in press, *Risk Analysis*

ADAPTIVE SPATIAL SAMPLING OF CONTAMINATED SOIL

Louis Anthony Cox, Jr.
Cox Associates, 503 Franklin Street, Denver, Colorado, 80218
November, 1998

ABSTRACT

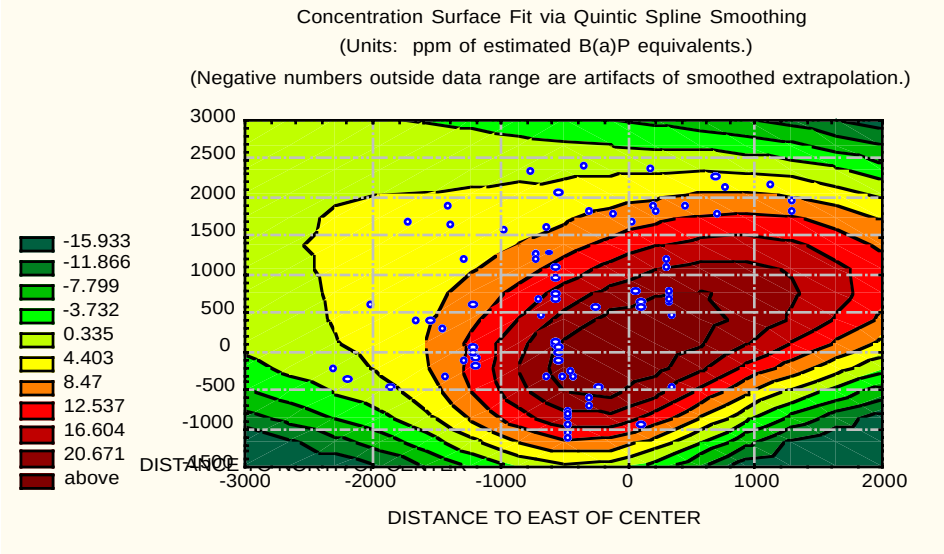
Suppose that a residential neighborhood may have been contaminated by a nearby abandoned hazardous waste site. The suspected contamination consists of elevated soil concentrations of chemicals that are also found in the absence of site-related contamination. How should a risk manager decide which residential properties to sample and which ones to clean? This paper introduces an *adaptive spatial sampling* approach which uses initial observations to guide subsequent search. Unlike some recent model-based spatial data analysis methods, it does not require any specific statistical model for the spatial distribution of hazards, but instead constructs an increasingly accurate nonparametric approximation to it as sampling proceeds. Possible cost-effective sampling and cleanup decision rules are described by decision parameters such as the number of randomly selected locations used to initialize the process, the number of highest-concentration locations searched around, the number of samples taken at each location, a stopping rule, and a remediation action threshold. These decision parameters are optimized by simulating the performance of each decision rule. The simulation is performed using the data collected so far to impute multiple probable values of unknown soil concentration distributions during each simulation run. This optimized adaptive spatial sampling technique has been applied to real data using error probabilities for wrongly cleaning or wrongly failing to clean each location (compared to the action that would be taken if perfect information were available) as evaluation criteria. It provides a practical approach for quantifying trade-offs between these different types of errors and expected cost. It also identifies strategies that are undominated with respect to all of these criteria.

Key words: Spatial statistics, optimal sequential search, adaptive sampling, simulation-optimization, multiple imputation

1. INTRODUCTION AND PROBLEM SETTING

Figure 1 shows a typical example of sampling locations and estimated soil concentration contours for polycyclic aromatic hydrocarbons (PAHs) in residential surface soils around an abandoned hazardous waste site. The site is located at the origin. Estimated average concentrations diminish with increasing distance from the site. Given such data and a desired maximum acceptable surface soil concentration level that is considered health-protective, how should a responsible risk manager decide which additional residential properties to sample, if any, and which ones to clean? This paper provides an answer based on *adaptive spatial sampling,* in which sampled soil concentrations are used to focus further sampling around the highest-concentration locations found so far.

FIGURE 1: Estimated P(a)H Soil Concentration Contours based on Sampled Data at Multiple Locations (500 - 2500 feet from Center)



This adaptive approach can also be applied with non-threshold criteria, such as expected utility or risk-cost-benefit comparisons, that may consider amounts of contaminant removed and remaining, costs of sampling and remediation, and likely sources (site or background) of the contamination at different locations. Decision rules that maximize expected utility or net benefit, or that minimize expected decision errors compared to the decisions that would be made with perfect information, can be identified by simulating the performances of different possible rules. Table 1 defines the decision problems addressed in this paper.

Optimal sampling requires choosing a sampling decision rule (element 4) to optimize an objective function (element 8). The decision rule maps available sample information into a decision about where to sample next and when to stop sampling. Applying it repeatedly, beginning with the initial set of sample data, creates an increasing set of sample observations that terminates with a final data set. A terminal act -- to clean or skip each house -- is then selected to maximize the objective function. A sampling decision rule is *optimal* if it maximizes the expected value of the objective function (compared to all other feasible sampling decision rules), taking into account that the terminal act will be chosen to attain this goal. Section 3 presents an approach to approximating optimal sampling decision rules. An approximation is needed because, in practice, expected values of different rules must be estimated from sample data.

