October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History

By

Mark C. Gemperline, U. S. Bureau of Reclamation, D-8340, PO Box 25007, Denver, CO 80225

Presented at the 15th Annual International Conference on Contaminated Soils and Water

October 18-19, 1999

University of Massachusetts at Amherst

Corresponding Author: Mark C Gemperline

U.S. Bureau of Reclamation, MC-D-8340

P.O. Box 25007

Denver, CO 80225-0007

Tel: 303-445-2334

Fax: 303 445-6341

mgemperline@do.usbr.gov

October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

ABSTRACT

It is difficult to know how many locations should be sampled when looking for contaminants at uncontrolled waste disposal sites. This is especially true at large sites which may have been subject to small releases of unknown chemicals at indefinite locations. A method was presented to estimate the minimum number of samples required to reasonably assure detection of the smallest contaminated area that may present an unacceptable impact to human health. The approach was applied to a 50 acre former municipal and industrial waste disposal site. Multiple contaminants having varying degrees of toxicity were considered. The results suggested using both composite and discrete samples to independently represent 16 areas of concern (AOCs). Using both methods provided a redundancy which yielded two separate estimates of mean contaminant concentrations for each AOC. The sampling plan was implemented. Mean chemical concentrations were calculated using composite sample data and compared to the means calculated using discrete samples. This comparison was a mechanism to evaluate the adequacy of site characterization. The distribution of PCBs in site surface soils was estimated utilizing point sampling data and graphically portrayed on a site map. Computer simulated composite sampling of this distribution, thus represented, was performed. The simulation produced data analogous to the site composite sampling. The computer simulation results were compared with actual field investigation results to indicate the adequacy of the representations of chemical distribution.

October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

INTRODUCTION

A site characterization is described that concurrently utilized both composite and discrete sampling to independently estimate the average chemical concentrations of 16 subdivided surface areas. The sampling plan was driven by the need to make risk-based decisions. Discussion is limited to surface soils, although air, surface water, ground water, and subsurface soil were also investigated. The average concentrations of 157 chemicals were estimated for each of the 16 subdivided areas. Lead and PCB measurements are used to compare the two investigation methodologies.

BACKGROUND

The site will remain anonymous due to pending litigation. It will be referred to simply as the Site. An intentionally distorted, however adequately representative, plan view of the 50 acre Site is shown on Figure 1. A nearby 30 acre area, selected to represent background conditions, is also portrayed. The Site was operated as a municipal and industrial dump and/or salvage yard from around 1950 until the early 1980s. Waste removal actions in the late 1980s and early 1990s resulted in the clearing of most surface debris and visible waste. Vegetation consisted primarily of native grasses; surface soil was predominantly clay; and access to nearly all locations was unobstructed.

Historical data suggested a potential for any of a large number of chemicals to have been released at any location on the Site in unknown quantities. It was assumed in developing the sampling plan that hot spots presenting unacceptable conditions could exist without visual traces. It was desired to reasonably assure that a threat to human health is not present if contamination is not discovered during the investigation. Consequently, detecting a surface hot spot that may present an unacceptable human health impact was established as an investigation goal. In addition, an unacceptable hot spot would need to be characterized adequately to meet human health risk assessment needs.

ESTIMATING DATA NEEDS FOR HUMAN HEALTH RISK ASSESSMENT

The level of effort invested in characterizing this Site was driven by data needs for human health risk assessment. Sampling needs were estimated by working backwards through a simplified risk assessment process (Gemperline, 1994). The following steps were involved: 1) developing a conceptual exposure model, 2) estimating the minimally acceptable average chemical concentrations for the Site, 3) identifying a minimally acceptable Site

chemical distribution model, 4) estimating the number of samples required to assure a chemical is detected if it presents an unacceptable condition, and 5) estimating the number of samples required to adequately estimate mean chemical concentrations. In order to simplify this process, a single, hypothetical contaminant is defined that has the characteristic of posing an unacceptable human health risk when its mean Site concentration equals 1. The concentration of this hypothetical contaminant at a single sample location is calculated as the sum of risk-normalized concentrations of all chemicals present in a sample. The subsequent section titled *Risk Normalized Concentration* presents how this concentration is calculated.

Subsequent discussion assumes the reader has a limited understanding of the human health risk assessment process. A detailed description of common risk assessment practice is contained in referenced U.S. EPA guidance documents (U.S. EPA, 1989, 1992).

Conceptual Human Exposure Model

A human exposure model describes how people might receive a dose of chemicals from the Site. A conceptual exposure model is needed to calculate the minimally acceptable average chemical concentration values for the Site. These values form the basis for quantifying data needs.

The conceptual model consisted of a child living one percent of a 70 year lifetime on the Site. While on the Site, the child receives a chemical dose from surface soil by dermal contact and ingestion. It was further assumed that his exposure would result from random encounters with contamination at a frequency equivalent to chemical spatial distribution on the Site. In other words, encounters with hot spots would occur at a frequency equal to the proportion of the Site area covered by the hot spot. Consequently, the arithmetic means of Site chemical concentrations are adequate to represent exposure concentrations. A carcinogenic risk greater than 1E-4 and hazard index greater than one were considered unacceptable.

Risk Normalized Concentration

The Hazard index (*HI*) is defined as the sum of the ratios of exposure dose to allowable dose for each chemical encountered in the exposure scenario.

$$HI = \sum_i \frac{Exd_i}{Ald_i} \qquad (1)$$

Where:

Exd_i = estimated daily exposure dose of chemical I

Ald_i = allowable daily dose of chemical I.

An *HI* greater than one is unacceptable.

The daily exposure dose, Exd_i, received by the hypothetical child represented in the conceptual exposure model is given by:

$$Exd_i = \overline{C_i} \times t \times (S \times Ss_i + D \times Ds_i) \quad (2)$$

The estimated exposure dose to the hypothetical child in the conceptual exposure model depends on the fraction (t) of his expected 70 year lifetime that is spent at the Site. One percent (t = 0.01) is consistent with the conceptual exposure model. This is appropriate for estimating doses of chemicals that exhibit adverse human health impacts under chronic exposure conditions, e.g., carcinogenic compounds. However, it is necessary to assume that the child spends all of his time on the Site (t=1.0) when considering short term daily exposure to acutely toxic chemicals. For the purpose of estimating data needs it was assumed that the child ingests 200 mg of soil per day (S) and has dermal contact with 10,000 mg of soil per day (D). The fraction of ingested chemical that is adsorbed into the body (Ss_i) and the fraction of dermal contacted chemical that is adsorbed into the body (Ds_i) were assumed to be 1 and 0.1 respectively for all chemicals except PCBs. For PCBs, Ss_i was assumed to be 0.3. These assumptions were made for computational simplicity and were considered adequate for the purpose of estimating data needs.

