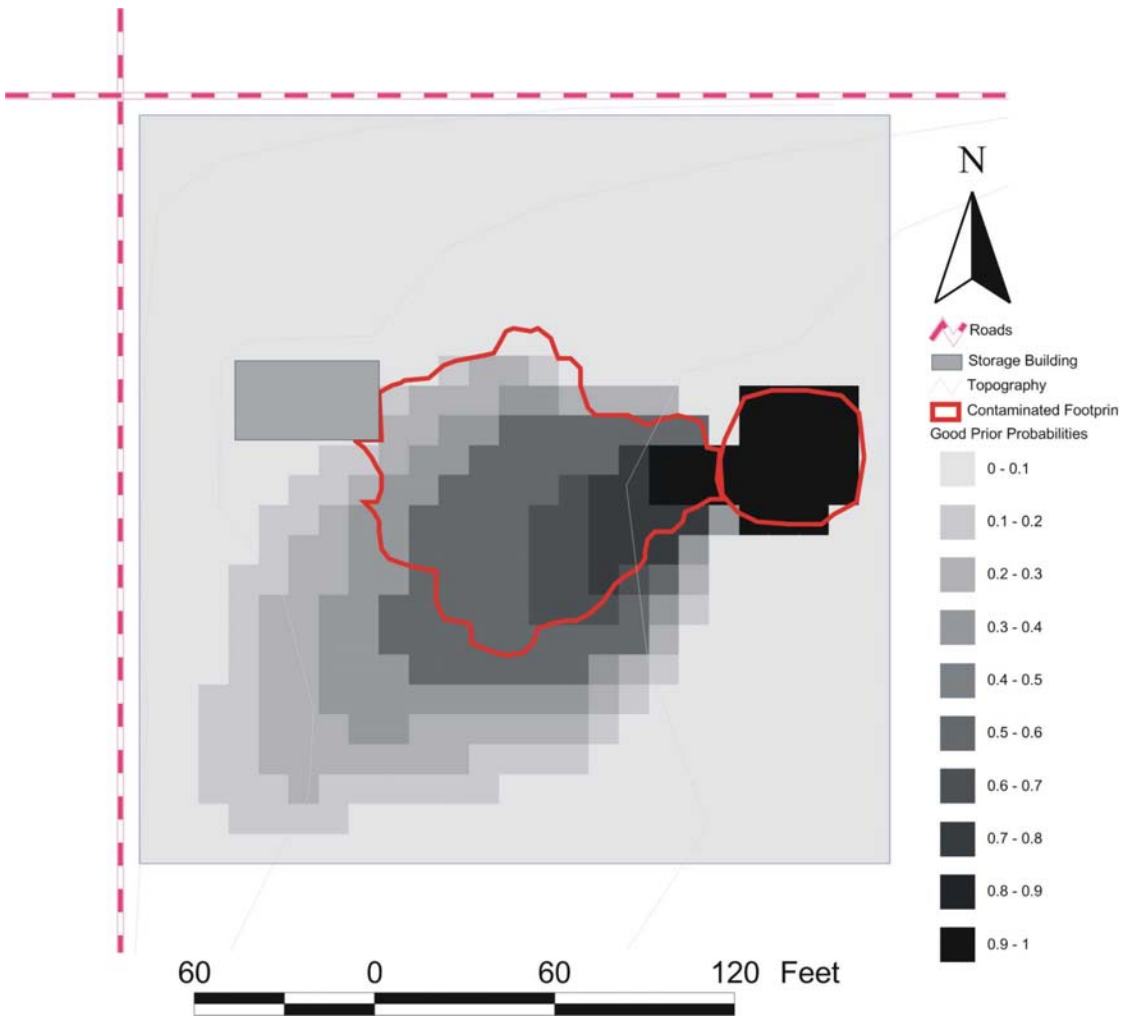

FIGURE 7 Initial Conceptual Site Model Using Prior Information

unlikely to be contaminated. Still, this prior is far from perfect. An error analysis of the decisions that would be made on the basis of this prior alone reveals that 4% of the contaminated area would have been missed, and we still would have remediated more than twice what was necessary.

Figure 8 shows the progression of the adaptive sampling program for this scenario at several key steps. The details resulting from this program are contained in Table 3. Table 3 includes the error rates observed at each step of the sampling program. With a "good" prior, BAASS consistently outperformed the 203-sample gridded program after only 22 samples. Compared with the noninformative prior scenario, the availability of a "good" prior allowed the adaptive sampling program to hone in on the actual contaminant boundary much faster than in the noninformative case. After 60 samples, the BAASS methodology can no longer identify additional sampling locations that are both likely to yield a clean result and increase the number of decision points classified as clean. At this stage of the adaptive sampling program, both false negative and false positive errors have been reduced to zero.

5.3 "BAD" PRIOR

There may be cases where initial, soft information is misleading (i.e., an incorrect initial CSM). The question in this instance is how badly an adaptive sampling program might perform given faulty information at the outset. Figure 9 shows a "bad" prior scenario. This prior misses almost 70% of the actual contamination footprint. In this particular case, if BAASS is allowed to pick sampling locations on the basis of this prior with the goal of maximizing clean areas, the prior is so misleading that BAASS will fail to identify the presence of contamination outside the footprint of the lagoon. However, if the adaptive sampling program is "seeded" with a sample that encounters contamination (alerting BAASS to the presence of contamination outside the lagoon), BAASS will proceed to delineate the contamination.

Figure 10 shows the progression of the adaptive sampling program for a "seeded" scenario at several key steps. In this case, the seeded location was selected to ensure that contamination would be encountered with the first sample. The details of the resulting adaptive sampling program are contained in Table 3. Compared with the "good" prior, the misleading prior in this case results in a sampling program that is not as efficient when delineating the contamination boundary. It takes almost twice as many samples to drive false negative errors to zero than for the "good" prior scenario. After 64 samples, the BAASS methodology can no longer identify additional sampling locations that are both likely to yield a clean result and increase the number of decision points classified as clean. At this stage of the adaptive program, there still remains 300 ft^2 of false positive area. However, the resulting sampling program is still significantly more efficient than a traditional gridded program.