TABLE 1: ELEMENTS OF SPATIAL SAMPLING AND REMEDIATION DECISION PROBLEMS

1. A *spatial distribution* of soil contamination, initially unknown.
2. A list of feasible *sampling locations* at which samples may be collected.
3. An *initial data set* of measured (sampled) concentrations at various locations.
4. A set of feasible sampling *decision rules.* A decision rule specifies what to do next (e.g., where to sample next and whether to stop sampling), given the data observed so far.
5. A *terminal action decision rule,* specifying which locations to clean when sampling halts.
6. A *cost function* specifying the dollar cost of any sequence of sampling and terminal actions.
7. A set of *outcome evaluation criteria*. Possible criteria include the following:
   - Total amount of contaminant removed from residential soils
   - Total amount of contaminant remaining in residential soils
   - Amount of contamination *due to the waste site* that has been removed;
   - Amount and distribution of remaining contamination *from the waste site*
   - Number of properties cleaned and amount of contamination removed from and remaining at each
   - Dollars spent (sum of sampling costs and remediation costs).
8. A *criterion or objective function for choosing among probability distributions for outcomes,* e.g., a utility function for the outcome evaluation criteria.

2. ALTERNATIVE APPROACHES

Table 2 summarizes a traditional approach to residential soil sampling and remediation that might be called the *bright-line* sampling strategy. The parameters of this procedure (S, k, and T), are chosen to give high statistical confidence that a *trigger condition* will be satisfied (i.e., remediation will be undertaken) unless a desired *target condition* is met. A target condition often used is that mean concentrations at uncleaned locations should not exceed mean concentrations at "background" locations, defined as locations not detectably affected by the site. Other common bright-line target conditions require at least 99% confidence that at least 95% of measurements taken at a location will fall below a background-based Maximum Contaminant Level (MCL) or health-based Alternative Concentration Limit (ACL).

Confidence limit and tolerance limit formulas for calculating trigger conditions have been provided by EPA for soil and ground water contamination (e.g., 40 CFR Part 264 for RCRA Subtitle C hazardous waste sites). Gibbons (1994) concludes that these recommendations are prone to false positives and recommends methods for computing nonparametric tolerance intervals when the concentration frequency distribution being sampled is unknown. However, such methods treat concentrations at different locations as statistically independent. As noted by Gibbons, in reality, "Both temporal and spatial correlations violate the independence assumptions (p. 33). …[When] comparisons are needed to control for inherent spatial variability, nonparametric procedures are often simply not available (p. 76)." Section 3 of this paper fills this gap by proposing a class of nonparametric sampling procedures that exploit spatial correlations to improve sampling and remediation decisions.

TABLE 2: SUMMARY OF BRIGHT-LINE STRATEGY

(a) Draw a fixed number, S, of samples from each location (e.g., each residential property);

(b) Compute sample statistics (e.g., the arithmetic or geometric mean and standard deviation) from the sampled values at each location; and

(b) Clean a location if and only if its statistics satisfy a *trigger condition,* e.g., if and only if the sample mean minus k sample deviations exceeds *an action threshold,* T. (T is the "bright line" above which action is required.)

An alternative to the bright-line approach is the *Bayesian decision analysis framework,* in which sequential sampling and terminal actions are based on the likely spatial distribution of contamination, conditioned on all data collected so far. The terminal act is selected to maximize the expected value of a utility function. The procedures are typically sequential, reducing the average sample size needed to make a confident decision (DeGroot, 1970). Table 3 outlines the Bayesian strategy.

TABLE 3: OUTLINE OF BAYESIAN DECISION ANALYSIS STRATEGY

(a) Assume that the spatial distribution of contamination is described by a parametric model, say (x, y ; **q**), where **q** is a vector of parameters. Assume a prior probability density function for **q**, say, $f_{\mathbf{q}}(\mathbf{q})$.

(b) After obtaining any set of observed sample values for soil concentrations at different locations, update beliefs about **q** by conditioning on the observed values using Bayes' rule. (This requires knowing or assuming a likelihood function relating **q** to the observed values.)

(c) Choose the terminal decision, a*, to maximize the expected value of a utility function u(c, **b**), where the expectation is taken with respect to the posterior (conditioned on observed data) probability distribution for **q**.

(d) Recursively choose the sampling locations and stopping rule via backward dynamic programming to maximize expected utility.

The Bayesian decision analysis framework is, arguably, a uniquely coherent and powerful guide to "rational" individual decision-making. It may incorporate much sophistication in quantifying prior beliefs about complex quantities, such as spatial distributions, and in the dynamic programming algorithms used to decide where to sample next, how many samples to take, and when to stop (DeGroot, 1970). However, the framework suffers from practical limitations, especially when multiple stakeholders such as PRPs, the community, and regulators are involved. These include the following:

〈 *How to convince others about models?* Even if one decision-maker has clear beliefs about the model describing the spatial distributions of contamination, it may be difficult to convince anyone else that it is true or useful. Ripley (1981) discusses the limitations of many parametric models for spatial distributions, suggesting that doubts and disagreements can easily arise.

〈 *Unknown and ill-defined priors for spatial density functions.* It may not be easy for even one decision-maker to form or justify a coherent probabilistic model for the prior distribution of the spatial distribution of contaminants (e.g., the joint priod distribution of the model parameters).

〈 *Hard-to-calculate decision rules.* Dynamic programming is ofen computationally impractical unless the class of decision rules considered can be restricted.

