SPECIAL REPORT 96-9

Comparison Criteria for Environmental Chemical Analyses of Split Samples Sent to Different Laboratories

Corps of Engineers Archived Data

Clarence L. Grant, Thomas F. Jenkins and Anand R. Mudambi May 1996

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

DTIC QUALITY INSPECTED 3

Abstract

Data comparison criteria were developed for quality control (QC) and quality assurance (QA) chemical analyses obtained during environmental studies directed by the U.S. Army Corps of Engineers (USACE). Archived results for 124 sets of eight metals in soils, 69 sets of fourteen volatile organic compounds (VOCs) in soils, 163 sets of total petroleum hydrocarbons (TPH) in soils, 79 sets of six explosives in soils, and 153 sets of fourteen VOCs in groundwater were analyzed statistically. Concentration ratios (QC_1/QC_2 and QC/QA) were lognormally distributed and this was the model used for comparisons. For both metals in soils and VOCs in groundwater, duplicate QC results should yield ratios between 0.50–2.00; less than 4% of the archived results were outside these limits. For QC/QA ratios, the limits suggested are 0.40–2.50; only 10.2% of metals ratios in soils and 5.6% of the VOCs in groundwater exceeded these limits. Considering that both methods are multi-analyte, we find only 4.0% of the metals samples and 2.0% of the VOC samples had more than one offending ratio per sample. Application of these limits to recent analyses produced very similar results. For VOCs, TPH, and explosives in soils, temporary limits of 0.25–4.00 are suggested with the understanding that improvements leading to tightening should be energetically pursued. Even with these wide limits, approximately 42% of VOCs, 14% of TPH, and 11% of explosives QC/QA ratios must be considered outliers. Here, too, recent analyses yielded very comparable percentages. The situation for VOCs in soils requires immediate attention to improve sampling and preparation procedures, but all methods can and should be capable of producing improved agreement between laboratories. Numerous recommendations are offered with this goal in mind.

For conversion of SI units to non-SI units of measurement consult ASTM Standard E380-93, *Standard Practice for Use of the International System of Units,* published by the American Society for Testing and Materials, 1916 Race St., Philadelphia, Pa. 19103.

This report is printed on paper that contains a minimum of 50% recycled material.

Special Report 96-9

US Army Corps of Engineers
Cold Regions Research & Engineering Laboratory

Comparison Criteria for Environmental Chemical Analyses of Split Samples Sent to Different Laboratories

Corps of Engineers Archived Data

Clarence L. Grant, Thomas F. Jenkins and Anand R. Mudambi

May 1996

Prepared for
OFFICE OF THE CHIEF OF ENGINEERS

19960628 020

Approved for public release; distribution is unlimited.

PREFACE

This report was prepared by Dr. Clarence L. Grant, Professor Emeritus, University of New Hampshire, Durham, New Hampshire; Dr. Thomas F. Jenkins, Research Chemist, Geological Sciences Division, Research and Engineering Directorate, U.S. Army Cold Regions Research and Engineering Laboratory (CRREL), Hanover, New Hampshire; and Dr. Anand R. Mudambi, Chemist, Chemical Data Quality Management Branch, HTRW Center of Expertise, U.S. Army Corps of Engineers, Omaha, Nebraska.

Funding for this work was provided by Headquarters, U.S. Army Corps of Engineers, Washington, D.C., through the Missouri River Division (USAED-MRD), MIPR# GP95-11.

The authors acknowledge the useful comments provided by Martin H. Stutz, U.S. Army Environmental Center, Aberdeen Proving Ground, Maryland, and Ann B. Strong, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi, during review of this manuscript. They also acknowledge the initial guidance provided by Kevin Coats of the Missouri River Division, Corps of Engineers, and Dr. David Koran, Headquarters, Corps of Engineers, who helped formulate the goals of this study. Brian Condike, formerly of the New England Division Laboratory and currently with the Southwest Division Laboratory, and Dr. Bruce Heitke, Headquarters, Corps of Engineers, are also acknowledged for providing useful comments on the initial draft of this report. The authors also are indebted to Gerald Roy, Information Systems Branch, CRREL, for selection of the data management system used to assemble and process the data used in this work, and to Jane G. Mason for her help in assembling and editing the manuscript.

The contents of this report are not to be used for advertising or promotional purposes. Citation of brand names does not constitute an official endorsement or approval of the use of such commercial products.

CONTENTS

	Page
Preface	ii
Introduction	1
Data comparison methodology	2
Results and discussion	6
Metals in soils	6
VOCs in soils	8
TPH in soils	12
Explosives in soils	15
VOCs in groundwater	17
Conclusions and recommendations	18
General recommendations	18
Metals in soils	18
VOCs in soils	19
TPH in soils	19
Explosives in soils	19
VOCs in groundwater	20
Summary	20
Literature cited	20
Appendix A: Data and sample calculations for chromium	22
Appendix B: Data for metals, VOCs, TPH, and explosives	27
Abstract	32

ILLUSTRATIONS

Figure

1. Frequency histogram for soil Cr QC/QA concentration ratios 2
2. Lognormal probability plots for soil and for groundwater QC/QA concentration ratios 3
3. Distribution of soil As, Ba, Cr, and Pb QC/QA concentration ratios outside the acceptable range of 0.30–3.00 as a function of QA concentrations normalized to their respective mean concentrations 6
4. Frequency histogram for soil VOC QC/QA concentration ratios 9
5. Lognormal probability plot for QC/QA concentration ratios of all VOCs in soils with ratios between 0.10–10.0 9
6. Lognormal probability plots for QC/QA concentration ratios for soil VOCs 10
7. Correlation of log concentration ratios for total xylenes and ethylbenzene in soil 12

Figure Page

8. Comparison of frequency histograms of soil TPH QC/QA concentration ratios from NPD and MRD 12
9. Lognormal probability plot for soil TPH QC/QA concentration ratios between 0.25 and 4.00 for NPD results 13
10. Frequency histogram for soil TPH QC_1/QC_2 concentration ratios from NPD 13
11. Lognormal probability plot for soil TPH QC_1/QC_2 concentration ratios between 0.25 and 4.00 for NPD results 13
12. Correlation of soil TPH concentration ratios for CAS, Inc./NET Pacific, Inc. with soil TPH concentration estimates reported by NET Pacific, Inc. 14
13. Frequency histogram for soil explosives QC/QA concentration ratios from MRD 15
14. Lognormal probability plot for QC/QA concentration ratios of explosives in soils with ratios between 0.25–4.00 16

Comparison Criteria for Environmental Chemical Analyses of Split Samples Sent to Different Laboratories

Corps of Engineers Archived Data

CLARENCE L. GRANT, THOMAS F. JENKINS, AND ANAND R. MUDAMBI

INTRODUCTION

In conjunction with hazardous waste remedial activities, the U.S. Army Corps of Engineers (USACE) requires that quality control (QC) and quality assurance (QA) samples be collected and analyzed by a contract laboratory and a USACE laboratory, respectively. Procedures for these activities are prescribed in *Chemical Data Quality Management for Hazardous Waste Remedial Activities* (USACE 1990). The QC and QA samples represent about 10% of the field samples and are splits or co-located samples. Splits are used for analyses such as metals in soils where bulk samples can be homogenized and subdivided in the field without losses or contamination. Co-located samples are required for volatile organic compounds (VOCs) in soil and water samples to reduce losses in handling, and are often preferred for other analytes in groundwater and surface water samples. Since co-located samples are not homogenized, the contribution of natural sample variability can increase differences compared to splits. Field personnel also collect rinsate blanks, trip blanks for VOCs, and background soil and groundwater samples. The identity of all QC samples is kept blind to analysts and laboratory personnel until data are in deliverable form.

Acceptability of contractor laboratory data is determined by a USACE QA laboratory. Guidelines exist for this assessment except for the numerical comparison of QC and QA analytical results from split or replicate samples. Currently, limits for these comparisons are not based on technical criteria and vary from one QA laboratory to another.

It was decided that an analysis of archived QC/QA data might be useful in establishing some realistic numerical comparison criteria. The major shortcoming of the archived data is the absence of any basis for assuming that the QA results are more accurate than the QC data being evaluated. In fact, a contractor laboratory generating QC results on one project may serve as the QA laboratory on another project. Lacking evidence to the contrary, we must assume similar uncertainties associated with both data sets.

The archived data received were as follows:

a) Forty-nine data reports for metals in soils from the New England Division (NED) laboratory, labeled M1-1 to M1-49, and 75 reports from the North Pacific Division (NPD) laboratory, labeled M3-1 to M3-75. These results were from 37 project locations in eight states and involved 16 laboratories.

b) Sixty-six data reports of volatile organic compounds (VOCs) in soils supplied by the Missouri River Division (MRD), labeled VS2-01 to VS2-66 (VS2-60 to VS2-66 used GC methods instead of GC/MS), and three reports from the NPD, labeled VS3-1 to VS3-3. These results were from 26 project locations in 15 states and involved 21 laboratories.

c) One hundred and two data reports for total petroleum hydrocarbons (TPH) in soils supplied by the NPD, labeled TS3-01 to TS3-102, and 61 reports from the MRD, labeled TS2-01 to TS2-61. Analysis was by Method 8015M, GC-FID or equivalent. These results were from 17 project locations in three states and involved eight laboratories.

d) Seventy-one data reports for explosives in soils supplied by MRD, labeled ES2-01 to ES2-71, and eight reports from NPD, labeled ES3-01 to ES3-08. Analyses were by method 8330. These results were from eight project locations in eight states and involved eight laboratories.

e) One hundred and thirty-four data reports for VOCs in groundwater from the MRD laboratory, labeled V2-1 to V2-134, and 23 from the NPD laboratory, labeled V3-1 to V3-23. These results were from 39 project locations in more than 21 states and involved more than 15 laboratories.

DATA COMPARISON METHODOLOGY

The first approach considered was a plot of QC concentration estimates vs. associated QA estimates. This is a very useful comparison technique. A linear regression line can be fitted by minimizing the residual sum of squares for the QC estimates while assuming zero error in the QA estimates. For an ideal system with perfect agreement between QC and QA values, the fitted model would be linear through the origin with a slope of 1.00. The extent of departure from this ideal can be used to calculate confidence limits for individual measurements at some selected probability level. Confidence intervals for individual measurements are called tolerance bands.

After reviewing the data sets submitted, we decided that the concentration ranges were much too wide to be fitted to individual models. Furthermore, there is no basis for assigning referee status to the QA values, and the data are not normally distributed, which is the underlying assumption of this approach. It is also likely that the more complex computations required by this approach might impede routine usage. Consequently, we decided to examine QC/QA concentration ratios for between-laboratory comparisons and QC_1/QC_2 ratios for comparison of within-laboratory replicates. This approach is similar to the first one considered except that no regression model is fitted. When duplicate QC values were given, the first listed value was always used for the QC/QA computation. For metals in soils and VOCs in groundwater, ratios below 0.30 and above 3.00 were designated outliers and excluded from further computations. Because of the larger unavoidable uncertainties attached to the estimation of low concentration organic analytes in soils, these limits were expanded to 0.25–4.00 for TPH and explosives and to 0.10–10.0 for VOCs. While this practice may seem arbitrary, exclusion of a very few highly extreme values is necessary to avoid unreasonable increases of standard deviations and the associated confidence bands for a data set. Except for VOCs in soils and one set of

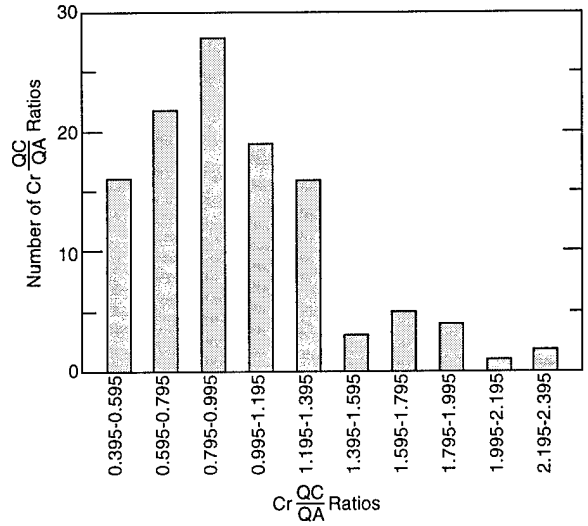


Figure 1. Frequency histogram for soil Cr QC/QA concentration ratios.

TPH results, the percent exclusions were always less than 13% and often well below 10%.

We should not expect the QC/QA ratios to be normally distributed using linear coordinates. A factor of 2 difference in QC and QA concentrations would lead to a low side ratio of 0.50 and a high side ratio of 2.00. Obviously these two values are not symmetrically distributed around the ideal value of 1.00 for perfect agreement. A histogram of the soil Cr QC/QA results (Fig. 1) shows that the typical shape of these distributions is skewed toward the high end. This histogram suggests that a lognormal distribution should describe the results. Simply stated, this means that the logs of the ratios will form a normal distribution when plotted as shown in Figure 1.

An effective method of testing the hypothesis that the ratios are lognormally distributed is to use Lognormal Probability graph paper in which the ordinate is a % probability scale and the abscissa is a log scale. If the lognormal model is correct, a straight line will result. To plot the Cr ratios shown in Figure 1, we convert the number of ratios in each cell to a cumulative probability. From Figure 1 there are 16 ratios between 0.395 and 0.595. Since the total number of ratios was 116, the probability of values in this cell is

$$\frac{(16)(100)}{116} = 13.8\%.$$

This value is plotted on the probability scale vs. the upper boundary of the cell, 0.595. The next cell (0.595–0.795) contained 22 ratios and the cumulative percent was

$$\frac{(16+22)(100)}{116} = 32.8\%,$$

which is plotted vs. 0.795, etc. (see Fig. 2c). Figures 2a, 2b, 2c, and 2d are log-cumulative probability plots of QC/QA ratios for soil As, Ba, Cr, and Pb respectively. Although a few points deviate slightly from the linear ideal for lognormal results, the fits are quite good and justify analyzing the data as being lognormally distributed. Plots for QC_1/QC_2 ratios behaved similarly, as expected.

