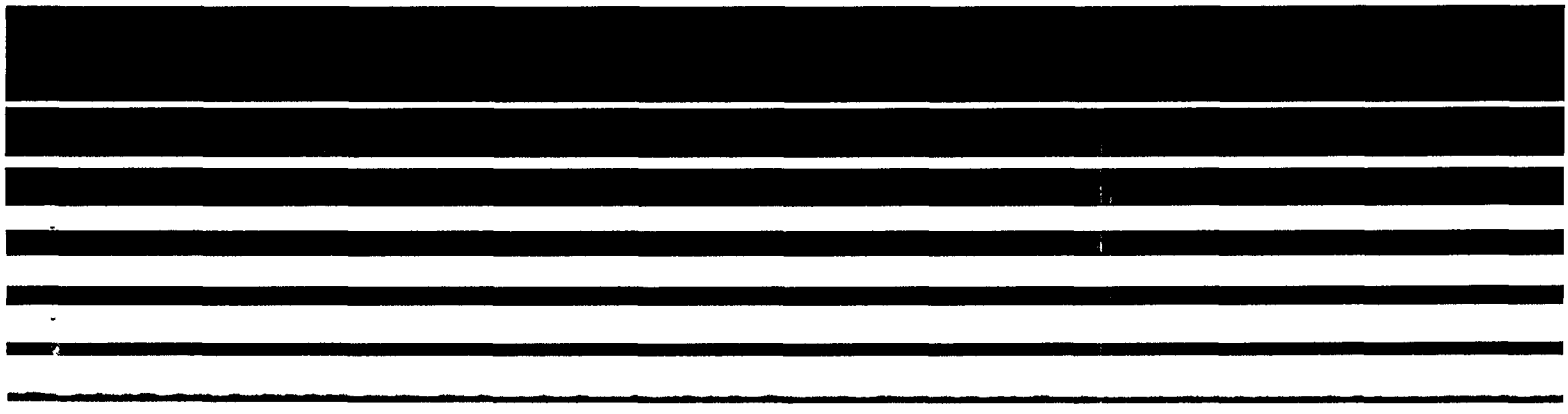
United States Environmental Protection Agency
Office of Air Quality Planning and Standards Research Triangle Park NC 27711
EPA-451/R-93-002 March 1993
Air/Superfund

EPA

AIR/SUPERFUND NATIONAL TECHNICAL GUIDANCE STUDY SERIES

Air Emissions from Area Sources: Estimating Soil and Soil-Gas Sample Number Requirements



DISCLAIMER

This report was prepared for the U.S. Environmental Protection Agency under Contract Number 68D00124, Work Assignment No. II-147. The contents are reproduced as received from the contractor. The contents do not necessarily reflect the views and policies of the Agency, neither does mention of commercial products or trade names constitute endorsement or recommendation for use.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION 1

INTRODUCTION

Soil sampling and soil-gas surveys are frequently used techniques to estimate air emissions from Superfund sites such as landfills and spill areas. In performing these surveys, it is important that the sampling strategy generate data that are an adequate statistical represention of the area source. These data are necessary in performing Air Risk Assessments at the confidence level stipulated in the Risk Assessment Guidance for Superfund (EPA 1989).

The purpose of this Manual is to provide guidance as to the necessary number of soil gas or soil samples needed to estimate air emissions from area sources. The Manual relies heavily on statistical methods discussed in Appendix C of Volume II of Air/Superfund National Technical Guidance Study Series (EPA 1990) and Chapter 9 of SW-846 (EPA 1986). These methods utilize the arithmetic mean as is specified in EPA Publication 9285.7-081, "Supplemental Guidance to RAGS: Calculating the Concentration Term".

If samples are taken over the entire surface of an area source using random sampling techniques, EPA experience shows that most of these data sets will be lognormally distributed. That is, some of the sample results will be either so much higher or lower than the mean of all samples that the plot of number of samples versus concentration will be significantly skewed from the normal or "bell shaped" distribution. Stated another way, a lognormal distribution suggests the area has one or more "hot or cold spots", or zones, in which the pollutant concentration differs significantly from the average concentration throughout the rest of the area.

The techniques in this manual are based on recognizing this inhomgeniety in the area, by observation or screening samples, before samples are taken. Each of the identified zones are then sampled, using random sampling techniques, and statistics calculated separately for each zone before combining the statistics to provide an estimate for the entire area. The assumption is that zoning can be effected such that each zone is reasonably normally distributed even when the overall area is lognormally distributed. If the zoning does not satisfy this assumption, the methods in this manual will fail.

It must be recognized that lognormal data cannot be "zoned" after the fact and analyzed by the techniques in this manual. It is extremely unlikely that such data would represent random samples for each zone, and, thus, the statistics would be biased. The techniques, and computer software, can however, be used to assist with the analysis of lognormally distributed data by the procedures given in the above referenced EPA publication. This application is discussed in Sections 3.3 and 4.3.

The statistical techniques presented may also be used to analyze other types of data and provide measures such as mean, variance, and standard

deviation. The methods presented in this Manual are based on small sample methods. Application of the methods to data which are appropriately analyzed by large sample methods or to data which is not normally distributed will give erroneous results.

Section 2 provides a brief overview of concepts and statistical techniques used in this document. Section 3 provides step-by-step procedures which can be used for desk-top calculations. Section 4 provides an overview of the computer software which accompanies this Manual. The software can be used to perform all procedures described in Section 3.

SECTION 2

BASIC CONCEPTS

2.1 DIVIDING THE AREA INTO ZONES

The number of samples that must be obtained to estimate the mean concentration of an area is strongly dependent on the heterogeneity of chemical distribution (for constant confidence and confidence interval). Thus, for an area with uniform chemical distribution, very few samples would be needed to provide good characterization. Conversely, areas with widely varying concentrations could require a great number of samples.

For areas with non-uniform distribution of chemical contamination, the total number of samples required for adequate characterization can be dramatically reduced by subdividing the area into zones with similar contamination levels. This situation is commonly encountered at Superfund sites. Such areas may be identified by variations in vegetation stress, area source records, or results of preliminary screening.

The maximum benefit in sample number reduction is obtained by defining zones within the area such that the concentrations within any particular zone are as uniform as possible. As many zones as is practical may be defined to accomplish this objective. Zones do not have to be of similar size or shape. The area of each zone must be determined.

2.2 RANDOM SAMPLING

To use the statistical methods in the Manual, it is necessary that the locations to be sampled within each zone be selected in a random manner. Random does not imply haphazard. One way haphazard sampling may occur is when sampling points are simply selected based on personal judgement that the points selected are random. There is no assurance with this procedure that sampling points are not selected with a conscious or subconscious bias. Samples collected from points selected haphazardly may not be statistically representative of the area.

Random sampling requires a plan to ensure that each potential sampling location has an equal chance of being selected. One method (but not the only method) to accomplish this is as follows. Define an imaginary square grid for each zone. The grid may be marked off in feet, yards, meters, or whatever unit is convenient so long as the number of points where grid lines intersect exceeds the estimated number of samples required by at least a factor of two (allowing additional samples to be collected if necessary). Neither the directional orientation of the grid nor the selection of the reference point from which all grid lines are measured are significant (grids should be established independently for each zone.) Number the grid intersections sequentially from 1 to X. The actual points on the grid to be sampled are selected using a table of random numbers (available in any book on statistics). No grid point may be selected for sampling more than once.

2.3 VARIANCE, STANDARD DEVIATION, AND COEFFICIENT OF VARIATION

The variance is simply the average of the squared deviations from the mean of the data. For the small sample methods used in area source sampling, the sum of the squared deviations are divided by the total number of samples minus one to obtain this average.

The standard deviation is the square root of the variance. Approximately two-thirds of all sample data will fall within a range defined by the mean ± standard deviation. Large standard deviations are indicative of highly variable concentrations within the area sampled and/or an inadequate number of samples.

The coefficient of variation (sometimes referred to as relative standard deviation or precision) is the standard deviation divided by the mean of the samples and multiplied by 100%. By expressing the standard deviation as a percentage of the mean, it is generally easier to grasp just how variable the data are. For example, if the coefficient of variation is 20 percent, two-thirds of the sample data fall within 20 percent of the mean.

The variance is calculated from:

$$S_K^2 = \frac{1}{n_K - 1} \left[\sum_{i=1}^{n_K} (X_i^2) - n_K \overline{X_k}^2 \right] \qquad (2\text{-}1)$$

where,

S_K^2 = variance for zone K
n_K = number of samples from zone K
X_i = value of sample i from zone K

$\overline{X}_k$ = mean of sample values from zone K

The standard deviation is the square root of the variance:

$$S_K = \sqrt{S_K^2} \qquad (2\text{-}2)$$

The coefficient of variation is calculated from

$$CV = \frac{100\% \, (S_K)}{\overline{X_k}} \qquad (2\text{-}3)$$

2.4 CONFIDENCE LIMITS AND THE CONFIDENCE INTERVAL

By definition confidence limits are the limits between which the true mean will fall with a specified probability (or confidence). For example, if calculations are made on a particular set of data from an area source at the 95 percent confidence level, a range is established within which the true mean of the area source concentration will fall 95 percent of the time. The upper 95 percent confidence limit is thus simply the highest mean value from this range. Note that because the true mean could just as likely be below the lower confidence limit as above the upper limit, the probability that the true mean might exceed the 95% UCL is only 5%/2 = 2.5 percent.

The confidence interval is the range of possible values for the true mean which lie between the upper limit and the lower limit. It shows how small or wide is the range of possible values for the true mean based on the sampling data collected. The confidence interval becomes smaller as the variability in the sample data becomes less and as the number of samples increases. Note that the value of the upper confidence limit is dependent on both the mean of the sample data and the confidence interval.

The impact of this relationship is simple and straight forward:

- The 95% upper confidence limit can be calculated for <u>any</u> data set containing more than one data point.
- The value at the 95% UCL is dependent on the confidence interval.
- The confidence interval is dependent on the variability in the sample data and the number of samples.

The number of samples to collect cannot be meaningfully specified without also specifying the confidence interval that is acceptable for the particular purpose. If it is only of interest to know that the true mean area concentration is no more than an order of magnitude below the 95% UCL, a large confidence interval is appropriate and few samples need be collected. Conversely, if it is of interest to know the true mean area concentration is no more than 10 percent below the 95% UCL, a small confidence interval is appropriate and many more samples would be required.