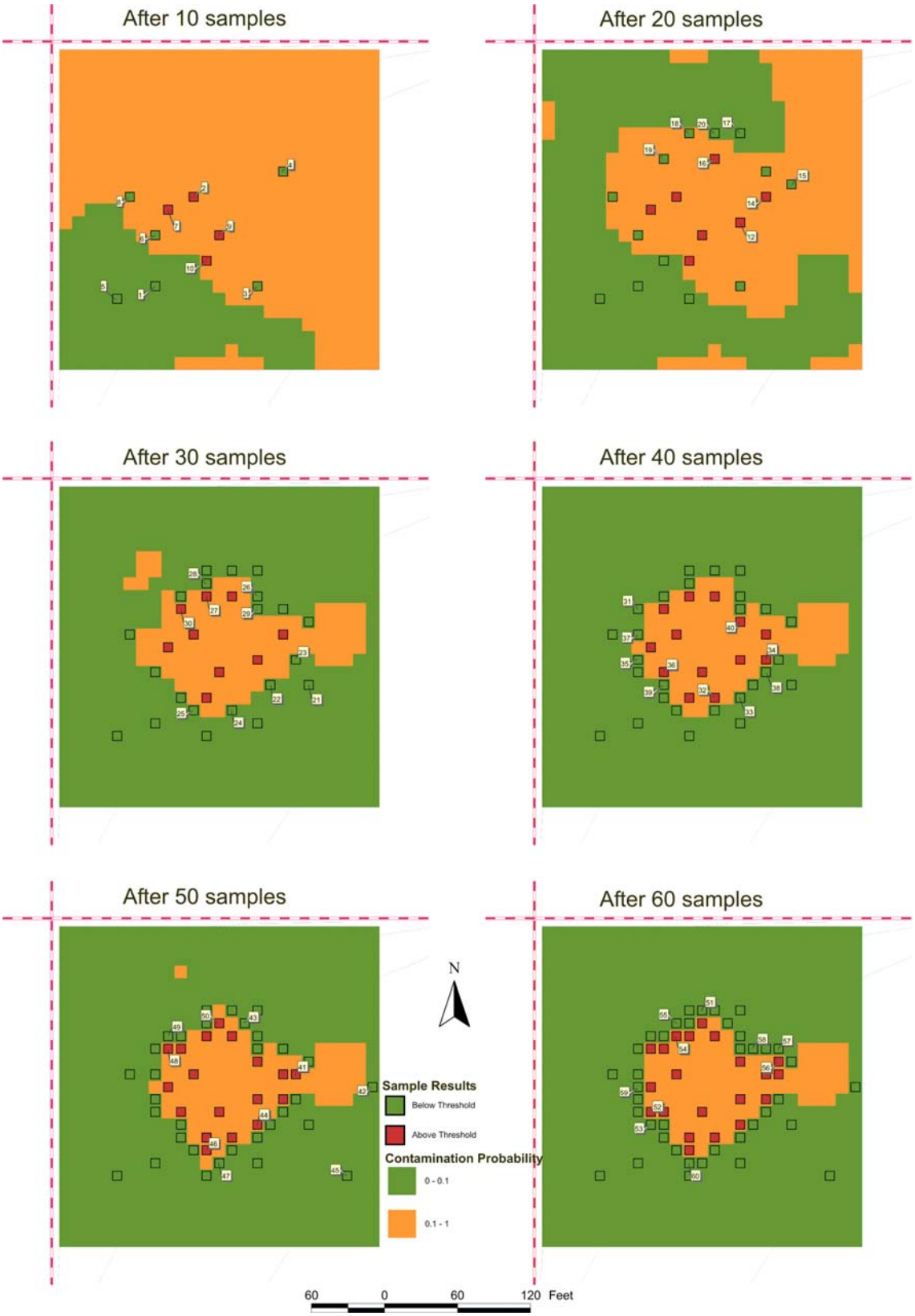

FIGURE 8 Adaptive Sampling Progression Based on "Good" CSM

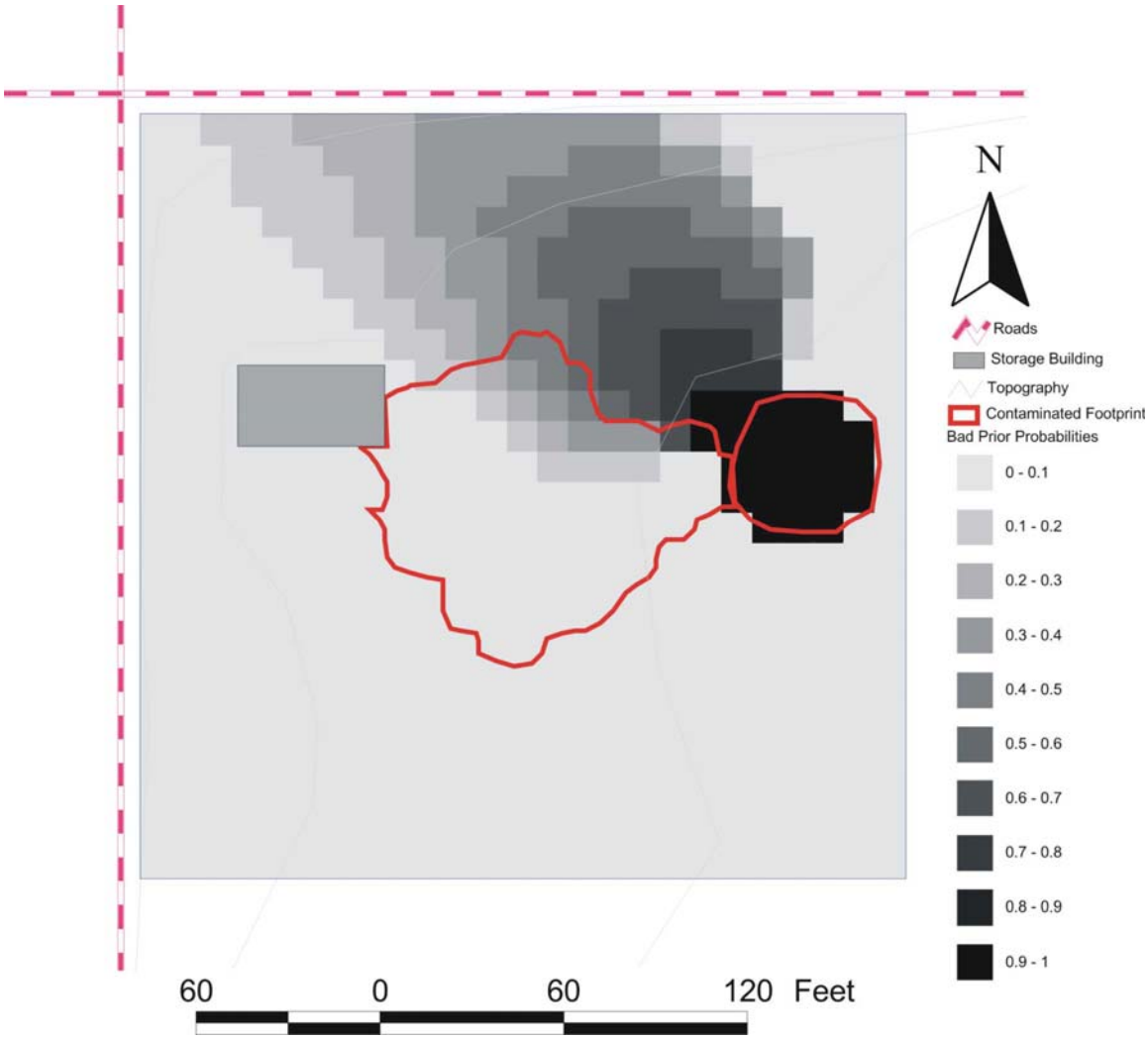

FIGURE 9 Initial Conceptual Site Model Based on Misleading Information

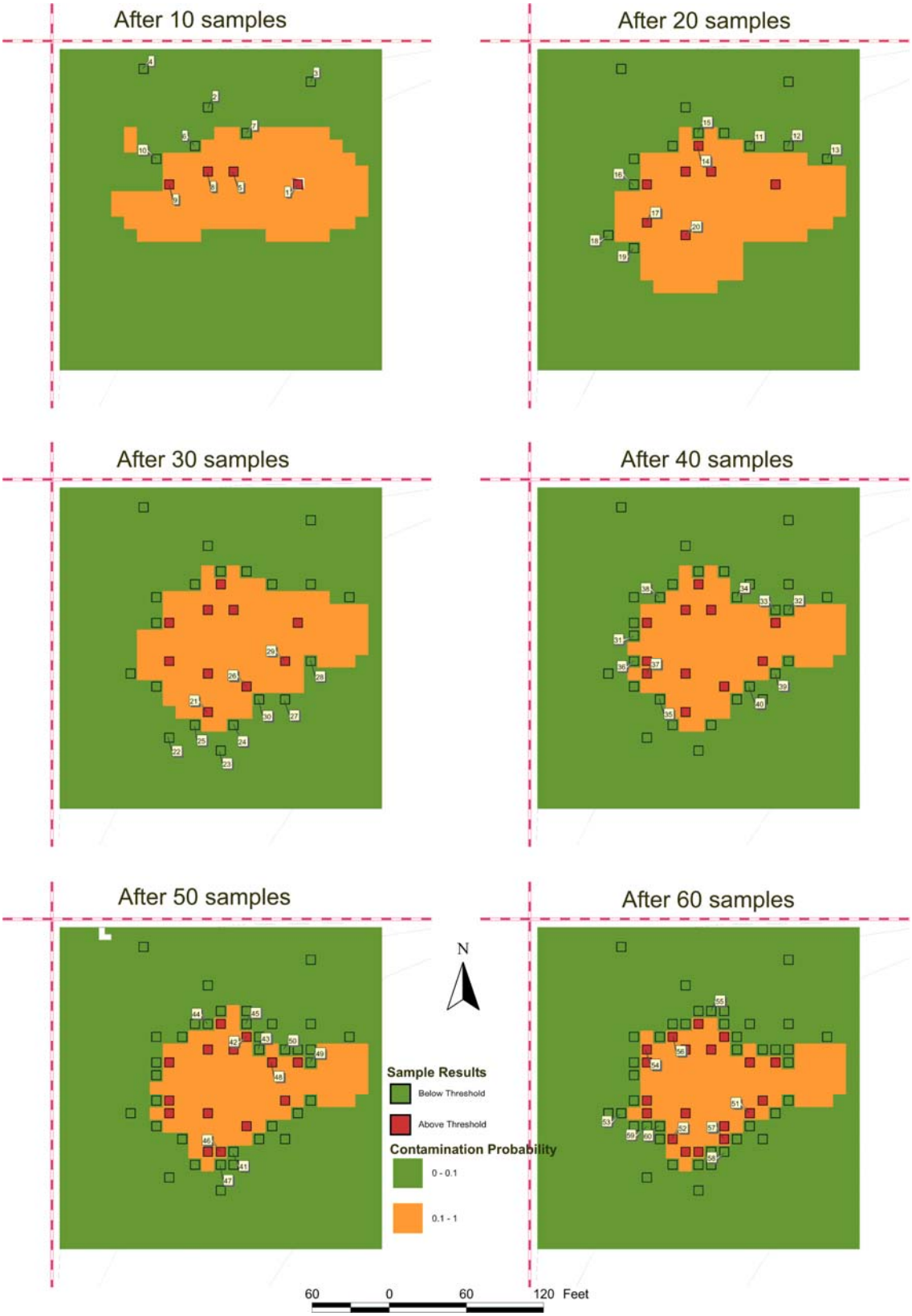

FIGURE 10 Adaptive Sampling Progression Based on "Bad" CSM

5.4 SPATIAL AUTOCORRELATION RANGE ASSUMPTIONS

BAASS requires an assumption about the range of the spatial autocorrelation presumed present. For sites where no samples exist, the spatial autocorrelation range assumption represents an educated guess. An assumed spatial autocorrelation range that is smaller than reality means that the resulting sampling density will likely be greater than actually necessary. An assumed spatial autocorrelation range that is larger than reality means that one will be estimating probabilities of exceeding guidelines that are far too certain (e.g., close to 1 or 0) than reality would warrant. For the simple example presented, the actual indicator autocorrelation range is 110 ft, or a little more than twice the range used for the scenarios presented thus far.

Figure 11 shows the progression of the adaptive sampling program if one uses the "correct" spatial autocorrelation range and the "good" prior described in Section 5.2. The details of this adaptive sampling program are contained in Table 3. Using the correct autocorrelation range reduces the number of samples required to completely delineate the contamination footprint by about 10% in this case. Setting the spatial autocorrelation range to 150% of its true value produces a sampling program that, with only one "clean" sample located close to the actual contamination, concludes that there is no contamination present. The repercussions of overestimating the spatial autocorrelation range are significant — potentially missing contamination completely.

The spatial autocorrelation range assigned to any particular problem to a large degree determines initial sample spacing for BAASS applications. The implications are that when the presence or absence of contamination within a large area (i.e., an area whose scale is much greater than any presumed spatial autocorrelation range) is unknown, the number of samples BAASS will allocate to that area in search of contamination will be determined by the assumed spatial autocorrelation range. A rule of thumb to use in this case is to set the autocorrelation range to the width of a contaminated area that would be important to find, if it existed.