Lognormal plots for individual values rather than groups are illustrated using four VOCs in groundwater (Fig. 2e–2h). Individual rather than grouped probabilities were used because of the relatively small number of ratios. Individual probabilities are calculated from

$$\frac{(i-0.5)(100)}{n}$$

where n is the total number of ratios in a set and i is the rank after the ratios are arranged from lowest to highest. Thus for 20 ratios ranging from 0.40 to 2.75, the first ranked one (0.40) would be plotted on the abscissa vs. a probability of

$$\frac{(1-0.5)\ (100)}{20} = 2.50\%$$

and the next higher ratio would be plotted vs.

$$\frac{(2-0.5)(100)}{20} = 7.50\%,$$

etc. Benzene (Fig. 2e) has several points at each extreme that deviate a good deal from linearity. However, a linear probability plot was even less satisfactory. Chloroform and ethylbenzene (Fig. 2f and 2g) fit the lognormal model with only minimal scatter while toluene (Fig. 2h) provides a fair fit. Other VOCs not shown behaved similarly to the ones shown. In view of these VOC plots, those for other analytes, and the expected distribution of ratios, we decided to treat all ratios by the same computational procedures. Since the lognormal model will be used, we must compute geometric means and geometric standard deviations as shown below.

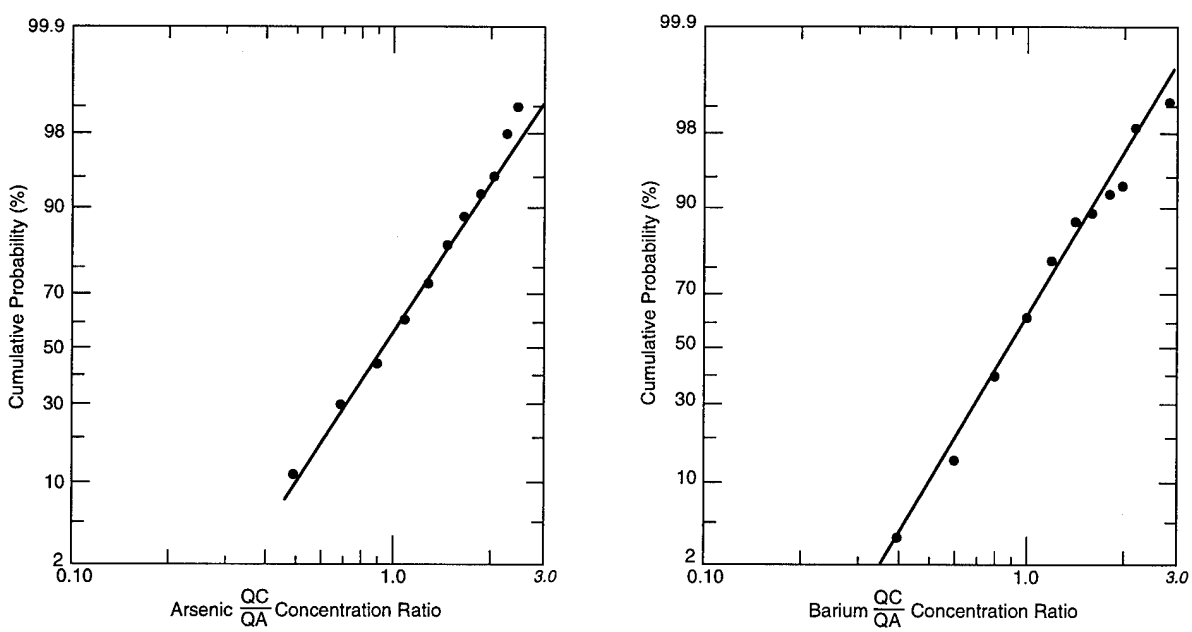
1. Compute QC_1/QC_2 ratios and their logarithms when QC_1/QC_2 results are available.

2. Compute mean and standard deviation of the logs.

3. Compute 95% confidence limits for the mean of the logs according to the equation

$$\bar{x} \pm \frac{t_{0.95}\, S_x}{\sqrt{n}}$$

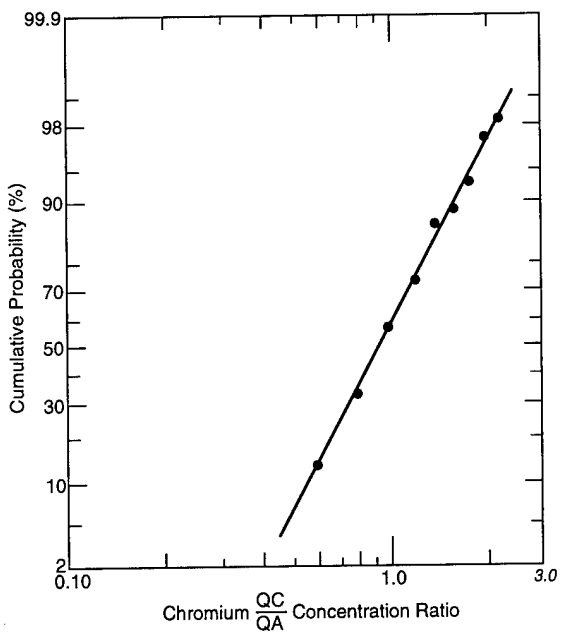
where $\bar{x}$ is the mean of the logs, $t_{0.95}$ is student's t for n–1 degrees of freedom (d.f.) and a two-tail alpha risk of 5% (α = 0.05), S_x is the standard



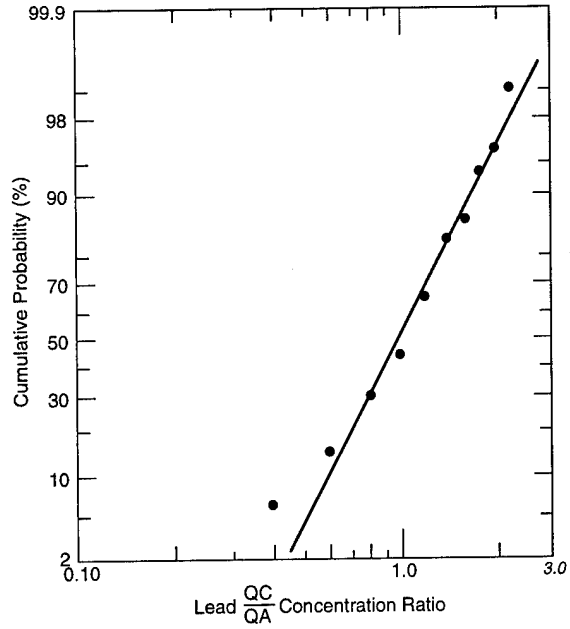
a. Soil/arsenic ratio.

b. Soil/barium ratio.

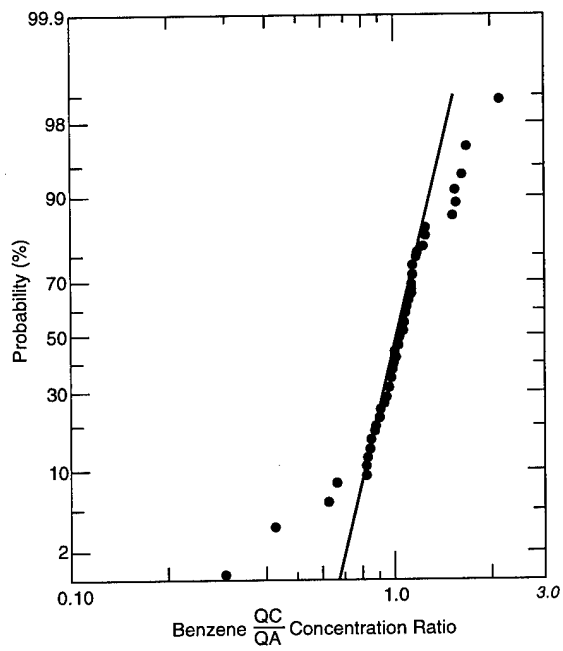
Figure 2. Lognormal probability plots for soil and for groundwater QC/QA concentration ratios.



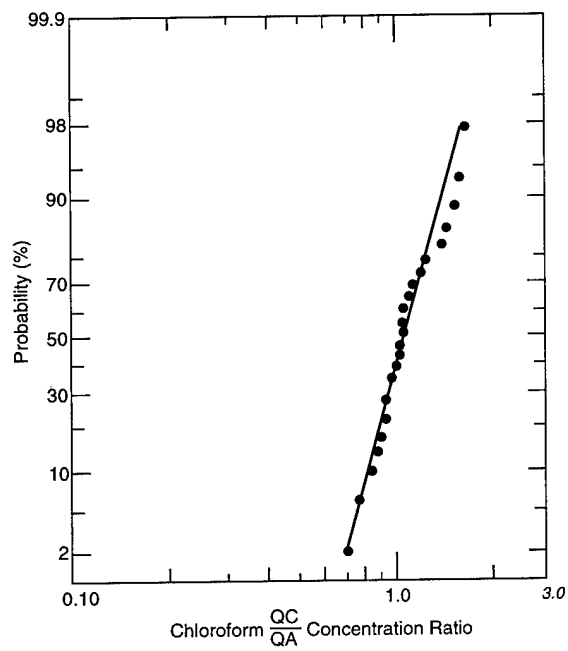
c. Soil/chromium ratio.



d. Groundwater/lead ratio.



e. Groundwater/benzene ratio.



f. Groundwater/chloroform ratio.

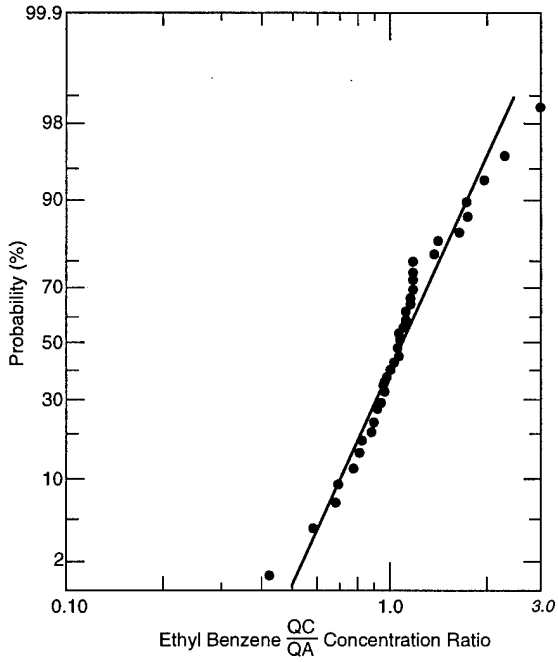
Figure 2 (cont'd). Lognormal probability plots for soil and for groundwater QC/QA concentration ratios.

deviation of the logs, and n is the number of logs in the set.

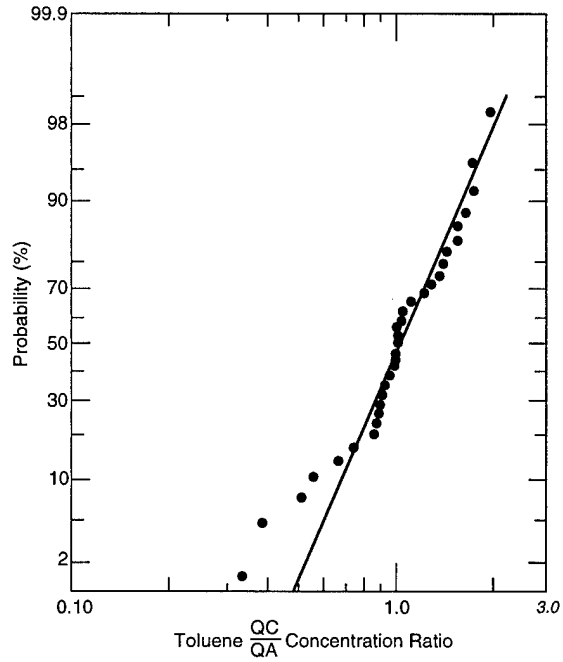
4. Obtain the antilogs of the mean and the upper and lower 95% confidence limits to see if the limits bracket the expected mean of 1.00. These confidence limits define a range of values within which we may reasonably assume (95% probability in this case) the true value lies if there were no systematic error present. Since the expected true mean for each set of ratios is 1.00, confidence limits that do not bracket 1.00 suggest some systematic bias in the experimental results.

5. Compute 99% tolerance limits for individual ratios from $\bar{x} \pm t_{0.99}(S_x)$ for the logs.

6. Obtain antilogs for these tolerance intervals to see if there are any values outside the limits.



g. Groundwater/ethylbenzene ratio.



h. Groundwater/toluene ratio.

Figure 2 (cont'd).

These tolerance intervals represent the lower and upper limits within which we may assume (99% probability here to give minimal probability of an outlier) that all individual ratios in a set should fall. If an individual ratio is outside these limits, it is assumed to be an outlier and it can be rejected. This decision is based on the assumption that the lognormal model is correct.

7. Assuming the expected (ideal) mean log of 0.00, obtain antilogs for $-t_{0.99}(S_x)$ to $+ t_{0.99}(S_x)$. Determine how many values are outside of these limits. NOTE: Do not count the outliers excluded before these calculations were started.

8. Repeat steps 2 through 7 for QC/QA ratios. An example of these computations is presented in Appendix A for the Cr results.

