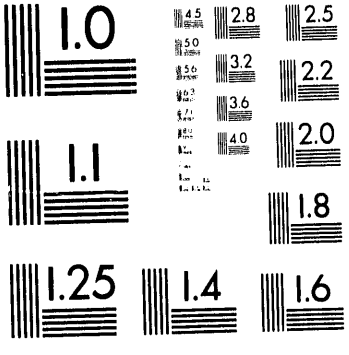

1.0
2.8
2.5
3.2
2.2
3.6
2.0
4.0
1.1
1.8
1.25
1.4
1.6

1 of 1

WHC-SA-1792-FP

Statistical Application of Groundwater Monitoring Data at the Hanford Site

C. J. Chou
V. G. Johnson
F. N. Hodges

Date Published
September 1993

To Be Presented at
Geological Society of America
Annual Meeting
Boston, Massachusetts
October 25-28, 1993

Prepared for the U.S. Department of Energy
Office of Environmental Restoration
and Waste Management

Westinghouse Hanford Company P.O. Box 1970 Richland, Washington 99352

Hanford Operations and Engineering Contractor for the
U.S. Department of Energy under Contract DE-AC06-87RL10930

Copyright License By acceptance of this article, the publisher and/or recipient acknowledges the U.S. Government's right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

MASTER

Approved for Public Release

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

ABSTRACT

Effective use of groundwater monitoring data requires both statistical and geohydrologic interpretations. At the Hanford Site in south-central Washington state such interpretations are used for (1) detection monitoring, assessment monitoring, and/or corrective action at Resource Conservation and Recovery Act sites; (2) compliance testing for operational groundwater surveillance; (3) impact assessments at active liquid-waste disposal sites; and (4) cleanup decisions at Comprehensive Environmental Response Compensation and Liability Act sites. Statistical tests such as the Kolmogorov-Smirnov two-sample test are used to test the hypothesis that chemical concentrations from spatially distinct subsets or populations (e.g., background area versus contaminated or suspect site) are identical within the uppermost unconfined aquifer. Experience at the Hanford Site in applying groundwater background data indicates that background must be considered as a statistical distribution of concentrations, rather than a single value or threshold. The use of a single numerical value as a background-based standard ignores important information and may result in excessive or unnecessary remediation. Appropriate statistical evaluation techniques include Wilcoxon rank sum test, Quantile test, "hot spot" comparisons, and Kolmogorov-Smirnov types of tests. Application of such tests is illustrated with several case studies derived from Hanford groundwater monitoring programs. To avoid possible misuse of such data, an understanding of the limitations is needed. In addition to statistical test procedures, geochemical, and hydrologic considerations are integral parts of the decision process. For this purpose a phased approach is recommended that proceeds from simple to the more complex, and from an overview to detailed analysis.

CONTENTS

1.0 INTRODUCTION 1

2.0 STATISTICAL ANALYSIS TECHNIQUES 1
- 2.1 RANDOM EFFECTS MODELS -- VARIANCE COMPONENTS 4
- 2.2 ATTAINMENT OF A BACKGROUND-BASED CLEANUP STANDARD 8

3.0 CASE STUDIES 13
- 3.1 CASE STUDY 1 - HANFORD S-10 FACILITY 14
- 3.2 CASE STUDY 2 - WELCOXON RANK TEST USING ARSENIC TEST CASE 16
- 3.3 CASE STUDY 3 - QUANTILE TEST USING THE ARSENIC TEST CASE 19

4.0 SUMMARY AND CONCLUSIONS 21

5.0 REFERENCES 22

FIGURES:

1-1 Locations of the RCRA Groundwater Monitoring Projects and Landmarks on the Hanford Site 2

1-2 Locations of Active or Recently Active Liquid Waste Disposal Sites 3

2-1 A Shift Case Where the Distribution of Measurements for a Contaminant of Concern in the Remediated Waste Site is Shifted Two Units to the Right of the Background Distribution 10

2-2 A Mixture Case Where 10% of the Distribution of Measurements for a Contaminant of Concern in the Remediated Waste Site Has Not Cleaned to the Background Level 12

3-1 Monitoring Well Locations Outside of Known Contamination Plume Areas 17

4-1 A Phased Approach to Contamination/Remediation Decisions 23

TABLES:

1 Analysis of Variance Table for a Two-Way Crossed Classification With a Random Effects Model 5

2 A Nested Random Effects Analysis of Variance Table 7

3 Background Specific Conductance Data[a] for the S-10 Facility 14

4 ANOVA for a Two-Way Crossed Classifications Random Effect Model 15

5 ANOVA for a Two-Way Nested Classifications Random Effect Model 15

CONTENTS (cont)

TABLES:

6 Arsenic Data From 10 Wells in the Rattlesnake Mountain Corridor and 32 Wells Across the Hanford Site 18
7 Results of the WRS Test--Arsenic Test Case 20

1.0 INTRODUCTION

Groundwater contamination problems at the U.S. Department of Energy (DOE) Hanford Site, in south-central Washington State, range from local groundwater contamination associated with the individual waste disposal sites to the overall impact of a 40-yr period of nuclear materials production and associated contaminants released to the groundwater beneath the 1,450 km^2 site. Statistical analysis has been applied to groundwater contamination problems, which include (1) determining local backgrounds to ascertain whether or not an individual disposal site is affecting the quality of groundwater, and (2) determining a "pre-Hanford" groundwater background for the Hanford Site to allow formulation of background-based cleanup standards. From these studies it is apparent that background for a specific analyte of concern should be considered as a statistical distribution, not as a single value or threshold. Additionally, it is important to distinguish different sources of variation in the data and to handle them in the proper manner. If due care is not exercised when determining background distributions there is a high probability that false determinations of contamination will occur, resulting in unnecessary remediation and/or cleanup expense.

Effective use of groundwater monitoring data requires an integration of statistical and geohydrologic approaches. This combination is much more effective than either approach used in isolation. At the Hanford Site such interpretations are used for (1) detection monitoring, assessment monitoring, and/or corrective action at Resource Conservation and Recovery Act (RCRA) sites (Figure 1-1), (2) compliance testing for operational groundwater surveillance, (3) impact assessments at active liquid waste disposal sites (Figure 1-2), and (4) cleanup decisions at Comprehensive Environmental Response Compensation and Liability Act (CERCLA) sites.

Statistical tests, such as the Kolmogorov-Smirnov two-sample test, have been used to test the null hypothesis that chemical concentrations from spatially distinct subsets or populations (e.g., background area versus hazardous waste site) are identical within the uppermost unconfined aquifer at the Hanford Site (WHC 1993). The primary purpose of this paper is to supplement statistical methods discussed in WHC (1993) by (1) demonstrating how to model spatial, temporal, and analytical variability of background measurements, (2) showing that the variance estimate s^2 (that assumes equal spatial and temporal variability) is biased when multiple upgradient wells comprise the background wells, (3) illustrating several statistical techniques that may be used in verifying the attainment of a background-based cleanup standard (e.g., Wilcoxon rank sum test, quantile test, etc.); and (4) recommending an approach that represents some of the initial efforts in establishing statistical guidance for evaluation of groundwater at the Hanford Site. Three case studies are provided.

2.0 STATISTICAL ANALYSIS TECHNIQUES

Groundwater quality information is needed by regulators and environmental managers for making decisions on assessing the impact of a facility on groundwater quality and/or the effectiveness of groundwater remediation efforts. Over 250 RCRA-compliant groundwater monitoring wells have been installed around various facilities on the Hanford Site since 1987,