The allowable daily dose of a chemical (Ald_i) that the child of the conceptual model may receive is given by:

$$Ald_i = Rfd_i \times M \quad (3)$$

Where:

Rfd_i = reference dose obtained from published toxicological data

M = mass of conceptual model child (16 kg)

Values of Rfd_i for acutely toxic chemicals were obtained from documented toxicological data and are typically presented in units of mass of chemical per mass of receptor per day. Terms analogous to the reference dose were calculated for carcinogenic compounds by dividing the acceptable risk by commonly reported values of

slope factors. However, care was taken to consider carcinogenic chemicals and acutely toxic chemicals separately.

Substituting Equations 2 and 3 into Equation 1 yields the following expression for the hazard index.

$$HI = \sum_i \left[\overline{C_i} \times \frac{t \times (S \times Ss_i + D \times Ds_i)}{M \times Rfd_i} \right] \quad (4)$$

The second term in the brackets weights the mean Site chemical concentrations ($\overline{C_I}$) with respect to the risk to human health. Hence, *HI* is the sum of risk-normalized exposure concentrations and represents all chemicals present. It was desired to sample adequately to assure that the *HI*s are less than 1 for carcinogens and noncarcinogens potentially left undiscovered at the Site. The sampling effort was driven by the smallest hot spot that could reasonably exist that would cause an *HI* greater than one.

Since hot spots were a primary concern, an equation for *HI* was sought that related chemical concentrations at individual Site locations. Since an unbiased sampling effort was planned, the mean Site concentration of each chemical would be the sum of concentrations measured at all locations divided by the total number of measurement locations (n). Equation 4 can therefore be written as:

$$HI = \frac{1}{n} \sum_j \sum_i \left[C_{i,j} \times \frac{t \times (S \times Ss_i + D \times Ds_i)}{M \times Rfd_i} \right] = \frac{1}{n} \sum_j \mathrm{C}_j \quad (5)$$

Where:

$C_{i,j}$ = The concentration of chemical I at Site location j.

$\mathbf{C}_j$ = The sum of all risk-normalized chemical concentrations at location j.

n = The total number of unbiased sampling locations.

The bracketed expression can be viewed as representing the risk-normalized concentration of a single chemical. The sum of all risk normalized chemical concentrations at a given location represents the concentration of a hypothetical chemical. Its average over the site is the value of HI. Therefore, this hypothetical chemical has the property that it presents an unacceptable threat to human health when its average Site value is 1. This

hypothetical chemical, identified as **C** with bold font, is used in subsequent discussion to represent the sum of risk normalized chemical concentrations.

Conceptual Contaminant Distribution

A conceptual hot spot, portrayed on Figure 2, was created to represent the potential distribution of chemicals on the Site. It is circular and exhibits a maximum concentration in its center and linearly decreasing concentration with increasing radial distance. The hot spot average concentration, expressed as a risk normalized concentration ($\overline{\mathbf{C}}_{HS}$), is one third its maximum concentration ($\mathbf{C}_{max}$).

$$\overline{C}_{HS} = \frac{1}{3} C_{max} \tag{6}$$

The hot spot contribution to the area average concentration, *HI* by definition, is related to the Sites area (A_{Site}), the area of the hot spot (A_{HS}), and $\overline{\mathbf{C}}_{HS}$.

$$HI = \overline{C}_{HS} \times \frac{A_{HS}}{A_{Site}} \tag{7}$$

The level of effort invested in the Site investigation was driven by the size of smallest hypothetical hot spot that causes the *HI* to equal 1.

The consequential hot spot area, A_{CHS}, is introduced to describe the smallest single hot spot that would cause an *HI* of 1. Setting the *HI* to 1 in Equation 7 gives:

$$A_{CHS} = \frac{A_{Site}}{\overline{\mathbf{C}}_{\mathbf{HS}}} \tag{8}$$

Substituting Equation 6 into Equation 8 results in an expression for the consequential hot spot area in terms of the expected Site maximum concentration.

$$A_{CHS} = \frac{3 \times A_{Site}}{\mathbf{C}_{\max}} \tag{9}$$

Historical information pertaining to the Site, chemical concentration measurements obtained during previous removal actions, and experience regarding chemical fate, were utilized to estimate the expected maximum surface soil chemical concentrations which were then normalized and summed to estimate C_{max}. This is demonstrated later in example 1.

The actual detectable area (A_{DHS}) of the consequential hot spot is smaller than the total hot spot area and is given by:

$$A_{DHS} = A_{CHS} \times (1 - \frac{\mathbf{C}_d}{\mathbf{C}_{\max}})^2 \quad (10)$$

Where $\mathbf{C}_d$ is the risk-normalized value of the analytical limit of detection. A risk-normalized detection limit is superficial since each analysis detection limit is chemical dependent and unrelated to Site risk. However this normalization is appropriate for comparison to $\mathbf{C}_{max}$. The largest normalized detection limit for all chemicals being considered should be used to estimate the detectable region of the consequential hot spot. This is demonstrated later in example 1.

Number of Samples Required to Detect Unacceptable Contamination

A mathematical expression for the number of random locations that must be sampled to be confident that at least one is within the detectable region of the consequential hot spot is discussed in this section.

The conceptual hot spot represents a hypothetical distribution of chemicals that results in an unacceptable risk on the Site. It is this minimally acceptable distribution that dictates the required sampling effort. Spacial continuity of contamination will not be relied upon. Although the following derivation portrays only one hot spot, the site normalized concentration distribution it represents may be the result numerous hot spots of various chemicals.