When both QC and QA laboratories reported "less than" values, the ratios of these values were checked to see if they were within a factor of 3.0 or 4.0, depending on the analyte. When one laboratory reported a value above the reporting limit and the comparison result was below the reporting limit, they were still compared using the appropriate factor. For example, if one laboratory reported 0.60 μg/g and the other reported not detected (ND < 1.0 μg/g), they were in agreement. But if one reported 0.60 μg/g and the comparison value was ND < 5.0, they disagreed because the "less than" value was unacceptably high. The most obvious disagreement would be if the two values were 5.0 μg/g and ND < 0.80 μg/g, as an example. Such comparisons were often unsatisfactory because the ranges of reporting limits were sometimes as wide as a factor of 1000 or more.

Inconsistency with significant figures was a problem throughout. For example, the QA laboratory might report a result of 10.7 μg/g and the QC laboratory might give the matching result as 8 μg/g. The result 10.7 μg/g implies uncertainty at the tenth of a μg/g whereas the result of 8 μg/g implies an uncertainty of ±1 μg/g. We will address this problem with a recommendation in a later section, but for the purpose of our computations we will treat 8 μg/g as though it is 8.0 μg/g. Similar problems were present for the reporting limit values.

Some analytes were determined by more than one method in different laboratories (inductively coupled plasma and graphite furnace atomic absorption in the case of lead, for example). We originally intended to segregate results according to method, but there were far too many data reports lacking method information to permit this to be done. In reality, we expect properly calibrated methods to yield comparable accuracy, so when numerical concentrations are reported using accepted methods, there should not be systematic bias in the results. Of course, reporting limits will vary for different methods.

The original plan called for reporting limits to be specified on all data reports for all analytes. However, many data reports showed either no reporting limits or only a few. The lack of complete data limited the comparisons of reporting limits to some extent.

For VOCs in groundwater concentrations, numerical surrogate recovery results were sparse. Often nothing was specified or the recoveries were simply labeled "OK." We had planned to use this information to aid in identifying outliers (unacceptable results), but too little data were present to permit this application.

A few values were flagged with "J" to designate that the value was estimated (below the reporting limit). However, these values yielded ratios that appeared to be about as reliable as the bulk of the results and, therefore, they were used in the comparison.

RESULTS AND DISCUSSION

The discussion that follows is based on the assumption that the lognormal model is adequate for these sets of ratios. The null hypothesis then is that geometric means of the ratios have an expected value of 1.00 and that 95% confidence limits based on logs should bracket 1.00 if there is no bias between the QC and the QA laboratories. Furthermore, 99% tolerance intervals based on similar assumptions should include nearly all individual ratios. Any ratios that fall outside of the tolerance limits should be flagged for further review.

Metals in soils

Results for As, Ba, Cr, and Pb are summarized in Table B1. We see that only 1.3% of the duplicate QC ratios were outside the arbitrarily chosen 0.30–3.00 range for classification as outliers. For Ba, Cr, and Pb, the expected mean ratio of 1.00 was bracketed by the 95% confidence limits, but the lower limit for As was slightly above 1.00, suggesting a slight bias. The 99% tolerance limits were all very similar, with only six values outside these boundaries. If we choose 0.50–2.00 as the regulatory limits for all four metals, the outlying Pb ratio is within tolerances and the total number of excluded ratios is 8(3+5), or 3.4%. This small percentage of unacceptable values seems reasonable, and the 0.50–2.00 range of ratios is recommended for duplicate QC samples. Incidentally, for the other four metals with few values above reporting limits, all ratios (12 Cd, 5 Hg, 1 Ag) were within this recommended range. When this range was used to analyze the most recent results from 104 samples (32 project locations), only 2% of the QC_1/QC_2 ratios were outside of these boundaries.

As expected, the QC/QA ratios were less reproducible than the duplicate QC ratios. In Table B1 we note that 29 values (6.3%) were designated outliers because they fell outside of 0.30–3.00. These 29 outliers are well distributed among the four metals, As, Ba, Cr, and Pb. The distribution of outliers relative to QA concentrations is shown in Figure 3. In order to plot results from four metals with very different concentration means, the QA concentration for each outlier was normalized to the mean QA concentration for that metal. Log-log scales were necessary to accommodate the wide ranges of values. Although there are a few more ratios below 0.30 than above 3.00 (16 vs. 13) and more ratios for concentrations below the normalized mean QA concentration of 1.00 than above (19 vs. 10), there is little basis to suggest that outliers occur primarily in extreme concentration regions. In fact, most of the outliers fall between one-half and two times the mean concentrations. For the 23 ratios available for Cd, Hg, and Se, three were outliers (all on the high

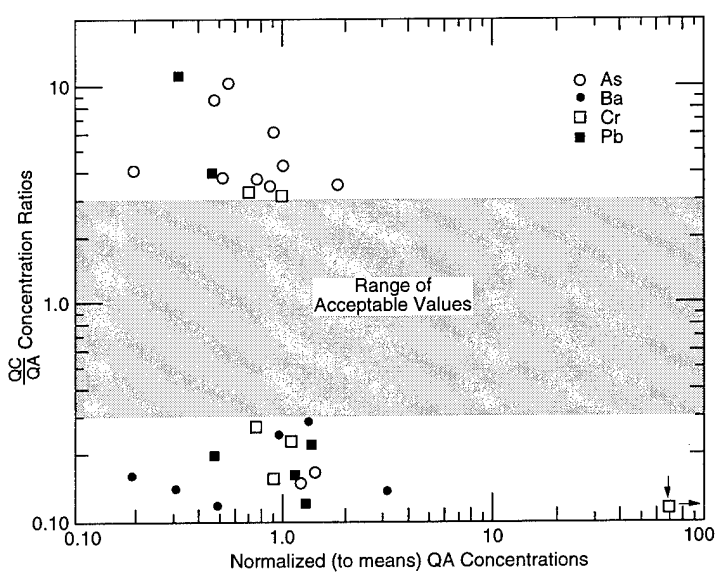


Figure 3. Distribution of soil As, Ba, Cr, and Pb QC/QA concentration ratios outside the acceptable range of 0.30–3.00 as a function of QA concentrations normalized to their respective mean concentrations.

side). There were no samples with Ag concentration above reporting limits for both QA and QC laboratories.

We also determined whether the outliers were for several metals in a few samples and whether they occurred with certain laboratories. However, the outliers were distributed among 25 different samples. Two samples had outliers for three metals and three samples had two metals outside of the limits. None of the latter five samples was from the same project. We lacked necessary information to state with certainty that different laboratories were associated with each of the five samples. In any case, we feel that outliers are occurring more or less randomly rather than as any systematic problem. Sample handling and/or analysis errors are likely causes.

Geometric means for all four metals were below the expected mean ratio of 1.00, but the 95% confidence limits bracketed 1.00 for all but Ba. The tendency of all the ratios to be below 1.00 indicates that QC laboratories are reporting slightly lower concentrations than QA laboratories, but this pattern is neither marked nor clearly established. The 99% tolerance intervals for Ba, Cr, and Pb were very similar. The somewhat wider limits for As is a reflection of its large standard deviation. Only five additional values beyond the original outliers are outside of these tolerance intervals.

It is instructive to consider the effects of imposing three possible sets of logarithmically distributed rejection limits. In Table B2 we see that the arbitrarily chosen limits of 0.30–3.00 produced 6.3% exclusions with nearly equal numbers of outliers below 0.30 and above 3.00 and reasonably similar numbers for each metal. If the limits are relaxed to 0.25–4.00, only 4.4% of the ratios are excluded. Two-thirds of the exclusions are on the low side, i.e., cases where the QC concentration is much lower than the QA concentration. In the other direction, if we use limits of 0.40–2.50, 10.2% of the ratios would be rejected. Once again two-thirds are on the low side, but the pattern is quite different for As than it is for the metals determined largely by plasma emission. The tendency for low QC results is absent for As but quite pronounced for Ba, Cr, and Pb. The most recent results (again, 32 project locations) contained 11.9% of the ratios outside 0.40–2.50 with similar percentages of low and high outliers.

Clearly the 0.40–2.50 limits would put pressure on laboratories to improve the accuracy of their analyses. Nonetheless, we believe that this range is reasonable for metals in soils and we recommend that it be used. Rejection of 10% of contract laboratory results would likely trigger demands to provide proof that the QA concentrations are sufficiently reliable to be used as referee values. We believe that this position is justified regardless of what regulatory limits are chosen.

A further very practical question remains: would an entire set of results for eight metals be rejected just because the QC/QA ratio for one metal was outside tolerance limits? We doubt it, especially if the results for the offending metal were low compared to concentrations of concern. We suspect that rejection would occur only when two or more ratios are outliers for a sample and only when concentrations are high. For the archived data, only five samples out of 124 (4.0%) had more than one ratio per sample outside of tolerance bands. Consequently, we believe ratios should be flagged if they are outside 0.40–2.50 and examined by a QA manager to determine the fate of the results. Better yet, an artificial intelligence computer program could probably be devised to do this task. With so few samples having multiple outliers, it seems that the 0.40–2.50 range for ratios is justified.

Although most results for Ag, Cd, Hg, and Se were less than the reporting limit (ND<), some insight can be gained by tabulating the reporting limits (Table B3). A few values were unusually high (27 of 722, or 2.4%) and were deleted without further consideration. Of the remaining values, eleven of the 17 unacceptable reporting limits were associated with only three samples, none of which were samples with excluded As, Ba, Cr, and Pb ratios. The concentrations of other metals in these three samples were not unusual. While matrix effects may be involved, we continue to believe that an occasional faulty analysis is the major culprit. After these exclusions, means and medians were in reasonable agreement, indicating that the distributions were not seriously skewed. Furthermore, the ranges and the means for QA and QC laboratories were also in good agreement. Even under these conditions, the ratios of QC/QA reporting limits were often outside 0.30–3.00 for specific samples. Whether this is a problem depends on the relationship between reporting limits and regulatory concentration values. When reporting limits are well below regulatory limits, there is less need for the reporting limits to be in close agreement.

When one laboratory reported ND< and the

other gave a concentration above the reporting limit, the ratios were often outside the factor of 3.00 criterion. By far the poorest metal in this respect was Cd, where 25 of 40 such ratios failed this test. The failure rate for Ag was 6 in 16 pairs, for Hg it was 4 in 27, and only 2 of 21 Se pairs failed to be within a factor of 3.00. In general, most values above reporting limits were from the QA laboratories, although several were flagged with "J," meaning they were estimates but below reporting limits. The QA laboratories may be providing a slightly more reliable analysis than the QC laboratories, but there still is insufficient real evidence to defend this statement.

VOCs in soils

For the 69 sets of VOCs in soils, there was a total of 134 analyte pairs having both a QC and QA concentration above reporting limits. Ethylbenzene with 25 pairs, toluene with 35 pairs, and total xylenes with 36 pairs represented the bulk of the data. The remaining 38 pairs were distributed among benzene (eight pairs), chloroform (one pair), total 1,2-dichloroethenes (five pairs), methyl ethyl ketone (two pairs), tetrachloroethene (11 pairs), and trichloroethene (11 pairs). Carbon tetrachloride, chlorobenzene, 1,2–dichloroethane, 1,1-dichloroethene, and vinyl chloride yielded no pairs. Although samples VS2–60 through VS2–66 used GC methods rather than GC/MS, there was no apparent reason to treat those results separately.

Most analytes also had examples where either the QC or the QA result was a real number but the other was below reporting limits. In many cases it was impossible to judge how well (or badly) these pairs compared because reporting limits were based on dilutions made to accommodate a high concentration analyte. For example, benzene in sample VS2–17 was reported as 59.7 μg/kg by the QA laboratory (reporting limit of 6.1 μg/kg) while the QC laboratory reported < 2500 μg/kg. These results may agree, but it is impossible to tell. Possibly this issue is unimportant given the very high concentration of xylene in this sample. One thing is clear: reporting these concentrations to three significant figures is a gross misrepresentation! In other cases, such as methyl ethyl ketone in sample VS2–18, the QA laboratory reported 271 μg/kg and the QC laboratory reported <12 μg/kg. Although the disagreement of these results could be represented by a ratio if we assigned the QC result a value of 12 μg/kg, we decided not to include such results in the statistical evaluation. Suffice it to say that such results occur with disturbing frequency.

Another common extreme is represented by several analytes in VS2–33 where the QA laboratory reported < 1.1 μg/kg and the QC laboratory reported < 7000 μg/kg. Such examples suggest the need for standardization of reporting procedures for samples requiring dilution to permit determination of one or more high concentration analytes. If such guidelines are already in place but are misunderstood or ignored by laboratories, a training program might improve the situation. Alternatively, if low concentration analytes really are of no interest when one or more analytes are present at very high concentration, data compilations could be greatly simplified.

For the 134 QC/QA ratios with concentrations above reporting limits, ratios ranged from 0.015 to 2830! The distribution of these ratios by analyte and for the total is shown in Table B4. We note that the distributions are very similar for individual analytes and for the group collectively. This suggests that statistical analysis applied to the entire group would adequately represent results for VOCs in soils. It is also apparent that extreme values are excessive. Over 40% of the ratios are outside the limits 0.25–4.00 and 24% are outside limits of 0.10–10.0. While we know of no standard for acceptability of analytical results on split or co-located samples sent to different laboratories, it stretches the boundaries of credibility to think that analysts would describe agreement within a factor of 10 as acceptable quantitation. Nonetheless, we will base statistical characterization on the 76% of ratios within 0.10–10.0.

A histogram of the logs of all 134 ratios is shown in Figure 4. Logs of the 102 ratios between 0.10 and 10.0 form a tolerable approximation to a normal distribution. Further support for the hypothesis of a lognormal distribution is provided by the linear cumulative probability plot of these logs on normal probability paper (Fig. 5). When separate lognormal plots were prepared for ethylbenzene, toluene, total xylenes, and the remaining analytes collectively (Figs. 6a–6d), all provided reasonable fits to this model.