5.5 MODIFYING ACCEPTABLE UNCERTAINTY LEVELS

In the scenarios presented thus far, sample locations were selected on the basis of their expected ability to increase the number of decision points deemed clean. The definition of cleanup to this point has been an expected probability less than 0.1. This section presents the impact from changing acceptable uncertainty levels on the evolution of the adaptive sampling program. For this scenario, the clean cutoff is raised to 0.4, that is, decision points that have an expected probability of contamination less than 0.4 are presumed clean. The goal of the sampling program is to maximize the number of decision points that satisfy that criterion. The "good" prior CSM is used as the basis for initiating the program.

Figure 12 shows the progression of the adaptive sampling program using this uncertainty level. The details of this adaptive sampling program are contained in Table 3. Surprisingly, relaxing the uncertainty requirements for decision making actually appears to make the sampling approach initially more efficient for this example in delineating the contamination footprint.

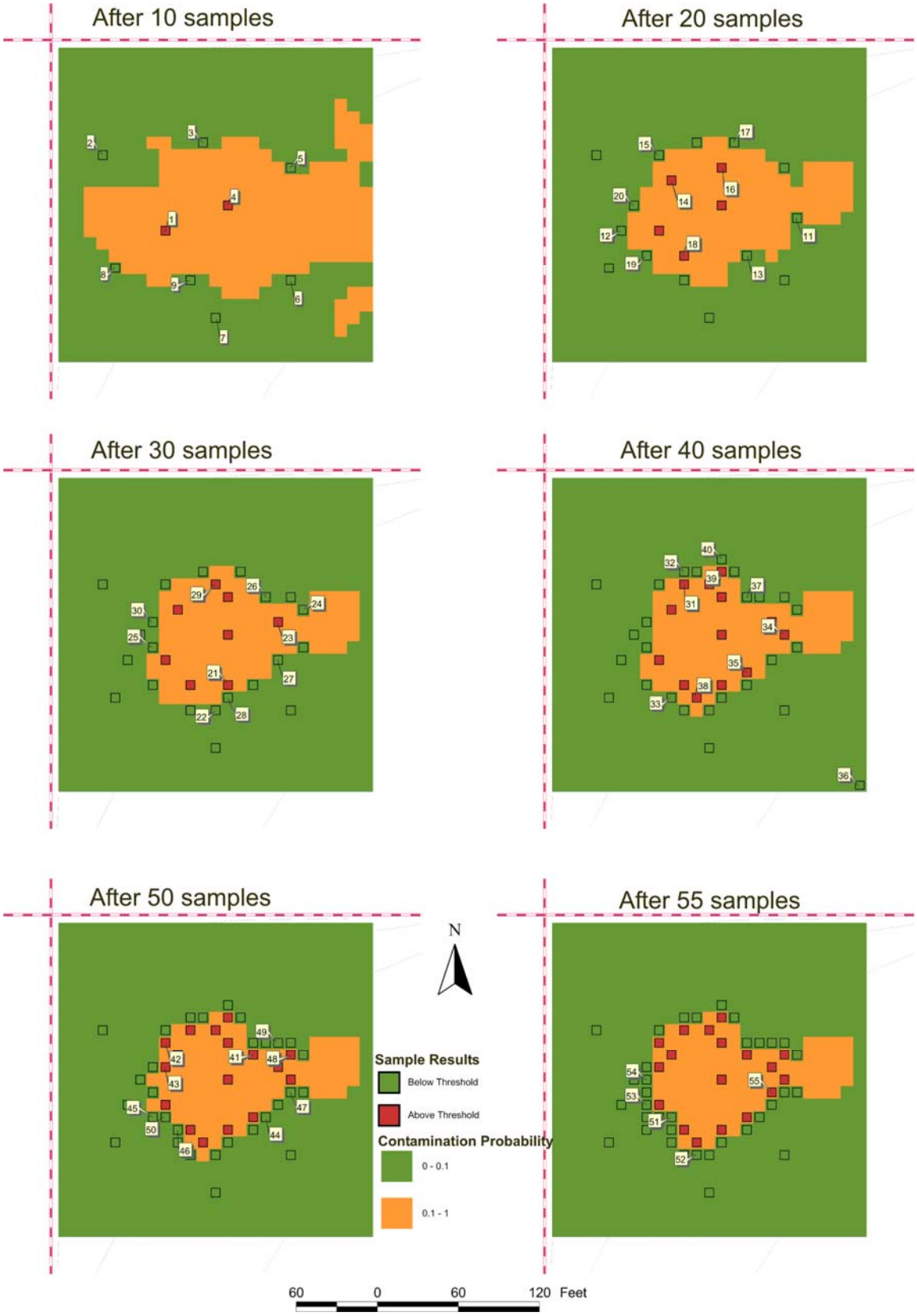

FIGURE 11 Adaptive Sampling Progression with Correct Range and "Good" CSM

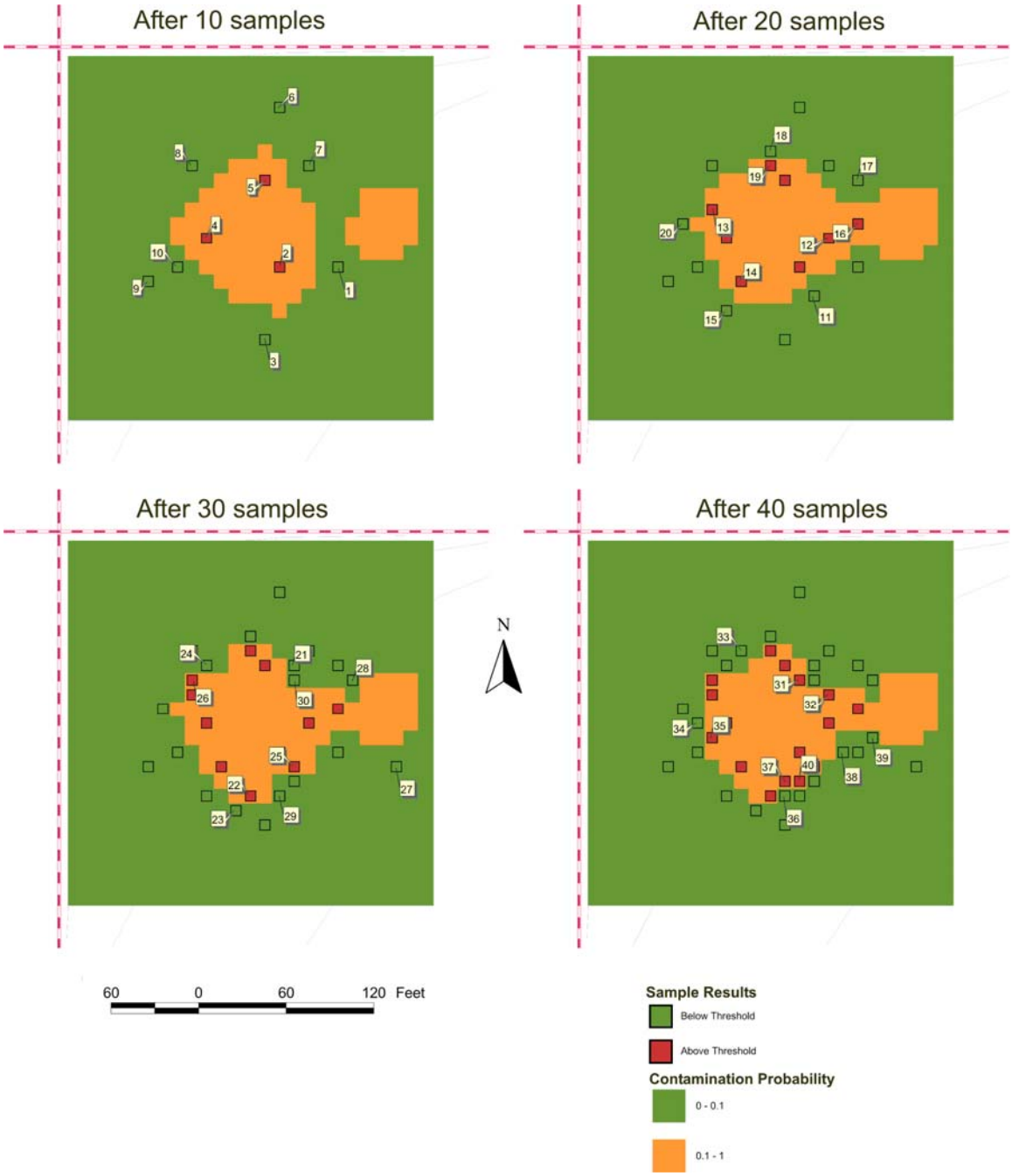

FIGURE 12 Adaptive Sampling Progression with Relaxed Confidence Level Requirements

While a lingering false negative problem remains through 40 samples, the false positive rate was significantly lower, particularly early in the program. The likely explanation is that relaxing the required confidence level allows a wider initial spacing for the sampling program. This, in turn, allows more rapid coverage of the overall site. After 40 samples, the BAASS methodology can no longer identify additional sampling locations that are both likely to yield a clean result and increase the number of decision points classified as clean using a 0.4 probability level.