The probability (q) that any single randomly selected sample location will fall within the detectable area of the consequential hot spot is given by:

$$q = \frac{A_{DHS}}{A_{Site}} \quad (11)$$

The probability (α) that no samples will be collected from within the detectable region of the hot spot in n randomly located sampling events is given by:

$$\alpha = (1 - q)^n \quad (12)$$

Using equations 9 through 12, the minimum number of samples N required to assure with 1-α confidence that at least one sample will be collected from with in the detectable area of the consequential hot spot determined to

be:

$$N = \frac{\log(\alpha)}{\log\left[1 - \frac{3}{\mathbf{C_{max}}} \times (1 - \frac{\mathbf{C}_d}{\mathbf{C}_{\max}})^2\right]} \qquad (13)$$

Equation 13 is valid for conditions where N decreases as $\mathbf{C}_{max}$ approaches one. It has been shown in a previous paper that this condition is satisfied when the normalized detection limit is less than 4/9 (Gemperline, 1994). The detectable hot spot area must also be less than or equal to the Site area.

Example 1.

The following Site-specific example illustrates the use of the Equation 13. For ease of presentation, only a few of the actual Site chemical compounds are used in this demonstration. Also, toxicological based values of Rfd and slope factors have changed for many chemicals since this analysis was originally performed to identify Site sampling needs. The original values are used herein to illustrate how sampling needs for this specific Site were determined.

It is desired to have 95 percent confidence that contamination will be detected if a consequential hot spot is present (i.e., $\alpha = 0.05$). The variable values $t = 0.01$, $S = 200$ mg/day , $D = 10000$ mg/day , $Ss_i = 1$, $Ds_i = 0.1$ and $M = 16$ kg have been previously defined for long term exposure to carcinogens. The t value for short term exposure to acutely toxic chemicals is 1. The value of Ss_i for PCBs was assumed to be 0.3. The chemicals listed in Table 1 were discovered during early debris removal actions at the Site. Estimates of the expected maximum concentrations, detection limits, Rfds for acutely toxic chemicals, slope factors for carcinogenic compounds, representative *Rfd*s calculated for carcinogenic compounds, normalized detection limits, normalized maximum concentrations, and the sum of the normalized maximum concentrations are also presented in Table 1.

The bolded values in Table 1 were used to represent the variables $\mathbf{C}_d$ and $\mathbf{C}_{max}$ in Equation 13. $\mathbf{C}_d$ was selected as the maximum normalized detection limit for the listed chemicals and $\mathbf{C}_{max}$ was calculated as the sum of the individual chemical normalized expected maximum concentrations. Acutely toxic chemicals were considered independently of the carcinogenic chemicals. Using Equation 13, the minimum number of locations to be sampled, N, is calculated to be 52 for acutely toxic chemicals and 3630 for carcinogens. If a chemical is both acutely toxic and carcinogenic then the larger of the two values would represent the required minimum number of randomly

selected samples.

The expense of testing samples from this many locations for a large suite of chemicals made the calculated requirement impractical. A review of the information in the table reveals that it is the need to find a small hot spot of PCB that forces extensive sampling. It was decided that a low-cost screening method, having a detection limit of 10 mg/kg, would be used to find a small, high-concentration, consequential PCB hot spot. A less extensive program would be used to look for a larger consequential hot spots of other chemicals and low-concentration PCB's. To determine the sampling needs for the screening effort, the 10 mg/kg detection limit for PCBs was used to calculate the normalized detection limit, $\mathbf{C}_d$. $\mathbf{C}_{max}$ was calculated using just the expected maximum PCB concentration. With these values, Equation 13 yields an N of 3618. Sampling needs for the lesser program were determined using the values in Table 1, except that the maximum expected concentration for PCBs was set at 30 mg/kg (three times the screening detection limit[1]). For this program N is calculated to be 52 for acutely toxic chemicals and 37 for carcinogens. These values represent a much more manageable number of sample locations than originally estimated.

Example 2.

It was desired to make independent decisions for subdivided Site areas. The investigation was to focus on specific drainage areas and areas thought to have the greatest level of contamination. Consequently, the Site was subdivided into 13 AOCs as shown on Figure 1. The background site was similarly subdivided. Furthermore, it was desired to put a greater investigation effort into AOCs prefixed by H (high probability of contamination) than those designated by L (low probability of contamination). The objective, to collect adequate data for risk assessment, remained paramount. Weighting the sampling effort towards the H-AOCs, in effect, results in less risk of missing a consequential hot spot in these areas. This increased confidence that contamination that presents a threat to human health will not be missed in the H-AOCs permits a greater risk to be taken for missing a consequential hot spot in the L-AOCs. A slight modification in the derivation of Equation 13 (retaining HI as a variable rather than setting it to 1) permits appropriate calculations.

$$N = \frac{\log(\alpha)}{\log(1 - \frac{3 * HI}{C_{max}} \times (1 - \frac{C_d}{C_{max}})^2)}$$

[1]The author demonstrates in reference (Gemperline, 1994) that the use of three times the higher detection limit to represent the maximum concentration of the less aggressive phase of a two phased program is appropriate.

(14)

The minimum number of locations required to be sampled in each subdivided area was determined using this equation. In doing this, care was taken to assure the *HI* for the entire Site remained 1. This was accomplished such that the sum of the *HI*s for individual AOCs, weighted to reflect their portion of the total area, equals 1. This limitation is described by:

$$\sum_{k} (HI_k \times \frac{A_K}{A_{Site}}) = 1 \qquad (15)$$

Where:

HI_K = a selected value that is inversely proportional to the desired level of sampling effort for the Kth AOC (or Kth group of AOCs).

A_K = area of the Kth AOC (of Kth group of AOCs).

It was decided that, more often than not ($\alpha = 0.5$) a consequential hot spot should not be missed when utilizing a sampling effort in the H-AOCs that is four times greater than the L prefixed. The same analysis as used in example 1 was performed, except Equation 14 was used to calculate N instead of Equation 13. The values of HI_k which were selected by trial and error, as well as the resulting calculated values of N, are presented in Table 2. Similar calculations were made for the PCB screening and less extensive programs described in the first example. These calculations indicated the need to screen for PCBs at 5293 locations in H-AOCs and 983 locations in L prefixed AOCs. The minimum number of locations to be sampled in the less extensive program was calculated to be 57 in the H-AOCs and 10 in the L- prefixed AOCs.

The following plan , which has fewer PCB screening locations than desired, was formulated and implemented with subsequently described contingencies. PCB screening was conducted in H-AOCs at points on a 25-ft grid. Figure 3 illustrates these sampling locations in AOC H1. This resulted in approximately 1030 screening samples. Similarly, PCB screening was conducted at locations on a 50-ft grid in L prefixed AOCs. This resulted in an additional 570 screening samples. It was also decided that samples collected at 100-ft grid point locations in H-AOCs and 200-ft grid point locations in H-AOCs would be analyzed for a large suite of chemicals, including those in Table 1. Figure 4 illustrates these sampling locations in AOC H1(61 locations for H-AOCs and 35 for L-AOCs). Samples were collected in similar size background areas at the same density as the H-AOCs.