〈 *Unknown or ambiguous utility function.* The utility function which roots the dynamic programming recursion with an evaluation for the terminal act may be unknown or disputed. For example, different stakeholders (the PRP, the residential property owners, and regulators) may have different degrees of risk aversion for uncertain health benefits, or may weight the rights and interests and assess the duties and responsibilities of the different stakeholders differently. They may therefore disagree about what should be done. More generally, Bayesian decision theory is not designed to show why the public should accept its prescriptions when these depend sensitively on subjective preferences and risk attitudes.

These challenges must be overcome by practical sampling and decision-making approaches. The fact that different people have different preferences and risk attitudes may perhaps best be addressed by bypassing subjective utility-theoretic calculations in favor of simpler criteria that most stake-holders accept as fair. The other challenges can be addressed by technical devices such as data analysis methods that do not require parametric modeling or prior beliefs These are developed next.

3. OPTIMIZED ADAPTIVE SPATIAL SAMPLING

This section introduces a new heuristic for identifying approximately optimal adaptive sampling decision rules. It is based on the following steps:

〈 *Bound the area(s) of greatest concern* by statistical analysis of spatial sample data.

〈 *Restrict the class of decision rules considered*. This step seeks to identify a set of high-performance decision rules that can be described by a small number of decision parameters. This paper considers *adaptive cluster sampling rules based on order statistics* (Thompson and Seber, 1996, Chapter 6). Such rules have several desirable statistical properties. They are intended to address the fact that the spatial distribution of soil contamination concentrations is unknown, while helping to identify properties with the highest concentrations. They can be indexed by a small set of decision parameters that define specific adaptive sampling procedures.

〈 *Evaluate decision rules in the restricted set using simulation-optimization.* Estimate the expected performance of each decision rule by simulating its application starting from the current data set. Assign multiple imputed values to unknown quantities encountered during the simulation (e.g., contaminant concentrations at locations that have not yet been sampled) by repeatedly sampling the available data from nearby locations. The performance of the decision rule is averaged over the multiple complete data sets, formed from the available data and the imputed values of currently missing data, to estimate its expected performance value. The neighborhood sampled from to impute missing values is selected to minimize the average error of the imputed values, using the currently available data points as a training set. We call this entire process *simulation-optimization,* i.e., simulation with decision parameters optimized with respect to performance criteria estimated from multiple imputations of missing data. It is motivated in part by computational Bayesian methods for missing data (Tanner, 1996; O Ruanaidh and Fitzgerald, 1996; Schafer, 1997).

Simulation-optimization yields the estimated expected value of the decision rule corresponding to each choice of decision parameter values. Searching a grid of parameter values yields an approximation to the optimal choice of the sampling decision rule. Expected values of multiple criteria emphasizing different goals or making different value trade-offs can be estimated simultaneously and used to examine the sensitivity of the optimal decision rule to changes in the evaluation criterion. These ideas are developed in the following paragraphs and illustrated in Section 4.

Bounding The Area of Greatest Concern

Non-parametric methods for establishing the distances and directions where site-related contamination is most likely include the following.

1. *Gradient methods:* Create several "cells" of neighboring sample points (e.g., via a Voronoi tessellation of sampled locations (Ripley 81)). Estimate the direction of steepest increase in concentration for each cell, e.g., by fitting a plane to its sampled concentration values. Then likely source(s) of contamination can be identified as sinks in the estimated gradient field, while the boundary of site-related contamination is roughly indicated by where the gradient directions become randomly distributed or no longer biased toward the site. Steepest ascent methods such as stochastic approximation (Kushner and Yin, 1997) or Evolutionary Operations (Box and Draper, 1969) can also be used to seek the location of the maximum value of the contamination density function based on sequentially sampled values.

2. *Nonparametric regression.* If sampled values are adequately spaced, then the spatial distribution of contamination may be approximated using nonparametric surface-fitting technique such as loess, splines, or kernel regression, as in Figure 1 (Venables and Ripley, 1994). Geostatistical techniques such as kriging may also be used to estimate iso-concentration contours (Isaaks and Srivastava, 1989; Venables and Ripley, 1994). Source(s) of contamination should be located approximately at the centers of concentric contours. Such methods may give unrealistic values outside the spatial range of the data, however (e.g., the negative numbers in Figure 1.)

3. *Nonparametric regression estimates of boundaries.* In each octant (or smaller angular sector, if there are enough data points) fit a nonparametric smooth curve to the relation between distance from the site and concentrations at sampled locations, e.g., using nonparametric regression. The distance at which this curve levels out (if any) indicates an approximate practical boundary for site-related contamination in each direction. An alternative that corrects for multiple hypothesis testing is to apply *recursive partitioning* (or "classification tree") algorithms to divide the sampled data points into relatively homogeneous sectors (Biggs 1991). Although such algorithms are not inherently optimized for spatial data, they can be applied directly to sample data (eastings, northings, and sample concentrations) to obtain an empirical histogram approximation to ρ. If sample locations are expressed in polar coordinates, recursive partitioning will create a histogram from annular segments (defined by inner and outer radii and subtending angular directions). Section 4 applies this technique to determine a boundary in each angular direction dividing apparent "background" data points from closer-in points with concentrations that are detectably elevated compared to the apparent background. (See Table 5.)