The 95% confidence limits for a zone (i.e., k) are calculated from:

$$95\%\ LCL_k = \overline{X_k} - T_{0.05}\ \frac{S_k}{\sqrt{n_k}} \qquad (2\text{-}4)$$

$$95\%\ UCL_k = \overline{X_k} + T_{0.05}\ \frac{S_k}{\sqrt{n_k}} \qquad (2\text{-}5)$$

Note that tabulated values for the Student's t-distribution are for the sum of the probabilities that the true mean could be either greater than the upper confidence limit or less than the lower confidence limit (i.e.,a two-tailed test). Thus, the tabulated t-values for a 95% confidence interval (0.05 probability) are the same as those for a one-tailed test with a 0.025 probability. Thus, if is only required to know with 95% confidence that the true mean does not exceed the sample mean (i.e., a one-tailed test), tabulated values for $t_{0.1}$ should be used.

The confidence interval is:

$$CI_k = 95\%\ UCL_k - 95\%\ LCL_k \qquad (2 - 6)$$

2.5 COMBINING ZONE DATA INTO AREA DATA

After adequate data have been obtained for each of the zones in the source area, the zone data must be combined to represent the overall area. These data are combined based on the weighted areas of the zones as follows:

The overall mean is calculated from:

$$\overline{X} = \sum_{K=1}^{Z} W_K\ \overline{X_K} \qquad (2\text{-}7)$$

where,

W_K = area zone K divided by total area

The overall variance is calculated from:

$$S^2 = \sum_{K=1}^{z} W_K \, S^2{}_k \qquad (2\text{-}8)$$

The overall standard deviation is calculated from:

$$S = \sqrt{S^2} \qquad (2\text{-}9)$$

And the overall 95% UCL is calculated from:

$$95\% \; UCL = \bar{X} + T_{0.05} \frac{S}{\sqrt{n}} \qquad (2\text{-}10)$$

where,

n = total number of samples from all zones.

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 3

PROCEDURES SUITABLE FOR DESK-TOP CALCULATIONS

The procedures in this Section assume that the area source has been divided into zones within which it is expected that concentrations are reasonably uniform and that a plan to randomly select sample locations has been formulated as discussed in Section 2. All calculations in this Section are predicated on the assumption that the 95% UCL on the zone and area mean concentrations are desired. It is quite simple to make calculations at other confidence levels as will be indicated in the text.

3.1 ESTIMATING SAMPLE NUMBER REQUIREMENTS

3.1.1 Preliminary Estimate

An estimate of sample number requirements can be made if a reasonable estimate can be made for the zone variabilities. The number of samples required for any zone will depend on the variability of concentrations within that zone and on the confidence interval desired. This relationship can be expressed mathematically as:

$$\frac{n_K}{T^2_{0.05}} \geq \frac{CV_K^2}{P^2} \qquad (3\text{-}1)$$

where,

- n_K = samples required for zone K
- $T_{0.05}$ = 0.05 percentage point of a Student's t-distribution with n_K-1 degrees of freedom
- CV_K = coefficient of variation of data from zone K
- P = acceptable percent variation between sample mean and true mean at stated confidence level

This equation can be readily derived by substituting equation 2-3 into Equation 2-5 and defining P as:

$$P = 100\frac{(95\%\ UCL_k - \overline{X_k})}{\overline{X_k}}$$

[Note that $T_{0.05}$ indicates we are calculating the 95% confidence limits. By simply replacing $T_{0.05}$ with appropriate values from the Student's

t-distribution, calculations can be made at any desired confidence. Thus, for 90% confidence limits, use tabulated values for $T_{0.10}$, for 99% confidence limits use tabulated values for $T_{0.01}$. Tabulated values for the Student's t-distribution can be found in any general statistics textbook.] Table 1 contains values of n, $T_{0.05}$ and $n/T^2_{0.05}$ for use with Equation 3-1.

Because the parameter P specifies how close the 95% UCL should be to be to the sample mean, it can, and should, be specified before any sampling is done. Thus, if P is given a value of 20 percent, enough samples must be taken that the sample mean will be no more than 20 percent below the 95% upper confidence limit.

The value of the coefficient of variation, CV, will not be known until the samples are analyzed. If preliminary screening sample data are available, these data can be used to calculate CV and a preliminary estimate of n_K can be made using Equation 3-1.

Estimates of CV can also be made based on experience with similar sites. That is, if a previously investigated site has a waste disposal pattern similar to that suspected for the site to be investigated, it would be reasonable to expect similarities in the relative spatial distribution of chemical concentrations. Because CV is a measure of that distribution, it would be reasonable for the CVs of the sites to also be similar. If the CV has been calculated, or could be readily calculated from the data using the relationships of Section 2.3, it could be used in Equation 3-1 for planning purposes.

If no preliminary data for the site are available and no estimate for CV can be made, a crude estimate of the number of samples required in each zone can be obtained from:

$$N_k = 6 + y\sqrt{Zone\ area} \qquad (3\text{-}2)$$

where,
y = 0.15 for zone areas in square meters, or
y = 0.046 for zone areas in square feet.

This arbitrary relationship, which appears in Appendix C of Volume II of the NTGS document referred to in Section 1, assumes that CV increases as a function of the size of the contaminated zone. This assumption may or may not be true and significant over or under forecasts of sample number requirements may result from its use. For example, for a 20 percent difference between the 95% UCL and the sample mean, the corresponding coefficients of variation calculated using this relationship range from 19% for areas less than ~10 m^2 (~110 ft^2), i.e., those requiring no more than 6 samples, to 48% percent for areas up to ~28,000 m^2 (300,000 ft^2), i.e., those requiring up to 25 samples.

TABLE 1. RELEVANT VALUES BASED ON STUDENT'S t-DISTRIBUTION

Degrees Freedom	A = 0.05	n/T^2	n = No. Samples
1	12.706	0.012	2
2	4.303	0.162	3
3	3.182	0.395	4
4	2.776	0.649	5
5	2.571	0.908	6
6	2.447	1.169	7
7	2.365	1.430	8
8	2.306	1.692	9
9	2.262	1.954	10
10	2.228	2.216	11
11	2.201	2.477	12
12	2.179	2.738	13
13	2.160	3.001	14
14	2.145	3.260	15
15	2.131	3.523	16
16	2.120	3.782	17
17	2.110	4.043	18
18	2.101	4.304	19
19	2.093	4.566	20
20	2.086	4.826	21
21	2.080	5.085	22
22	2.074	5.347	23
23	2.069	5.606	24
24	2.064	5.868	25
25	2.060	6.127	26
26	2.056	6.387	27
27	2.052	6.650	28
28	2.048	6.914	29
29	2.045	7.174	30
30	2.042	7.434	31

NOTES:

- Degrees freedom is maximum number of variates that can be freely assigned before the rest are determined; generally, one less than the total number of samples.
- Values listed under A = 0.05 are $T_{0.05}$ values for a two-tailed test at 95 percent confidence for the listed degrees of freedom.
- Values for n/T^2 are provided for convenience when using Equation 3-1. They were obtained by dividing the square of the values in the column headed A = 0.05 into one more than the associated degree of freedom.
- Values under column headed n = number samples are the values used for n in the column headed n/T^2.

Regardless of the method used to estimate the number of samples, it is recommended that a few extra samples be collected and stored for later analysis just in case the sample variability is greater than estimated.

3.1.2 Example Applications

Example 1. Assume the mean concentration of zone 1 (1,000 m²) of the area source is needed within 20 percent of the UCL at the 95 percent confidence level and no preliminary site data is available. Based on past experience with similar sites, we guess most of the samples will be within 40 percent of the mean.

The parameters for use in the analysis are:

Zone area = 1,000 m^2
CV_k = 40
P = 20

Using Equation 3-1

$$\frac{n}{T^2_{0.05}} \geq \frac{(40)^2}{(20)^2} = 4$$

From Table 1, the smallest number of samples for which $n/T^2_{0.05}$ is greater than 4 is n = 18.

Using Equation 3-2:

$$N_k = 6 + 0.15\sqrt{Zone\ area,\ m^2}$$

the number of samples required for this zone is estimated as:

$$n = 6 + 0.15\sqrt{1,000\ m^2}$$

$$n = 11$$

Thus, experience indicates at least 18 samples should be taken from this zone.

Example 2. Make the same assumptions as example one except that four preliminary screening data points are available for the zone. These data are 30, 33, 41, 37 units.

The zone mean is:

$$\bar{X} = \frac{30 + 33 + 41 + 37}{4} = 35$$

The variance is:

$$S_1^2 = \frac{1}{4-1}\left[(30^2+33^2+41^2+37^2) - 4\left(\frac{30+33+41+37}{4}\right)^2\right]$$

$$= \frac{1}{3}\left[5{,}039 - 4{,}970\right]$$

$$= 23$$

The standard deviation is:

$$S = \sqrt{23}$$

$$= 4.8$$

The coefficient of variation is:

$$CV = \frac{100\ (4.8)}{35}$$

$$= 13.7 \text{ percent}$$

From Equation 3-1

$$\frac{n}{T^2} = \frac{(13.7)^2}{(20)^2}$$

$$= 0.469$$

and from Table 1,

$$n = 5$$

The preliminary data indicate the zone is fairly homogeneous and only five samples are required. Several additional samples should be collected and stored for later analysis in case the site is not as homogeneous as the preliminary data indicate.

Example 3. Assume an area source has been divided into two zones. Zone 1 is 100 m^2 and expected to contain high concentrations. Zone 2 is 400 m^2 and expected to have much lower concentrations. In each

zone 4 preliminary screening samples had been taken haphazardly (i.e., not using random selection techniques). The project required that we determine the 95% upper confidence limit for the area such that the mean concentration was within 20 percent of the 95% upper confidence limit.