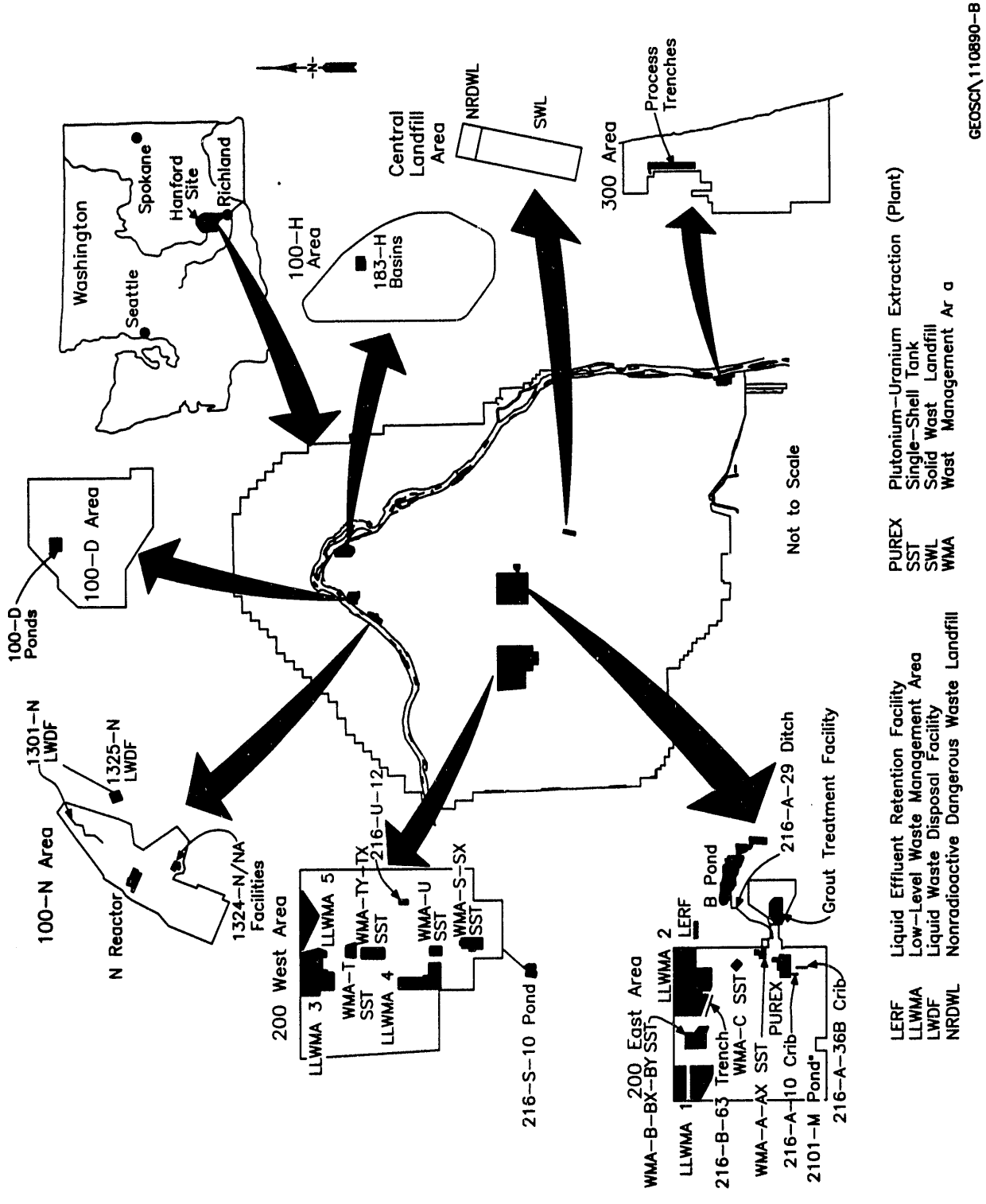


Figure 1-1. Locations of the RCRA Groundwater Monitoring Projects and Landmarks on the Hanford Site.

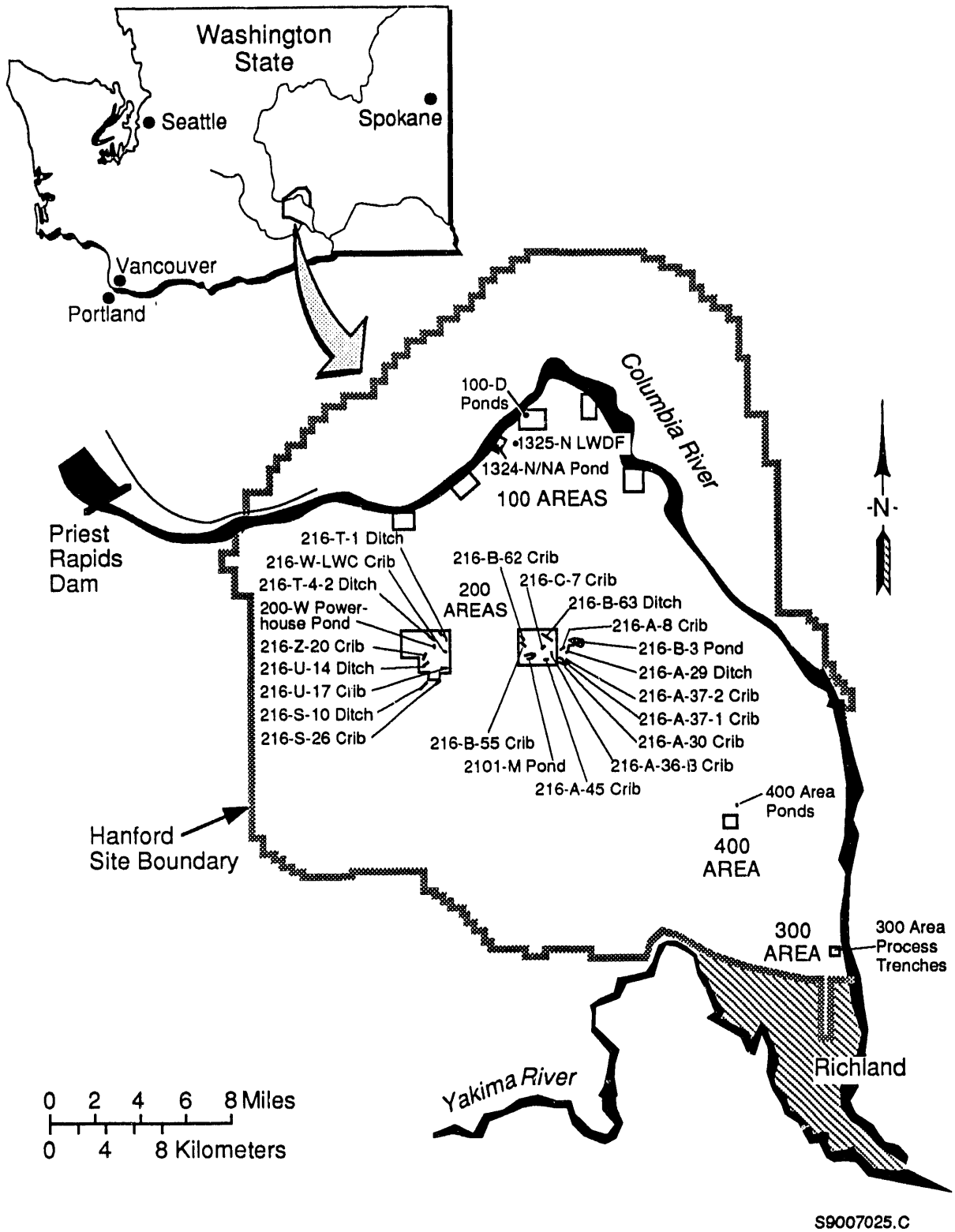


Figure 1-2. Locations of Active or Recently Active Liquid Waste Disposal Sites.

in addition to numerous older wells. At least 60,000 analytical results are generated annually for data evaluation and reporting.

From a statistical perspective, the null hypothesis of interest is that there is no difference in the chemical compositions of groundwater upgradient (background) and downgradient of a facility. However, the geohydrologic conditions at each site may be different, resulting in groundwater quality data that are highly site specific. Several statistical analysis techniques that model the background measurements (multiple upgradient wells, quarterly sampling events, and quadruplicate analysis results) into factors due to spatial, temporal, and analytical variabilities are provided. If background wells consist of several upgradient wells, the variance estimate s^2 (which assumes equal spatial and temporal variability) will underestimate the true variance. The formula to calculate this bias is given for each model under consideration. Finally, non-parametric test procedures that are useful when assessing the attainment of a background-based cleanup standard are presented.

2.1 RANDOM EFFECTS MODELS -- VARIANCE COMPONENTS

2.1.1 Crossed Classifications

In studying the variability that is evident in the data, the main interest is to attribute that variability to various data classifications. Classifications that identify the source of each datum are called *factors* (or independent variables). The individual classes of a factor are called *levels* of that factor. A factor is *random* if its levels consist of a random sample of levels from a population of all possible levels, otherwise, it is *fixed*. In the context of groundwater monitoring, the background chemical composition for a particular analyte of concern is usually obtained from replicate analysis of samples collected from multiple (e.g., quarterly) sampling events of wells upgradient of the facility. When the locations of the background wells and the sampling times are assumed to be random variables, the background concentration can be modeled by the following general equation:

$$Y_{ijk} = \mu + W_i + T_j + WT_{ij} + \epsilon_{(ij)k},$$

$i = 1, 2, \ldots a$ (number of background wells),
$j = 1, 2, \ldots b$ (number of sampling times),
$k = 1, 2, \ldots r$ (number of replicate analyses),

where Y_{ijk} denotes the kth analysis on the jth sampling time from the ith upgradient well. Terms used in the model are defined as follows:

μ = the true mean level for background measurements.

W_i = the true effect of the ith well. W_i is assumed to be a random observation from a population with mean 0 and with variance σ^2_W.

T_j = the true effect of the jth sampling time. T_j is assumed to be a random observation from a population with zero mean and with variance σ^2_T.

WT_{ij} = the interaction between the ith well and j^{th} sampling time. WT_{ij} is assumed to be a random observation from a population with mean zero and with variance σ^2_{WT}.