4. *Nonparametric changepoint detection and image segmentation techniques.* A more sophisticated algorithm approximates the spatial density function with a histogram formed from polygonal prisms obtained by iterative improvement of randomly generated Voronoi tessellations (Green, 1995) using Markov Chain Monte Carlo methods (Gamerman, 1997).

For the case study described in Section 4, different approaches gave similar estimated boundaries for the contaminated area. However, no statistical approach can preclude the possibility that *some* site-related contamination extends beyond the statistically identified boundaries. Rather, boundary-setting helps to focus attention and resources on the most contaminated areas.

Adaptive Sampling: A Restricted Class of Decision Rules

The *adaptive cluster sampling* rules (Thompson and Seber, 1996) considered in this paper use four decision parameters to determine where to sample next and when to stop sampling based on the data obtained so far. They are defined as follows:

- *N = initial number of random samples* to be drawn from the area of concern. These N samples are in addition to those that have already been obtained, if any, when adaptive sampling begins. N is the *number of seeds* for the adaptive cluster sampling algorithm. They may be selected by listing all residential properties in the area of concern and drawing N of them at random. Where to sample within a selected property is determined by a standard field protocol (e.g., a randomly selected point with no surface discoloration, evidence of oil spills, or other obvious non-site-related sources of contamination.)

- *S = number of soil samples* to take from each newly sampled property.

- *K = maximum allowed length of a "search list".* This is a list of properties with the highest average sample concentrations seen so far; see below.

- *T = action threshold.* A property is cleaned if and only if the average of its S sample values exceeds T.

The adaptive sampling procedure for parameter vector (N, S, K, T) is as follows. Call a property *fathomed* if S samples have been taken from it and from each of its immediate neighbors. (The "immediate neighbors" of a property may be defined in various ways without changing the algorithm. A common choice is to define them as the four properties that are closest to it from the north, south, east, and west directions.) Initially, place the K properties with the highest average sampled values seen so far (among the initial samples + N added samples) on the search list. Fathom each property on the list (by sampling), and remove it from the list when it has been fathomed. Whenever a property is found that has an average sample value greater than the K + 1^{st} initial value (i.e., that is higher than any of the values for properties not included on the initial search list), place it on the search list. Continue until the search list becomes empty (i.e., all properties previously placed on it have been fathomed and no new, unfathomed properties are added to it.) This ends the adaptive sampling phase. Then, clean all and only those properties for which the sample average concentrations meet or exceed the action threshold, T. Table 4 summarizes the procedure.

TABLE 4: WORST-FIRST ADAPTIVE SAMPLING WITH PARAMETERS N, K, S, T

1. Randomly sample N previously unsampled locations in the area of concern.
2. Place on a search list the K properties with the highest sample average concentrations observed so far.
3. Fathom each property on the search list by taking S samples from it and from each of its previously unsampled neighbors. Remove each property from the list when it has been fathomed.
4. Place onto the search list any property with a sufficiently high sample average concentration, where "sufficiently high" means higher than any of the initial sample values not included on the initial search list in step 2.
5. Continue steps 3 and 4 until the search list is empty.
6. Clean those properties with sample average concentrations of at least T.

This procedure may be called *worst-first adaptive search* for seeking properties with high soil contamination levels. As shown by Thompson and Seber (1996, p. 163), such sampling allows unbiased estimates of the average soil concentration remaining in unsampled locations at any stage. The terminal decision rule specified by T is motivated by the Neyman-Pearson Lemma (DeGroot, 1971, p.146).

Simulation-Optimization of the Adaptive Search Decision Parameters

To find an approximately optimal adaptive sampling strategy, we first restrict the strategies considered to worst-first adaptive search rules parameterized by the decision parameter vector **p** = (N, K, S, T). Seeking an optimal strategy in this class is straightforward. It is often practical to perform an exhaustive grid search over all values of the integer variables N, K, and S of practical interest, e.g., with N, K, and S each ranging from 0 to 20. (In the case study of Section 4, performance is optimized by combinations well within the interior of this region, as discussed later.) For each combination of values for the integer variables, a line search is conducted to optimize the value of T (M. Box, 1969).

Evaluating each trial value of **p** is less straightforward. Given a proposed parameter vector, say $\mathbf{p}_0 = (N_0, K_0, S_0, T_0)$, its expected performance on the selected criteria (e.g., utility, type 1 and type 2 error rates, amounts of contamination removed and remaining, etc.) is estimated by simulation. During the simulation runs, missing data values are imputed from available data multiple times and the average performance of $\mathbf{p}_0$ is calculated by averaging over these multiple imputations. In more detail, $\mathbf{p}_0$ is evaluated as follows.

1. IMPUTE CONCENTRATION DISTRIBUTIONS: For each location, impute a number, C, of values for its estimated concentration distribution by drawing C times at random with replacement from among the values of the M nearest neighboring locations for which sample data are available (possibly including the location itself). (Alternatively, values could be imputed from all sampled values within R feet of the location, if there are enough of them.) More elaborate methods of imputation, such as distance-weighted sampling probabilities or kriging among nearby values, could also be used in this step (Ripley, 1981), but sampling from nearby data values is simple and reflects the local distribution of concentrations. The parameters C and M (or R) are *meta-parameters*, and are set via another round of optimization (Step 5).