The initial screening data was:

ZONE 1	Zone 2
16	5.2
18	3.6
15	4.4
13	3.3

Using Equation 3-2 we estimate the number of samples needed as:

Zone 1: $n = 6 + 0.15\ (100)^{1/2} = 7.5$
$n = 8$

Zone 2: $n = 6 + 0.15\ (400)^{1/2} = 9$
$n = 9$

Using Equation 3-1 we estimate the number of samples needed as follows:

Zone 1: Mean, $\bar{X} = (16 + 18 + 15 + 13)/4 = 15.5$

$$Variance, S^2 = \frac{1}{4-1}\left[\Sigma x^2 - 4\,\bar{X}^2\right]$$

$$= \frac{1}{3}\left[974 - 961\right]$$

$$S^2 = 4.3$$

$$Standard\ Deviation,\ S = \sqrt{4.3}$$

$$S = 2.1$$

$$Coefficient\ of\ Variation,\ CV = \frac{100\,(2.1)}{15.5}$$

$$CV = 13.5 \text{ percent}$$

$$\frac{n}{T^2_{0.05}} = \frac{(13.5)^2}{(20)^2} = 0.46$$

$$n \text{ (from Table 1)} = 5$$

Thus, we opt to collect and analyze 5 samples and collect and store three (3)extra samples just in case the samples are more variable than estimated.

Zone 2:

$$\text{Mean} = 4.1$$
$$S^2 = 0.73$$
$$S = 0.85$$
$$CV = 20.8$$

$$\frac{n}{T^2_{0.05}} = \frac{(20.8)^2}{(20)^2}$$

$$= 1.08$$

$$n = 7$$

Thus, for Zone 2 we opt to collect 7 samples for analysis and two extras in case the zone is more variable than we estimate.

3.2 ANALYZING COLLECTED DATA

In the proceeding section methods for making initial estimates were presented. In this section methods to verify an adequate number of samples were collected and calculation of the 95 percent upper confidence limit are given.

3.2.1 Analyzing The Data

As indicated in the previous Section, because we do not know before samples are collected and analyzed just how variable site concentrations are, the best we can hope for in making preliminary estimates of required sample numbers is that we estimated enough samples so that the final confidence interval will be small enough to satisfy the particular need. After the data are available calculations can be made to determine if, in fact, an adequate number of samples were collected and, if not, how many more are needed. In either case, the confidence interval and 95% upper confidence limit can be calculated for that data set.

Let us continue with the project described in Example 3 above. After collecting random samples from each zone, the analytical results are:

Zone 1	Zone 2
11.3	3.4
21.2	4.4
15.6	4.8
17.1	3.6
14.6	2.9
	3.1
3 samples held in reserve	3.9
	2 samples held in reserve

Calculations for Zone 1

Mean	=	15.96
S^2	=	13.11
S	=	3.62
CV	=	22.7 percent

$$\frac{n}{T^2_{0.05}} = \frac{(22.7)^2}{(20)^2} = 1.288$$

Thus, from Table 1, 8 samples should be analyzed. The three samples held in reserve are analyzed and yield results of 17.5, 14.3, 15.1.

For this set of 8 data points

Mean	=	15.8
S^2	=	8.31
S	=	2.88
CV	=	18.2

Because the new CV is less than for the set of 6 samples (i.e., 22.7), we know that no more samples are needed.

We can now calculate final statistics for Zone 1. The 95% upper confidence limit can be found, using Equation 2-5, from:

$$T_{0.05} = \frac{(UCL - \bar{X})\ (n)^{1/2}}{S}$$

where UCL is the highest value for the true mean for Zone 1 at the 95% confidence level. From Table 1 for 8 samples (7 degrees of freedom), $T_{0.05}$ = 2.365. Thus,

$$2.365 = \frac{(UCL - 15.8)\ (8)^{1/2}}{2.88}$$

$$95\%\ UCL = 15.8 + 2.4 = 18.2$$

This value is the 95% upper confidence limit for a data set with the limit within 20% or less of the mean. The mean for Zone 1 is within

$$(100)(2.4)/15.8 = 15.2 \text{ percent.}$$

Calculations For Zone 2

Mean = 3.7
S^2 = 0.47
S = 0.69
CV = 18.6 percent

$$\frac{n}{T^2_{0.05}} = \frac{(18.6)^2}{(20)^2} = 0.865$$

Since n from Table 1 is 6 and we analyzed 7 samples, no additional sample analysis is required.

The 95% upper confidence limit is then ($T_{0.05}$ for 7 samples, 6 degrees of freedom):

$$2.477 = \frac{(UCL - 3.7)(7)^{1/2}}{0.69}$$

UCL = 3.7 + 0.65
UCL = 4.4

and the UCL is within 100 (0.65)/3.7 = 17.6 percent of the mean for Zone 2.

Calculations for the Area Source

Using the relationships of Section 2.5, the mean value for the entire area is calculated as the weighted sum of the Zone means:

$$\text{Overall mean} = \frac{\text{area Zone 1}}{\text{Total area}} \times \text{Zone 1 mean} + \frac{\text{area Zone 2}}{\text{Total area}} \times \text{Zone 2 mean}$$

$$= \frac{100}{500} \times 15.8 + \frac{400}{500} \times 3.7$$

$\bar{X}$ area = 6.12

The overall variance is calculated similarly as

$$S^2_{area} = 0.2\ (8.31) + 0.8\ (0.47)$$

$$S^2_{area} = 2.038$$

The overall standard deviation is

$$S_{area} = \sqrt{S^2_{area}} = 1.428$$

The overall 95% upper confidence limit is ($T_{0.05}$ for the 15 total samples, 14 degrees of freedom):

$$2.145 = \frac{(UCL - 6.12)(15)^{1/2}}{1.428}$$

$$95\%\ UCL = 6.12 + 0.79$$
$$= 6.91$$

and the UCL is within 100 (0.79)/6.12 = 12.9 percent of the area mean.

Therefore, we report the statistics for the area source as:

mean	=	6.12 ± 0.79
variance	=	2.04
standard deviation	=	1.43
95% UCL	=	6.91

3.2.2 Analyzing Multi-Component Data

In the preceding sections, simple cases for one or two zones and only one contaminant were considered. In this section, a more complex case, representative of situations likely to be encountered, is presented.

Example 4. A landfill, approximately 12 acres, is divided into 5 zones based on an inspection that revealed leachate seeps, eroded covers, fracture traces, vegetation breaching the cap, etc. There are three compounds of interest at the site: vinyl chloride, TCE, and PCE. A varying number of preliminary samples have been taken in each zone for each of the compounds. We wish to determine whether these samples are adequate to calculate the 95% UCLs for each zone and the overall area under the constraint that the mean determined for each zone be within 20 percent of the 95% UCL for that zone. If they are not, we must determine the number of additional samples needed for each compound in each zone, collect and analyze those samples and determine final statistics for the zones and the landfill.

Given in Table 2 are the areas, estimated CVs based on prior investigations of similar sites, and the preliminary sample data.

Vinyl Chloride

For Zone 1, there is no sample data; only an estimate of the CV. Thus the number of samples required is estimated from Equation 3-1 as:

$$\frac{n}{T^2_{0.05}} = \frac{(20)^2}{(20)^2} = 1$$

and from Table 1, 7 samples should be taken (Equation 3-2 would estimate 17 samples should be taken).

Table 2. Preliminary Data for Example 4

Contaminant	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
Area, m^2	5,000	7,000	10,000	12,000	15,000
Estimated CV	20	20	35	25	30
Vinyl Chloride, ppm	No data	11.5	11.5 12.1	101.5 112.1 93.1	51.5 52.1 53.1 52.5
TCE, ppm	101 98 120 105	38 26 30	15 19	1,175 860	480 390 530 615
PCE, ppm	101 105 98 103	55 50 46 52	1,000 975 1,050 1,025	10.1 11.5 9.9 10.4	1.1 1.3 0.9 1.0

For Zone 2 there is one data point. As this is statistically meaningless (degrees of freedom equal zero), the number of samples required are also calculated from Equation 3-1. Because the estimated CV for Zone 2 is the same as Zone 1, 7 samples are also estimated to be required for this Zone. Thus, if the preliminary sample is considered a valid sample, only six more should be needed.

For Zone 3, there are two data points. For these data, the mean is 11.8, the standard deviation is 0.42, and the CV is 3.56%. From Equation 3-1, $n/T^2_{0.05}$ is 0.0317, and, from Table 1, 3 samples, or only 1 more than we already have, are required based on this preliminary data.

For Zone 4, there are 3 data points. For these data, the mean is 102.2, the standard deviation is 9.52, and the CV is 9.3%. From Equation 3-1, $n/T^2_{0.05}$ is 0.216, and, from Table 1, 4 samples, or only 1 more than we already have, are required based on this preliminary data.

For Zone 5, there are 4 data points. For these data, the mean is 52.3, the standard deviation is 0.673, and the CV is 1.3%. From Equation 3-1, $n/T^2_{0.05}$ is 0.004, and, from Table 1, 2 samples are required based on this preliminary data. Because we already have 4 samples, no additional samples are needed.

TCE and PCE

Following the same procedures, the estimates shown in Tables 3 and 4 can be obtained for these compounds.

Table 3. Preliminary Estimates for TCE

Parameter	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
Mean	106.0	31.33	17.0	1017.5	503.75
SD	9.76	6.11	2.83	222.7	94.11
CV	9.2	19.5	16.6	21.9	18.7
$n/t_{0.05}$	0.212	0.951	0.689	1.199	0.874
Samples Needed	4	7	6	8	6
Samples Taken	4	3	2	2	4
Additional Needed	0	4	4	6	2

Table 4. Preliminary Estimates for PCE

Parameter	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
Mean	101.75	51.25	1012.5	10.475	1.15
SD	2.99	2.99	32.27	0.714	0.129
CV	2.9	5.8	3.2	6.8	11.2
$n/t_{0.05}$	0.021	0.084	0.026	0.116	0.314
Samples Needed	3	3	3	3	4
Samples Taken	4	4	4	4	4
Additional Needed	0	0	0	0	0

Note that no additional samples are required in Zone 1 for TCE and no additional samples are required in any Zone for PCE.

Because in this case, the preliminary samples are also considered valid for final statistics, only the number of additional samples required are obtained and analyzed. The sample analytical results and Zone statistics are given in Table 5. Note that the precision for all zones and all compounds equals or is better than the stated goal of 20 percent. Thus, no additional samples are required and the Zone statistics meet our objective and can now be combined into area statistics. These statistics are calculated as described in Section 2.5 and Example 3 above. The results are given in Table 6.