$\epsilon_{(ij)k}$ = the analytical error associated with the kth analysis of samples collected from well i at time j. $\epsilon_{(ij)k}$ is assumed to be a random observation from a population with mean zero and with variance σ^2_A.

The above model assumes that all wells are measured at the same time. Therefore, the two factors (well and time) are referred to as "crossed" factors. It is also assumed that the W_i, T_j, WT_{ij}, and $\epsilon_{(ij)k}$ are mutually uncorrelated. Thus, each observation Y_{ijk} is from a population with mean μ and with variance

$$Var(Y_{ijk}) = \sigma^2_W + \sigma^2_T + \sigma^2_{WT} + \sigma^2_A.$$

The objective is to estimate the various variance components, σ^2_W, σ^2_T, σ^2_{WT}, and σ^2_A. The following analysis of variance (ANOVA) table (Table 1) is given to provide guidance in computing the estimates of these variance components.

Table 1. Analysis of Variance Table for a Two-Way Crossed Classification With a Random Effects Model.

Source	Degrees of Freedom df	Sum of Squares SS	Mean Squares MS	Expected Mean Squares E(MS)
Well	(a-1)	SS_W	MS_W	$\sigma^2_A + r\sigma^2_{WT} + rb\sigma^2_W$
Time	(b-1)	SS_T	MS_T	$\sigma^2_A + r\sigma^2_{WT} + ra\sigma^2_T$
Well x Time	(a-1)(b-1)	SS_{WT}	MS_{WT}	$\sigma^2_A + r\sigma^2_{WT}$
Analytical	ab(r-1)	SS_A	MS_A	σ^2_A
Total	abr-1	SS_{Total}		

Each mean square (MS) is obtained by dividing the sum of squares (SS) with the respective degrees of freedom (df). For example, $MS_W = SS_W/(a-1)$, and

$$SS_W = br\Sigma_i(\overline{y}_{i..} - \overline{y}_{...})^2,$$

$$SS_T = ar\Sigma_j(\overline{y}_{.j.} - \overline{y}_{...})^2,$$

$$SS_{WT} = r\Sigma_i\Sigma_j(\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...})^2,$$

$$SS_A = \Sigma_i\Sigma_j\Sigma_k(y_{ijk} - \overline{y}_{ij.})^2,$$

$$SS_{Total} = \Sigma_i\Sigma_j\Sigma_k(y_{ijk} - \overline{y}_{...})^2.$$

Note that the "." used in the above expression denotes the summation over a particular subscript. For example, $y_{i..} = \Sigma_j\Sigma_k\, y_{ijk}$ and $\overline{y}_{i..} = y_{i..}/br$. Most commercial statistical software packages will generate the quantities needed in the above table except for the expressions given in the expected mean squares column. A set of simple rules to determine these expressions quickly (without recourse to their derivation) is given in Hicks (1982).

Estimates of the variance components σ^2_W, σ^2_T, σ^2_{WT}, and σ^2_A are obtained by equating the MS quantities with the corresponding E(MS) expressions (Table 1) and then solving the resulting four equations. This way of comuting the variance components is called the method of moments. This yields the estimators

$$\hat{\sigma}_A^2 = MS_A,$$

$$\hat{\sigma}_{WT}^2 = (MS_{WT} - MS_A)/r,$$

$$\hat{\sigma}_T^2 = (MS_T - MS_{WT})/ar,$$

$$\hat{\sigma}_W^2 = (MS_W - MS_{WT})/br.$$

Variance of Y_{ijk} [$Var(Y_{ijk})$] is estimated by $\hat{\sigma}_W^2 + \hat{\sigma}_T^2 + \hat{\sigma}_{WT}^2 + \hat{\sigma}_A^2$, which is an unbiased estimator of total variance (Milliken and Johnson, 1984). The usually computed variance estimate s^2 ($s^2 = \Sigma_i\Sigma_j\Sigma_k(y_{ijk} - \bar{y}_{...})^2/(abr-1)$) that assumes equal spatial and temporal variability is biased because the expectation of s^2 [$E(s^2)$] is:

$$\begin{aligned} E(s^2) &= E[SS_{Total}/(abr-1)] \\ &= E[(SS_W + SS_T + SS_{WT} + SS_A)/(abr-1)] \\ &= [(a-1)E(MS_W)+(b-1)E(MS_T)+(a-1)(b-1)E(MS_{WT})+ab(r-1)E(MS_A)]/(abr-1) \\ &= [(n-br)/(n-1)]\sigma^2_W + [(n-ar)/(n-1)]\sigma^2_T + [(n-r)/(n-1)]\sigma^2_{WT} + \sigma^2_A, \end{aligned}$$

where n = abr (the total number of background measurements). This equation reveals that s^2 will underestimate the true variance of Y_{ijk} because the coefficients associated with σ^2_W, σ^2_T, and $^2_{WT}$ are all less than 1. The amount of bias is:

$$Var(Y_{ijk}) - E(s^2) = [(br-1)/(n-1)]\sigma^2_W + [(ar-1)/(n-1)]\sigma^2_T + [(r-1)/(n-1)]\sigma^2_{WT}.$$

Additionally, an estimator for the background mean μ is given by $\bar{y}_{...} = y_{...}/abr$, which is the overall average of the background observations. The variance of $\bar{y}_{...}$ is given by

$$\begin{aligned} Var(\bar{y}_{...}) &= (br\sigma^2_W + ar\sigma^2_T + r\sigma^2_{WT} + \sigma^2_A)/abr \\ &= \sigma^2_W/a + \sigma^2_T/b + \sigma^2_{WT}/ab + \sigma^2_A/abr, \end{aligned}$$

and is estimated by $\hat{\sigma}_W^2/a + \hat{\sigma}_T^2/b + \hat{\sigma}_{WT}^2/ab + \hat{\sigma}_A^2/abr$. Or, it can be calculated by $(MS_W + MS_T - MS_{WT})/abr$. The first is important especially when one needs to redesign the sampling plan to obtain a more precise estimate of $Var(\bar{y}_{...})$ (see Case Study 1). Finally, it is important to know that if $r=1$, σ^2_A and σ^2_{WT} cannot be separately estimated; however, it is still possible to estimate σ^2_W and σ^2_T.

2.1.2 Nested Classifications

When background wells are sampled at different times, then the time factor is said to be "nested" within the well factor. Therefore, the nested model is appropriate and is given by the following equation:

$$Y_{ijk} = \mu + W_i + T_{(i)j} + \epsilon_{(ij)k},$$
$$i = 1, 2, \ldots a, \; j = 1, 2, \ldots b, \; k = 1, 2, \ldots r,$$

where μ, W_i, and $\epsilon_{(ij)k}$ are as before and $T_{(i)j}$ is the true effect of the jth sampling time within the i^{it} well. This nested model has no interaction term present, as the wells are not crossed with the times. It is also assumed that the W_i are uncorrelated, the $T_{(i)j}$ are uncorrelated, and the $\epsilon_{(ij)k}$ are uncorrelated, and there are no correlations among the W_i, $T_{(i)j}$, and $\epsilon_{(ij)k}$. Thus, y_{ijk} is an observation from a population with mean μ and with variance $Var(Y_{ijk}) = \sigma^2_W + \sigma^2_T + \sigma^2_A$. The ANOVA table for a nested random effects model is displayed in Table 2.

Table 2. A Nested Random Effects Analysis of Variance Table.

Source	Degrees of Freedom df	Sum of Squares SS	Mean Squares MS	Expected Mean Squares E(MS)
Well	(a-1)	SS_W	MS_W	$\sigma^2_A + r\sigma^2_T + rb\sigma^2_W$
Time	a(b-1)	SS_T	MS_T	$\sigma^2_A + r\sigma^2_T$
Analytical	ab(r-1)	SS_A	MS_A	σ^2_A
Total	abr-1	SS_{Total}		

where equations for calculating SS_W, SS_A, and SS_{Total} are same as before, except for SS_T which is given below:

$$SS_T = r\Sigma_i\Sigma_j(\overline{y}_{ij.} - \overline{y}_{i..})^2.$$

Estimates of the variance components σ^2_W, σ^2_T, and σ^2_A are obtained by equating the computed MS to their expected values and then solving the resulting equations for the unknown variance components. This leads to the unbiased estimators

$$\hat{\sigma}_A^2 = MS_A,$$
$$\hat{\sigma}_T^2 = (MS_T - MS_A)/r,$$
$$\hat{\sigma}_W^2 = (MS_W - MS_T)/br.$$

Variance of Y_{ijk} is estimated by $\hat{\sigma}_W^2 + \hat{\sigma}_T^2 + \hat{\sigma}_A^2$, which is an unbiased estimator of the total variance (Milliken and Johnson, 1984). Using the method described earlier, it can be shown that s^2 is biased and underestimates the variance of Y_{ijk}. The amount of bias is:

$$Var(Y_{ijk}) - E(s^2) = [(br-1)/(n-1)]\sigma^2_W + [(r-1)/(n-1)]\sigma^2_T, \text{ where } n=abr.$$