One option for leveraging the apparent initial lower false positive rates that the relaxed confidence level provides is to initiate an adaptive sampling program with a relaxed level of uncertainty until good sampling coverage for the area of concern has been achieved, then tighten confidence level requirements to drive out any remaining false positive and false negative errors. Applying this logic, one could drive the first 30 samples using a 0.4 confidence level. Tightening this to a 0.1 confidence level for the balance of the program would yield exactly the same total number of samples (60) as the program that uses the 0.1 confidence level throughout, but with significantly improved false positive error rate performance at the outset of the program.

5.6 INSIDE-OUT VERSUS OUTSIDE-IN STRATEGIES

All of the adaptive sampling examples presented to this point drive sample selection by maximizing the area believed to be clean at some probability level. This translates practically into an outside-in approach to boundary delineation. An alternative approach would be to maximize the area believed to be contaminated at some probability level. This would translate into an inside-out approach to boundary delineation. This section presents the impact of changing from an outside-in to an inside-out program. The objective is to maximize the number of decision points believed to be contaminated at a 90% level. The “good” prior CSM is used as the basis for initiating the program.

Figure 13 shows the progression of the adaptive sampling program. The details of this adaptive sampling program are contained in Table 3. As Table 3 indicates, for this example an inside-out approach yields consistently lower false negative rates for most of the samples collected than the outside-in approach exemplified by maximizing the error defined as clean. This is at the expense, however, of false positive errors, which are consistently and significantly higher throughout the course of the program than the outside-in alternative.