Based on the lognormal model, the geometric mean of the 102 ratios between 0.10 and 10.0 was 0.95 and the 95% confidence limits on this mean were 0.77 to 1.16. The 99% tolerance intervals for individual ratios were 0.062 to 14.4. The fact that these intervals are considerably wider than the boundaries used for editing suggests that some values outside the range of 0.10 to 10.0 may also

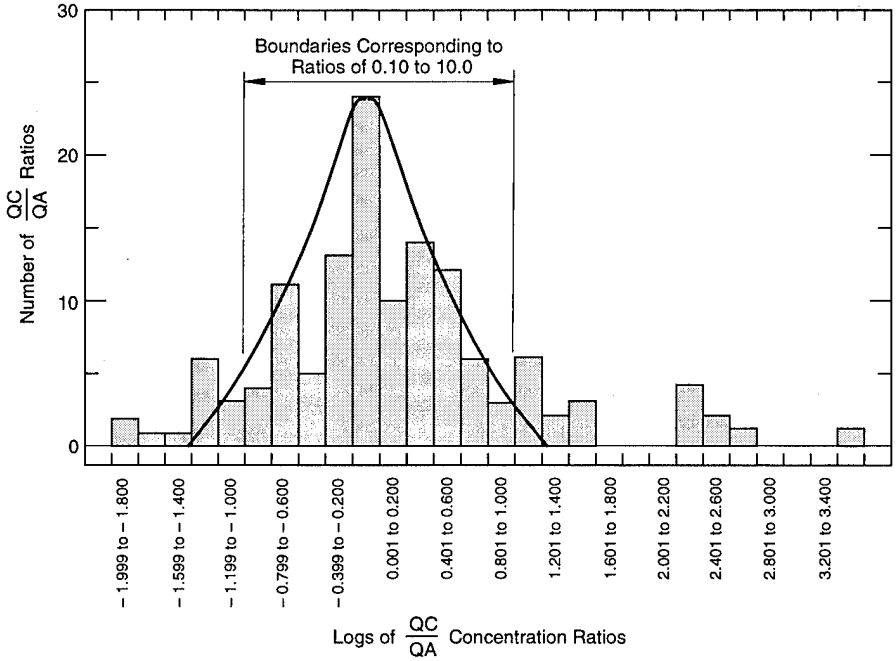


Figure 4. Frequency histogram for soil VOC QC/QA concentration ratios.

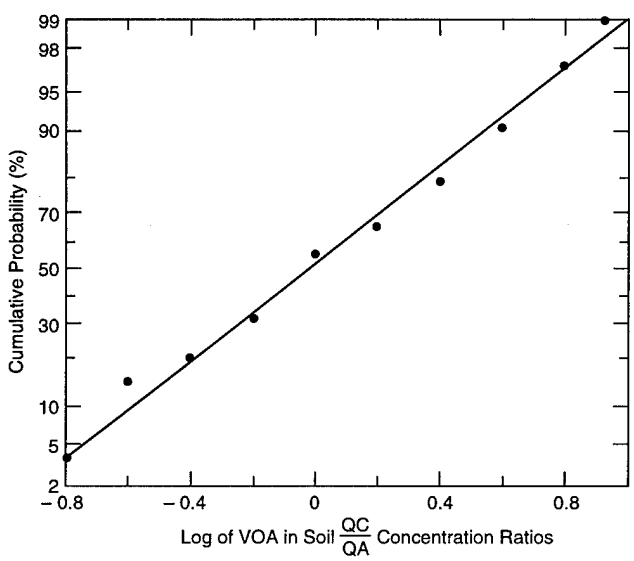
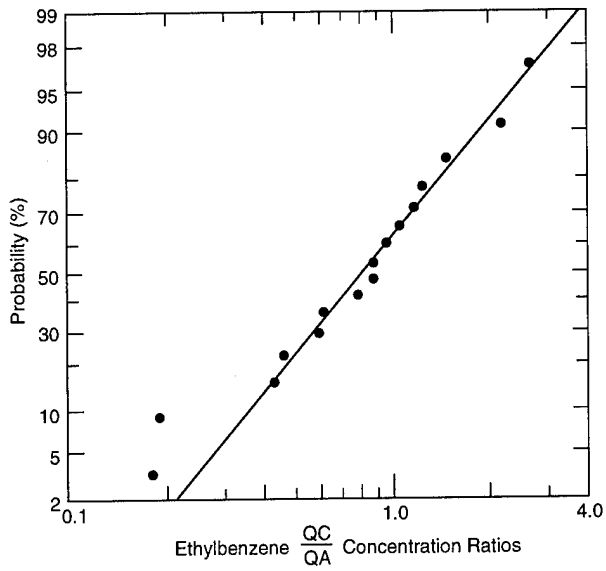


Figure 5. Lognormal probability plot for QC/QA concentration ratios of all VOCs in soils with ratios between 0.10–10.0.

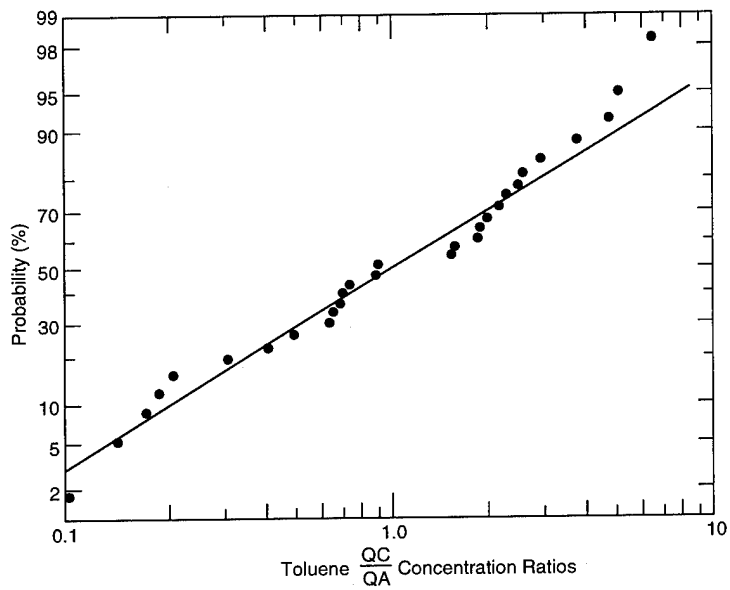
arise as a consequence of random rather than systematic errors. The standard deviation is larger than expected for a normal distribution and the tolerance intervals are excessively wide.

What can we conclude from these computations? Even after excluding 24% of the ratios representing the most divergent values, the agreement between laboratories for the determination of VOCs on split soil samples is so poor that it would be inappropriate to establish acceptability limits based on this data. In a recent article dealing with methods for regulatory analytical environmental chemistry, Kimbrough and Spinner (1994) wrote, "If year after year certain methods produce wildly inaccurate results, then this should be a signal that the method ought to be evaluated and hopefully improved. *The acceptance of very wide control limits basically sends the message that if everybody is wrong, then everyone is right.*" (Italics are ours.) We heartily concur with these statements and we believe that it is time to make concerted efforts to improve the situation. If the Corps wishes to use temporary limits while procedures are being brought under better control, we suggest 0.25–4.00. Clearly this will result in many values being questioned, but wider limits will not encourage the needed improvement.

Obviously, it is not a trivial task to significantly improve the procedures used to estimate VOC concentrations in soils. Before considering our suggestions for improvement, let's first ask if there is more to learn from the archived data. We failed to associate any specific laboratories with either good or bad results. In the absence of standards and with numerous laboratories participating (including some that serve either the QC or the QA function on different projects), it was impossible to establish a pattern. Clearly, the more laboratories that submit results, the more



a. Ethylbenzene.

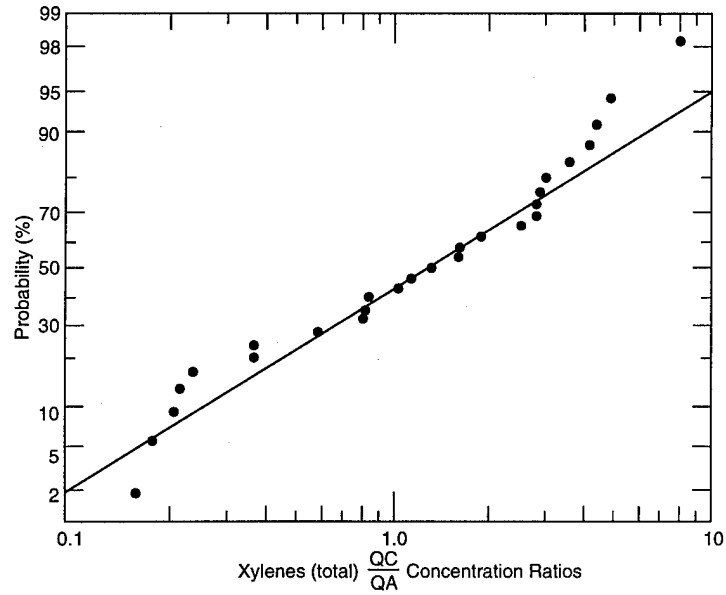


b. Toluene.

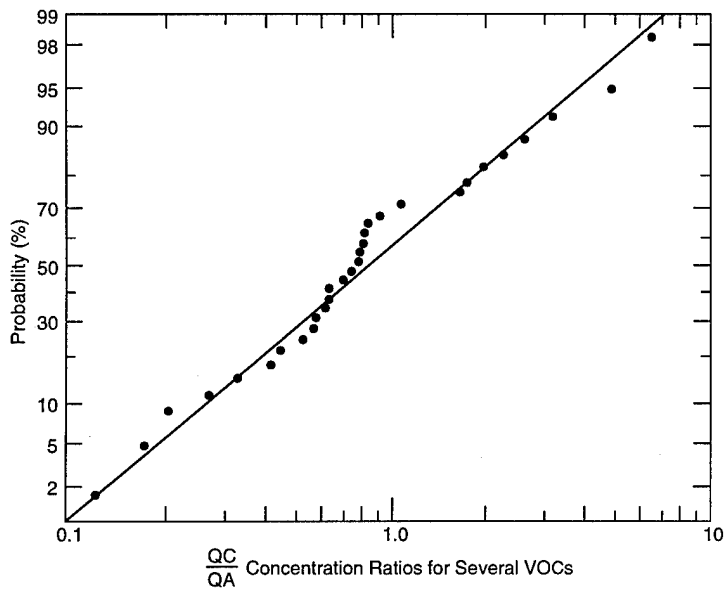
Figure 6. Lognormal probability plots for QC/QA concentration ratios for soil VOCs.

likely we are to experience biases and the more difficult it becomes to identify them.

What about the QC results from matrix spike/matrix spike duplicates (MS/MSD)? We could find no meaningful correlation between these results and the degree of agreement between QC and QA laboratories. Similar failure of MS/MSD results as a quality indicator for VOCs was recently reported by Robertson et al. (1995). Generally speaking, the MS/MSD results appeared to be quite satisfactory and suggestive of much greater reliability than was actually obtained on co-located samples. However, we must remember that MS/MSD results are derived by spiking extracts and, therefore, they omit all error associated with sample collection, transport, storage, and extraction. Assuming that the solution used for spiking is prepared independently from the calibration standards (frequently untrue), MS/MSD results can serve to validate calibration and to reveal specific interferences from the matrix under analysis. Does this imply that calibration variations do not contribute to the disagreement between laboratories? The answer is no. Significant bias may be introduced from calibration errors, but we do not believe it is the major source of error. In contrast to the soil results, we found that most QC/QA ratios for GC-MS determination of VOCs in



c. Xylene (total).



d. Several VOCs in soils (benzene [seven], 1,2-dichloroethene [total] [five], methylethylketone [two], tetrachloroethene [eight], trichloroethene [eight]).

Figure 6 (cont'd).

groundwater fell between 0.40–2.50. Nonetheless, we recommend that the Corps start to use vapor-fortified secondary soil standards of the type described by Hewitt (1994) and Hewitt and Grant (1995). These fortified samples sealed in glass ampoules are stable for extended periods and they can be distributed to participating laboratories. By using such secondary standards in an organized program of performance evaluation prior to qualification and later as periodic QC samples, the laboratory portion of the analytical procedure (extraction plus analysis) can be evaluated.

We believe that the major losses are caused by volatilization during sample collection and laboratory subsampling for analysis. Additional volatilization loss can occur during transport and storage, and biodegradation can also be a problem. Recent articles (Siegrist and van Ee 1994, Hewitt et al. 1995) document losses of a factor of 1000 or more. This postulation is consistent with our observation in this study that samples containing multiple analytes produce QC/QA ratios that are consistently high or low for all analytes present. In Figure 7 we illustrate this effect by plotting logs of total xylene QC/QA ratios vs. logs of ethylbenzene ratios for corresponding samples. The correlation coefficient of 0.906 provides strong evidence for the relationship between these re-

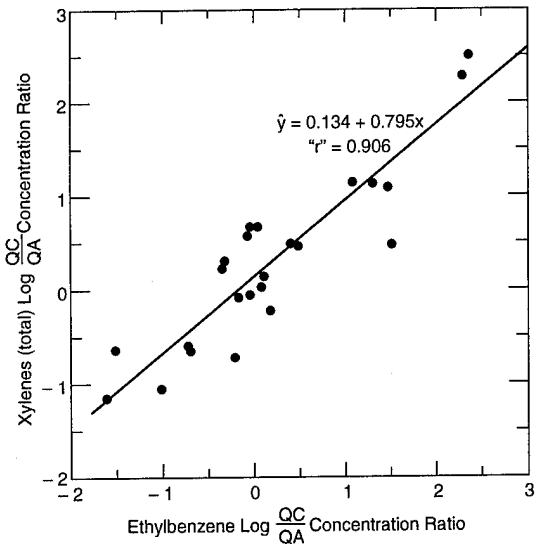


Figure 7. Correlation of log concentration ratios for total xylenes and ethylbenzene in soil.

sults. Hewitt et al. (1995) state, "Only those sampling methods that maintain soil structural integrity, use a single soil transfer step, and preserve against biological degradation should be used when establishing VOC concentrations." A tipless plastic syringe is suggested for obtaining soil samples, which are immediately transferred to VOC vials containing either methanol or water acidified with $NaHSO_4$. Details of this approach, which satisfies SW-846, Method 8260A, are available from Hewitt. We believe that it is imperative for the Corps to consider this or alternative methods to improve the quality of analytical results for VOCs in soil.