Just like the "crossed" case, an estimator for the overall background mean μ can be obtained by $\bar{y}_{...} = y_{...}/abr$, which is the overall average of the background observations. The variance of $\bar{y}_{...}$ is given by

$$\begin{aligned} Var(\bar{y}_{...}) &= (br\sigma^2_W + r\sigma^2_T + \sigma^2_A)/abr \\ &= \sigma^2_W/a + \sigma^2_T/ab + \sigma^2_A/abr \end{aligned}$$

and is estimated by $\hat{\sigma}_W^2/a + \hat{\sigma}_T^2/ab + \hat{\sigma}_A^2/abr$. Or, it can be estimated by MS_W/abr. This first result is important especially when one needs to redesign the sampling plan to obtain a more precise estimate of $Var(\bar{y}_{...})$. If spatial variability is the most dominant factor in the total variance components, the most effective way to reduce the uncertainty in estimating the background mean is to increase the number of background wells. That is, to make $Var(\bar{y}_{...})$ smaller, one has to make $\hat{\sigma}_W^2/a$ smaller. To make $\hat{\sigma}_W^2/a$ smaller, one has to increase the denominator a (note 'a' denotes the number of background wells). To offset the cost due to sampling of more wells, one can decrease the number of sampling times and reduce the number of replicate analyses, especially when the cost associated with each chemical analysis is high.

2.1.3 Impact on Groundwater Monitoring

Regardless of whether the model is a "crossed" or "nested" model, s^2 is a biased estimator when multiple background wells exist. In calculating tolerance intervals, prediction intervals, and/or confidence intervals, $Var(Y_{ijk})$ should be used rather than s^2. Otherwise, the resulting calculated intervals will be smaller. In groundwater monitoring, an upper $(1-\alpha)100\%$ tolerance bound, calculated on the basis of background measurements, is often required by regulators for use as a threshold value where $1-\alpha$ denotes the level of confidence. The threshold value is used to determine the presence or absence of contamination. Without considering spatial and temporal variabilities in the calculations, the resulting bounds will be too conservative (i.e., too low). Decisions based on such bounds may lead to too many false positive conclusions, and as a result, unnecessary or excessive remediation.

2.2 ATTAINMENT OF A BACKGROUND-BASED CLEANUP STANDARD

After the groundwater at a RCRA/CERCLA site has been remediated, it is necessary to determine whether the remediation effort has been successful (i.e., verification of cleanup). This determination should be statistically based, using appropriate statistical sampling designs and tests. Appropriate statistical tests may include the Wilcoxon Rank Sum (WRS) test and/or the Quantile test depending on the type of residual contamination scenarios (Gilbert and Simpson, 1990). If the remedial action has "uniformly" reduced contamination levels (i.e., the shift alternative), but not to background levels, the WRS test should be used because it has greater power than the quantile test. However, if most of the cleanup unit has been remediated to background levels and only a few "hot spots" remain (i.e., the mixture alternative), the quantile test is preferred because it has more power than the WRS test (Gilbert and Simpson, 1990).

2.2.1 The Wilcoxon Rank Sum Test

The Wilcoxon rank sum test may be used to test for a shift in location (Figure 2-1) between two independent populations (e.g., chemical concentration from the background area and waste site). In the case of cleanup verification, the null hypothesis H_o to be tested is the attainment of the background-based cleanup standard and the alternative hypothesis H_a is the non-attainment (i.e., a one-tailed test). The WRS test is performed in the following steps:

1. Obtain two random samples of sizes m (from background area) and n (from waste site). Let N = m + n.

2. Order the N data values as though they were obtained from the same population.

3. Assign ranks to the ordered observations. Assign rank 1 to the smallest observation and N to the largest. When several data values are exactly equal to each other (i.e., tied), assign to each the average of the ranks they would have received had there been no ties.

4. If some of the data values are less-than values, assume these values are tied at a value less than the smallest detected value in the combined data set and follow the procedure for handling ties.

5. Total the ranks given to the n samples from the waste site. Denote this total as W_{rs}.

6. If m and n are less or equal to 10, compare W_{rs} to the appropriate critical value (T_U) in Appendix 18 of Ostle and Malone (1988). Reject H_o if $W_{rs} \geq T_U$, where T_U is the upper critical value for the selected one-tailed α value (note, α is the level of significance of the test, usually α is set to be 0.05 or 5%).

7. If both m and n are larger than 10 and no ties are present, compute the large sample test statistic

$$Z_{rs} = \frac{W_{rs} - n(N+1)/2}{[mn(N+1)/12]^{1/2}}$$

8. If both m and n are larger than 10 and ties are present, compute Z_{rs} based on the following equation:

$$Z_{rs} = \frac{W_{rs} - n(N+1)/2}{\{(nm/12)[(N+1-\Sigma_j t_j(t_j^2-1)/N(N-1)]\}^{1/2}}$$

where g is the number of tied groups and t_j is the number of tied data in the jth group.

9. Reject H_o and accept H_a if $Z_{rs} \geq Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution.

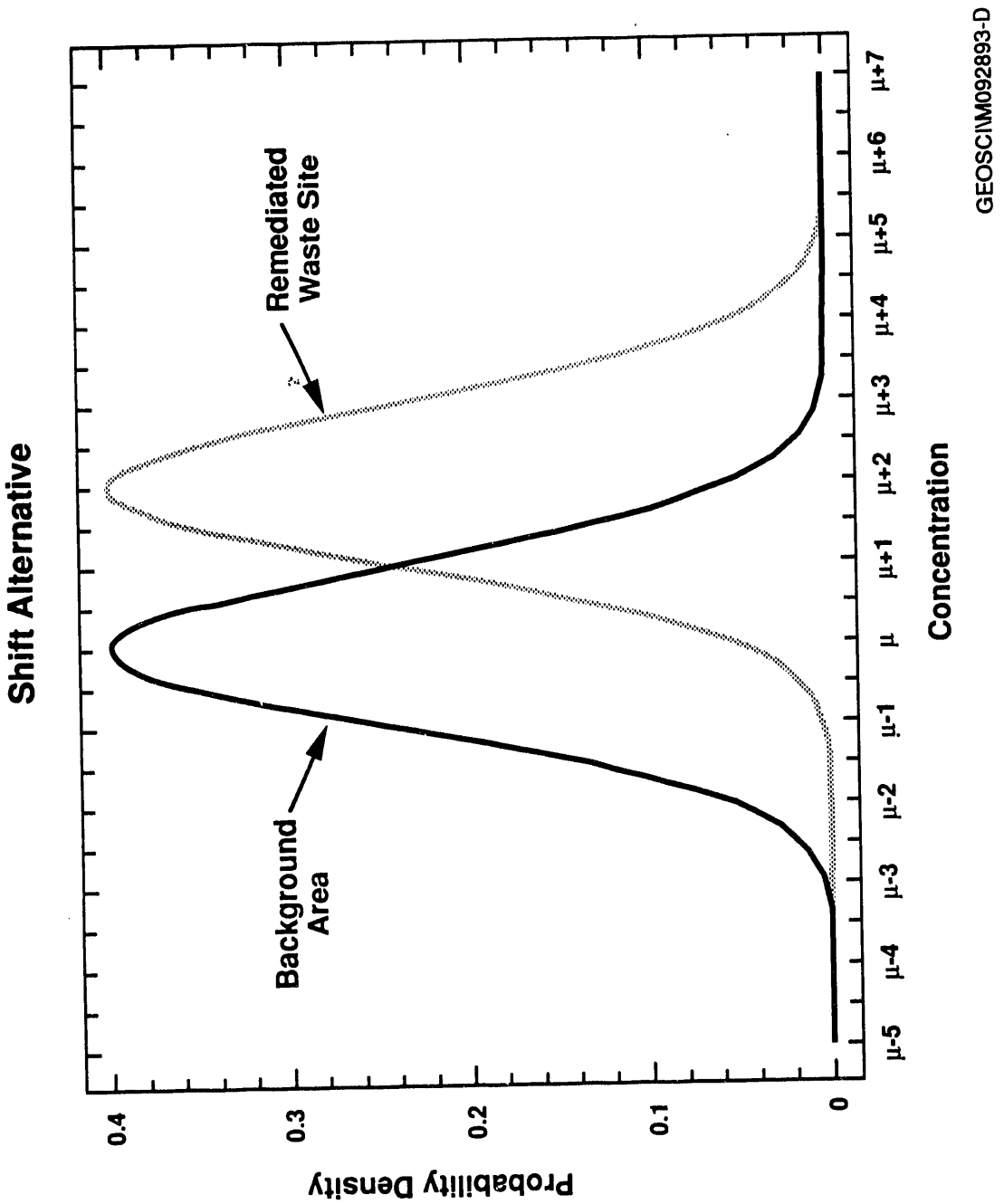


Figure 2-1. A Shift Case Where the Distribution of Measurements for a Contaminant of Concern in the Remediated Waste Site is Shifted Two Units to the Right of the Background Distribution.