Table 5. Final Zone Data for Example 4

Contaminant	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5
Vinyl Chloride, ppm	1076 982 1117 991 1215 905 1036	11.5 10.5 12.6 13.1 10.9 12.2 9.9	11.5 12.1 11.9	101.5 112.1 93.1 105.1	51.5 52.1 53.1 52.5
Mean	1046	11.53	11.83	102.95	52.3
SD	101.3	1.17	0.31	7.91	0.67
CV	9.7	10.1	2.6	7.7	1.3
95% UCL	1139	12.6	12.6	115.5	53.4
Precision	9.0	9.4	6.4	12	2
TCE, ppm	101 98 120 105	38 26 30 33 29 37 35	15 19 21 16 18 15	1,175 860 1050 976 990 1125 890 945	480 390 530 615 450 555
Mean	106	32.57	17.3	1001.4	503.3
SD	9.76	4.43	2.42	109.6	15.9
CV	9.2	13.6	14.0	10.9	15.9
95% UCL	121.5	36.7	19.9	1093	587.4
Precision	15	13	15	9.2	17
PCE, ppm	101 105 98 103	55 50 46 52	1,000 975 1,050 1,025	10.1 11.5 9.9 10.4	1.1 1.3 0.9 1.0
Mean	101.8	51.3	1012.5	10.48	1.15
SD	2.99	2.99	32.3	0.71	0.13
CV	2.9	5.8	3.2	6.8	11.2
95% UCL	106.5	56.0	1063.8	11.6	1.36
Precision	4.7	9.3	5.1	11	18

Table 6. Area Statistics for Example 4

Statistic	Vinyl Chloride	PCE	TCE
Mean	152.0	418.3	222.3
SD	32.6	70.2	14.7
CV	21.5	16.8	6.5
95% UCL	165.5	444.1	234.1
Precision	8.9	6.2	3.0

3.3 ANALYZING LOGNORMALLY DISTRIBUTED DATA

EPA Publication 9285.7-081 provides guidance for calculating statistics for lognormally distributed data. The methods previously illustrated for calculating the mean and standard deviation may be used to assist with calculating those statistics.

The first step is to calculate the natural logarithm of each data point (i.e., calculate ln(x), where x is the data point). The resulting values are referred to as transformed data.

Calculate the mean and standard deviation of the transformed data. Determine the H-statistic for the number of data points being analyzed, the standard deviation of the data, and the confidence level desired. Table 7 provides tabulated H-statistics for the 95 percent confidence level for a number of combinations of sample numbers and standard deviations. Note that it is necessary to use the H-statistic for the standard deviation calculated from the transformed data.

Calculate the upper confidence using equation 3-3:

$$UCL = e^{(\overline{x} + 0.5\, s^2 + sH/\sqrt{n-1})} \qquad (3\text{-}3)$$

where:

e = constant (base of natural log, equal to 2.718)

$\overline{x}$ = mean of the transformed data
s = standard deviation of the transformed data
H = H-statistic from Table 7

Table 7. H-statistics for Use With Lognormal Distributions and 95 Percent Confidence Limits*

S	DEGREES OF FREEDOM																		
	10	12	14	16	18	20	22	24	27	30	35	40	45	50	60	80	100	200	400
0.1	1.787	1 763	1.749	1.738	1.729	1.722	1 716	1.711	1.706	1.701	1.695	1 690	1 687	1.684	1.679	1.674	1.670	1.662	1.658
0.2	1.860	1.830	1.809	1.793	1.781	1 771	1.763	1.756	1.749	1.742	1.734	1.727	1.722	1.718	1.711	1.703	1.697	1.685	1.677
0.3	1.949	1.909	1 882	1.861	1 845	1.833	1.822	1.813	1.802	1.793	1.783	1.773	1.766	1.761	1.752	1.740	1.733	1.716	1.705
0.4	2.054	2.003	1.968	1.942	1.921	1.905	1.892	1 881	1.867	1.856	1 841	1.830	1.821	1.813	1.802	1.787	1.777	1.755	1.740
0.5	2 176	2.112	2 068	2 035	2.009	1.989	1.973	1 959	1.942	1.928	1 910	1 896	1 884	1.876	1 861	1.842	1.830	1 802	1.784
0.6	2 314	2.235	2.181	2 141	2.110	2 085	2 065	2.048	2.027	2 010	1.988	1.971	1 957	1.946	1.929	1.906	1.891	1.857	1.835
0.7	2.466	2.371	2 306	2.258	2.221	2.191	2.167	2.147	2.122	2.102	2 075	2 055	2.038	2.025	2.005	1.977	1.960	1.919	1.892
0.8	2.632	2.520	2.443	2.386	2.342	2.307	2.279	2 255	2.225	2.202	2.171	2.146	2 127	2.112	2.088	2.056	2.035	1.988	1.957
0.9	2 810	2 679	2 589	2.523	2.472	2.432	2 399	2.371	2.337	2.310	2.273	2 246	2.224	2.206	2.178	2.141	2.117	2 062	2.027
1.0	2.998	2 848	2 744	2.669	2 611	2 564	2 526	2.495	2.456	2.423	2.383	2 352	2.327	2.306	2.275	2.232	2 205	2 143	2.102
1.25	3.500	3 300	3 163	3 062	2 984	2 923	2.873	2.830	2.779	2 737	2 682	2 641	2 607	2.580	2.538	2.483	2.447	2.364	2 310
1.50	4.033	3.784	3 612	3.485	3 388	3.311	3.248	3.195	3.130	3.077	3.008	2 956	2.915	2.881	2.838	2.758	2.713	2 609	2.542
1.75	4 587	4.288	4 081	3 929	3.812	3.719	3.643	3.579	3.501	3 437	3.355	3.292	3.241	3.200	3.136	3.052	2.997	2.872	2.791
2.00	5.154	4.805	4 564	4.387	4 251	4.141	4.052	3 977	3.886	3.812	3 715	3 640	3 582	3.533	3.458	3 359	3.295	3.148	3 053
2.50	6 312	5.866	5.557	5.328	5.153	5.013	4 898	4 802	4.683	4 588	4.463	4.367	4.290	4.228	4.131	4.003	3.920	3.729	3.605
3.00	7.489	6.947	6.570	6.293	6 078	5.907	5.766	5.649	5.504	5.388	5.234	5 117	5.023	4.947	4.828	4.671	4 569	4 334	4.183
3.50	8.677	8.039	7.596	7.269	7.016	6.815	6 649	6 510	6.340	6 201	6 020	5 881	5.771	5.681	5 540	5.354	5.233	4.956	4.776
4.00	9.872	9.140	8 630	8.254	7.963	7.731	7.540	7.380	7.184	7.024	6.816	6.656	6.528	6 424	6.262	6.047	5.908	5.588	5.380
4.50	11 07	10.24	9 669	9 244	8.916	8.652	8 437	8 257	8.034	7.854	7 618	7.437	7 293	7.174	6 991	6.747	6.590	6.227	5.991
5.00	12 27	11.35	10 71	10.24	9.872	9 579	9.338	9.137	8.889	8.688	8 424	8 222	8.061	7.929	7.725	7.453	7.277	6 871	6.608

* - Adapted from Abramowitz, M and I.A Stegun, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical tables, Applied Mathematics Series 55, National Bureau of Standards, U.S. Government Printing Office, Washington (1964). Similar tabulations may be found in many standard textbooks, including: Gilbert, R.O., Statistical Methods for Environmental Pollution Monitoring, New York, NY, 1987.

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 4

USING THE COMPUTER SOFTWARE

The calculations of Section 3 may also be made using the computer software accompanying this manual. The algorithm was developed using IF® database management system. It is not necessary to have this software to run the program. The algorithm has been compiled and should be executable on any IBM compatible computer. The reports generated are written directly to the printer in an ASCII format at the time of creation. Disk copies of the reports are not created. The reports should print on all printers.

The software has been named Area Source Analysis Program (ASAP). There are three files basic to the program:

ASAP.EXE
FOXPRO.ESL
FOXPRO.ESO

These files must not be deleted or altered. ASAP.EXE is fairly small but contains all the algorithm code. The two IF® files (FOXPRO.ESL and FOXPRO.ESO) are larger but contain only file structure and display information. Because ASAP is provided with the examples in this manual already programmed to give the first time user an easy orientation, several additional files have been created. The size of these files will change as additional data is entered or deleted. There should be three Site_Cat files, two Zones files and two Samples files. As supplied, the software occupies 1.4 kilobytes of disk space.

ASAP may be run from either the floppy disk or from a hard drive. However, because the execution from the floppy disk is quite slow (due to display screen generation), it is strongly recommended that it be installed on a hard drive.

4.1 INSTALLATION AND EXECUTION

4.1.1 Using a Floppy Disk

Place the read/write protect tab in the unprotect mode (tab on 3½ inch disks should be placed so that the hole is covered). Place the disk in the disk drive and change the computer prompt to that drive. Type in ASAP and press the Enter or Return key. It will take 30 seconds or more for the first screen to appear. The program will read and write to the floppy disk.

4.1.2 Installing on a Hard Drive

ASAP will perform most satisfactorily if installed on a hard drive. The files should be installed in a separate directory. (Consult with your system manager for installation on other than a personal use computer.)

The following assumes your hard drive is designated "C" and the floppy disk drive is "A".

At the C prompt enter:
MKDIR ASAP (It is not necessary to name the directory ASAP)
CD ASAP (To place the computer in the newly created directory)

Place the floppy disk protect tab in the PROTECT mode. Place the floppy disk containing the ASAP files in drive A. Enter **A:** so the computer will read the floppy disk placed in drive A.

Enter **Copy *.* C:** to copy all files from the floppy disk into the ASAP directory on the hard drive. Enter **C:** to transfer execution to the hard drive. Remove the floppy disk and put it in a safe place.

ASAP can now be run from the hard disk by typing in **ASAP** and pressing the enter or return key.

4.2 USING THE SOFTWARE

4.2.1 Functions Available at the Main Menu

The first screen that appears upon execution of ASAP is the Main Menu (see Figure 1). The options may be executed by pressing the number of the option, clicking on it with a mouse, or highlighting it with the up and down arrow keys and pressing the enter key. All options except number 6, Review Sampling Plan, are also available at different screens within the program.