The planned number of PCB screening locations was considerably less than the desired number of sampling locations. Testing all locations was cost inhibiting. Also, the number of locations receiving a detailed analysis in the less aggressive program was marginally adequate to detect any consequential contamination. The planned program also did not assure that the average Site concentration would be known with the confidence needed for risk assessment. Therefore, it was decided that a redundant program utilizing composite sampling would be designed to assure that consequential hot spots are adequately represented in the Site average concentration. This redundancy also provided a mechanism to evaluate the appropriateness of chemical distributions hypothesized from data collected at discrete locations.

Estimating Composite Sampling Needs for Human Health Risk Assessment

Representing a consequential hot spot in estimates of the average concentration for the Site was the goal of the composite sampling program. The hypothesized distribution of chemicals within the hot spot was previously discussed and is illustrated on Figure 2. Its size is calculated using Equation 9.

It was decided that the average concentration of each AOC would be determined independently. This would be accomplished by collecting specimens at grid points within an AOC and compositing them into a single sample. Quadruplicate or octuplicate composite samples were created for each AOC and the background areas. This permitted the calculation of a confidence interval for the expected mean and allowed for statistical comparison to background. Using multiple samples also resulted in more manageable sample sizes.

A simple computer simulation was developed that modeled the consequential hot spot and the composite sampling process. This was used to determine the number of specimens required to reasonably assure adequate representation of the hot spot in composite samples. Adequate representation requires that enough specimens are collected from within the consequential hot spot to appropriately represent it in estimates of the average concentration. The simulation included 1000 repetitions of the composite sampling event, with the hot spot randomly relocated between each repetition. Histograms displaying the distribution of possible results provided a mechanism to evaluate the ability of the modeled specimen collection density to represent the hot spot in the average.

A statistical approach was also developed to estimate the required number of composite specimen collection locations (Gemperline, 1994). The derivation is similar to that leading to Equations 13 and 14. The only

difference is that Equation 12 was replaced with a more general binomial cumulative distribution function. This calculated the probability that a specified number of randomly selected locations would be within the hot spot area.

The implemented composite sampling program is described as follows. Four or eight composite samples were created to represent each AOC. Each composite was comprised of four times as many specimen collection locations as were proposed for discrete sampling locations. Specimens were collected from grid locations throughout the AOC, as illustrated for AOC H1 on Figure 5. Each symbol on this figure represents locations where a specimen was collected. All specimens represented by like symbols (e.g., all circles) were combined to form a composite sample. This resulted in quadruplicate representation of the AOC . A second set of four composite samples also was created for most AOCs. Hence, octuplicate representations were often created. Specimen collection locations for the second set were the midpoints of the previously used specimen collection locations. Cumulatively, the number of specimens collected for composites in most AOCs is either 16 or 32 times greater than the total number of discrete samples depending on the number of composite samples collected (i.e., 4 or 8).

Expectations

If less than half of an area is contaminated, any single randomly selected sampling location is more likely to be outside the area of contamination than within. For this reason, it is expected that the average chemical concentration, when calculated from data representing a few discrete locations, will under represent the average concentration more frequently than over represent it. This is demonstrated by the following computer simulation.

Assume the hot spot portrayed in Figure 2 covers 6 percent of the Site. Also assume that the maximum concentration of the hot spot is 10,000 units. This makes the conceptual hot spot average concentration 3333. The corresponding Site average concentration would be 200. Since 6 percent of the area is contaminated, it is expected that a single sample, collected at random, has only a 6 percent chance of encountering any contamination. Consequently, if a sampling episode consisting of collecting one randomly located sample is repeated a large number of times, then contamination is expected to be encountered 6 percent of the time. A computer algorithm that emulated this scenario was executed 1000 times and resulted in a frequency distribution of concentration measurements shown on Figure 6. The high frequency of samples that exhibited concentrations less than 40 supports the expectation that approximately 94 percent of the modeled samples missed the hot spot. The histogram also correctly portrays that nearly all of the hypothetically contaminated area (hot spot) exhibited concentrations

greater than 360.

The computer simulation was repeated again, this time using 10 samples per sampling episode rather than 1, and plotting the distribution of the calculated average instead of the individual concentrations. A thousand episodes were simulated and the frequency distribution of results is shown on Figure 7. Statistical calculations suggest that 54 percent of the time the hot spot will remain undetected. This explains why most of the occurrences are represented in the 0 to 40 range of the histogram. Note that a majority of the sampling episodes result in calculated averages that are less than the modeled Site average concentration of 200 mg/kg.

The scenario was repeated with sampling episodes consisting of 30, 100, and 500 samples per episode and the resulting histograms of all these modeling events are shown on Figure 8, along with the previous two examples. It is observed that, as the number of samples increase, the distribution of the mean approaches the normal distribution. This is a demonstration of the Central Limit Theorem of Statistics. Because of this, the frequency of underestimating the mean approaches the frequency of overestimating the mean as the number of samples increases. This can also be seen on Figure 8. The frequencies never become equal until the entire Site surface is sampled. Therefore it is concluded that the estimate of the mean will more likely be low than high when less than half the Site area is contaminated. The converse would be true if less than half the Site is clean.

In general, at the Site, it was expected that less than half of the surface area of each AOC would exhibit chemical contamination. Consequently, based on the above discussion, it was expected that the average chemical concentrations calculated from the results of composite sampling would typically be greater than the average concentration calculated from the results of the much more limited discrete sampling. This was observed and is discussed in the following section.

Observations

The discrete and composite samples collected as described in the previous section represent redundant AOC characterization efforts. The results of both programs can be used to independently estimate mean chemical concentrations. Lead and PCBs were two contaminants commonly found in AOCs. These will be used to demonstrate the usefulness of the redundant programs.

The previously described expectations are demonstrated in the results of the Site investigation. Tables 3 and 4 list the calculated average concentrations for lead and total PCBs. Also provided on these tables in

parentheses and brackets are, respectively, the number of samples used in the calculation of the averages, and the total number of specimens used in composite samples. Not all AOCs exhibited measurable PCB concentrations and values are only presented for AOCs in which PCBs were detected. PCBs detection limits were typically less than 1 mg/kg. Half the detection limit was used to represent concentrations in samples with immeasurable levels of PCBs. Lead concentrations were measurable in all samples.