Gilbert and Simpson (1992) give detailed procedures on how to determine the total number of samples needed (N) for the WRS test. It is calculated based on specified values of α, β, and the amount of shift (in units of standard deviation), δ/σ, that is important to detect with power $1-\beta$. [Note, α denotes the Type I error rate and β denotes the Type II error rate, where Type I error is the error of rejecting a true null hypothesis (false positive) and Type II error is the error of accepting a false null hypothesis (false negative)].

2.2.2 The Quantile Test

The nonparametric Quantile test was developed by Johnson et al. (1987) to detect changes in a small proportion of a treated population. The test is simple to use and a locally most powerful test for the mixture alternative (Figure 2-2). The null hypothesis is attainment of background-based cleanup standard. The test statistic is merely a count of the number of site measurements (k) that are among the largest r measurements of the combined data set. If k is sufficiently large then the test indicates the remediated waste site has not attained the background-based cleanup standard. Most importantly, the test statistic has a hypergeometric distribution when the null hypothesis is true. Hence, its probability can be calculated exactly. The Quantile test can be performed as follows:

1. Specify the required Type I error rate, α.

2. Assume there are m measurements from the background area and n measurements from the waste site, and let $N = m + n$. Choose a value of q that is greater than 0.5 and less than 1.0, where q is the proportion of the remediated waste site that has been cleaned to the background level. Therefore, 1-q is the proportion of the remediated waste site that has not cleaned to the background level. Note when $q = 0.5$, the median test (Conover 1980) is obtained.

3. Compute $r = N(1-q)$, where r is the number of largest measurements among the N combined measurements that must be examined. When less-than values are present in either data set, assume that their value is less than the rth largest measured value in the combined data set. If fewer than r measurements are greater than the detection limit, then the Quantile test cannot be performed.

4. Order the combined data set from smallest to the largest. Count the number, k, of measurements from the waste site that are among the r largest measurements from the combined data set.

5. If the rth largest measurement (count down from the largest measurement) is among a group of tied measurements, then increase r to include the entire set of tied measurements. Also increase k by the same amount.

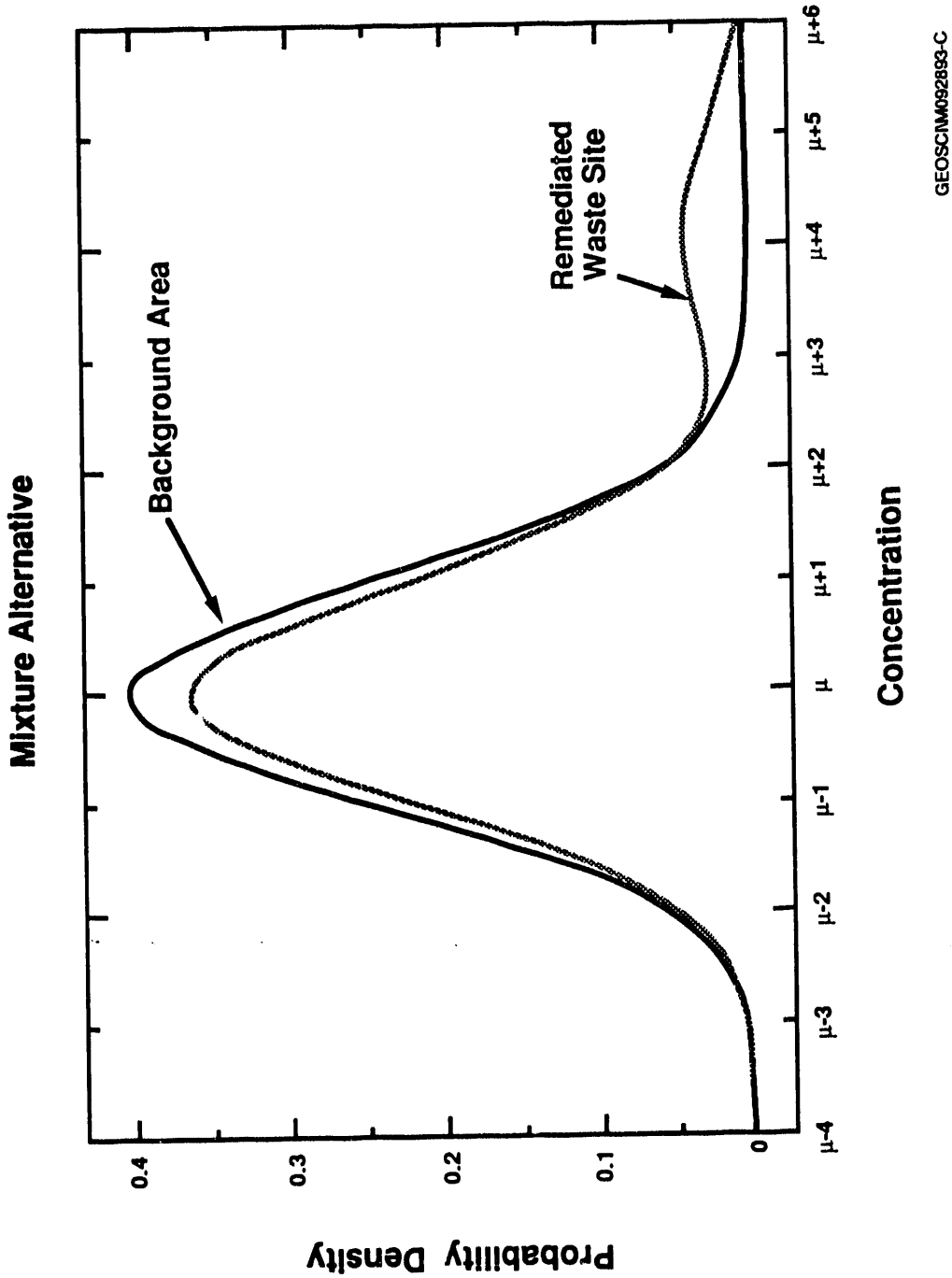


Figure 2-2. A Mixture Case Where 10% of the Distribution of Measurements for a Contaminant of Concern in the Remediated Waste Site Has Not Cleaned to the Background Level.

6. If $r \leq 20$, calculate the probability, P, of obtaining a value of k as large or larger than the observed k, if H_o is true

$$P = \frac{\Sigma_i \binom{m+n-r}{n-i}\binom{r}{i}}{\binom{m+n}{n}},$$

where $\binom{a}{b} = \frac{a!}{b!\,(a-b)!}$ and $a! = a(a-1)(a-2) \ldots 1$ and the summation over subscript i is from k, k+1, ..., r.

7. If $r > 20$, use the following equations to determine P, the probability of obtaining a value of k as large or larger than the observed k, if H_o is true

$$\bar{X} = \frac{nr}{m+n}$$

= mean of the hypergeometric distribution

$$SD = \left[\frac{mnr(m+n-r)}{(m+n)^2\,(m+n-1)}\right]^{1/2}$$

= standard deviation of the hypergeometric distribution, and

$$Z_p = \frac{k - 0.5 - \bar{X}}{SD},$$ where Z_p is a standard normal variable.

Use a standard normal distribution table with the computed value of Z_p to determine the corresponding p value and let P = 1 - p.

8. Reject H_o and accept H_a, if $P \leq$ specified α. If H_o is rejected, conclude that the remediated waste site has not attained the background standard and additional remedial effort is needed.

Gilbert and Simpson (1992) gives a detailed description on how to determine the number of samples based on computer simulations for the case where the residual contamination is assumed to be distributed at random throughout the cleanup site and the background and waste site measurements are assumed to be normally distributed. Look-up tables for conducting the Quantile test are also provided.

3.0 CASE STUDIES

Three case studies are provided in this section. The first case study is to demonstrate the variance components analysis techniques described earlier. The second and the third case studies illustrate the verification of cleanup efforts through the use of the WRS test and the Quantile test, respectively. Unless otherwise specified, the statistical software package STATGRAPHICS (Version 4.2) (a trademark of Statistical Graphics Corporation) was used to generate results presented in the ANOVA tables.

3.1 CASE STUDY 1 - HANFORD S-10 FACILITY

Facility Background--The S-10 Facility is a RCRA-regulated treatment, storage, and disposal facility located south-southwest of the 200 West Area of the Hanford Site (see Figure 1-1). The S-10 Facility consists of a pond and a ditch. In the past, it received waste water that contained dangerous waste and radioactive materials from the Reduction-Oxidation Plant. The effluent stream to the S-10 Facility was permanently deactivated in October 1991. Currently, this facility is operated under the RCRA interim-status regulations (EPA 1989).