Option 1. List all Available Sites -- Selection of this option takes you directly to the Site Catalog screen (Figure 2). This screen lists, in alphabetical order, all sites that have been entered into the program, the number of zones in each site, and the total area of all zones. Note that if the number of zones is zero for any site, it means either that no zone information was added or that the entered data has not been analyzed (Option 7) to generate a sampling plan or analyze screening or actual data. This screen is updated only when the analyze routine is executed. Thus, if site zone descriptions are added, edited or deleted, the screen will be incorrect until the new data is analyzed. Many functions are available directly from this screen by holding down the ALT key and pressing a second key. Use the up and down arrow keys to select the site for which information is desired.

Option 2. Add a New Site -- Selection of this option takes you to the "Add a New Site" screen (Figure 3). At this screen, enter the descriptive name for the site, the confidence level (e.g., 95%) desired, and the precision desired. As described in Sections 2 and 3, the latter two are key drivers for sample number requirements. Note that the program does not check for duplication and will accept multiple sites with the same name. The screen also requests whether or not you plan to input screening data. Neither selection prevents you from changing your mind later. The last entry on the screen requests the units for data entry. ASAP does not make unit

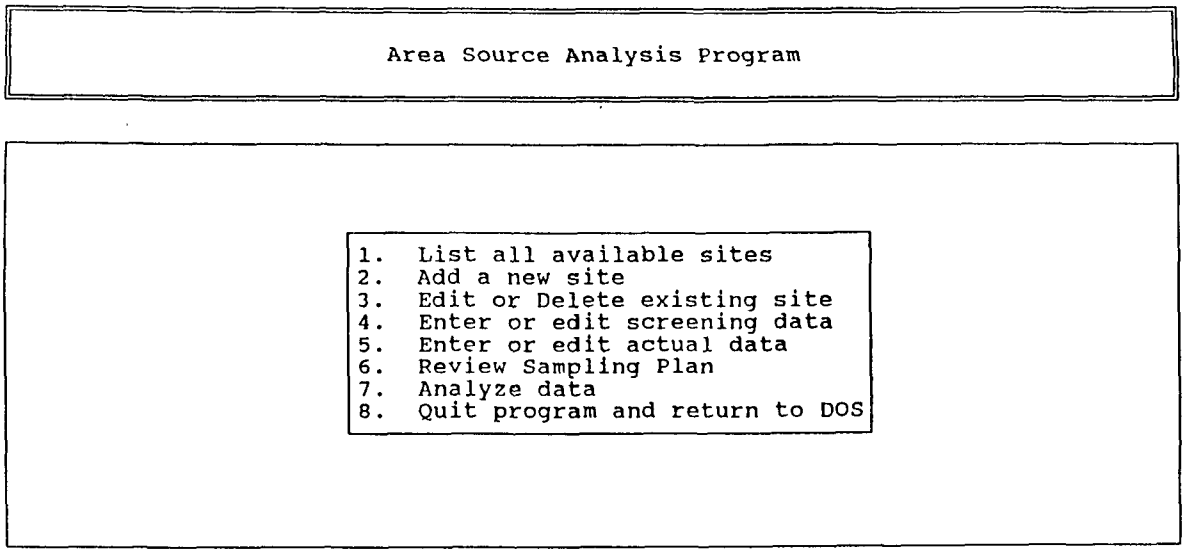


Figure 1 Main Menu Screen

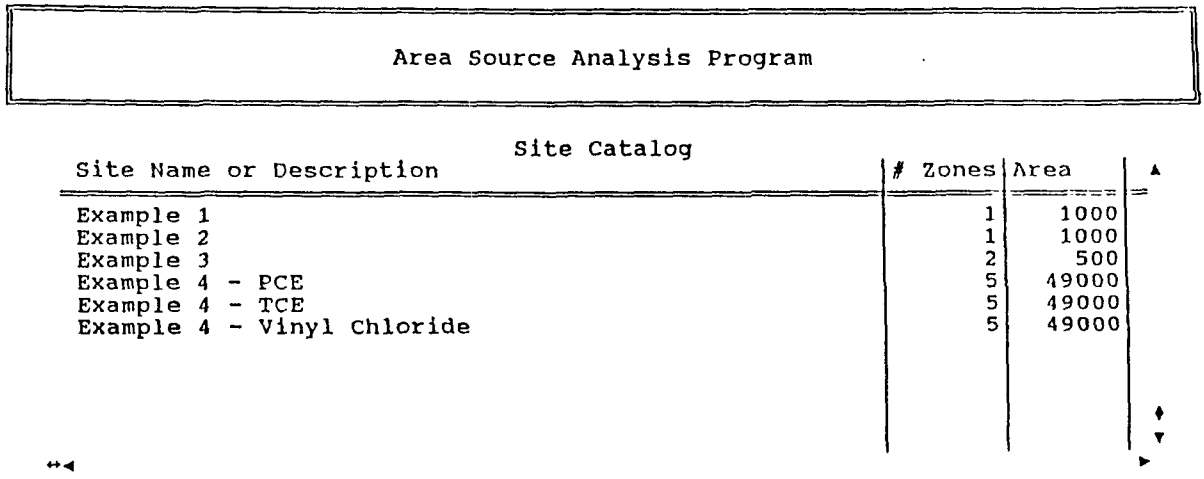
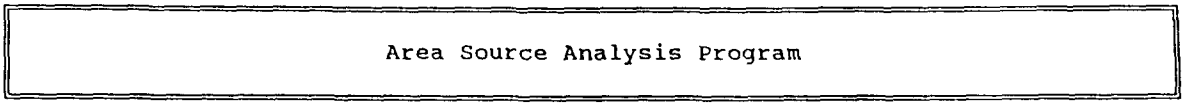


Figure 2 Site Catalog Screen



Add a New Site

Site Name or Description

Desired Confidence Level (80%, 90%, 95%, 99%) 0

Desired Precision (%) 0

Is screening data available? (Y/N) N

Enter the concentration units for all samples (ppm, ug/m², etc.)

Esc=Cancel Enter=Accept

Figure 3 Add a New Site Screen

conversions. It uses this field only for reporting purposes. After the last entry, a new screen (Figure 4) will appear for entry of zone information.

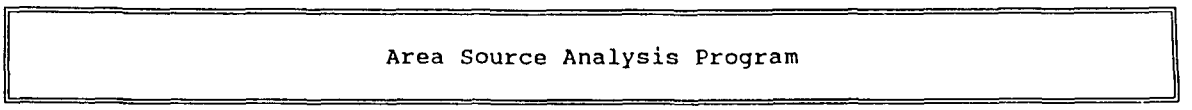
At the "Add a New Zone" screen, enter a descriptive label for one of the site zones and the area, in square meters, for that zone. Enter an estimate of the coefficient of variation for this zone if desired. If you leave this field blank, sample number requirements will be estimated based on area of the zone, or screening data if entered in the next steps. This screen will continue to appear until you press the ESC(ape) key indicating you have entered descriptions for all zones.

A Zone List screen (Figure 5) will now appear and will list all zones and zone areas for the new site. You may edit the entries, if incorrect, by using the arrow keys to select the zone and pressing the ALT and E keys. If you wish to enter screening data now for one or more zones, select the zone and press ALT-S. [**DO NOT** enter actual data (defined as the sample results you want to use for final site statistics) at this time even if available. Data entry will be much easier after a sample plan is generated.] If you do not wish to enter screening data, press ALT-C to return to the Site Catalog screen or ALT-M to return to the Main Menu.

Pressing the ALT-S keys takes you to the screening data entry screen (Figure 6). By default, there are 15 numbered positions for data entry. Type in the screening data and, using either the arrow keys or enter (return) key, move down the list until all screening data for that zone has been entered. You may move up and down the list to review or change entered data. Editing works in the insert mode rather than overwrite mode, therefore incorrect data must be deleted with the backspace or delete keys. Use the ALT-Z keys to return to the Zone List and select additional zones for screening data entry. Screening data does not have to be entered for all zones. After you have completed all entries, press ALT-C to return to the Site Catalog or ALT-M to return to the Main Menu. Data processing may be done from either screen (Option 7 returns directly to the Site Catalog).

Data processing to generate a sampling plan is addressed under Option 7.

Option 3. Edit or Delete Existing Site -- Selection of this option enters the Site Catalog screen. Select the site to be edited or deleted and press the appropriate key combination shown at the bottom of the screen. The delete selection removes all site data, including samples and zones. Before completing the deletion, the program prompts for confirmation. Selection of the edit option transfer to an edit screen (Figure 7) which allows you only to edit the overall site information (description, confidence level, precision desired, and units). However, individual zone descriptions can be added, edited, or the entire zone deleted, after pressing ALT-Z at the Site Catalog screen to access the zone list. Note that the site catalog does not automatically update if zones are added or deleted. Changes will be made to that screen only after data analysis (Option 7 or ALT-A in the Site catalog) has been executed. Also, individual screening or actual data for zones can edited from the zone list.