In contrast to the 42% of archived QC/QA ratios outside 0.25–4.00, only 23% of the most recent results (44 samples from 11 project locations) were outside of these limits. It is interesting to note that 20% of the QC_1/QC_2 ratios were also outside 0.25–4.00 (no QC_1/QC_2 values were available in the original archived data). Of the 20%, 17% of the ratios were above 4.00, i.e., the first value was much larger than the second value. Perhaps VOCs were being lost during storage, thus causing the second values to be low.

TPH in soils

The 102 data reports for TPH in soils submitted by the North Pacific Division (NPD) contained 83 pairs of replicate QC results with concentrations above reporting limits, and 95 pairs with both QC and QA results above reporting limits. For these data, 6.0% of QC_1/QC_2 ratios and 13.7% of QC_1/QA ratios were outside 0.25–4.00 (Table B5). We concede that these limits are arbitrary, but they form a reasonable starting point and the rejection rate for QC/QA is only moderately greater than 10%, which we consider an acceptable goal. As noted in the VOC section, we believe that it is imperative to seek improvements in procedures that yield widely divergent results.

In contrast to the results from contractors reporting to NPD, 28 of the 61 pairs (45.9%) of QC/QA ratios from the Missouri River Division (MRD) databank were outside 0.25–4.00. No replicate QC values were included on these data reports. The histograms of log QC/QA ratios shown in Figure 8 provide a very clear depiction of the difference in dispersion for the two data sets. The lognormal model appears to be a reasonable fit to the NPD QC/QA ratios between 0.25–4.00. This hypothesis is reinforced by the probability plot of

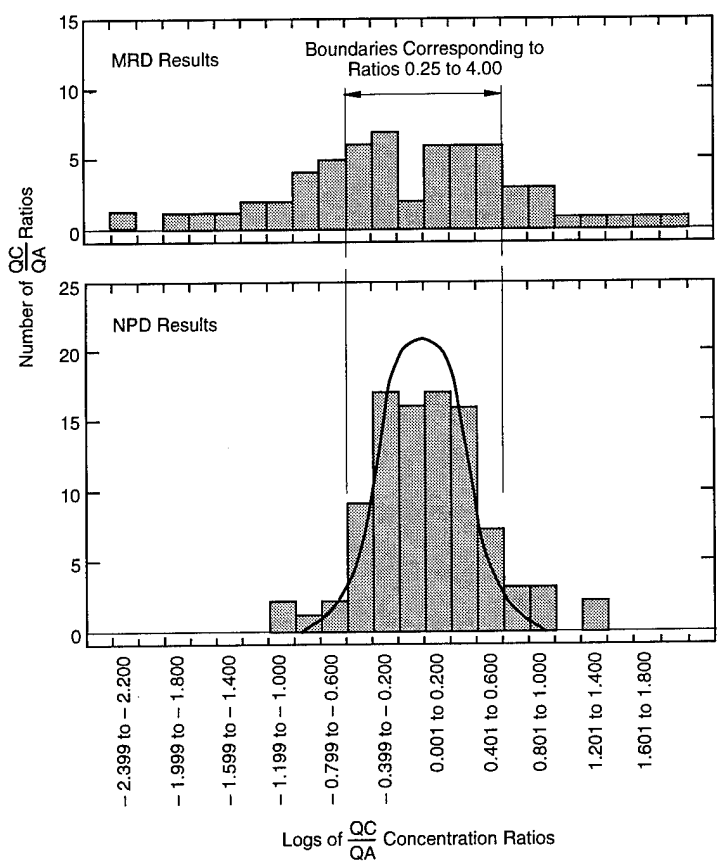


Figure 8. Comparison of frequency histograms of soil TPH QC/QA concentration ratios from NPD and MRD.

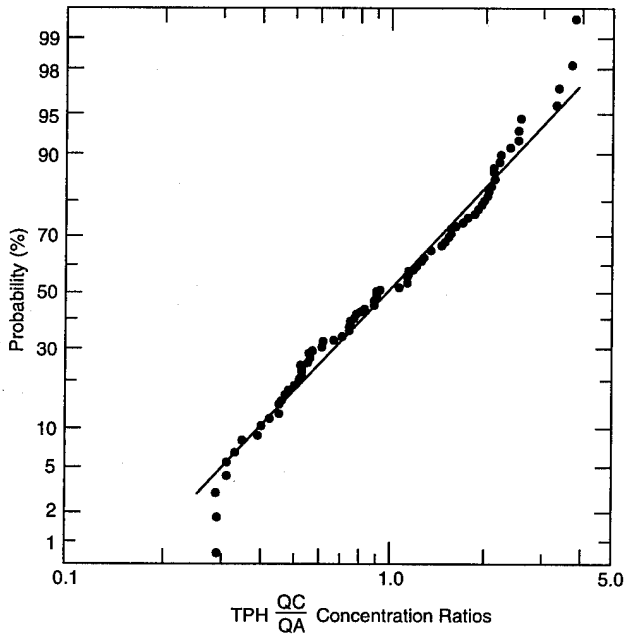


Figure 9. Lognormal probability plot for soil TPH QC/QA concentration ratios between 0.25 and 4.00 for NPD results.

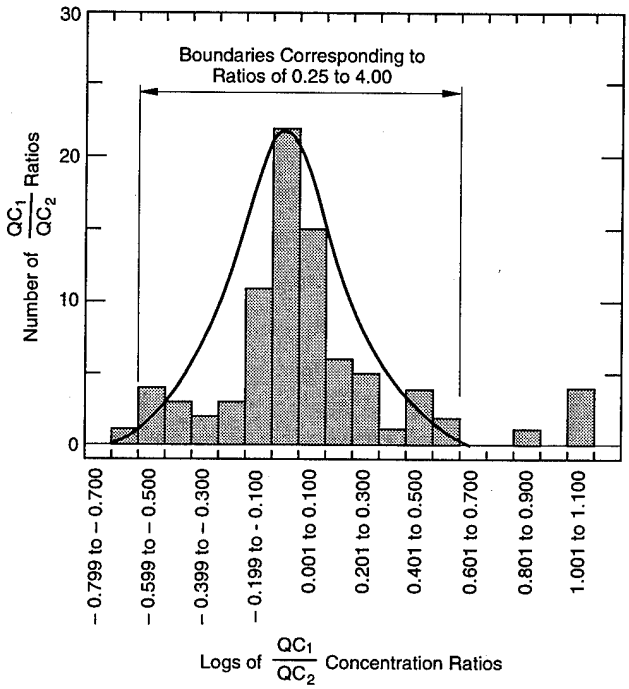


Figure 10. Frequency histogram for soil TPH QC_1/QC_2 concentration ratios from NPD.

these ratios on lognormal paper (Fig. 9). However, the histogram of ratios for MRD results shows extreme dispersion, and the values between 0.25–4.00 form a rectangular distribution, thereby precluding statistical analysis using a lognormal model. The NPD results were probably more reliable than the MRD results because the former tried to ensure that both QC and QA laboratories used the same state-approved method. There is a wide variation in both the extraction and analysis procedures for TPH as a function of various state requirements. We chose to analyze only the more reliable NPD results; no further analysis was conducted on the MRD results. This decision was rooted in the conviction that future aspirations should at least be based on the best of past performances.

A histogram of the log QC_1/QC_2 results is shown in Figure 10 and the lognormal probability plot appears in Figure 11. Both plots provide only fair fits to the lognormal model and there is some uncertainty about the correct location of the straight line in Figure 11. Nonetheless, the expectation of lognormality and the absence of a better model convinced us to proceed with this model.

The geometric mean of the 78 QC_1/QC_2 ratios between 0.25–4.00 was 0.95 and 95% confidence limits on this mean were 0.84–1.07 (Table B5). The 99% tolerance intervals for individual ratios were 0.22–4.07, which is in good agreement with the arbitrary limits used to edit the data. The geometric mean of the 82 QC/QA ratios between 0.25–4.00 was 0.98 with 95% confidence limits of 0.84–1.15 and 99% tolerance intervals of 0.16–6.15.

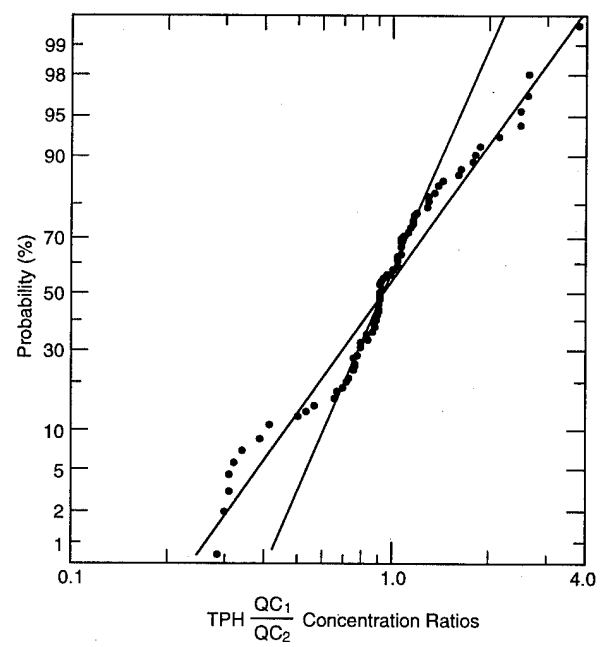


Figure 11. Lognormal probability plot for soil TPH QC_1/QC_2 concentration ratios between 0.25 and 4.00 for NPD results.

The poor agreement of the tolerance intervals with the editing limits implies that the standard deviation of the log ratios was larger than expected for a good-fitting lognormal distribution. A look at the histogram (Fig. 8) shows that there are more ratios than expected toward the extremes and fewer than expected in the center. Deficiency of ratios near the ideal value of 1.00 (log = 0) may be due to systematic bias as suggested in later paragraphs.

As with the VOCs, we found no meaningful correlation between either surrogate recoveries or MS/MSD results with corresponding QC/QA ratios. In many cases MS/MSD results were absent or were indicated to be unavailable because they were diluted out of range. In other examples exhibiting very poor agreement between QC and QA results, surrogate recoveries were excellent and MS/MSD results were quite good. We must ask whether MS/MSD results are providing enough useful information to justify the cost. Perhaps the use of independently prepared secondary standards would offer greater cost/benefit. More will be said on this later.

Some of the random error undoubtedly arises as a consequence of sampling and splitting soils with heterogeneously distributed petroleum hydrocarbons. Possibly contractor guidance can be revised to reduce this source of error. Volatilization seems unlikely to be a major source of error, but that too could be examined. Biodegradation is another potential source of error. A well-planned nested experimental design could be used to determine the magnitude of these errors prior to instituting corrective measures.

The GC-FID method may be a significant source of error for TPH determinations. Besides inevitable random errors, the unique arrangement followed by the Corps of Engineers may allow systematic errors such as calibration or signal integration biases to appear as random error. We refer to the use of a given laboratory for the QA function on one project and for the QC function on another project. Thus, if a laboratory reports consistently low results, they would inflate QC/QA ratios when serving as a QA laboratory but they would depress ratios when acting as a QC laboratory.

We believe that GC-FID for TPH is susceptible to systematic error, especially in the integration of peaks on chromatograms with broad undifferentiated background. This is particularly true for weathered residues where the normal hydrocarbons, which produce a very recognizable pattern in fuels, have been reduced relative to the branched chain components. Evidence of bias is found in Figure 12. NET Pacific, Inc., Santa Rosa, California, was the QA or the QC laboratory for all 95 QC/QA ratios computed from NPD results. In 52 cases the other laboratory was CAS, Inc., Kelso, Washington. NET Pacific was the QA laboratory 28 times and the QC laboratory 24 times. When we plotted the log of the concentration ratios for CAS, Inc./NET Pacific, Inc. vs. the log of

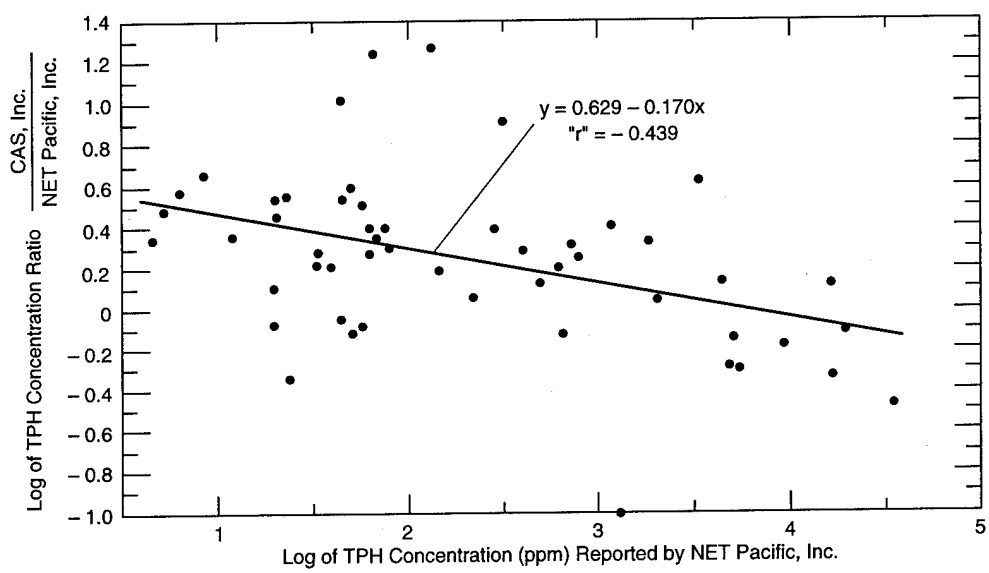


Figure 12. Correlation of soil TPH concentration ratios for CAS, Inc./NET Pacific, Inc. with soil TPH concentration estimates reported by NET Pacific, Inc.

the concentration reported by NET Pacific, Inc. (Fig. 12), the correlation was significant at the 99.8% probability level. The fitted linear equation indicates no bias (log ratio = 0.0) at a TPH concentration of 5600 ppm. At lower TPH concentrations, CAS, Inc. reports concentrations on sample splits that exceed those reported by NET Pacific, Inc., and the difference increases as concentration decreases. For example, when NET Pacific, Inc. reports a TPH concentration of 100 ppm, CAS, Inc. reports, on average, a concentration of 200 ppm. Bias of this magnitude is unacceptable. Because of the numerous laboratories used to analyze samples, this problem is magnified several times. In the case of the NPD results, the 43 ratios not shown in Figure 12 contained results from five other laboratories.