As expected, averages calculated from composite samples are typically greater than those calculated from discrete samples. Sometimes this is a large difference. Large differences imply that the unbiased discrete sampling effort did not appropriately represent contamination within the area. Occasionally only one composite sample in an AOC would encounter contamination. This indicates that contamination is spatially noncontinuous, i.e., located in hot spots.

The PCB screening data was not used in the previous comparison, yet provides considerable insight into the distribution of PCBs on the Site. The spatial distribution of total PCBs was estimated by cokriging the soil screening and the laboratory data from the discrete sampling program. A typical result is portrayed graphically on Figure 9. The dark areas fringed by lighter areas indicate the greatest PCB concentrations. To evaluate the appropriateness of this representation, a computer algorithm was created to simulate composite sampling of this modeled Site contamination. Similarity between the computer simulated measurements and direct measurements would bolster confidence of a good geostatistical representation of PCB distribution. The computer estimated means and variances were calculated and compared with the actual field composite test results.

When comparison revealed that the simulated average concentrations were higher than actual field composite values, it was concluded that the map generated by cokriging was portraying more surface PCB contamination than probably exists. Also, when the variance of the simulated composite samples were considerably less than the variance of real field composite test results, it was interpreted as the model indicating a more uniform distribution of contamination than truly exists.

Unlike field composite samples, the computer simulation provides estimates of chemical concentration for individual composite specimens. It was observed that occasionally only one or two simulated specimens accounted for nearly all PCBs encountered in AOC composites.

Summary and Conclusions

The objective of the Site surface soil sampling program was to, 1) reasonably assure that if no contamination is discovered at the Site then unacceptable contamination was not present, and 2) to reasonably assure that a hot spot presenting a threat to human health could be adequately represented for risk assessment purposes. Redundancy in the Site investigation provided a unique opportunity to evaluate the effectiveness of this sampling effort. Two independent programs, one using composite samples, and the other using discrete samples, were used to characterize the distribution of chemicals. It was demonstrated that the greater sampling density associated with composite samples typically yielded higher area average concentrations than was calculated using discrete sample data. Compositing multiple specimens into a single sample reduced the analytical effort typically associated with programs that utilize high discrete sampling density while appropriately representing potentially unacceptable contamination in estimates of the mean. It was demonstrated that composite samples can be used to evaluate the appropriateness of hypothetical contaminant distributions. Redundancy in the sample plan provided valuable insight into the effectiveness of sampling effort and promotes the continued use of composite samples in Site characterization.

REFERENCES

Gemperline, Mark. 1994. Surface Sampling to Detect Unacceptable Human Health Risk. *First International Congress on Environmental Geotechnics*. International Society for Soil Mechanics and Foundation Engineering and Canadian Geotechnical Society, 233-240.

U.S. EPA. May 1992. *Supplemental Guidance to Risk Assessment Guidance for Superfund (RAGS): Calculating the Concentration Term.* Publication 9285.7-081.

U.S. EPA. December 1989. *Risk Assessment Guidance for Superfund (RAGS), Volume I, Human Health Evaluation Manual (Part A). Interim Final.* EPA/540/1-89/002.

Table 1. Information Used in Example 1.

Chemical Name	Expected Maximum Concentration (mg/kg)	Analytical Detection Limit (mg/kg)	Rfd for acutely toxic compounds (mg/kg/day)	Slope factor for carcinogenic compounds (Unit risk per mg/kg/day)	Representative Rfd for Carcinogenic Compounds (mg/kg/day)	Normalized Detection Limit (Acutely Toxic Chemicals)	Normalized Detection Limit (carcinogens)	Normalized Expected Maximum Concentrations C_i (Acutely Toxic Chemicals)	Normalized Expected Maximum Concentrations C_i (Carcinogens)
PCBs	4000	0.05		7.70	1.30E-05		2.55E-01		3.52E+03
Beryllium	10	0.05	4.00E-04	4.40	2.27E-05	9.46E-03	1.65E-01	1.88E+00	9.43E+00
Antimony	100	0.05	4.00E-04			9.46E-03		1.88E+01	
Mercury	10	0.05	1.00E-04			3.79E-02		7.50E+00	
Aldrin	10	0.05	3.00E-05	0.17	5.88E-01	1.26E-01	6.38E-03	2.50E+01	2.44E+00
(C_d)*						**1.26E-01**	**2.55E-01**		
(C_{max})**								**5.31E+01**	**3.53E+03**

*conservatively selected as the maximum.

**the sum of the C_i.

Table 2. Tabular Presentation of Example 2 Variables and Results.

AOC	A_k (ft^2/1000)	A_{site} (ft^2/1000)	HI_k	Weighted HI	N (acutely toxic chemicals)	N (carcinogens)
H1 through H7	643	2062	0.5	0.16	24	5234
L1 through L7	1419	2062	1.22	0.84	10	972
Totals	2062			1.00	34	6206

Table 3. Comparison of Lead Subunit Averages from Point and Composite Samples

Subunit	**Area (ft^2/1000)**	**Average Concentration calculated from Point Samples** (mg/Kg) **with Corresponding Number of samples shown in Parentheses.**	**Average Concentration calculated from Composite Samples** (mg/Kg) **with Number of Samples in Parentheses and Total Number of Specimens in Brackets.***
H1	200	**2789** (18)	2145 (8) [640]
H2	266	1128 (27)	**1217** (8) [851]
H3	51	20 (3)	**37** (4) [82]
H4	49	122 (5)	**136** (8) [157]
H5	24	970 (4)	**6116** (8) [77]
H6	41	132 (4)	**171** (4) [66]
L1	167	415 (3)	**736** (8) [134]
L2	203	71 (5)	**83** (4) [81]
L3	245	132 (6)	**506** (4) [98]
L4	240	366 (6)	**372** (4) [96]
L5	203	194 (5)	**1463** (4) [81]
L6	198	239 (6)	**727** (8) [158]
L7	163	**857** (4)	856 (4) [65]
B1	368	42 (8)	**47** (4) [147]
B2	475	56 (10)	56 (4) [190]
B3	413	49 (9)	**59** (4) [165]

* The number of specimens is approximate.