SITE: Example 4 - Vinyl Chloride

Add a New Zone

Zone Description

Area (square meters) 0

Estimated coefficient of variation (%) 0
Leave blank if using screening data

Esc=Cancel Enter=Accept

Figure 4 Add a New Zone Screen

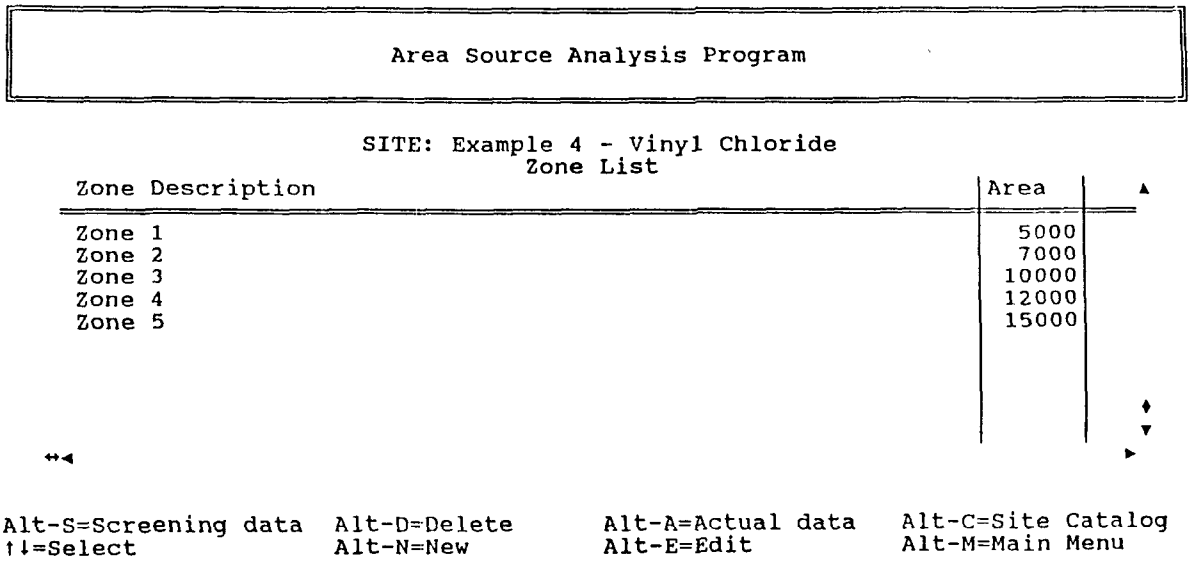


Figure 5 Zone List Screen

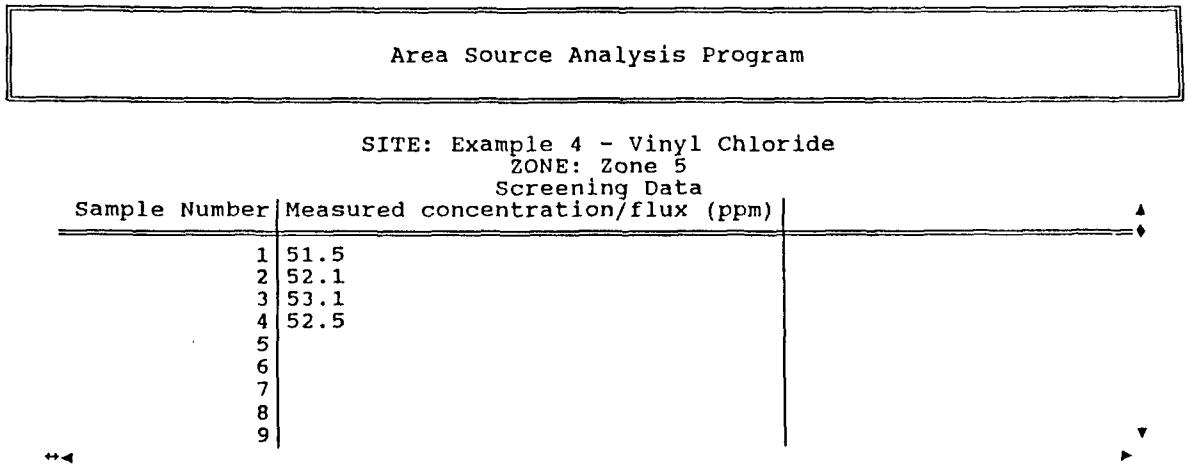


Figure 6 Screening Data Entry Screen

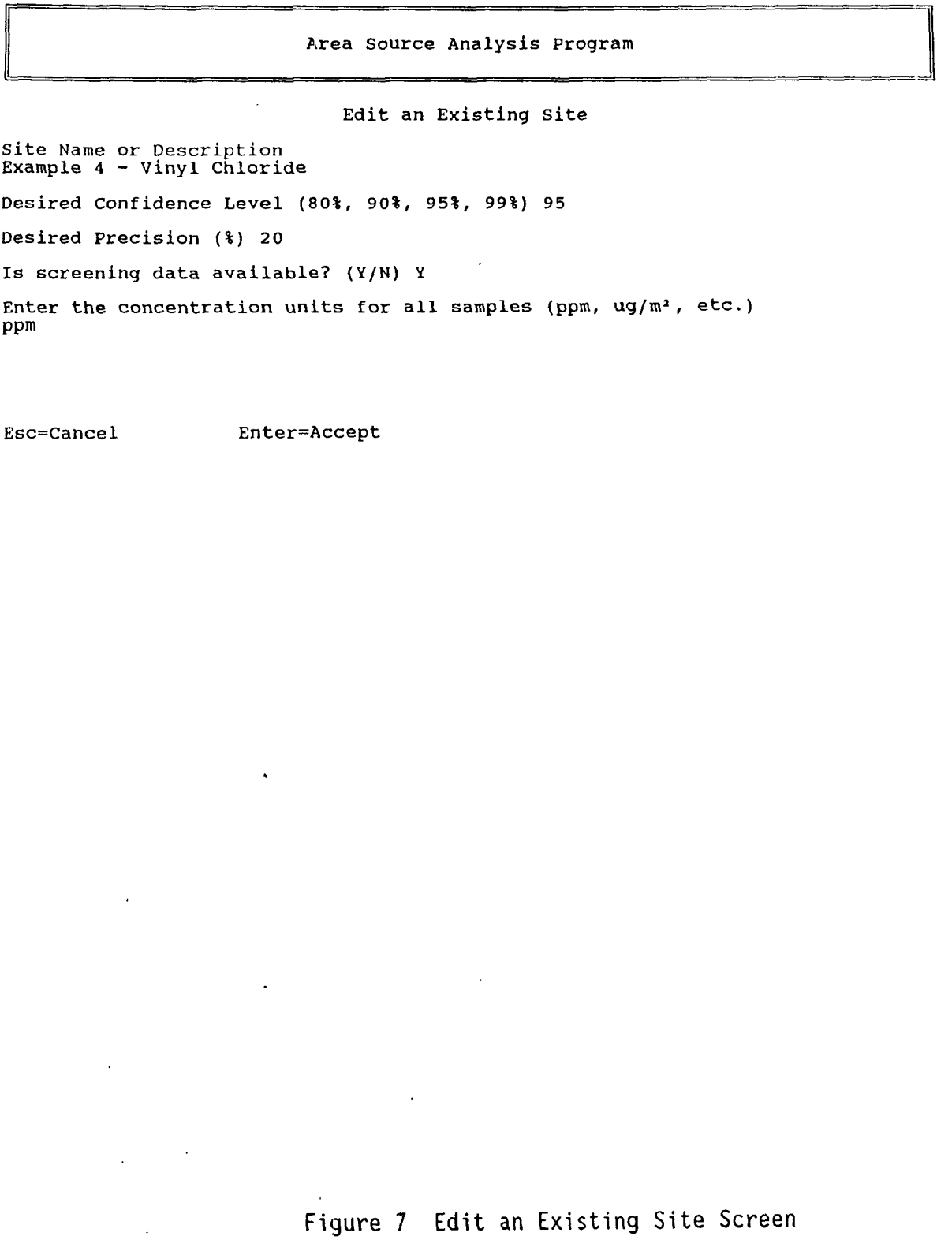


Figure 7 Edit an Existing Site Screen

Option 4. Enter or Edit Screening Data -- Selection of this option enters the Site catalog screen. Select the site for which data is to be entered or edited and press the zone list keys (ALT-Z). At the zone list screen, select the zone for which data is to be entered or edited and press ALT-S to access the screening data entry screen. Entering and editing data at this screen was discussed under Option 2, above.

Option 5. Enter or Edit Actual Data -- Selection of this option enters the Site catalog screen. If there is a zero in the Zones column, no sampling plan has been generated. In this case, select the site and press ALT-A to analyze the data previously entered (minimum required is zone areas). Within a few seconds, a sampling plan will appear which you may review and print or simply press the D(one) key and n(o) for printout. This returns directly to the Site Catalog with the site highlighted. Press the zone list keys (ALT-Z) and, at the zone list screen, select the zone for which data is to be entered or edited and press ALT-A to access the Sample Measurements data entry screen (Figure 8). First entry to the screen is at the bottom (subsequent zones accessed while in this mode are at the top of the data list). Press the page-up or use the arrow keys to move to the top of the listed data. Editing is similar to that for screening data. If additional data points, beyond those for which the analyze function created blank records, are needed, press the ALT-N keys. After data entry is complete, press ALT-Z to enter data for additional zones, or ALT-C or ALT-M to return to the Site catalog or Main Menu, respectively.

Option 6. Review Sampling Plan -- Selection of this option allows the review and printing of previously created sampling plans. Plans may be reviewed both before and after final sample data have been entered and final site statistics have been calculated. Selection of the option displays a screen (Figure 9) similar to the Site catalog except that the options available are limited to reviewing a plan or returning to the Main Menu. To review a plan, select a site using the arrow keys and press ALT-R.

ASAP does not retain the actual plan in memory, but recreates it when the option is executed. Therefore, if site data such as confidence and precision requirements, zone names, zone areas, estimated CVs, or screening data have been altered using the edit functions, a different sample plan may be generated. Because of this feature, a partial sampling plan can be generated using this option, rather than Option 7, that gives the estimated number of samples needed based on areas, estimated CVs or screening data. However, Option 6 does not generate a suggested sampling grid and does not update the Site Catalog listing for number of zones and zone area.