2. SIMULATE ADAPTIVE SAMPLING. Simulate the procedure corresponding to $\mathbf{p}_0$. Thus, select N_0 new locations at random from among those not yet sampled. Take S_0 samples from each new location (drawing S_0 times at random with replacement from the C available values where imputed data values must be used). Form a search list of length K_0, recursively fathom its members, and identify the locations that would be cleaned, i.e., whose average sample concentrations exceed T_0.

3. EVALUATE PERFORMANCE. Evaluate how well the procedure $\mathbf{p}_0$ has done at each location, using the location's imputed concentration distribution to evaluate the decision made there (skip or clean). (For example, the value of the true average concentration would be imputed by taking the average of the C imputed values.)

4. ITERATE. Repeat Steps 1-3 a total of I times, where I = number of iterations. (I = 20 is a typical value in our applications.) Average the performance of $\mathbf{p}_0$ over the I iterations. Thus, though Step 3 treats imputed values as if they were correct within each iteration, evaluating $\mathbf{p}_0$ for multiple imputations (as well as for multiple random seeds) builds up a frequency distribution of evaluations. The mean of this distribution is an estimate of the expected value of $\mathbf{p}_0$ given the available data used for imputations (Schafer, 1997).

5. OPTIMIZE META-PARAMETERS. Steps 1-4 involve the meta-parameters C, M (or R), and I. The final step is to set the values of these meta-parameters. The value of I is set by satisficing, with 20 or 30 iterations generally proving adequate. The criterion for adequacy is that increasing the number of iterations further (e.g., doubling or quadrupling it) should make no significant difference in the expected values estimated. The values of C and M, by contrast, are set by a separate round of optimization based on cross-validation (Gamerman, 1997). Values of C and M are selected to minimize the average error in imputed means *vs.* empirical means at locations where sample data are available. (In the case studies reported below, we used C = 10 and M = 5.)

Applying the evaluation loop 1-4 to each trial value of **p** identifies the procedures giving the best estimated average performance with respect to any performance criterion desired (e.g., error rates or expected utilities) that can be calculated from the imputed values in Step 3. Assessing multiple criteria increases the computational burden very little: it is only necessary to store a performance vector for each **p** value instead of a single value during each iteration. The stored performance data then allow correlations and trade-offs among criteria to be displayed.

In decision analysis, "optimal" decisions are typically made based only on the *probabilities* of consequences induced by the decisions (DeGroot, 1970). The actual consequences of alternative risk management acts are seldom known when a decision among them must be made. In the soil remediation context, the actual consequences may remain unknown, since the amounts of contaminant removed and remaining are usually never learned. To the extent that multiple imputations produce a probability distribution for consequences that is close to the true one, however, simulation-optimization provides the information needed to make optimal (e.g., expected utility-maximizing) statistical decisions, given the available data.

4. A CASE STUDY

Application Setting: Statistical Challenges and Policy Goals

In 1995, the regional Environmental Protection Agency had reviewed the surface soil sample data underlying Figure 1 with two PRPs. All parties had agreed on the following policy goals:

〈 *Health-protection/efficiency:* Remove as much as possible of the residential surface soil contamination created by the site, for any level of resources applied. (This motivated the "worst-first" approach.)

〈 *Fairness:* Avoid having the PRPs clean soil contamination not due to their site. Since the hypothesized mechanism of contamination was airborne deposition, the exact extent of site-related contamination in different directions was uncertain.

〈 *Clarity:* Be able to explain and justify to members of the community which locations were selected for sampling and/or remediation, and why.

Both the EPA Project Manager and the PRPs recognized that these goals partially conflict. The statistical methodology developed to pursue them also had to deal with the following realities. First, *it is impossible to determine whether (or by how much) any specific individual residential property has been contaminated by the site.* For example, some background locations have higher sample soil concentrations than some of the locations closer to the site. Second, soil concentrations are so variable, even on the same residential property, that *it is impractical to determine which locations have the highest true*

mean soil concentrations. Finally, the spatial distribution of soil contamination concentrations is unknown.

To reach a pragmatic solution, the EPA Project Manager, in agreement with the PRPs, made the following simplifying assumptions:

1. The objective is to *minimize the expected number of decision errors*, defined as the sum of type 1 and type 2 errors ("false cleans" and "false skips" of residential locations, respectively). This objective considers only the number of errors made and not their sizes. Yet, it was considered simple and clear enough to be understood by the community (unlike utility-based calculations) and likely to produce reasonably fair, efficient, and politically acceptable outcomes. This objective function was later generalized to a *weighted sum* of type 1 and type 2 errors.

2. *No location should be cleaned that has not first been sampled.* Remediation is sufficiently disruptive for home-owners so that it should not be undertaken without strong evidence of elevated PAHs. In addition, sampling a location is orders of magnitude less expensive than cleaning it, making it worthwhile to sample extensively before allocating remediation resources.

To define the minimum expected-errors criterion, a risk-based health protective concentration level for benzo(a)pyrene [B(a)P] in residential soils was first estimated via probabilistic risk assessment, taking into account likely soil ingestion, inhalation, and dermal contact scenarios and bioavailability factors. This set an action threshold such that it was deemed to be a "correct" decision to clean a residential location if and only if the average concentration of PAHs there (expressed as B(a)P equivalents, using conversion tables provided by the EPA) exceeded the action threshold.