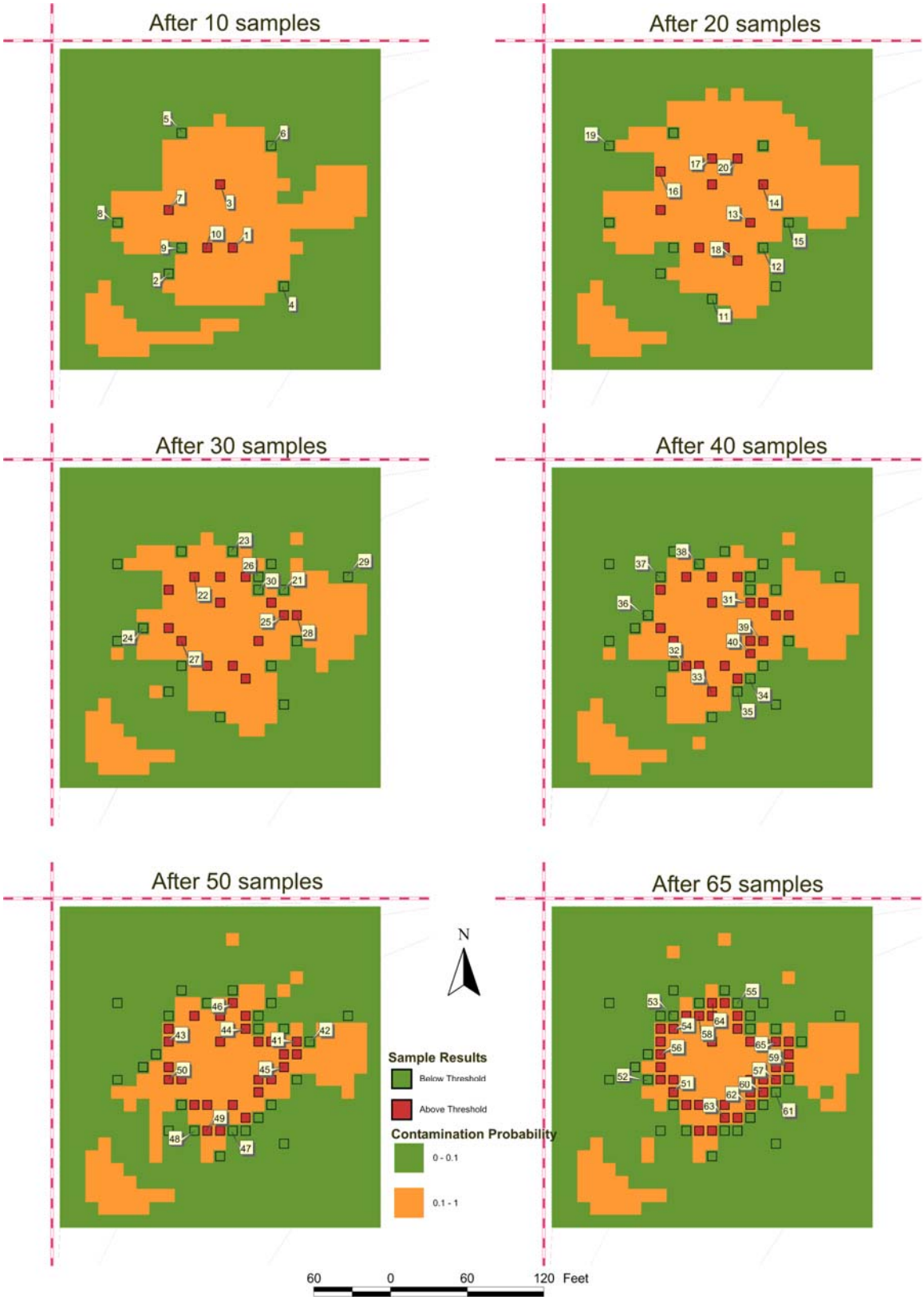

FIGURE 13 Inside-Out Approach to Contamination Delineation

6 DISCUSSION AND CONCLUSIONS

6.1 ADAPTIVE SAMPLING PROGRAM PERFORMANCE

For this simple example, in all adaptive cases considered, the adaptive sampling program converged on the actual footprint with far fewer samples than a traditional gridded program. Convergence is measured by the false positive and false negative error rates observed. In the best case, using a "good" prior and the correct spatial autocorrelation range, after only 20 samples the adaptive program achieved consistently lower false positive and false negative error rates than a gridded program with 203 samples. Even in the worst case presented, with a noninformative prior and an incorrect range, the adaptive program consistently outperformed the 203-sample gridded program after only 42 samples had been collected.

For hazardous waste site or emergency response decision making, the health risk repercussions of false negative errors (i.e., missing contamination of concern) are significantly greater than those for false positive errors. The adaptive programs presented were almost all driven by the same goal: maximize the area categorized as clean or safe at a prescribed confidence level (which is directly equivalent to establishing an acceptable false negative rate). Because the driver was the false negative rate, in general, the false negative rates observed as adaptive sampling progressed were significantly lower than the false positive rates. The one alternative considered, an inside-out approach that maximized the area known to be contaminated at the 90% confidence level, yielded slightly lower false negative rates at the expense of significantly higher and lingering false positive rates. The possibility of completely missing contamination because of a faulty CSM suggests that rather than pursuing a purely inside-out or outside-in approach, a more effective option would be to "search-and-bound," that is, to invest initial data collection in verifying that contamination is present at specific locations where it is suspected, and then to switch to a program that finds boundaries by maximizing the area verified as clean.

For all but one of the adaptive sampling program scenarios considered, the confidence level for determining that a decision point was clean was set to 90% (i.e., if the expected probability of contamination at that decision point was determined to be less than 0.1, the decision point was considered clean). If this probability was accurately estimated, and all of the clean decision points had an expected probability of contamination equal to 0.1, then one would expect that 10% of those points would in fact turn out to be contaminated. Figure 14 shows the expected and observed false negative rate for several of the adaptive sampling program scenarios as a function of sample number. Here the expected rate is simply calculated as the expected or average probability value across the set of decision points determined to be clean. Figure 14 underscores several key points:

- In general, the expected false negative rates were significantly below 10% throughout the sampling sequence for most of the programs. This is not surprising when one compares the expected false negative rates with the 10% criteria. In fact, requiring that all decision points categorized as clean have a