Variance Components Analysis: Crossed Classification--The monitoring network has six wells; two upgradient, three downgradient, and one perched water zone well. Eight quarters of quadruplicate measurements of indicator parameters (field pH, field specific conductance, total organic carbon, and total organic halogen) have been collected. For illustrative purposes only, specific conductance data taken from two upgradient wells during March 1992 through December 1992 (four quarters of data) are used. Quadruplicate measurements are required for this parameter (40 CFR 265.(c)(2)). The input data is presented in Table 3. The two upgradient wells were sampled about the same time, therefore, a "crossed" model is used.

Table 3. Background Specific Conductance Data[a] for the S-10 Facility.

Background Well	Sample Date	Duplicate Sample Number	Specific Conductance (µmho/cm)
2-W26-7	03/12/92	1	262
		2	260
		3	259
		4	258
	06/09/92	1	253
		2	255
		3	254
		4	254
	09/11/92	1	268
		2	261
		3	261
		4	260
	12/18/92	1	256
		2	252
		3	256
		4	255
2-W26-8	03/13/92	1	234
		2	236
		3	235
		4	234
	06/09/92	1	242
		2	242
		3	241
		4	241
	09/11/92	1	248
		2	246
		3	244
		4	242
	12/22/92	1	243
		2	241
		3	241
		4	240

[a]Data used here is for illustrative purpose only.

Applying techniques described in Section 2.1.1 to the input data, the following analysis variance table (Table 4) is obtained.

Table 4. ANOVA for a Two-Way Crossed Classifications Random Effect Model.

Source	Degrees of Freedom	Sum of Squares	Mean Squares
Well	1	2,346.125	2,346.125
Time	3	224.375	74.792
WellxTime	3	193.375	64.458
Analysis	24	91.000	3.792
Total	31	2,854.875	

Using formulas given in Section 2.1.1, the following estimates of variance components and percentages (%) are obtained (note, numbers are rounded to two decimal places).

		Variance Components	%
$\hat{\sigma}_A^2 = MS_A$	=	3.79	2.33
$\hat{\sigma}_{WT}^2 = (64.4583-3.7917)/4$	=	15.17	9.31
$\hat{\sigma}_T^2 = (74.7917-64.4583)/8$	=	1.29	0.79
$\hat{\sigma}_W^2 = (2,346.125-64.4583)/16$	=	142.60	87.57
Total		162.85	100.00

Hence, spatial variability (due to multiple background well locations) accounts for more than 87% of the total variance and analytical error accounts for less than 3% of the total variance. The biased estimator s^2 is calculated to be 92.08 ($SS_{Total}/31$ = 2,854.875/31 = 92.09), and the amount of bias is equal to 70.76 (162.85 - 92.09). Hence, the confidence intervals, prediction intervals, and/or tolerance intervals calculated using the biased estimator s^2 will be smaller than they should be had the unbiased estimator been used.

Variance Components Analysis: Nested Classifications--The nested random effects model is used when background wells are not sampled at the same time. To illustrate how to use this model, input data in Table 3 are used except to change the sample year from 1992 to 1991 for one of the background well 2-W26-7. The ANOVA table is presented in Table 5.

Table 5. ANOVA for a Two-Way Nested Classifications Random Effect Model.

Source	Degrees of Freedom	Sum of Squares	Mean Squares
Well	1	2,346.125	2,346.125
Time	6	417.750	69.625
Analysis	24	91.000	3.792
Total	31	2,854.875	

Using formulas given in Section 2.1.2, the following estimates of variance components and (%) are obtained.

	Variance Components	%
$\hat{\sigma}_A^2 = MS_A =$	3.79	2.33
$\hat{\sigma}_T^2 = (69.625-3.792)/4 =$	16.46	10.13
$\hat{\sigma}_W^2 = (2{,}346.125-69.625)/16 =$	142.28	87.54
Total	162.53	100.00

Just like the "crossed" case, spatial variability accounts for more than 87% of the total variance and only a small percent (less than 3%) of the total variability is due to analytical error. The total unbiased variance of background measurements is estimated to be 162.53 $(\mu mho/cm)^2$. A background standard deviation of 12.75 (µmho/cm) ($\sqrt{162.53} = 12.75$) should be used when calculating confidence intervals, prediction intervals, and/or tolerance intervals.

Let us use the more general "nested" model to show how to use information gained in the variance components analysis to design a future sampling plan that will reduce the uncertainty associated with the estimate for the overall background level. The overall background mean µ is estimated by

$$\bar{y}_{...} = \Sigma_i \Sigma_j \Sigma_K y_{ijk}/abr \quad = 7{,}974/32 \approx 249 \ \mu mho/cm, \text{ and}$$

$Var(\bar{y}_{...})$ is estimated by $MS_W/abr = 2{,}346.125/32 \approx 73.3$ $(\mu mho/cm)^2$. Or, it can be estimated by

$$\hat{\sigma}_W^2/a + \hat{\sigma}_T^2/ab + \hat{\sigma}_A^2/abr = \frac{142.28}{a} + \frac{16.46}{ab} + \frac{3.79}{abr}$$

This expression indicates that a more precise estimate of $Var(\bar{y}_{...})$ will be gained by increasing the denominator 'a' (the number of background wells) because spatial variability is the most dominant factor in calculating the total variability of the data.

3.2 CASE STUDY 2 - WILCOXON RANK TEST USING ARSENIC TEST CASE

To illustrate the WRS test for the purpose of verifying cleanup, a test case is provided. The background data set for this test consists of arsenic concentrations from 10 wells (m=10) in the Rattlesnake Mountain Corridor. Another 32 (n=32) wells across the Hanford Site that lie outside of known contamination plumes are selected to simulate a waste site that has been remediated. The locations of these 42 (m+n or N=42) wells are presented in Figure 3-1. The raw data (Table 6) consists of U.S. Geological Survey results from various wells sampled at the Hanford Site during 1977 through 1984. The primary criteria used in the selection of wells include site geohydrology, well characteristics, and distribution of indicator contaminants in the groundwater (i.e., tritium less than 1,000 pCi/L). The Kolmogorov-Smirnov two-sample test procedure has been applied to this data set in WHC (1993).

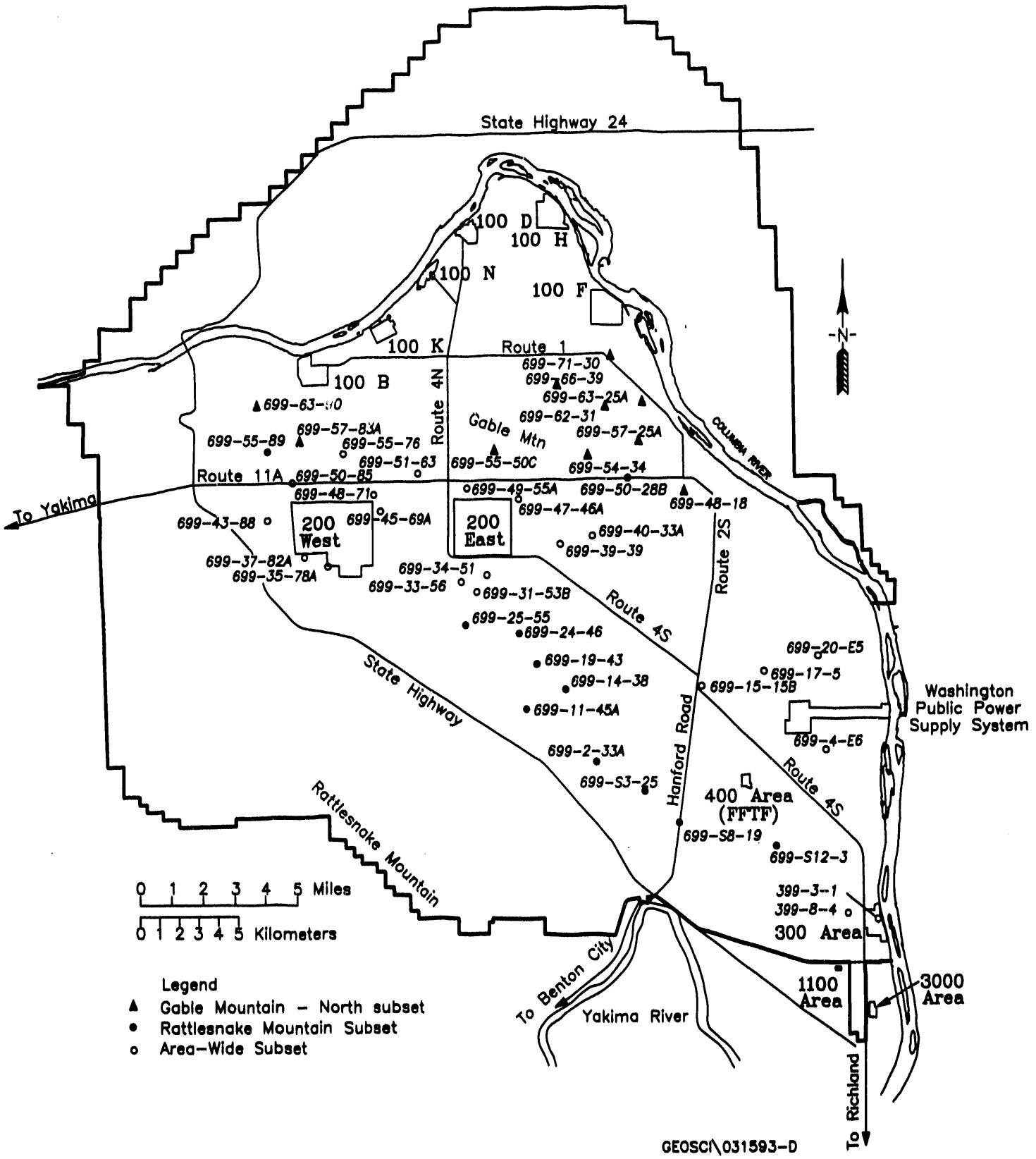


Figure 3-1. Monitoring Well Locations Outside of Known Contamination Plume Areas, Source: WHC (1993).