Table 4. Comparison of Aroclor 1254 Subunit Averages from Point and Composite Samples

Subunit	**Area (ft^2/1000)**	**Average Concentration calculated from Point Samples** (mg/Kg) **with Corresponding Number of samples shown in Parentheses.**	**Average Concentration calculated from Composite Samples** (mg/Kg) **with Number of Samples in Parentheses and Total Number of Specimens in Brackets.***
H1	200	**12.9** (18)	5.3 (8) [640]
H2	266	16.4 (22)	**27.7** (8) [851]
H3	51	0.6 (3)	**1.3** (4) [82]
H4	49	1.2 (5)	**68.6** (8) [157]
H5	24	4.7 (2)	**43.2** (8) [77]
L1	167	0.6 (4)	**17.0** (8) [134]
L2	203	0.1 (5)	**1.0** (4) [81]
L3	245	1.5 (6)	**2.7** (4) [98]
L4	240	3.0 (6)	**3.1** (4) [96]
L5	203	1.9 (5)	**2.1** (4) [81]
L6	198	0.7 (5)	**1.7** (8) [158]
L7	163	**0.8** (4)	0.7 (4) [65]

* The number of specimens is approximate.

October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

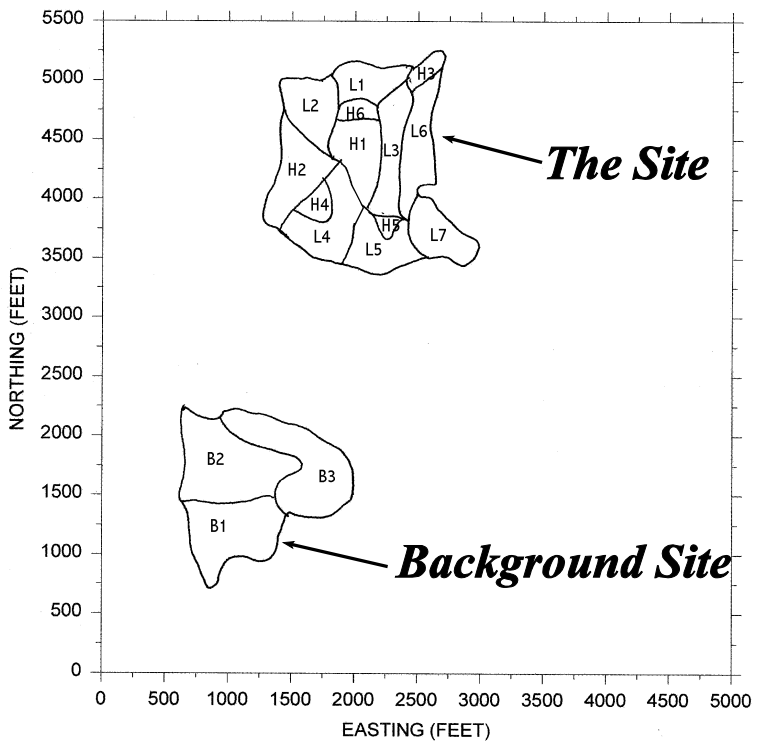

Figure 1. Representations of the Site and its associated Background Site.

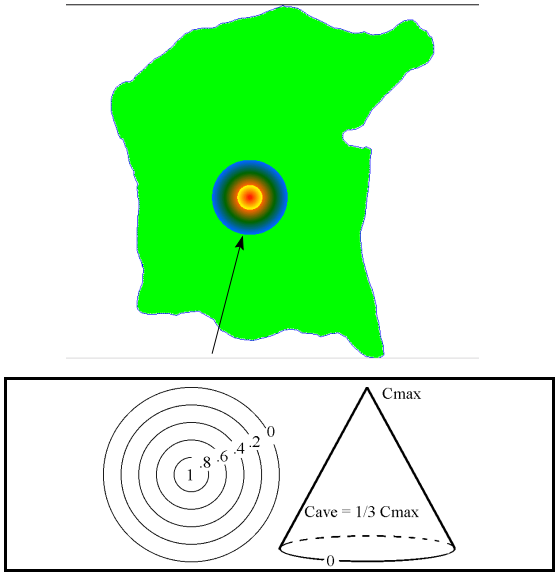

Figure 2. Conceptual Hot Spot Model Superimposed on Representation of the Site. Contours of Concentration Normalized with respect to C_{max}.

October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

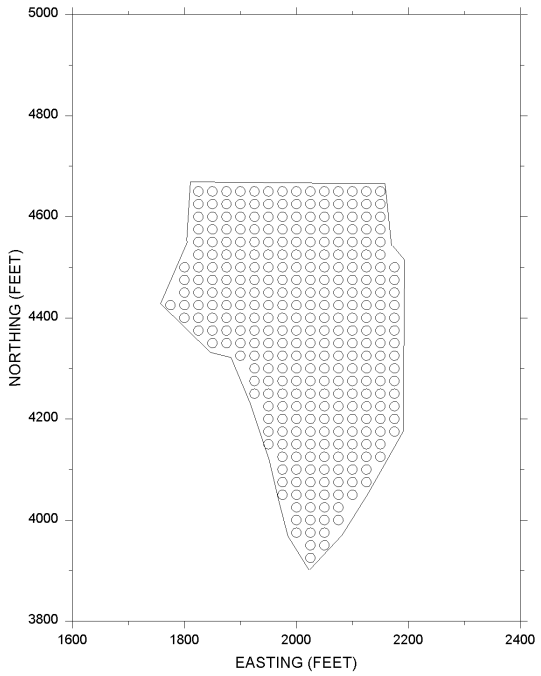

Figure 3. PCB screening locations in AOC H1.