One possible way to identify the source of bias and, hopefully, to eliminate it, is to prepare several typical soils fortified with known concentrations of TPH. To minimize potential error sources such as volatilization and biodegradation, we suggest weighing portions of homogenized dry soils into ampoules. Each ampoule would be spiked with a known amount of TPH and sealed as described by Hewitt (1994). Perhaps three soils with different clay and organic matter content could be used. After participating laboratories analyzed these secondary standards and submitted the results for statistical analysis, a general meeting could be held to resolve causes of large differences. Ideally the number of participating laboratories would be minimized. Unfortunately, there is no obvious way to compensate for the effects of variable degrees of weathering, and that may prove to be one of the largest sources of error.

In summary, we think it is inappropriate to suggest acceptability limits until the method as currently practiced is brought under better control. If temporary limits are desired for the interim, we feel they should be no wider than 0.25–4.00. When these limits were used for the most recent results for 87 samples from 15 project locations, 5.5% of the QC_1/QC_2 ratios and 16% of the QC/QA ratios were outside of these limits. These results agree well with the 6.0% and 14% estimates obtained for archived NPD data.

As soon as possible, secondary TPH standards should be prepared for use in a study to define and minimize bias. Such standards could also become part of the routine QC program. A study of sampling, sample splitting, and storage should also be conducted. It is also important to ensure that both QC and QA laboratories follow the same extraction and analysis procedures. Finally, we recommend stopping the dual use of laboratories for QA and QC functions.

Explosives in soils

The 71 data reports for explosives in soils submitted by the MRD contained a total of 65 pairs with both QC and QA as real numbers. The distribution of these pairs among several analytes is shown in Table B6. The MRD, Omaha, Nebraska, laboratory served the QA function in all of these results. Seven of the 65 pairs (10.8%) yielded QC/QA ratios outside 0.25–4.00. However, clustering around the ideal ratio of 1.00 was much better than for VOCs or TPH; 72% were between 0.50–2.00. Bias was suggested because 41 of the 65 ratios were less than 1.00. Since there was no apparent distinction in the distributions or geometric means of the ratios for individual analytes, they were treated as a single group. Figures 13 and 14 show a histogram and a probability plot, respectively, for the logs of the 58 ratios between 0.25–4.00. The lognormal model once again provides a reasonable fit.

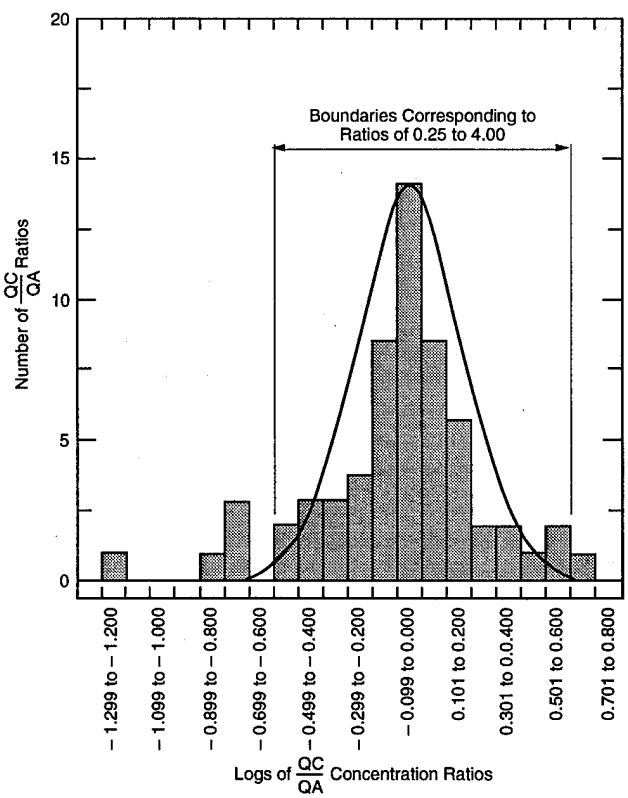


Figure 13. Frequency histogram for soil explosives QC/QA concentration ratios from MRD.

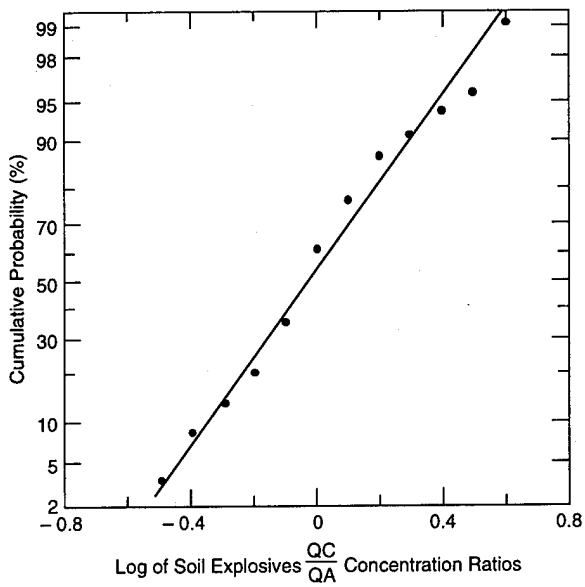


Figure 14. Lognormal probability plot for QC/QA concentration ratios of explosives in soils with ratios between 0.25–4.00.

The geometric mean of the 58 ratios between 0.25–4.00 was 0.91 with 95% confidence limits of 0.78 to 1.06. If we isolate the 40 ratios with Weston, West Chester, Pennsylvania, as the QC laboratory, the geometric mean for QC/QA is again 0.91. Consequently, despite the high percentage of ratios below 1.00, we are unable to demonstrate statistically significant bias between the contracted QC laboratories and the MRD QA laboratory. The 99% tolerance limits are 0.20–4.18, slightly wider than the editing limits. As before, this minor discrepancy is attributed to a slightly imperfect fit of the lognormal model.

Because a very high proportion of the concentrations are close to method reporting limits, it is not surprising that there were 84 cases where either the QC or the QA laboratory reported a real value while the other reported "less than." Of these, 47 were cases where the "less than" value was smaller than the real value, but many of these were in close agreement, e.g., < 0.25 vs. 0.29. Whenever the real value was within a factor of 4 of the "less than" value, we considered the ratio to be acceptable. For 36 of the 47 pairs in this group, this criterion was met.

The "less than" value was greater than the real value in 37 cases, e.g., < 0.50 vs. 0.39. We used the same factor of 4 here although it is clear that this factor could be exceeded without disagreement of the values, e.g., < 5.0 vs. 0.27. The problem here is that the "less than" value is too high. Of the 37 pairs in this group, 27 were within a factor of 4. Overall then, 63 of 84 pairs with only one real value were in acceptable agreement. This situation could obviously be improved with better interlaboratory agreement on reporting limits, but current agreement is far superior to that for VOCs and TPH.

MS/MSD results were provided in some cases, especially those from the MRD laboratory. Unfortunately, we were unable to correlate those results with QC/QA ratios. We remain unconvinced that the MS/MSD effort is cost-effective in this program.

There were also eight data reports for the Umatilla Washout Lagoon, submitted by NPD. All samples showed extremely high RDX and TNT concentrations requiring high dilution prior to analysis of extracts. Consequently, they were examined separately. All QA results were generated by Maxwell S^3, San Diego, California, and all QC results came from Precision Analytecs, Pullmann, Washington. Concentration estimates for TNT were missing for two QA samples due to high dilution. Of the 14 pairs of results (eight RDX, six TNT), 13 gave QC/QA ratios well under 1.00. The geometric mean was 0.58 with 95% confidence limits of 0.39 to 0.86. The substantial systematic error between these two laboratories could arise in various ways, including losses due to holding time differences. Our research has also shown that nitroaromatic compounds are susceptible to biodegradation (Grant et al. 1995). Sample heterogeneity, incomplete extraction, and calibration differences are additional potential sources of error.

A recent experiment involving the MRD laboratory and the Cold Regions Research and Engineering Laboratory (CRREL) provides evidence to support the potential contribution of such error sources (Solsky pers. comm.*). Seventeen soil samples were split in the usual fashion and analyzed by MRD and CRREL. A total of 53 pairs of results was generated for 2,4,6-TNT, 1,3,5-TNB, 2,4-DNT, RDX, and 1,3-DNB. Seven of the 53 ratios for MRD/CRREL were outside the limits 0.25–4.00. The geometric mean of the 46 ratios inside of the editing limits was 1.08, with 95% confidence limits for the mean of 0.95–1.23, and 99% tolerance limits of 0.33–3.51. These limits are somewhat narrower than the ones reported for the previously described MRD results, but they exhibit similar characteristics.

* J. Solsky, U.S. Army Corps of Engineers, Missouri River Division, Omaha, Nebraska.

When five of these samples were dried, ground, and split prior to analysis by both labs, the 20 pairs of results for the same analytes yielded only one ratio outside 0.25–4.00. In fact, the range of the other 19 ratios was 0.44–1.16; the geometric mean was 0.79, 95% confidence limits for the mean were 0.70–0.89, and 99% tolerance intervals were 0.39–1.61. It is unclear why a significant bias existed between the laboratories when none was present for the samples split in the normal way. However, it is most important that the standard deviation of the ratios for the dried and ground samples was considerably less than for the original splits (significant at the 99% confidence level). Interestingly, if the bias between laboratories is corrected to a mean ratio of 1.00, the 99% tolerance intervals would become 0.49–2.04. This provides considerable evidence for the benefit of sample drying and grinding prior to splitting.

In summary, there is considerable support in these data for temporary acceptance limits of 0.25–4.00 with the expectation of tightening in the future as procedures are improved. For the most recent results for 19 samples from three project locations, 13% of the QC/QA ratios were outside 0.25–4.00. This is similar to the original results and indicates a need for further improvement. Secondary standards prepared similarly to our earlier recommendation for TPH should be used throughout projects as a means of minimizing bias between laboratories. We cannot isolate the contributions of error sources such as sample heterogeneity and biodegradation in these results, but there clearly is a potential for improvement in this aspect of the program.

VOCs in groundwater

For the 157 sets of VOCs in groundwater, we discovered that four sets (V2-63 through V2-66) were repeats of V2-56 through V2-59. Thus, there were actually only 153 sets of results. For carbon tetrachloride, chlorobenzene, 1,2-dichloroethane, 1,1-dichloroethene, and methyl ethyl ketone, there were fewer than five pairs of values where each was above the reporting limit. The number of pairs above reporting limits for the other nine analytes ranged from nine to 63. Twenty-three of the QC laboratory samples included concentrations for field duplicates, but results for most analytes were below reporting limits.

Only 1,2-dichloroethene and trichloroethene, with 11 and 14 pairs of duplicate QC results, respectively, provided enough evidence to judge the limits to be expected for VOCs in groundwater. One of the 25 ratios was excluded because it exceeded 3.00. The 95% confidence limits on the geometric means bracketed the expected mean of 1.00 very nicely in both cases. However, the low number of degrees of freedom in the standard deviation estimates caused the 99% tolerance intervals to be wider than desired, especially for 1,2-dichloroethene (0.39–2.42). If one uses the acceptability criterion suggested previously for metals (0.50–2.00), all 24 ratios are within these limits. Although there are insufficient data to permit evaluations of the other VOCs, there is nothing to indicate that this range would be unsuitable for them. On the contrary, the fact that these data arise from a single multianalyte procedure suggests that the precision should be quite similar for each of the analytes. This differs from the situation with metals where several methods are employed. Therefore, we recommend 0.50–2.00 as the acceptability limits for duplicate QC ratios of VOCs in groundwater. When these limits were applied to the most recent results for 67 samples from 17 project locations, only 3% of the ratios were outside of this range.

The QC/QA ratio characteristics are summarized for nine VOCs in Table B7. Eleven of 321 ratios (3.4%) were excluded because they were outside 0.30–3.00. Outliers were spread among several VOCs and they occurred over a wide spectrum of concentrations. Geometric means for each of the nine VOCs were close to the expected value of 1.00, and the 95% confidence limits included 1.00 in every case. The 99% tolerance bands are in reasonable agreement with each other except for vinyl chloride, which offered insufficient data for close tolerances (too few degrees of freedom). Only three of the 310 ratios remaining after the initial 11 exclusions were outside 99% tolerances.