Initial Soil-Sample Data and Identification of Area of Greatest Concern

Table 5 shows the results of applying a recursive partitioning algorithm (Biggs *et al.,* 1991) to the initial sample data underlying Figure 1 to help bound the area of greatest concern and automatically detect any potential "hot spots". It automatically partitioned the available soil concentration data (B(a)P equivalent concentrations measured at different polar coordinates) into five distinct "sectors", represented by the five rows in Table 5. Three of these sectors lie in the north-northeast octant, at different distances from the center of the site. The recursive partitioning algorithm creates sectors with statistically significantly different frequency distributions of sampled soil concentrations, such that no sub-sectors (identifiable by subdividing the angle or distance coordinates) are statistically significantly different. The greatest mean concentration occurs in Sector 1, which lies immediately to the north-northeast of the site, from the fence line out to 765 feet. Seven distinct residential locations had been sampled in this sector, some more than once, suggesting that a confident decision about the true mean concentration at a residential property will require many samples.

TABLE 5: Sample Soil Concentration Data Partitioned Into Sectors

Sector	Angle Range (degrees)	Distance Range	Mean Conc. (B(a)P Equiv)	Std Deviation	N
1	48 - 89	< 765 feet	36.6	15.0	10
2	48 - 89	765 - 1193	16.8	6.5	12
3	48 - 89	> 1193	4.9	5.1	21
4	89 - 185	All (< 2500)	4.7	5.8	56
5	> 185	All (< 2500)	10.3	10.4	26

Analysis of the correlation between concentrations at different locations, expressed as a function of the distance and direction between them, revealed that concentrations are positively correlated at distances on the order of 100 feet throughout

the sample. Concentrations at the nearest sampled neighbor of a location (typically on the order of 100 feet away) is a better predictor of its concentration than are concentrations at more distant locations. This observation motivated the need for an approach that uses observed high concentrations to guide the search for other locations with high concentrations.

Implementation Experience and Results

The PRPs proposed the adaptive spatial sampling approach and simulation-optimization method to the regional EPA team in 1996. Several rounds of discussion showed that the approach seemed suitable to address the needs to learn from data, allocate sampling and remediation resources effectively, and maximize the probability of finding high soil concentrations within the area of concern. Refinements and extensions made during these discussions included using traditional statistical methods such as linear regression to confirm results of the spatial data analysis and the recursive partitioning algorithm (Table 5).

During 1997, it became clear that the relative value weights for different policy goals might change, depending in part on the effects of local environmental activists in shaping community perceptions and preferences. In particular, the relative priorities to be given to type 1 and type 2 errors were identified as uncertain and labile. Therefore, the EPA Project Manager identified a need to be able to minimize any weighted sum of type 1 and type 2 errors with arbitrary non-negative relative value weights (summing to 1). In addition, the PRPs wished to use a decision method that could also be applied to make normatively "rational" decisions, i.e., to maximize the expected value of a single- or multi-attribute utility function. These needs led to the requirements addressed by the final methodology. Both the regional EPA team and the PRPs judged worst-first adaptive spatial sampling with optimized decision parameters to be the best method proposed to date for balancing the competing policy goals and making effective use of existing data in sampling and remediation decisions. The EPA Project Manager approved the method for use in 1997 and it was implemented in 1998.

Figure 2 shows the expected type 1 and type 2 error rates, assessed via simulation-optimization, for many different parameter vectors **p** = (N, K, S, T). Table 6 shows example outputs from the 20 vectors having the lowest total error rates, sorted in order of increasing total error rate. (The strategies used in these illustrations are a subset of over 5000 evaluated by grid search.) Each strategy takes several CPU-seconds to evaluate via simulation-optimization. The two superimposed scatter plots in Figure 2 show the estimated total error rate (upper scatter plot, square points) and type 2 error rate (lower scatter plot, circles) for each strategy on the vertical axis, while the horizontal axis shows the corresponding type 1 error rates. The regression lines

FIGURE 2: Error Trade-Offs Among Adaptive Sampling Strategies, Estimated by Simulation-Optimization

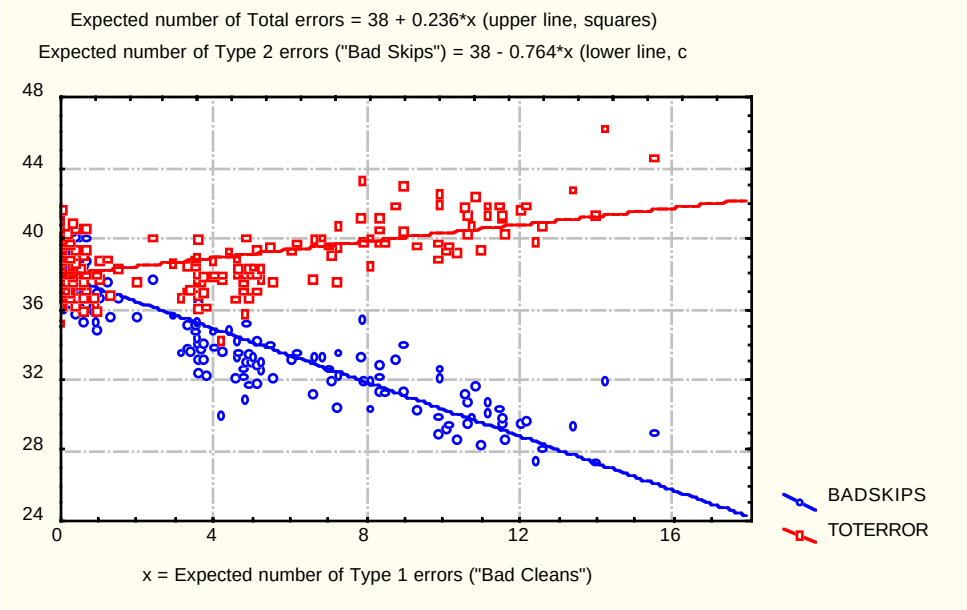


TABLE 6: Simulation-Optimization Outputs: Evaluations of 20 Adaptive Sampling Strategies with the Smallest Total Errors (out of