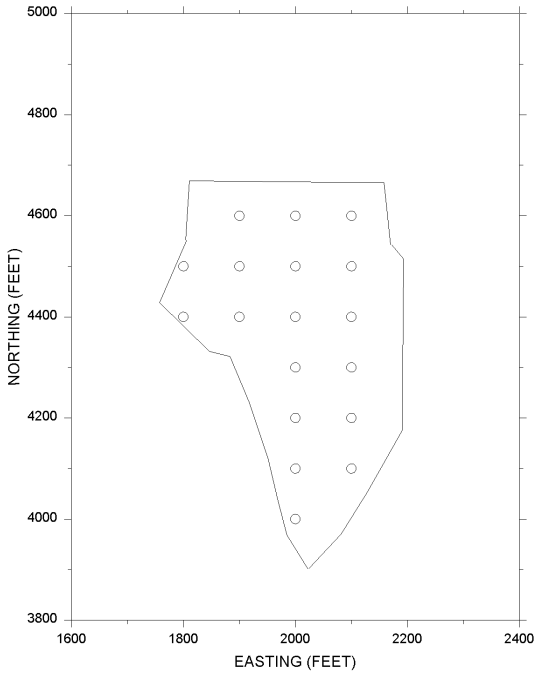

Figure 4. Discrete Surface Soil Sampling Locations in AOC H1.

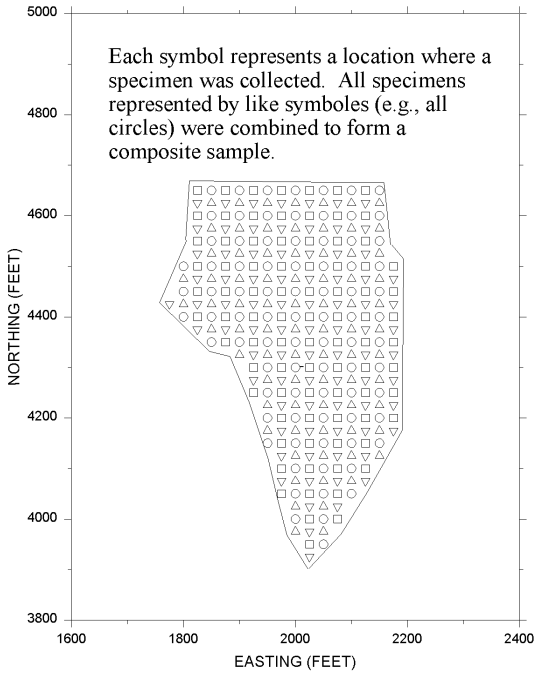

Figure 5. Composite Samples Specimen Collection Locations in AOC H1.

October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

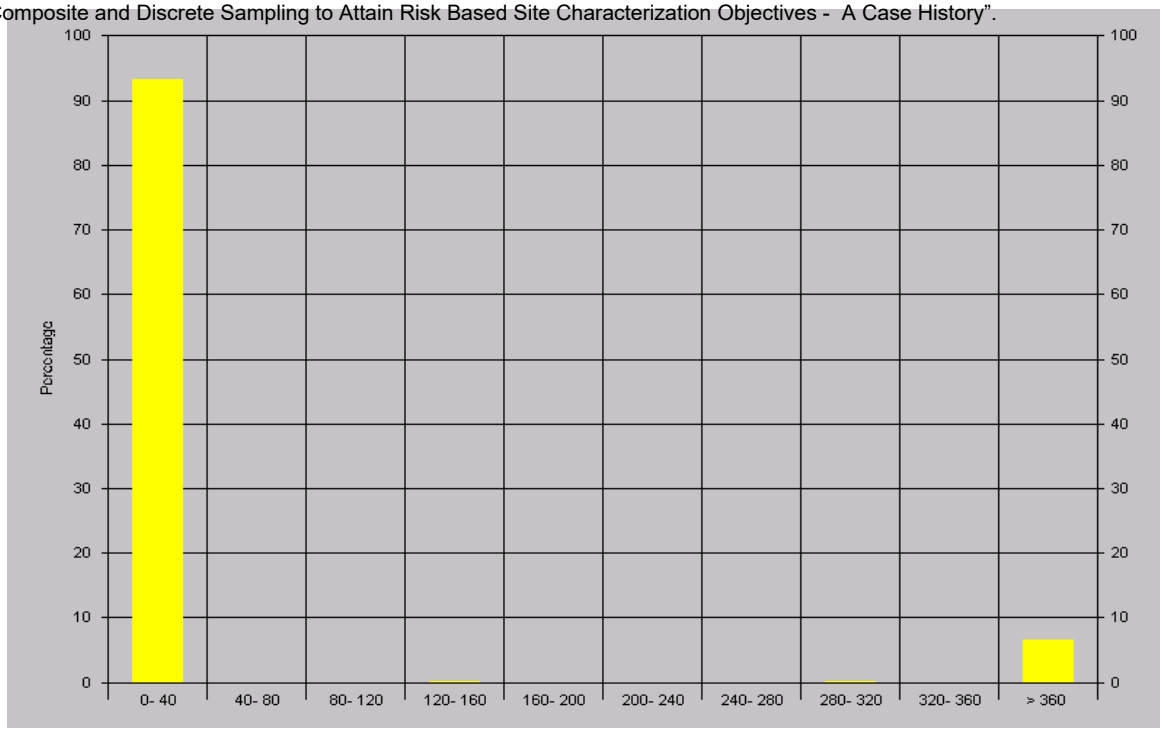

Figure 6. Percentage distribution of measured concentrations from 1000 sampling episodes consisting of 1 sample per episode.