As with metals in soils, let us examine the effect of using 0.40–2.50 as the acceptability criterion for QC/QA ratios. We find seven additional exclusions beyond the 11 originally rejected values for 0.30–3.00 limits. Of these 18 unacceptable ratios (5.6%), 11 are for ratios below 0.40 and seven are for ratios above 2.50. Further insight can be obtained from Table B8, which shows the distribution of ratios outside 0.40–2.50 with respect to the number of analytes above concentration reporting limits for both QC and QA per sample. Three samples have two outlier ratios each and two of these three have no other analytes present above reporting limits. The other 12 rejected ratios occur one per sample even though each of these samples has acceptable ratios for one or

more other analytes. The three samples with two outliers clearly require close scrutiny for likely rejection of the entire analysis, but the 12 samples with only one outlier may well be acceptable after examination of the whole set of analyses. The pleasing result is that 90% of the samples had no unacceptable ratios according to the suggested limits. Further support for using limits of 0.40–2.50 for QC/QA ratios is gleaned from the less than 3% of outliers found in the most recent results for 67 samples from 17 project locations. The much greater homogeneity of groundwater compared to soil and the reduced amount of manipulation during preparation and analysis must certainly account for the vastly superior VOC results for groundwater.

Some VOC reporting limits were absent on the data sheets and this reduced the effectiveness of comparisons. The "practical quantitation limits" for groundwater tabulated in Table D-4 of ER 1110-1-263 (USACE 1990) are 5 μg/L for 12 of the VOCs, 10 μg/L for vinyl chloride, and 100 μg/L for methyl ethyl ketone. About three-fourths of the samples apparently contained no analytes at concentrations requiring dilution prior to GC/MS. For those samples, the ER 1110-1-263 criteria were usually equaled or bettered. In general, however, the reporting limits for QA laboratories were lower by about a factor of 2 than corresponding values from the QC laboratories. Thirty-five samples listed one or both reporting limits above the ER 1110-1-263 recommendations, often by orders of magnitude. More often than not the high reporting limits were from the QC laboratory while the QA value was acceptable. Twenty-seven of these samples contained one or more VOCs at concentrations high enough to have required substantial dilution. It appears that the QC reporting limits were derived from diluted samples while QA values were for undiluted samples. From a practical perspective, the question is whether high reporting limits decrease the acceptability of concentration estimates when one or more VOCs are present at very high concentrations. Greater concern about reliability would attach to the eight samples where high reporting limits cannot be readily explained.

CONCLUSIONS AND RECOMMENDATIONS

General recommendations

1. To improve future evaluations, both QC and QA laboratories should be required to submit the method of estimation and numerical values of reporting limits for all analytes, actual surrogate recoveries ("OK" is unacceptable), and to designate method of analysis for each analyte.

2. A convention for significant figures should be adopted for all concentration reports. One possibility would be to use two significant figures for all values below 100 (e.g., 0.72, 6.9, and 54) and a maximum of three significant figures for values of 100 or greater. This would alleviate the problem of comparing pairs such as 6 and 6.25 (are they different?).

3. Standard practice should be adopted for reporting limits where dilutions are required. The target reporting limits given in ER 1110-1-263 (USACE 1990) might be specified as the highest acceptable values, assuming that these are low enough for the intended use of the data. Special exceptions could be made for a few samples with unusual matrix problems. At the very least, both QC and QA should specify reporting limits on the same basis rather than one on the original sample and one on a diluted portion.

4. A concerted effort is needed to reduce uncertainty in the QA values if they are to serve a referee purpose. This effort might combine the use of blind duplicates of selected samples, some spike recoveries, and the regular analysis of certified samples that are as similar in matrix as possible to regular samples. Such an expanded internal QC effort by the QA laboratories would help justify the use of their results for determining acceptability of analyses by other laboratories. Only those laboratories with demonstrated excellence should serve the QA function.

5. A policy for dealing with outlier ratios and unacceptable reporting limits must be promulgated by USACE. Such a policy could likely be incorporated into an artificial intelligence computer program to greatly reduce the time and cost required to administer the policy.

6. Consideration should be given to using as few contract laboratories as possible and to avoid using a given laboratory for the QA function on some projects and the QC function on others.

Metals in soils

Ratios of archived results for metals in split soil samples were lognormally distributed for those metals with sufficient numbers above concentration reporting limits (As, Ba, Cr, Pb). For duplicate QC ratios, only 3.4% were outside the limits of 0.50–2.00 while 10.2% of the QC/QA ratios were outside limits of 0.40–2.50. Only five samples out of 124 (4.0%) had more than one metal

QC/QA ratio per sample outside of the 0.40–2.50 limits.

We recommend that

a) Acceptance limits for duplicate QC ratios be set at 0.50–2.00.

b) Acceptance limits for QC/QA ratios be set at 0.40–2.50. Application of these limits to recent results confirms that these are workable comparison criteria for both QC_1/QC_2 and QC/QA.

c) Future results for multianalyte procedures should be carefully analyzed with regard to the number of rejects/sample. From such analysis, a policy for dealing with such analyses should be developed.

VOCs in soils

Archived results for VOCs in split soil samples produced QC/QA ratios that were distributed lognormally as expected. However, over 40% of the QC/QA ratios were outside the range 0.25–4.00 and 24% were outside 0.10–10.0. The magnitude of this scatter is so large that it is impossible to recommend effective limits of acceptability. Instead, we believe that steps are urgently needed to improve data quality.

We recommend that

a) Temporary acceptance limits of 0.25–4.00 be set, despite the possibility of up to 40% unacceptable data. Analysis of recent results using these limits indicated some improvement, but more is needed.

b) Secondary standards sealed in ampoules (see Hewitt 1994, Hewitt and Grant 1995) should be used for performance evaluation during laboratory qualification and as periodic QC samples.

c) The change with the greatest potential for data quality improvement would be to start a program of improved techniques for sample collection, storage, and laboratory subsampling in preparation for analysis. Specific recommendations can be found in Hewitt et al. (1995). An instructional videotape might prove to be useful.

TPH in soils

Ratios of archived results for TPH in split soil samples analyzed by QC and QA laboratories were provided by the North Pacific Division (NPD) and also by the Missouri River Division (MRD). Nearly half of the QC/QA ratios for the MRD results were unacceptably wide, so only the NPD data was statistically analyzed. Although 86% of these ratios were between 0.25–4.00, the standard deviation of the logarithms was larger than expected for a lognormal distribution.

We recommend that

a) As with VOCs in soils, temporary QC/QA acceptance limits of 0.25–4.00 be adopted while procedures are being brought under better control, especially those at MRD. It is especially important to ensure that both the QC and the QA laboratories use the same extraction and analysis protocols depending on the requirements of the applicable state method (if present).

b) Secondary standards sealed in ampoules should be prepared and used to identify sources of systematic error so they can be reduced to a minimum.

c) A nested experimental design should be used to evaluate the magnitude of errors in sampling, sample homogenization and splitting, and storage. This information would be used to focus attention on any large random error sources.

Explosives in soils

Ratios of MRD archived results for explosives in split soil samples analyzed by QC and QA laboratories were distributed lognormally, with 89% of the 65 QC/QA ratios between 0.25–4.00 and 72% between 0.50–2.00. Because many concentrations were close to reporting limits, there were 84 cases where either the QC or the QA laboratory reported "less than" while the other provided a real number. For these 84 pairs, 63 (75%) were considered to be in acceptable agreement.

The NPD submitted a series of analyses from the Umatilla Washout Lagoon where concentrations of RDX and TNT were very high. There was a large bias (nearly a factor of 2) between the QC and QA laboratories, possibly due to either calibration differences or incomplete extraction by the QC laboratory.

We recommend that

a) Temporary acceptance limits for QC/QA ratios be set at 0.25–4.00. Application of these limits to recent results yielded 13% of the ratios outside this range. We believe this percentage can be significantly reduced.

b) Secondary standards containing known concentrations of explosives in dry soils and sealed in ampoules should be used by participating laboratories on a regular basis in order to reduce interlaboratory bias.

c) Sample collection, splitting, and storage procedures should be reviewed with the objective of reducing variations, especially biodegradation losses of nitroaromatics.

VOCs in groundwater

Ratios of archived results for VOCs in co-located groundwater samples were vastly superior to results for many of the same analytes in split soil samples. This finding reinforces the conclusion that the major problems with VOCs in soils occur during sampling and sample preparation because the determinative step is basically the same for each. For the 25 duplicate QC results reported, only one was outside of the limits 0.50–2.00 while 5.6% of the QC/QA ratios were outside limits of 0.40–2.50. Recent results provided even lower percentages (3%). Only three samples out of 153 (2.0%) had more than one VOC ratio per sample outside of the 0.40 to 2.50 limits.

We recommend that

a) Acceptance limits for duplicate QC ratios be set at 0.50–2.00, the same as for metals in soils.

b) Acceptance limits for QC/QA ratios be set at 0.40–2.50, the same as for metals in soils. The recent data confirm the workability of those ranges for both QC_1/QC_2 and QC/QA.

c) Like metals in soils, a policy must be promulgated for dealing with samples with more than one outlier.

d) More attention needs to be given to consistency in the procedure for specifying concentration reporting limits.

SUMMARY

The objective of chemical environmental measurements is to provide valid characterization of sites under investigation so that regulatory decisions can be made with confidence. To obtain the accurate analyses required to realize this objective, the process of analysis must be considered to start at the time of sampling. Procedures must minimize volatilization and biodegradation losses and contamination gains. This includes sampling, sample transport, and sample storage. Analytical validation must deal with the entire procedure, i.e., extraction, determination, and data manipulation. These requirements are doubly important in programs where many different laboratories contribute to the databank.

In our judgment, ratios of chemical analyses on split samples that differ by more than a factor of 4, reported by two laboratories, are useless for decision making whenever concentrations are near regulatory limits. In reality, we dislike having limits greater than 2.5 and we believe that such agreement is well within reach with proper sampling. Obviously there is no fundamental basis for either factor, but consider the dilemma when using wider limits, such as 0.10–10.0. Assume a regulatory value of ten and one laboratory reports six while the second reports 58. The analyses agree within a factor of 10, but a valid regulatory decision is impossible. Admittedly, the problem is minimized when results depart greatly from regulatory limits, but it is around these limits that disputes most often arise. Of course, this problem is not eliminated with tighter limits, but it certainly reduces the frequency of occurrence.

We reluctantly recommended temporary limits of 0.25–4.00 for VOCs in soils because this criterion may cause up to 40% of results to be unacceptable. However, we believe the results can be dramatically improved by making fairly simple changes in the sampling procedures. For TPH, where the same criterion was recommended, it is clear that the NPD laboratories have this analysis under better control than the MRD laboratories. Resolution of this difference should be the first objective, to be followed by further improvements in the procedures.

We believe that the U.S. Army Corps of Engineers should commit a significant portion of its effort in chemical environmental measurements toward the realization of improved data quality. Some specific suggestions are enumerated in this report. Clearly, close interaction with participating laboratory personnel is also needed. To do otherwise and maintain the status quo seems certain to validate the Kimbrough and Spinner (1994) quote, "The acceptance of very wide control limits basically sends the message that if everybody is wrong, then everyone is right."

LITERATURE CITED

Grant, C.L., T.F. Jenkins, K.F. Myers, and E.F. McCormick (1995) Comparison of fortification vs. field contamination for holding time estimates of soils containing explosives residues. *Journal of Environmental Toxology and Chemistry*, **14**: 1865–1874.

Hewitt, A.D. (1994) Vapor-fortified QA/QC soil subsamples for the analysis of volatile organic compounds. *American Environmental Laboratory*, **3**(94): 1–8.

Hewitt, A.D. and C.L. Grant (1995) Round-robin study of performance evaluation soils vapor-fortified with volatile organic compounds. *Environmental Science and Technology*, **29**: 769–774.

Hewitt, A.D., T.F. Jenkins, and C.L. Grant (1995) Collection, handling and storage: The keys to improved data quality for volatile organic compounds in soil. *American Environmental Laboratory,* **2**(95): 25–28.

Kimbrough, D.E. and R. Spinner (1994) Performance-based methods for regulatory analytical environmental chemistry. *American Environmental Laboratory,* **11-12**(94): 1–10.

Robertson, G.L., F.C. Garner, D.W. Hewetson and G.A. Laing (1995) The effectiveness of matrix spike analysis in the contract laboratory program. Presented at the *9th Workshop of QAEM, April 1995, Scottsdale, Arizona.*

Siegrist, R.L. and J.J. van Ee, Eds. (1994) A symposium summary: *Measuring and Interpreting VOCs in Soils: State-of-the-Art and Research Needs.* EPA/540/R–94/506.

USACE (1990) Chemical data quality management for hazardous waste remedial activities, 1 October 1990, U.S. Army Corps of Engineers ER 1110-1-263.

APPENDIX A: DATA AND SAMPLE CALCULATIONS FOR CHROMIUM

Table A1. Cr concentrations, ratios, and logarithms.