N	S	K	T	Error Rank	Cost Rank	Type 1* (Bad Cleans)	Type 2* (Bad Skips)	Total Errors	Expected Total Cost*
5	16	1	3.3	**1**	**142**	**4**	**30**	34	966571
20	12	3	3.9	**2**	**97**	**0**	**35**	35	685,714
20	8	3	3.3	3	141	5	31	36	951333
0	8	3	3.6	**4**	**43**	1	35	36	410,190
0	12	3	3.6	5	63	**0**	**35**	36	512,095
5	20	1	3.3	6	161	4	32	36	1,066,667
0	16	1	3.9	7	56	0	36	36	474667
20	4	3	3.9	**8**	**18**	0	36	36	245,048
5	30	1	3.3	9	200	4	32	36	1414286
0	30	3	3.9	10	159	0	36	36	1060000
5	30	3	4.2	11	180	0	36	36	1232857
20	12	3	4.2	12	93	0	36	36	669143
10	16	3	3.6	13	114	0	36	36	789429
20	8	1	3.6	14	61	0	35	36	496286
10	20	3	3.3	15	201	5	32	37	1414762
10	8	3	3.6	16	49	0	36	37	439238
0	8	1	3.9	**17**	**17**	0	37	37	236,190
20	4	1	4.2	**18**	**12**	0	37	37	198,286
0	30	3	3.6	19	167	0	36	37	1123810
10	20	1	3.3	20	175	**3**	**33**	37	1,174,286

* Expected values, rounded to nearest integer, assessed by simulation-optimization. Bold numbers indicate strategies that are undominated with respect to the attributes in their columns.

indicate that:

(a) Strategies with the lowest type 1 error rates tend to have higher type 2 error rates (as expected); and

(b) Strategies that spend resources to reduce type 2 errors tend to create more than one type 1 error for each type 2 error removed. Thus, the total error line slopes up while the type 2 error line slopes down.

Intuitively, cleaning all locations would eliminate all type 2 errors, but at the cost of introducing a larger number of type 1 errors. If type 1 and type 2 errors are weighted equally, then the optimal strategy will lie toward the left in Figure 2. A contribution of the simulation-optimization method is to allow such trade-offs to be quantified. It also allows the sensitivity of the "best" strategy to the relative value weights placed on different goals or criteria to be assessed

In the absence of known, stable relative value weights for type 1 and type 2 errors, it is worthwhile to identify *undominated strategies*. A strategy is undominated with respect to any set of criteria if no other strategy is better on all of them. For example, strategies 8, 17, and 18 in Table 1 are undominated with respect to the criteria of "error rank" and "cost rank". Here, the "error rank" column is used to number the strategies. This column shows the ordinal ranking of each strategy based on total error rate. (Bold numbers in Table 6 indicate the strategies that are undominated with respect to the attributes in the corresponding columns.) The Type 1 and Type 2 error columns and the cost rank column show information for these criteria. "Cost" in Table 6 is defined as the sum of sampling plus remediation costs, using estimated unit costs of sampling and cleaning each location provided by the PRPs. Sensitivity analyses showed that the undominated strategies are robust to variations in these cost estimates.

Although expected type 1 and type 2 error rates were primary criteria for decision-making, some attention was also paid to cost-effectiveness. In part, this recognized that any remediation activity would remove some contamination that was not site-related. (Indeed, the site may not be the source of the elevated soil contamination levels. For example, Table 5 shows no apparent distance-related decline in soil concentrations around the site except in the north-northeast octant. This contrasts with the impression given by the contours in Figure 1, which reflect the assumption of a *smooth* relation.) In light of these considerations, undominated strategies were sought that ranked in the top 10% of all strategies on *both* total error rate and cost. Strategies 4, 8, 17, and 18 in Table 6 satisfy these requirements.

The adaptive sampling procedure corresponding to strategy 17 (N = 0, S = 8, K = 1, T = 3.9) was implemented in 1998. As anticipated based on previous multiple imputations, none of the four neighbors of the most-contaminated site (K = 1) identified in the initial sample had higher average concentrations than those already observed, suggesting that the locations with the highest concentrations may have already been identified. The adaptive sampling procedure therefore halted with the conclusion that additional sampling was unlikely to reveal additional locations in need of remediation, given the data collected already. By contrast, an originally discussed "bright line" sampling plan would have required sampling residential properties out to 2,000 feet in all directions, yet probably would have produced little additional information for improving the final decisions on where to skip and where to clean. Both the PRPs and the regional EPA felt that worst-first adaptive sampling had helped to focus the search and greatly reduce the time and effort needed to obtain information useful for resource-allocation decisions.