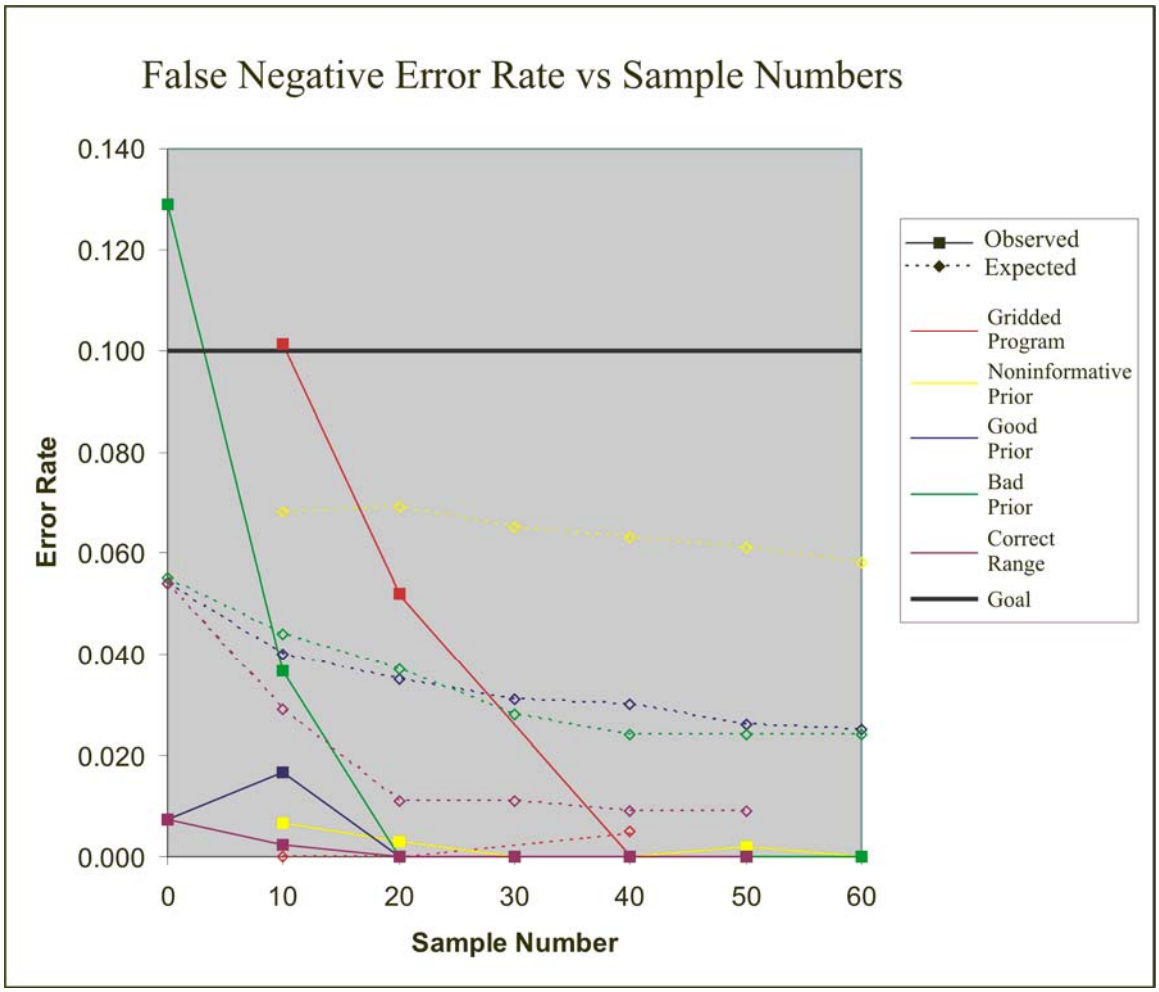

FIGURE 14 Expected and Observed Error Rates for Different Sampling Scenarios

probability of less than 0.1 means that, on average, the probability of contamination across that set is less than 0.1, as is illustrated by the expected false positive rates. Thus, if one's goal is to achieve a particular false positive rate on average across a set of decision points, the confidence level required for any particular point can be relaxed.

- The expected false negative rate is strongly linked to the assumed spatial autocorrelation range. The longer the range, the greater the influence sampling results have on the estimated probability of contamination for neighboring decision points, and the lower the expected false negative rates derived from a given set of samples. Ranges that are larger than reality result in estimated probabilities at decision points that are much more confident than the data would warrant. Conversely, ranges that are smaller than reality result in estimated probabilities whose confidence is underestimated.

- For the adaptive sampling programs, the observed false negative rates were even less than the expected false negative rates. There may be several reasons for this, but one likely explanation is the fact that the presumed range was less than the actual range. In contrast, for small sample numbers the actual false negative rate for the traditional gridded sampling program was larger than the expected false negative rate.

6.2 STOPPING CRITERIA

Adaptive sampling programs can be open-ended. In many cases, the number of samples that can be collected is resource constrained, that is, there is a fixed budget allocated to data collection, or a limited set of data acquisition assets have a limited window of time to capture data. In resource constrained settings, the question is not when to stop sampling, but how to make sure sampling resources are used most efficiently.

However, when resources are not constrained, the question is how to determine when sufficient samples have been collected. Figure 15 shows how the categorization of surface soils changes as an adaptive sampling program progresses. In Figure 15, soils have been broken into three categories: (1) those deemed clean (expected probability of contamination less than 0.1), (2) those deemed contaminated (expected probability of contamination greater than 0.9), and (3) those with expected probabilities falling between these two levels. As Figure 15 illustrates, the "gain" (as measured by increase in area deemed clean) per sample in general decreases as more samples are collected. The variability observed in Figure 15 is the result of individual

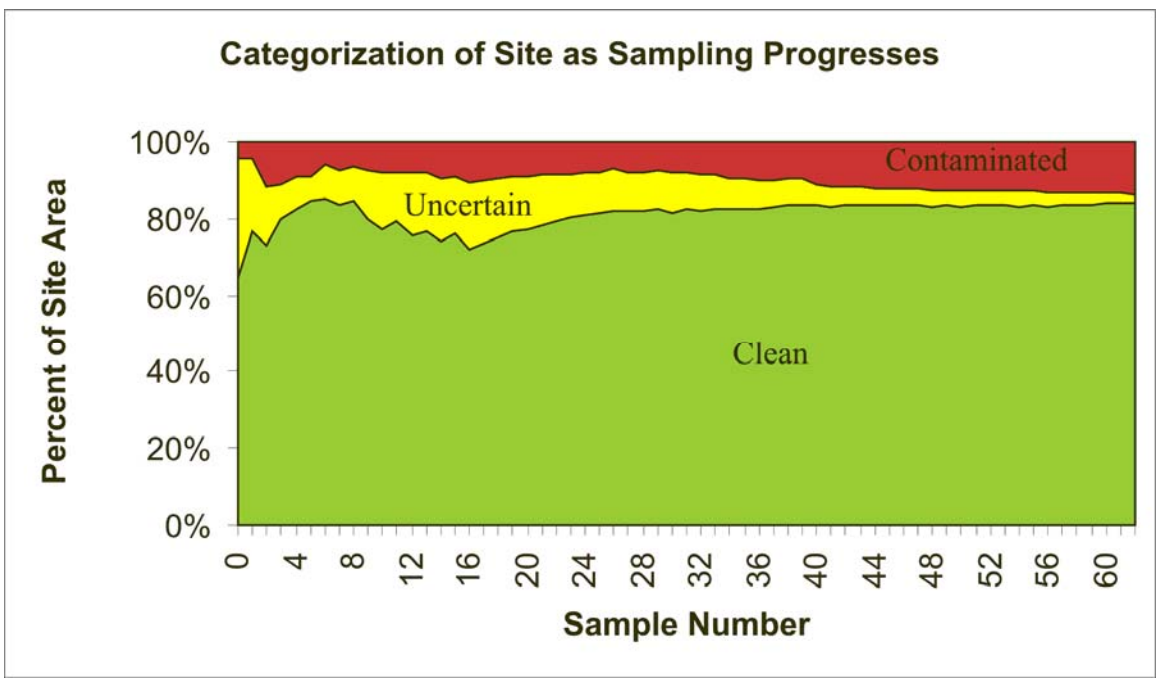

FIGURE 15 Site Categorization as a Function of Sample Numbers

samples unexpectedly encountering contamination (i.e., instead of reaping gains in the clean area as expected, the area considered contaminated grows and the clean area shrinks). Figure 16 presents the same conclusion a bit differently. Figure 16 plots the average gain in clean area per sample, where the average represents the average gain of that particular sample and the two previous samples. Average gain is used here to smooth out some of the variability observed in Figure 15. For remediation programs, sampling stopping criteria become a financial cost-benefit calculation. When the marginal cost of sampling exceeds the average remediation cost avoided by reclaiming an area as clean, then sampling should stop. In an emergency response situation, appropriate stopping criteria are not as obvious and would likely involve significant political input.

It is important to note that there is an ultimate sampling end point for the types of adaptive programs presented in these scenarios. This occurs when BAASS can no longer identify sampling locations that it expects will add to the "clean soils" category. However, this end point may represent a larger number of samples than would be justified on the basis of a stopping criterion.