Table 6. Arsenic Data From 10 Wells in the Rattlesnake Mountain Corridor and 32 Wells Across the Hanford Site.

Region[a]	Well Code	Well Name	Arsenic (ppb)	Region[a]	Well Code	Well Name	Arsenic (ppb)
1	1	6-S12-3	5.5	2	22	6-45-69	4.0
1	2	6-S8-19	11.0	2	23	6-71-30	6.0
1	3	6-S3-25	6.0	2	24	6-51-63	4.0
1	4	6-2-33	6.5	2	25	6-57-25A	13.5
1	5	6-11-45A	3.0	2	26	6-37-82A	2.0
1	6	6-14-38	4.0	2	27	6-43-88	0.5
1	7	6-19-43	3.0	2	28	6-48-18	1.0
1	8	6-24-46	1.0	2	29	6-50-85	2.0
1	9	6-25-55	3.0	2	30	6-55-50C	3.0
1	10	6-55-89	1.0	2	31	6-57-83	1.0
2	11	6-20-E5-0	5.0	2	32	6-62-31	1.0
2	12	6-33-56	4.7	2	33	6-63-25A	4.0
2	13	6-35-78	3.0	2	34	6-66-39	1.0
2	14	6-4-86	7.0	2	35	3-8-4	8.0
2	15	6-49-55	6.6	2	36	6-15-15B	8.0
2	16	6-48-71	1.0	2	37	6-17-5	3.0
2	17	6-63-90	3.0	2	38	3-3-1	1.0
2	18	6-50-28B	7.0	2	39	6-40-33	10.0
2	19	6-34-51	6.0	2	40	6-39-39	2.0
2	20	6-55-76	1.3	2	41	6-47-46	3.0
2	21	6-31-53B	6.0	2	42	6-54-34	1.0

[a]Region 1 = Rattlesnake Mountain Corridor,
Region 2 = Hanford Site.

The combined data set is ordered and ranked from smallest to largest. Average ranks are assigned to the ties, the result is shown in Table 7 below. The sum of ranks assigned to the waste cleanup unit is 675.5 ($=W_{rs}$).

Since m=10 and n=32, and ties are present, the formula given in step 8 of Section 2.2.1 is used to calculate Z_{rs}. There are g=7 groups of ties: 1 group with t_j=9; 1 group with t_j=3; 1 group with t_j=8; 2 groups with t_j=4; and 2 groups with t_j=2 (see Table 7, column t_j). Therefore,

t_j (1)	Number of Groups (2)	$t_j(t_j^2 - 1)$ (3)	Product of (2)x(3)
9	1	720	720
3	1	24	24
8	1	504	504
4	2	60	120
2	2	6	12
		Total =	1,380

and $\Sigma_j\ t_j(t_j^2-1) = 1,380$, and

$$Z_{rs} = \frac{675.5 - 32(42+1)/2}{\{(32*10/12)[42+1-1,380/42(42-1)]\}^{1/2}}$$

$$= \frac{-12.5}{33.5454}$$

$$= -0.373$$

Since Z_{rs} = -0.373 is less than $Z_{1-\alpha}=Z_{0.95}$ = 1.645 (obtained from a standard normal distribution table), we cannot reject the null hypothesis that background based cleanup standard is attained at this site with α = 0.05. Therefore, there is no statistical evidence that the cleanup unit has not attained the cleanup standard in this hypothetical post-remediation test case.

3.3 CASE STUDY 3 -- QUANTILE TEST USING THE ARSENIC TEST CASE

The Quantile test described in Section 2.2.2 is illustrated using the same set of data as used for case study 2. The following steps are performed:

1. Specify α = 0.05.

2. There are m=10 arsenic measurements from the background area and 32 measurements from the waste site. Let q=0.8.

3. Compute r = N(1-q) = 42*0.2 = 8.4 ≈ 9 (round up to 9). It means we have to examine the largest 9 observations from the combined data set.

4. Order the combined data set from smallest to largest (see Table 7). Count the number, k, of measurements from the cleanup unit that are among the largest 9 observations. In Table 7, we find k = 7.

Table 7. Result of the WRS Test--Arsenic Test Case. (sheet 1 of 2)

Region[a]	Well Code	Well Name	Arsenic (ppb)	Rank	New Rank[b]	Background Rank	Site Rank	t_j
2	27	6-43-88	0.5	1	1		1	
1	10	6-55-89	1.0	2	6	6		9
1	8	6-24-46	1.0	3	6	6		
2	38	3-3-1	1.0	4	6		6	
2	32	6-62-31	1.0	5	6		6	
2	34	6-66-39	1.0	6	6		6	
2	31	6-57-83	1.0	7	6		6	
2	42	6-54-34	1.0	8	6		6	
2	28	6-48-18	1.0	9	6		6	
2	16	6-48-71	1.0	10	6		6	
2	20	6-55-76	1.3	11	11		11	
2	29	6-50-85	2.0	12	13		13	3
2	40	6-39-39	2.0	13	13		13	
2	26	6-37-82A	2.0	14	13		13	
1	9	6-25-55	3.0	15	18.5	18.5		8
1	5	6-11-45A	3.0	16	18.5	18.5		
1	7	6-19-43	3.0	17	18.5	18.5		
2	13	6-35-78	3.0	18	18.5		18.5	
2	37	6-17-5	3.0	19	18.5		18.5	
2	17	6-63-90	3.0	20	18.5		18.5	
2	41	6-47-46	3.0	21	18.5		18.5	
2	30	6-55-50C	3.0	22	18.5		18.5	
1	6	6-14-38	4.0	23	24.5	24.5		4
2	22	6-45-69	4.0	24	24.5		24.5	
2	33	6-63-25A	4.0	25	24.5		24.5	
2	24	6-51-63	4.0	26	24.5		24.5	
2	12	6-33-56	4.7	27	27		27	
2	11	6-20-E5-0	5.0	28	28		28	
1	1	6-S12-3	5.5	29	29	29		
1	3	6-S3-25	6.0	30	31.5	31.5		4
2	19	6-34-51	6.0	31	31.5		31.5	
2	23	6-71-30	6.0	32	31.5		31.5	
2	21	6-31-53B	6.0	33	31.5		31.5	
1	4	6-2-33	6.5	34	34	34		
2	15	6-49-55	6.6	35	35		35	

Table 7. Result of the WRS Test--Arsenic Test Case. (sheet 2 of 2)

Region[a]	Well Code	Well Name	Arsenic (ppb)	Rank	New Rank[b]	Background Rank	Site Rank	t_j
2	14	6-4-86	7.0	36	36.5		36.5	2
2	18	6-50-28B	7.0	37	36.5		36.5	
2	36	6-15-15B	8.0	38	38.5		38.5	2
2	35	3-8-4	8.0	39	38.5		38.5	
2	39	6-40-33	10.0	40	40		40	
1	2	6-S8-19	11.0	41	41	41		
2	25	6-57-25A	13.5	42	42		42	
				Total		227.5	675.5	

[a]Region 1 = Rattlesnake Mountain Corridor (i.e., background data set)
Region 2 = Hanford Site (i.e., waste cleanup unit).
[b]Average ranks are assigned to the tied group.

5. Because $r \leq 20$, we calculate the probability of obtaining a value of k as large or larger than 7 using the formula for P (see Step 6 of Section 2.2.2)

$$P = \frac{\Sigma_i \binom{m+n-r}{n-i}\binom{r}{i}}{\binom{m+n}{n}}$$

$$= 0.6385$$

Note in calculating P, the summation over i is for i=k (=7) to i=r (=9). In other words, there are 3 terms (i=7, i=8, and i=9) in the above equation.

6. Do not reject H_o because P = 0.6385 > 0.05 (the specified value of α). Therefore, there is no statistical evidence that the cleanup unit has not attained the cleanup standard.