Despite this initial success, other value weights could lead to different strategies and justify further sampling. The next step is to incorporate the newly collected data and to re-rank the strategies according to different value weights for type 1 errors, type 2 errors, and costs. This process is currently underway.

5. DISCUSSION, EXTENSIONS, AND CONCLUSIONS

Simulation-optimization appears to be useful for identifying adaptive sampling plans to meet desired performance objectives. Unlike recent model-based spatial data analysis techniques (Byers 98, Dasgupta 98), it does not require specific assumptions about the spatial distribution of contamination. As sampling progresses, new data are used to obtain increasingly refined estimates of this distribution. The technique provides a computationally practical approach to deciding where to sample next, when to stop, and what to do to optimize an overall performance objective such as expected utility or number of decision errors. By imputing probability distributions for unknown concentrations, it allows expected values of objective functions that depend on these quantities to be estimated and allows quantitative trade-offs and sensitivity analyses to be performed.

Simulation-optimization can be applied whenever the set of decision rules considered is described by a vector of decision parameters and the expected performance of each rule can be accurately simulated. For example, suppose that a project manager decides that adaptive sampling is impractical, e.g., because it requires too many intrusive visits to the community, and instead elects to use a two-stage sampling procedure, in which an initial sample is used to screen each residential property for further investigation, and a subset of the screened properties is then investigated further to determine whether remediation is needed. No location may receive more than two sampling visits. The appropriate class of decision rules in this case might be indexed by a vector of decision parameters of the following form:

$$\mathbf{p} = (S_1, T_1, S_2, T_2),$$

where S_1 samples are first taken from each location, an additional S_2 samples are then taken from all and only those locations for which the arithmetic average of the first S_1 samples exceeds the screening threshold T_1, and then those locations are cleaned for which the average of the $(S_1 + S_2)$ samples exceeds the action threshold T_2. Although such two-stage sampling (or its n-stage generalization) are very different from the adaptive sampling rules considered in this paper, the simulation-optimization heuristic can be applied to optimize its decision parameters.

In summary, the mathematical forms of decision rules that can be implemented are often determined by pragmatic constraints rather than by purely statistical considerations. Once the form has been decided, however, simulation-optimization can be used to identify the best specific procedure, provided that data are available to support the required imputations. Thus, the simulation-optimization framework can potentially be applied to improve statistical decision-making in a wide variety of practical applications.

ACKNOWLEDMENT

I thank David Burmaster of Alceon for introducing this problem to me and sharing data sets and ideas on how to tackle it.

REFERENCES

1. Bell, D.E., 1995. A contextual uncertainty condition for behavior under risk. *Management Science,* **41**, 7, 1145-1150.

2. Biggs, D., B. de Ville, E. Suen, 1991. A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics.* **18**, 1, 49-62.

3. Box, G.E.P. and N.R. Draper, 1969. *Evolutionary Operations.* Wiley.

4. Box, M.J., D. Davies, W.H. Swann, 1969. *Non-linear Optimization Techniques.* Oliver & Boyd, Edinburgh.

5. Byers, S., and A. E. Raftery, 1998. Nearest-Neighbor Clutter Sampling for Estimating Features in Spatial Point Processes. *Journal of the American Statistical Association,* Volume **93**, Number 442, June 1998, pp. 577–584

6. Cox, L.A., Jr., 1989. Managing uncertain risks through 'intelligent' classification: A combined artificial intelligence/ decision analysis approach. Pages 473-482 in J.J. Bonin and D.E. Stevenson (eds), *Risk Assessment in Setting National Priorities.* Plenum Press.

7. Dasgupta, A., and A. E. Raftery , 1998. Detecting Features in Spatial Point Processes with Clutter via Model-based Clustering, *Journal of the American Statistical Association*, Volume **93**, Number 441, March 1998, pp 294-302.

8. DeGroot, M.H., 1970. *Optimal Statistical Decisions.* McGraw-Hill.

9. Feller, W., 1968. *An Introduction to Probability Theory and Its Applications. Volume 1.* Third Edition. Wiley.

10. Gamerman, D., 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman & Hall.

11. Green, P.J., 1995. Reversible jump Markov chain Monte-Carlo computation and Bayesian model determination. Biometrika, **82**. 711-32.

12. Gittins, J.C., 1989. *Multi-armed Bandit Allocation Indices.* Wiley.

13. Isaaks E.H. and R.M. Srivastava, 1989. *Applied Geostatistics.* Oxford University Press.

14. Kushner, H.J., and G.G. Yin, 1997. *Stochastic Approximation Algorithms and Applications.* Springer.

15. Okabe, A., B. Brooks, K. Sugihara, 1992. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* Wiley.

16. Ruanaidh, J.J.K. and W.J. Fitzgerald, 1996. *Numerical Bayesian Methods Applied to Signal Processing.* Springer.

17. Ripley, B.D., 1981. *Spatial Statistics.* Wiley.

18. Schafer, J.L., 1997. *Analysis of Incomplete Multivariate Data.* Chapman & Hall.

19. Tanner, M.A., 1996. *Tools for Statistical Inference, Third Edition.* Springert.

20. Thompson, S.K., and G.A.F. Seber, 1996. *Adaptive Sampling.* Wiley.

21. Venables, W.N., and B.D. Ripley, 1994. *Modern Applied Statistics with S-Plus.* Springer-Verlag.