After ALT-R is pressed, a message appears indicating the statistics are being recalculated. The program ignores any data entered as "actual" measurements. A Sample Plan then appears which can be reviewed. It gives the number of samples needed overall and for each zone and the basis for the estimate. Information for each zone can be displayed by pressing the M key. You cannot move backwards in the display. At any time, the D key

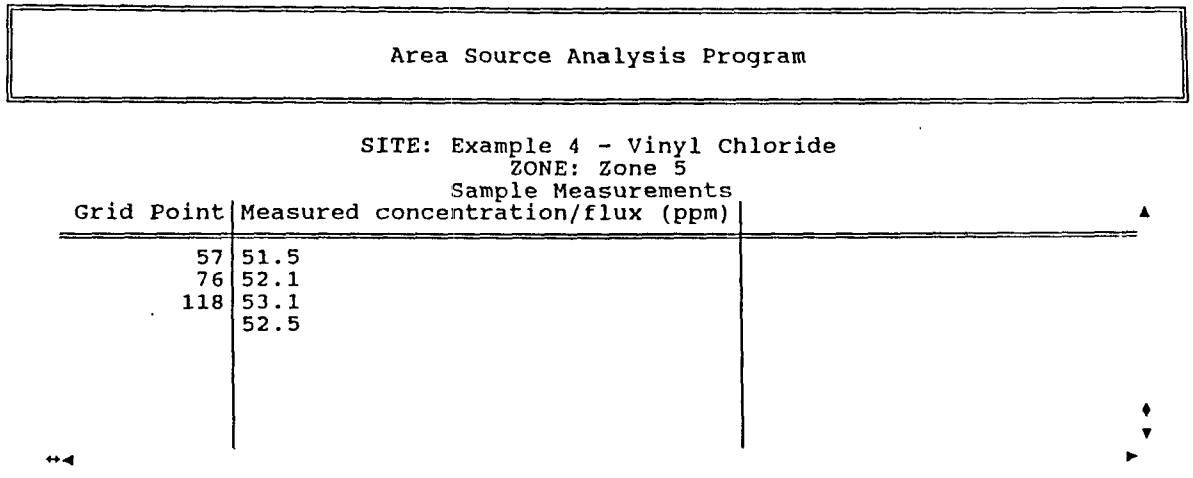


Figure 8 Sample Measurement Entry Screen

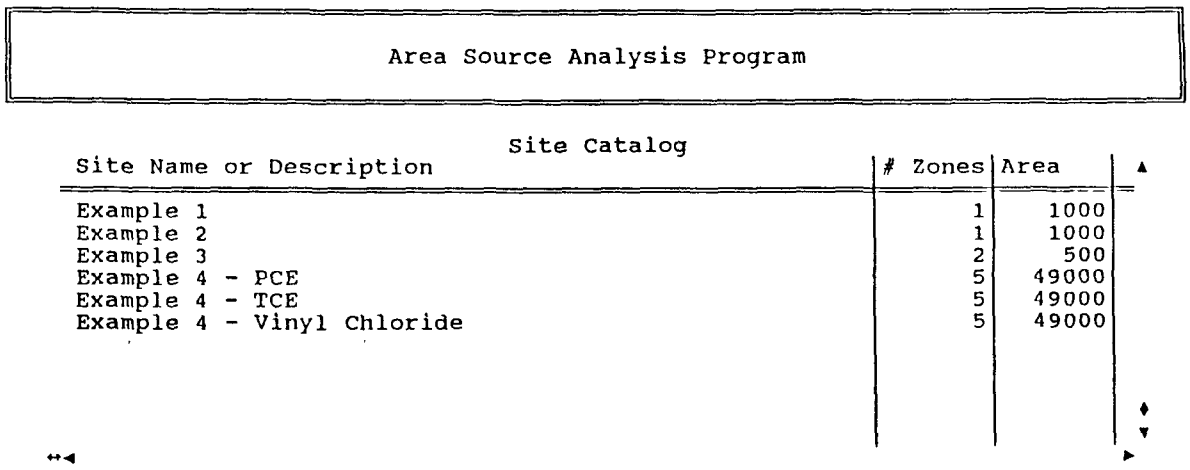


Figure 9 Review Sampling Plan Site Catalog Screen

(for Done) may be pressed to terminate display and access a prompt for printing the plan. If Y(es) is selected and no printer is available, a message will appear and the prompt repeated. The screen cannot be exited except by successfully printing the plan or selecting no print.

Option 7. Analyze Data -- Option 7 is used to analyze any information that has been input for a site. It is used to generate sampling plans, determine whether measurement data are adequate for site statistics, and to calculate final site and zone statistics. The program automatically selects the function(s) to execute depending on the type of data entered.

Selection of the option transfers to the Site Catalog display screen. At this screen, select the site for which analysis is desired and press the ALT-A keys. Sampling plans are automatically created if no sampling plan for the site has been previously created or if the only data inputted are screening data, coefficients of variation (CV) for zones, or area of zones. Priority for data to be used in creating the plan is actual measurement data, screening data, estimated CVs, and zone areas. Minimum data requirement are zone areas. If no information has been entered, the program will prompt "Nothing to Analyze" and return to the Site Catalog screen.

If no "actual" measurement data have been entered, the program generates a simple sampling plan (Figure 10) which gives the number of samples to collect and analyze for each zone and generates a suggested sampling grid. This screen will give the number of zones in the site, total area, requested confidence level and precision, and the number of samples that should be analyzed as well as the number of backup samples that should be collected just in case the zones are more variable than estimated.

Specific plan details for each zone may be accessed by pressing the M key (for more). You cannot move backwards in the screens. You will have to press this key two or three times to move from the summary to the first zone. The zone details include the basis for the estimated number of samples and suggestions for a sampling grid and points to samples on that grid. Grid points are selected using a random number generator and change each time you run the Analyze routine.

You may exit this screen at any time by pressing the D key (for done). A prompt then appears asking whether or not you want to print the report. If you are satisfied with the results (or just want a printout), select yes. [If no printer is available, the prompt will be repeated until you select no.] If you are not satisfied with the results, select no for print and you will be returned to the Site Catalog screen where you may initiate changes in previous input, as previously described, and reanalyze the data.

Selection of this Option is the preferred way to create blank data entry records for initial entry of "actual" data for analysis if "actual" measurement data, rather than screening data, are to be used for plan generation. Be aware that if the program detects even one "actual" data entry, it ignores all other screening and planning data for that zone. It

11/21/92 SUMMARY OF SAMPLING PLAN FOR SITE

SITE: Example 1

Number of Zones:	1	
Total Area:	1000	square meters
Desired Confidence Level:	95	%
Desired Precision:	20	%
Number of Samples to Analyze:	18	
Number of Extra Samples:	4	

Figure 10 Sampling Plan Screens

SITE: Example 1

ZONE: Zone 1

Desired Precision: 20 %
Zone area: 1000 square meters
Unit area: 25 square meters

Number of grid points: 40

Estimated coefficient of variation: 40 %
Number of Samples to Analyze: 18 based on estimated C.V.
Number of Extra Samples: 4

Number of samples would be: 11 if based on zone area.

Samples to Analyze

Grid Point	Concentration/Flux (ppm)
28	
40	
21	
39	
33	
20	
4	
14	
34	
5	
18	
29	
36	
15	
23	
6	
32	
2	

Extra Samples

Grid Point	Concentration/Flux (ppm)
12	
8	
31	
16	

Figure 10 Sampling Plan Screens (cont'd)

will calculate the number of samples needed for other zones based on the planning and screening data.

When using the program this way, follow the instructions above for creation of the plan. At the printout prompt, select no print. This returns you to the Site Catalog with the site already highlighted. Press the ALT-Z keys to move to the Zone List screen and select the Zone for data entry. Press the ALT-A keys to access the Sample Measurement screen and enter the data. Press the ALT-Z keys to return to the Zone List screen and repeat the above steps until all "actual" measurement data has been entered. Press the ALT-C keys to return to the Site Catalog and then press the ALT-A keys to generate a sampling plan based on the entered data.

Sampling plans generated this way provide actual statistics for the entered data at both the overall site summary level and for each zone. At the bottom of the summary and for each zone, a message appears if an inadequate number of samples were collected. The M(ore) key may have to be pressed to display the message. CAUTION: The "Actual Precision" on the summary may be within the tolerance specified even when an inadequate number of samples have been entered for one or more zones. Site statistics are not valid if the message indicates more samples are needed.

Once a sampling plan has been generated by either of the above methods and "actual" measurement data has been entered for all zones, the program analyzes the data but does not generate a new sampling plan. The program assumes planning is complete and analyzes and generates statistics based on the "actual data". All screening data previously entered are ignored. Screen display, messages, and prompts are the same as above for generating plans using "actual" data. A detailed printout is provided as part of the following example application.

4.2.2 Example Application

To illustrate the use of the program, the steps needed to analyze Example 4 of Section 3 are presented. In this case, it is assumed that the preliminary data are adequate for final measurements. Thus, the steps are somewhat different than when using only screening data. How to plan using only planning and screening data are illustrated, however, for the vinyl chloride data. This Example is already in the program and can reviewed.

At the Main Menu, select Option 2 - Add a New Site. At the Add a New Site screen, enter **Example 4 - Vinyl Chloride**, **95** for confidence desired, **20** for precision, **y** for screening data, and **ppm** for units. The Add A New Zone screen will appear. Enter **Zone 1**, **5000** for area, and **20** for CV. New screens will appear each time the CV value is entered. In sequence, enter

Description	Area	CV
Zone 2	7000	20
Zone 3	10000	35
Zone 4	12000	25
Zone 5	15000	30

When the next blank screen appears, press the **escape** key. This terminates automatic entry of new zones and returns to the Zone List screen.

For illustration, first treat the data as simple screening data. Because no data is available for Zone 1, select Zone 2 using the arrow keys and press **ALT-S** to enter screening data. For sample number 1, enter **11.5** and press **ALT-Z** to return to the Zone List. Select Zone 3 and press **ALT-S** to enter the data. In sequence, the following data should be entered:

Zone	Data
3	11.5, 12.1
4	101.5, 112.1, 93.1
5	51.5, 52.1, 53.1, 52.5

After all data has been entered, press **ALT-C** to go to the Site Catalog screen. The Example 4 - Vinyl Chloride site should be highlighted (if not, select it). Note that this screen shows zeros for number of zones and total area, indicating a sample plan has not been generated. Press **ALT-A** to analyze the data and generate a sampling plan. The first page (Summary) of a 6 page report appears. Pressing **D** terminates review and goes directly to the print prompt. By pressing **M** a couple of times, the plan for Zone 1 is displayed. The plan for each Zone can be reviewed by continuing to press **M**. You cannot move backwards in the screens. For each Zone, suggested grid size and grid points to sample are given. The plan for Zones gives the following information:

Zone	No. Samples Needed	Basis For Estimate	No. Needed Based on Area
1	7	CV	17
2	7	CV	19
3	3	Screening data	21
4	4	Screening data	22
5	2	Screening data	24

After review is complete, press the **D**(one) key and a print prompt appears. For the current example, **N**(o) print is selected and return to the Site Catalog is automatic. This completes the illustration using only planning and screening data. The balance of this example is completed assuming the preliminary data are valid "actual" samples.

First, re-enter the vinyl chloride data as "actual" measurements. Setting up the site and Zone descriptions is the same as above. At the Zone List screen, however, the data are entered using the **ALT-A** keys, rather than **ALT-S**, to designate the data as "actual" measurement data useable for emission determinations. Just as before, when data entry is complete, return to the Site Catalog and press **ALT-A** to analyze the data. A 6 page (Summary plus 5 for Zone details) Statistics report is now available for review and printing.