4.0 SUMMARY AND CONCLUSIONS

Results of the random effects model analyses indicate that spatial variability is the major contributing factor to the total variability. Only specific conductance values from a RCRA facility were used for demonstration purposes, however, experience suggests this statement is generally true for all of the constituents occurring naturally in the groundwater at the Hanford Site. Background levels calculated without considering spatial variability will be conservative (too low) and lead to unnecessary remediation because s^2 will underestimate the true variance. If upper confidence levels, tolerance intervals, and/or prediction intervals are to be used as threshold values for determining the presence of contamination, one should use the unbiased estimator in the calculations. In addition, to obtain a more precise estimate

precise estimate of Var($\overline{y}_{...}$), one should use the results of the variance components analysis as a guide. For example, if spatial variability is the most important factor in the total variance components, the most effective way to reduce the uncertainty in the estimation of background mean is to increase the number of background wells. To offset the cost due to sampling of more wells, one can decrease the number of sampling times and reduce the number of replicate analyses, especially when the cost associated with each chemical analysis is high.

To effectively collect and utilize groundwater monitoring data in the four programmatic areas at the Hanford Site, the background region should be selected from area(s) not influenced by the operations of the hazardous waste site and similar to the test site in physical, chemical, or biological characteristics. Furthermore, concentrations of chemicals in groundwater vary considerably depending on factors such as soil characteristics, proximity to recharge and discharge areas, and flow rates. Additionally, background should be considered as a statistical distribution of concentration levels, rather than a single concentration, so that statistical techniques discussed in this paper can be applied.

Finally, when making contamination and/or remediation decisions about a waste site, all available information must be used. In addition to statistical test procedures, geochemical and hydrologic considerations are integral parts of the decision making process. A phased approach, as shown in Figure 4-1, is recommended. The phases proceed from simple to the more complex, and from an overview to detailed analysis. All phases should be completed and evaluated before a decision is reached. Work is in progress toward this approach.

5.0 REFERENCES

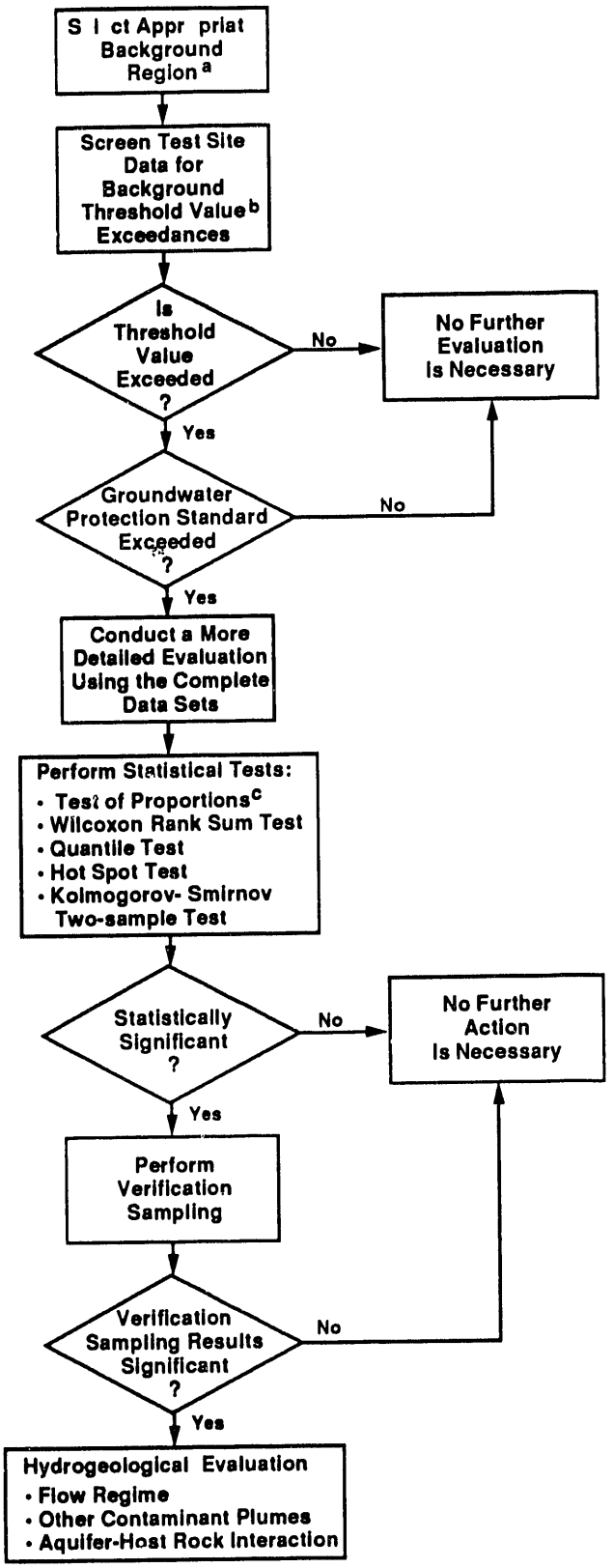
Conover, W. J., 1980, *Practical Nonparametric Statistics*, Second Edition, John Wiley and Sons, Inc., New York, New York, pp. 171, 216-223.

EPA, 1989, *Interim Status Standards for Owners and Operators of Hazardous Waste Treatment, Storage, and Disposal Facilities*, Title 40, Code of Federal Regulations, Part 265, as amended, U.S. Environmental Protection Agency, Washington, D.C.

Gilbert, R. O. and J. C. Simpson, 1990, *Statistical Sampling Analysis Issues and Needs for Testing Attainment of Background-Based Cleanup Standards at Superfund Sites*, PNL-SA-17907, Pacific Northwest Laboratory, Richland, Washington.

Gilbert, R. O. and J. C. Simpson, 1992, *Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media*, PNL-7409 Vol. 3, Rev.1, Pacific Northwest Laboratory, Richland, Washington.

Hicks, C. R., *Fundamental Concepts in the Design of Experiments*, Third Edition, CBS College Publishing, New York, New York, pp. 214-220.



[a]Concensus Among DOE, WHC, EPA, and Ecology (Work in Progress)

[b]Background Threshold Value Determined Based on Background Measurements. (In Most Cases, this is the Maximum Observed Value for a Particular Analyte of Concern.)

[c]If More than 50% of the Samples in the Background Data S t are B low th Limit of Detection

Figure 4-1. A Phased Approach to Contamination/Remediation Decisions.

Johnson, R. A., S. Verrill, and D. H. Moore II, 1987, "A Robust Two-Stage Multiple Comparison Procedure with Application to a Random Drug Screen." *Biometrics* 45:1281-1297.

Milliken, G. A. and D. E. Johnson, 1984, *Analysis of Messy Data*, Wadsworth, Inc., Belmont, California, pp. 212-273.

Ostle, B. and L. C. Malone, 1988, *Statistics in Research, Basic Concepts and Techniques for Research Workers*, Fourth Edition, Iowa State University Press, Ames, Iowa, pp. 158-160, 636.

WHC, 1993, *Hydrologic Characterization Plan: I. Principal Recharge Zone for the Hanford Site 200 Areas Waste Storage and Disposal Sites, Cold Greek Valley Area, Western Hanford Site*, WHC-SD-EN-AP-133, Westinghouse Hanford Company, Richland, Washington.

DISTRIBUTION

Number of copies		
ONSITE		
10	U.S. Department of Energy-Richland Operations Office	
	R. D. Freeberg	A5-19
	M. J. Furman (4)	A5-21
	J. M. Hennig	A5-21
	R. G. McLeod	A5-19
	R. P. Saget	A5-52
	Public Reading Room (2)	A1-65
4	Pacific Northwest Laboratory	
	J. V. Borghese	K6-96
	M. A. Chamness	K6-96
	S. S. Teel	K6-96
	Technical Files	K1-11
58	Westinghouse Hanford Company	
	D. J. Alexander	H6-06
	D. B. Barnett	H6-06
	D. J. Brown	S1-06
	J. A. Caggiano	H6-06
	C. J. Chou (15)	H6-06
	M. H. Edrington	H6-06
	K. R. Fecht	H6-06
	B. H. Ford	H6-06
	P. B. Freeman	H6-06
	M. J. Hartman	H6-06
	F. N. Hodges (5)	H6-06
	D. G. Horton	H6-06
	R. L. Jackson	H6-06
	V. G. Johnson (5)	H6-06
	J. C. Johnston	H6-06
	D. K. Jungers	H6-06
	G. L. Kasza	H6-06
	A. J. Knepp	H6-06
	A. G. Law	H6-06
	R. B. Mercer	H6-06
	R. E. Peterson	H6-06
	S. P. Reidel	H6-06
	V. J. Rohay	H6-06
	J. S. Schmid	H6-06
	J. A. Serkowski	H6-06
	L. C. Swanson	H6-06
	R. R. Thompson	L4-96
	D. K. Tyler	H6-06
	B. A. Williams	H6-06
	Information Release Administration (3)	H4-17
	Central Files (2)	L8-04
	EPIC (2)	H6-08

DATE
FILMED
12/8/93

END