6.3 SPATIAL AUTOCORRELATION RANGE ASSUMPTIONS

Geostatistics provides tools for determining the range of spatial autocorrelation present for a contamination event on the basis of sample data. However, an accurate assessment of the true spatial autocorrelation range present for a given situation requires a substantial data set. For many hazardous waste sites, and certainly in an emergency response situation, these data sets are

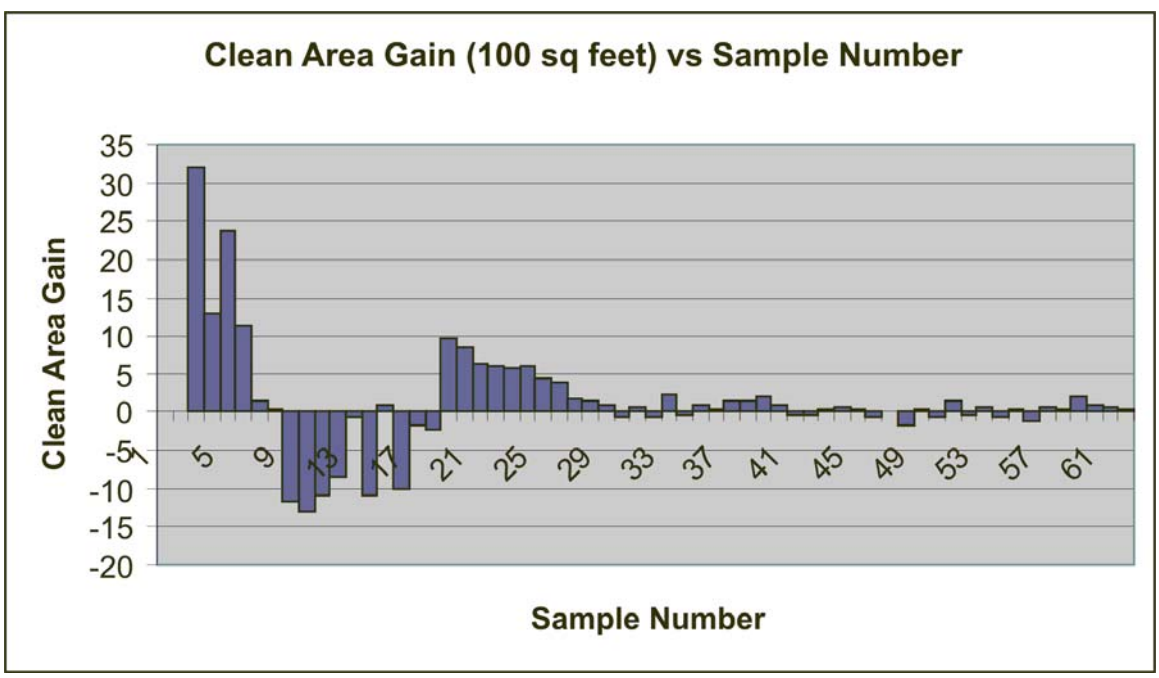

Figure 16 Clean Area Gain as a Function of Sample Numbers

not available before additional data collection takes place. The question then is what spatial autocorrelation range value should be used for BAASS. As already discussed, spatial autocorrelation ranges that are too small will lead to overly dense sampling program designs. Spatial autocorrelation ranges that are too large lead to overly confident decision making and the possibility of missing contamination that is present. The consequence of the first error is wasted resources (time and money). The consequence of the second error is the potential for unacceptable health risks.

Identifying an appropriate spatial autocorrelation range upon which to base BAASS analysis can be approached in several ways. The first is to consider whatever soft information is immediately available (e.g., fate and transport modeling results) that might suggest the size of a contamination footprint one would expect. The average width of this expected footprint could be a starting point. A second is to identify the size of a contamination footprint that would be critical to identify, if it did indeed exist. Again, the average width of this critical size could function as a first estimate at a spatial autocorrelation range. When using BAASS, the range can be adjusted and refined as data become available that provide support for a variogram calculation of the range value.

6.4 CERTAINTY LEVEL SPECIFICATIONS

The scenarios presented suggest that the certainty level specification is perhaps not as critical a factor in sampling program success as one might expect. In fact, from the perspective of quickly determining the approximate footprint of contamination, the scenario with the relaxed confidence level appears to have been more efficient than the scenario with a tightened confidence level. In both cases, the final error rates observed when the sampling programs were complete were significantly less than the confidence levels required. The primary risk of less restrictive confidence levels is overlooking contamination altogether. However, one hybrid approach would be to initiate an adaptive sampling program with a loose confidence level, and then subsequently tighten the level once all sampling under the loose confidence level was complete. Implementing this for the example problem suggests that initial false positive errors can be significantly reduced, while still achieving the same overall sampling efficiency.

6.5 INITIAL CONCEPTUAL SITE MODEL

The CSM represents one’s initial understanding of the likely distribution of contamination. It is captured in the prior beta distributions defined for each decision point. The scenarios indicate that a “good” CSM can lead to significant sample number reductions as compared with having no prior information, but that an adaptive sampling program can perform well even under a “bad” initial CSM. The primary danger with a “bad” CSM is the risk of missing contamination altogether because areas are identified as not requiring any sampling at all, when in fact contamination is present. To at least partially offset the potential negative effects of a “bad” CSM, an adaptive sampling approach that begins as an inside-out approach to first identify locations where contamination is present and that then switches to an outside-in

approach to bound the encountered contamination may make the most sense. This form of program could be termed a search-and-bound approach.

6.6 APPROACH LIMITATIONS

The implementation of BAASS and associated adaptive sampling programs, as described here, is specifically applicable to the identification and delineation of contamination when there are clearly defined thresholds that distinguish between contamination levels that are of concern, and those that are not. The efficiencies observed are predicated on the presence of spatial autocorrelation. If spatial autocorrelation is not present at a site, and/or if the range of spatial autocorrelation is less than the separation distance between decision points, then the value of BAASS is lost. Also, BAASS assumes that sampling produces results that are representative of the contamination status of the location where the sampling occurred. The implementation of BAASS, as described here, presumes that both measurement error and short-scale heterogeneity are not a significant concern for indicator (yes/no) decision making. The presence of measurement error and/or short-scale heterogeneity are problematic for sampling programs in general, be they adaptive programs or more traditional gridded programs.

7 REFERENCES

Johnson, R., 1996, "A Bayesian/Geostatistical Approach to the Design of Adaptive Sampling Programs," pp. 102–116 in *Geostatistics for Environmental and Geotechnical Applications*, ASTM STP 1283, R.M. Srivastava et al. (editors), American Society for Testing and Materials, Philadelphia, PA.

Thompson, S.K., and G.A.F. Seber, 1996, *Adaptive Sampling*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, NY.