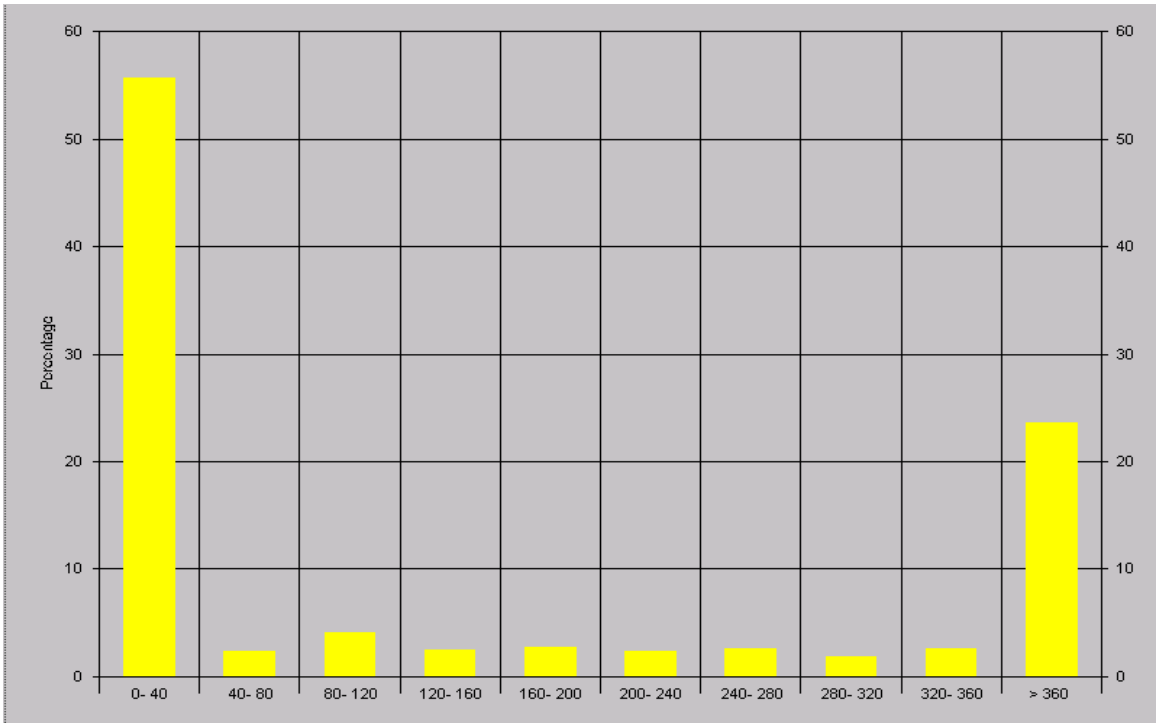

Figure 7. Percentage distribution of measured average concentrations from 1000 sampling episodes consisting of 10 samples per episode.

October, 1999: Amherst, Massachusetts, 15TH Annual International Conference on Contaminated Soils and Water, Paper and Presentation titled, "Composite and Discrete Sampling to Attain Risk Based Site Characterization Objectives - A Case History".

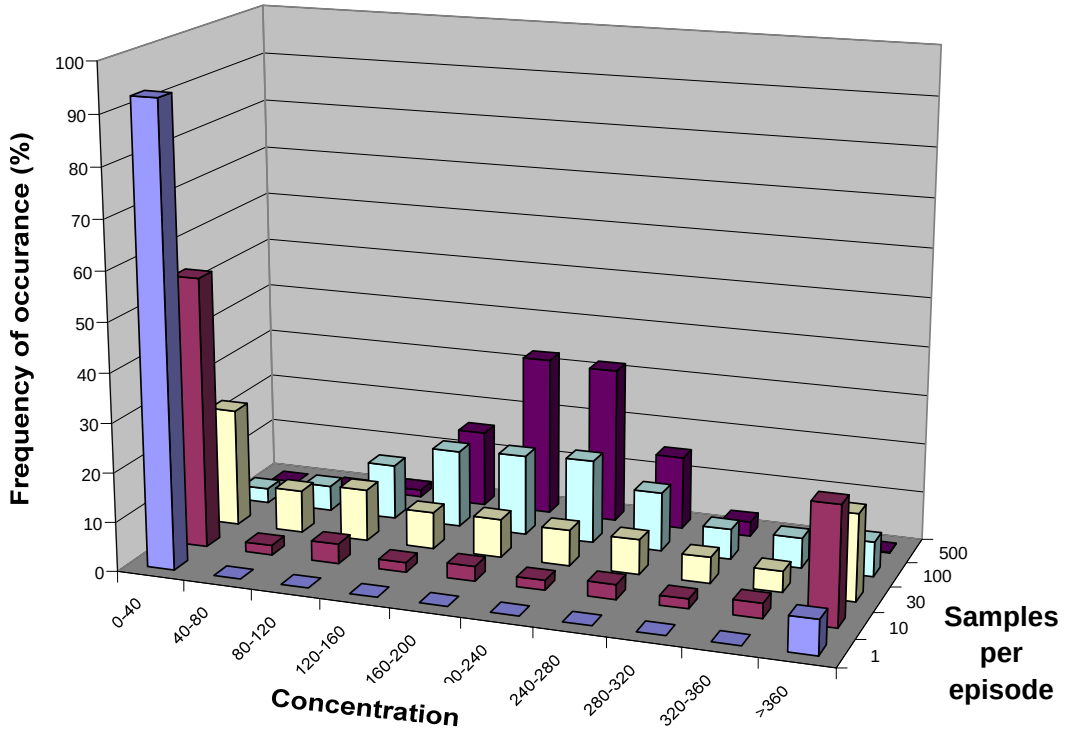

Figure 8. Percentage distribution of measured average concentration s from 1000 sampling episodes consisting of differing numbers of samples per episode.

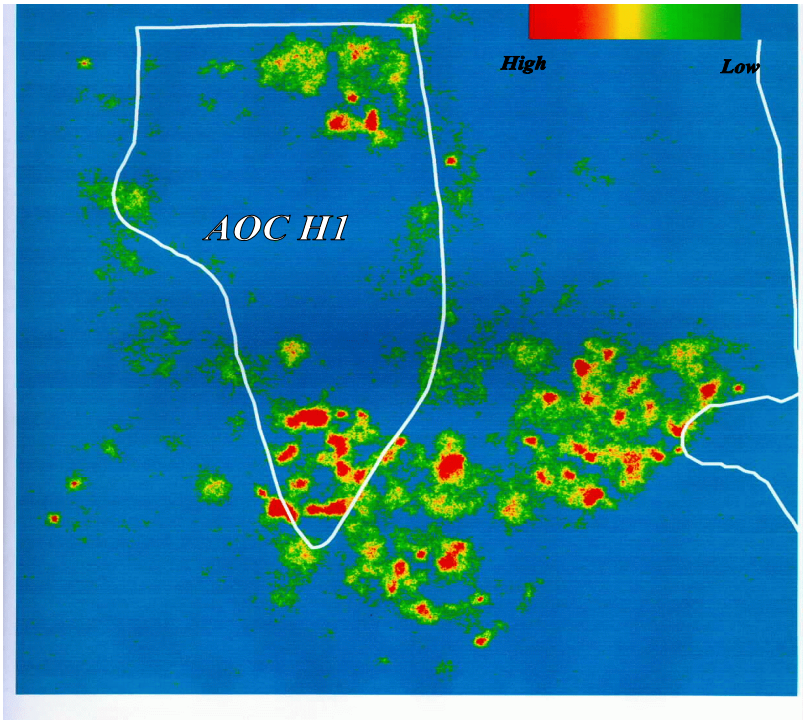

Figure 9. Hypothetical Distribution of PCBs.