This report gives the mean, standard deviation, actual CV, actual precision, and 95% UCL for the data entered. A message appears at the bottom

of the Summary page indicating additional samples are needed. A review of the individual Zone detail pages gives the following information:

Zone	Number Samples Needed	Mean	UCL	Actual Precision	Extra Samples Needed
1	7	-	-	-	-
2	7	11.5	11.5	0	6
3	3	11.8	15.6	32	1
4	4	102.2	125.9	23	1
5	2	52.3	53.4	2	0

The second column provides the same information generated previously. The last column shows how many additional samples are needed to supplement those already analyzed. It is obvious that no more samples are required for vinyl chloride in Zone 5.

A sampling plan for TCE must now be generated. Because ASAP has no ability to copy existing site or zone descriptions, these must be created again just as if a new site was being entered. At the Main Menu, select Option 2 and enter **Example 4 - TCE** for the site name. Enter **exactly** the same information as for vinyl chloride for the other entries on this screen and for the 5 new zones that must be created. After the last zone description is entered, press the **ALT-C** keys to return to the Site Catalog (or **ALT-M** to go to the Main Menu). At the Site Catalog screen, press **ALT-A** to analyze the planning data and create a preliminary sampling plan (this inserts blank records in the Sample Measurements screen for easy entry of "actual" data). When the plan appears, select **D**(one) and **N**(o) for printing. This returns to the Site Catalog. The site should be highlighted (note that it has been alphabetically arranged before the vinyl chloride site). Press **ALT-Z** to move to the Zone list screen. Select Zone 1 and press **ALT-A** to enter the data as "actual". After data has been entered for all zones, go to the Site catalog (press **ALT-C**) and analyze the data (press **ALT-A**). A sampling plan, complete with zone statistics, is generated similar to that for vinyl chloride. In this case, review of the zone detail reports indicates no additional samples are required in Zone 1, and that 4, 4, 6, and 2 additional samples are required in Zones 2, 3, 4, and 5, respectively.

To generate a sample plan for PCE, the same steps as for TCE need to be completed. The only difference is that **Example 4 - PCE** should be entered for the site description, and the appropriate "actual" data entered for the zones. Using the data presented in Section 3, the sampling plan generated will indicate that the data are, in fact, adequate for our objectives and that no additional samples are required. The report generated, thus, is the statistics report needed for PCE and should be printed. Note that if additional data for PCE is collected along with data needed for adequate statistics for vinyl chloride and TCE, it can be entered into the Sample Measurements screen using Option 5 and new statistics generated using Option 7.

After the additional samples indicated are collected (including several extras held in reserve) and analyzed, the new data can be entered and new statistics calculated to determine whether or not the data now meet the objective. At the Main Menu, select Option 5 - Enter or Edit Actual Data. This activates the Site Catalog screen. Highlight the appropriate site using the arrow keys and press **ALT-Z** to access the Zone List screen for that site. At the Zone List screen, select the Zone for data input and press **ALT-A** to enter the new data. It may be necessary to add new blank records, using **ALT-N**, to enter all the data. Return to the Zone List screen and repeat the data entry for each Zone. When all data has been entered, return to the Site catalog (**ALT-C**) and press **ALT-A** to analyze the data. The message at the bottom of the Summary page will indicate if additional samples are still required. If a message does appear calling for more samples, review the Zone detail reports to see which zone(s) need more data. If no message appears on the Summary report, no additional samples are required for any zone and the report provides all the statistics needed and should be printed.

Figure 11 is the printout that should be obtained for the site for vinyl chloride. It is comprised of a Summary Statistics Report for the site and Detailed Statistics Reports for each of the 5 zones. Similar reports would be generated for TCE and PCE.

4.3 ANALYZING LOGNORMALLY DISTRIBUTED DATA

In Section 3.3, the procedures for calculating statistics for lognormally distributed data were presented. If the data set is small, the mean and standard deviation may easily be calculated using desktop procedures. However, if the data set is large, these statistics may be more easily calculated using the computer software and the results used in Equation 3-3. The procedures to accomplish this follow.

At the Main Menu, select option 2 to add a new site. Enter the requested information as before. At the new zone screen, enter a coefficient of variation sufficiently large to ensure the program will generate a sufficient number of blank records to accomodate the number of data point to analyze. Enter data for only one zone. Return to the Site Catalog and execute the "Analyze data" option. If the first attempt does not indicate as many samples are required as there are data points, edit the entry as required.

At the "Zone List" screen for the site, select "Enter Actual Data". Enter the transformed data (i.e., enter the natural logarithm of the data point). Return to the Site Catalog screen and analyze the data. Ignore all screen outputs except the mean and standard deviation. Write these down, or obtain a printout as a check on entrys of the transformed data.

Using the mean and standard deviation calculated by the software, complete the calculation following the procedures in Section 3.3.

11/21/92 SUMMARY STATISTICS REPORT FOR SITE

SITE: Example 4 - Vinyl Chloride

Number of Zones:	5
Total Area:	49000 square meters
Desired Confidence Level:	95 %
Desired Precision:	20 %
Number of Samples Analyzed:	25
Mean:	152.01904762
Standard Deviation:	32.60894352
Coefficient of Variation:	21.5 %
95% Lower Confidence Limit:	138.55807573 ppm
95% Upper Confidence Limit:	165.48001950 ppm
Actual Precision:	8.9 %

Figure 11 Statistics Report for Vinyl Chloride

11/21/92 DETAILED STATISTICS REPORT FOR ZONES Page: 1

SITE: Example 4 - Vinyl Chloride

ZONE: Zone 1

Zone area: 5000 square meters
Unit area: 31 square meters

Number of grid points: 160
Number of samples needed: 4

Grid Point	Concentration/Flux (ppm)
114	1076
29	982
25	1117
116	991
107	1215
68	905
49	1036

Number of Samples Analyzed:	7
Mean:	1046.00000000
Standard Deviation:	101.32785073
Coefficient of Variation:	9.7 %
95% Lower Confidence Limit:	952.28399211 ppm
95% Upper Confidence Limit:	1139.71600789 ppm
Actual Precision:	9.0 %

Figure 11 Statistics Report for Vinyl Chloride (cont'd)

SITE: Example 4 - Vinyl Chloride

ZONE: Zone 2

Zone area: 7000 square meters
Unit area: 44 square meters

Number of grid points: 160
Number of samples needed: 4

Grid Point	Concentration/Flux (ppm)
123	11.5
101	10.5
32	12.6
43	13.1
73	10.9
150	12.2
135	9.9

Number of Samples Analyzed:	7
Mean:	11.52857143
Standard Deviation:	1.16721076
Coefficient of Variation:	10.1 %
95% Lower Confidence Limit:	10.44904263 ppm
95% Upper Confidence Limit:	12.60810023 ppm
Actual Precision:	9.4 %

Figure 11 Statistics Report for Vinyl Chloride (cont'd)

SITE: Example 4 - Vinyl Chloride

ZONE: Zone 3

Zone area: 10000 square meters
Unit area: 63 square meters

Number of grid points: 160
Number of samples needed: 3

Grid Point	Concentration/Flux (ppm)
56	11.5
40	12.1
17	11.9

Number of Samples Analyzed:	3
Mean:	11.83333333
Standard Deviation:	0.30550505
Coefficient of Variation:	2.6 %
95% Lower Confidence Limit:	11.07435546 ppm
95% Upper Confidence Limit:	12.59231120 ppm
Actual Precision:	6.4 %

Figure 11 Statistics Report for Vinyl Chloride (cont'd)

SITE: Example 4 - Vinyl Chloride

ZONE: Zone 4

Zone area: 12000 square meters
Unit area: 75 square meters

Number of grid points: 160
Number of samples needed: 3

Grid Point	Concentration/Flux (ppm)
158	101.5
36	112.1
80	93.1
96	105.1

Number of Samples Analyzed:	4
Mean:	102.95000000
Standard Deviation:	7.90506167
Coefficient of Variation:	7.7 %
95% Lower Confidence Limit:	90.37304688 ppm
95% Upper Confidence Limit:	115.52695312 ppm
Actual Precision:	12 %

Figure 11 Statistics Report for Vinyl Chloride (cont'd)

SITE: Example 4 - Vinyl Chloride

ZONE: Zone 5

Zone area: 15000 square meters
Unit area: 94 square meters

Number of grid points: 160
Number of samples needed: 2

Grid Point	Concentration/Flux (ppm)
57	51.5
76	52.1
118	53.1
0	52.5

Number of Samples Analyzed:	4
Mean:	52.30000000
Standard Deviation:	0.67330033
Coefficient of Variation:	1.3 %
95% Lower Confidence Limit:	51.22877917 ppm
95% Upper Confidence Limit:	53.37122083 ppm
Actual Precision:	2.0 %

Figure 11 Statistics Report for Vinyl Chloride (cont'd)

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-451/R-93-002		

4. TITLE AND SUBTITLE
Air/Superfund National Technical Guidance Study Series - Air Emissions From Area Sources: Estimating Soil and Soil-Gas Sample Number Requirements

5. REPORT DATE
March 1993

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Wayne Westbrook

8. PERFORMING ORGANIZATION REPORT NO

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Pacific Environmental Services, Inc.
560 Herndon Parkway, Suite 200
Herndon, Virginia 22070-5225

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.

12. SPONSORING AGENCY NAME AND ADDRESS
U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

13. TYPE OF REPORT AND PERIOD COVERED
Final

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This document provides guidance regarding the necessary number of soil gas or soil samples needed to estimate air emissions from area sources. The Manual relies heavily on statistical methods discussed in Appendix C of Volume II of Air/Superfund National Technical Guidance Study Series (EPA 1990) and Chapter 9 of SW-846 (EPA 1986).

The techniques in this manual are based on recognizing the inhomgeniety of an area, by observation or screening samples, before samples are taken. Each of the identified zones are then sampled, using random sampling techniques, and statistics calculated separately for each zone before combining the statistics to provide an estimate for the entire area.

The statistical techniques presented may also be used to analyze other types of data and provide measures such as mean, variance, and standard deviation. The methods presented in this Manual are based on small sample methods. Application of the methods to data which are appropriately analyzed by large sample methods or to data which is not normally distributed will give erroneous results.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Air Sampling Superfund Soil Gas Samples		

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
	20. SECURITY CLASS *(This page)*	22. PRICE