Sample number	Concentrations µg/g			Ratios and logarithms			
	QA	QC_1	QC_2	$\frac{QC_1}{QA}$	$log \frac{QC_1}{QA}$	$\frac{QC_1}{QC_2}$	$log \frac{QC_1}{QC_2}$
M1-1	21	3.4		*0.16	*		
M1-2	12	6.2		0.52	–0.287		
M1-3	8.8	4.6		0.52	–0.282		
M1-4	18	8.3		0.46	–0.336		
M1-5	26	17.7		0.68	–0.167		
M1-6	32	15.9		0.50	–0.304		
M1-7	30	14.2		0.47	–0.325		
M1-8	18	8.3		0.46	–0.336		
M1-9	34	14.2		0.42	–0.379		
M1-10	9.2	7.6		0.83	–0.083		
M1-11	2.6	2.9		1.12	0.047		
M1-12	8.0	9.3		1.16	0.065		
M1-13	33	17.5		0.53	–0.275		
M1-14	10	19		1.90	0.279		
M1-15	18	13.1		0.73	–0.138		
M1-16	14	22.7		1.62	0.210		
M1-17	2800	38.3		*0.014	*		
M1-18	J19	18.6		0.98	–0.009		
M1-19	J13	15.4		1.18	0.074		
M1-20	J18	28.4		1.58	0.198		
M1-21	J16	25.9		1.62	0.209		
M1-22	14	16.9		1.21	0.082		
M1-23	J17	30.5		1.79	0.254		
M1-24	12	12.6		1.05	0.021		
M1-25	27.8	34		1.22	0.087		
M1-26	22	24		1.09	0.038		
M1-27	11	7		0.64	–0.196		
M1-28	15	10.6		0.71	–0.151		
M1-29	16	14.4		0.90	–0.046		
M1-30	11	5.8		0.53	–0.278		
M1-31	17	4.8		*0.28	*		
M1-32	10	8.2		0.82	–0.086		
M1-33	16	8.2		0.51	–0.290		
M1-34	9.9	6.2		0.63	–0.203		
M1-35	12	10.8		0.90	–0.046		
M1-36	14	7.9		0.56	–0.249		
M1-37	18	15.5		0.86	–0.065		
M1-38	12	15.2		1.27	0.103		
M1-39	15	12.8		0.85	–0.069		
M1-40	12	13		1.08	0.035		

Continued

Table A1 (cont'd). Cr concentrations, ratios, and logarithms.

Sample number	Concentrations μg/g			Ratios and logarithms			
	QA	QC_1	QC_2	$\frac{QC_1}{QA}$	$log\frac{QC_1}{QA}$	$\frac{QC_1}{QC_2}$	$log\frac{QC_1}{QC_2}$
M1-41	18	16.3		0.91	–0.043		
M1-42	3.5	3.7		1.06	0.024		
M1-43	20	13		0.65	–0.187		
M1-44	14	11.5		0.82	–0.085		
M1-45	14	13.7		0.98	–0.009		
M1-46	350	199		0.57	–0.245		
M1-47	10	7		0.70	–0.155		
M1-48	43	24		0.56	–0.253		
M1-49	11	13		1.18	0.073		
M3-1	19	14.9	16.4	0.78	–0.106	0.91	–0.042
M3-2	ND<0.99	2.5	1.9	**	**	1.32	0.119
M3-3	19.9	24.6	24.2	1.24	0.092	1.02	0.007
M3-4	16	13	12	0.81	–0.090	1.08	0.035
M3-5	160	217		1.36	0.132		
M3-6	9.0	8.0		0.89	–0.051		
M3-7	11	7.8	10.9	0.71	–0.149	0.72	–0.145
M3-8	5.5	6.6	12.8	1.20	0.079	0.52	–0.288
M3-9	12	6.8	8.6	0.57	–0.247	0.79	–0.102
M3-10	25	6.0	6.0	*0.24	*	1.00	0.000
M3-11	6.0	8	6	1.33	0.125	1.33	0.125
M3-12	28	52	43	1.86	0.269	1.21	0.083
M3-13	24	35	31	1.46	0.164	1.13	0.053
M3-14	27	40	45	1.48	0.171	0.89	–0.051
M3-15	15.9	15		0.94	–0.025		
M3-16	32	41	38	1.28	0.108	1.08	0.033
M3-17	9.8	9	13	0.92	–0.037	0.69	–0.160
M3-18	44	50	36	1.14	0.056	1.39	0.143
M3-19	42	35	30	0.83	–0.079	1.17	0.067
M3-20	16.7	15	13	0.90	–0.047	1.15	0.062
M3-21	33.2	27	28	0.81	–0.090	0.96	–0.016
M3-22	9.3	15	13	1.61	0.208	1.15	0.062
M3-23	10	13	13	1.30	0.114	1.00	0.000
M3-24	10.2	14	14	1.37	0.138	1.00	0.000
M3-25	25	20	18	0.80	–0.097	1.11	0.04
M3-26	13	17	16	1.31	0.117	1.06	0.026
M3-27	6	8.1	10	1.35	0.130	0.81	–0.092
M3-28	13	6.9	7.8	0.53	–0.275	0.88	–0.053
M3-29	5	6.6	7.6	1.32	0.121	0.87	–0.061
M3-30	9	9.5	10	1.06	0.023	0.95	–0.022
M3-31	10.2	22	19	2.16	0.334	1.16	0.064
M3-32	42.4	31	30	0.73	–0.136	1.03	0.014
M3-33	53	44	48	0.83	–0.081	0.92	–0.038
M3-34	31	23	24	0.74	–0.130	0.96	–0.018
M3-35	24	17	20	0.71	–0.150	0.85	–0.071

Continued

Table A1 (cont'd).

Sample number	Concentrations μg/g			Ratios and logarithms			
	QA	QC_1	QC_2	$\frac{QC_1}{QA}$	$\log \frac{QC_1}{QA}$	$\frac{QC_1}{QC_2}$	$\log \frac{QC_1}{QC_2}$
M3-36	38	26	24	0.68	–0.165	1.08	0.035
M3-37	29	31	24	1.07	0.029	1.29	0.111
M3-38	48	32		0.67	–0.176		
M3-39	31	35		1.13	0.053		
M3-40	39.5	26		0.66	–0.182		
M3-41	50	36		0.72	–0.143		
M3-42	13	11	13	0.85	–0.073	0.85	–0.073
M3-43	5.6	10.8	12.7	1.93	0.285	0.85	–0.070
M3-44	15.1	16	18	1.06	0.025	0.89	–0.051
M3-45	8.9	7.5		0.84	–0.074		
M3-46	5.6	9.5		1.70	0.230		
M3-47	15.7	9.5		0.61	–0.218		
M3-48	16	12.9	12.9	0.81	–0.094	1.00	0.000
M3-49	21	16.2	15.6	0.77	–0.113	1.04	0.016
M3-50	13	11		0.85	–0.073		
M3-51	15	11	12	0.73	–0.135	0.92	–0.038
M3-52	11.5	6.9	8.2	0.60	–0.222	0.84	–0.075
M3-53	11.6	9.0	8.1	0.78	–0.110	1.11	0.046
M3-54	8.2	5.8	9.2	0.71	–0.150	0.63	–0.200
M3-55	16	18.8	10.8	1.18	0.070	1.74	0.241
M3-56	12	10	8	0.83	–0.079	1.25	0.097
M3-57	35.8	43	36	1.20	0.080	1.19	0.077
M3-58	24	11	12	0.46	–0.339	0.92	–0.038
M3-59	16	19.5	22.5	1.22	0.086	0.87	–0.062
M3-60	49	44	51	0.90	–0.047	0.86	–0.064
M3-61	13	15	16	1.15	0.062	0.94	–0.028
M3-62	21	22	23	1.05	0.020	0.96	–0.019
M3-63	26	26	26	1.00	0.000	1.00	0.000
M3-64	37.2	37	35	1.00	0.000	1.06	0.024
M3-65	16	51	38	*3.19	*	1.34	0.128
M3-66	ND<5	ND<10	ND<10	**	**	**	**
M3-67	25.4	59	23	2.32	0.366	2.57	0.409
M3-68	13	24	39	1.85	0.266	0.62	–0.211
M3-69	9.6	22.5		2.34	0.370		
M3-70	25	22	22	0.88	–0.056	1.00	0.000
M3-71	16.9	14	13	0.83	–0.082	1.08	0.032
M3-72	11.2	9.1		0.81	–0.090		
M3-73	8.6	9.7		1.13	0.052		
M3-74	22.6	68.7	31.5	*3.04	*	2.18	0.399
M3-75	33.6	42.5	53.7	1.26	0.102	0.79	–0.102

* Excluded from calculations because (<0.30 or >3.00).
** No ratio calculated because of ND< values.

SAMPLE CALCULATIONS FOR Cr DATA

Step 1. See Table A1 for ratios and logarithms.

Step 2. Mean of logs of $\frac{QC_1}{QC_2} = \frac{+0.304}{60} = +0.0051$ (antilog = 1.01).

Standard deviation $(S_x) = \sqrt{\frac{0.83787 - 0.00154}{59}} = 0.119$.

Step 3. 95% confidence limits on mean of logs are

$$+0.0051 \pm \frac{(2.00)(0.119)}{\sqrt{60}} = +0.0051 \pm 0.0307;\ \text{Prob}\left[-0.0256 \le \mu \le +0.0358\right] = 0.95.$$

Step 4. Taking antilogs, we get Prob $\lfloor 0.94 \le \mu \le 1.09 \rfloor = 0.95$.
The ideal or expected mean ratio of 1.00 is within the upper and lower boundaries for 95% confidence level.

Step 5. 99% tolerance limits for logs are

$$+0.0051 \pm (2.66)(0.119) = +0.0051 \pm 0.3165;\ \text{Prob}\left\lfloor -0.3114 \le x_i \le +0.3216 \right\rfloor = 0.99.$$

Step 6. Taking antilogs, we get Prob $\lfloor 0.49 \le x_i \le 2.10 \rfloor = 0.99$.
Two of the 60 $\frac{QC_1}{QC_2}$ ratios are outside of these boundaries for individuals;
$(M_3 - 67 = 2.57,\ M_3 - 74 = 2.18)$.

Step 7. For the ideal mean log of 0.00, Prob $\lfloor -0.3165 \le x_i \le +0.3165 \rfloor = 0.99$
antilogs give Prob $\lfloor 0.48 \le x_i \le 2.07 \rfloor = 0.99$.
The same two ratios are still outside these boundaries.

Step 8. Mean of logs of $\frac{QC_1}{QA} = \frac{-3.712}{116} = -0.0320$ (antilog = 0.93)

Standard deviation $(S_x) = \sqrt{\frac{3.355082 - 0.118784}{115}} = 0.168$

95% confidence limit on mean of logs

$$-0.0320 \pm \frac{(1.98)(0.168)}{\sqrt{116}} = -0.0320 \pm 0.0309.$$

Prob $\left[-0.0629 \le \mu \le -0.0011\right] = 0.95$; taking antilogs, we get
Prob $\lfloor 0.87 \le \mu \le 1.00 \rfloor = 0.95$.

The expected mean of 1.00 is barely within the upper boundary for 95% confidence level.
99% tolerance limits for logs

$$-0.0320 \pm (2.62)(0.168) = -0.0320 \pm 0.440.$$

Prob $\left[-0.472 \le x_i \le +0.408\right] = 0.99$; taking antilogs, we get

Prob $\left[0.34 \le x_i \le 2.56\right] = 0.99$.

None of the 116 ratios are outside of these boundaries.

For the ideal mean log of 0.000

Prob $\left[-0.440 \le x_i \le +0.440\right] = 0.99$ and, taking antilogs, we get

Prob $\left[0.36 \le x_i \le 2.76\right] = 0.99$.

Again, none of the 116 ratios are outside the boundaries.

APPENDIX B: DATA FOR METALS, VOCS, TPH, AND EXPLOSIVES

Table B1. Summary of As, Ba, Cr, and Pb concentration ratios.

(a) for duplicate QC samples

Element	*Total no. of ratios QC_1/QC_2*	*No. of ratios excluded (outside 0.30–3.00)*	*Percent exclusions*	*Geometric mean of retained ratios*	*95% confidence limits on means*	*99% tolerance limits for individuals*	*No. previously accepted ratios outside tolerances*
As	60	1	1.7	1.06	1.01–1.13	0.53–2.11	2
Ba	60	1	1.7	0.97	0.91–1.03	0.53–1.77	1
Cr	60	0	0	1.01	0.94–1.09	0.49–2.10	2
Pb	53	1	1.9	1.05	0.98–1.12	0.57–1.93	1
			Avg=1.3%	Avg=1.02			

(b) for QC/QA comparison

Element	*Total no. of ratios QC_1/QC_2*	*No. of ratios excluded (outside 0.30–3.00)*	*Percent exclusions*	*Geometric mean of retained ratios*	*95% confidence limits on means*	*99% tolerance limits for individuals*	*No. previously accepted ratios outside tolerances*
As	113	10	8.8	0.93	0.84–1.02	0.25–3.49	0
Ba	121	6	5.0	0.89	0.82–0.96	0.29–2.72	2
Cr	121	6	5.0	0.93	0.87–1.00	0.34–2.56	0
Pb	104	7	6.7	0.96	0.89–1.03	0.35–2.58	3
			Avg=6.3%	Avg=0.93			

Table B2. Percentages of unacceptable QC/QA ratios with three possible sets of rejection limits for As, Ba, Cr, and Pb.

Rejection limits	*0.25–4.00*		*0.30–3.00*		*0.40–2.50*	
	Exclusions		*Exclusions*		*Exclusions*	
Metal	*Low*	*High*	*Low*	*High*	*Low*	*High*
As	2	4	2	8	7	9
Ba	4	2	4	2	8	4
Cr	3	0	4	2	4	2
Pb	4	1	6	1	11	2
Subtotals	13	7	16	13	30	17
Grand totals	20		29		47	
Percent exclusions*	4.4%		6.3%		10.2%	

* Based on 459 total results.

Table B3. Summary of reporting limits for Ag, Cd, Hg, and Se.

Number of values	*Ag*		*Cd*		*Hg*		*Se*	
	QA 105	QC 93	QA 76	QC 60	QA 85	QC 101	QA 98	QC 104
Range µg/g	0.10–10	0.15–10	0.046–5.2	0.21–5.0	0.011–0.50	0.050–0.50	0.060–50	0.090–5.0
Exclusions	4	3	1	1	2	2	3	1
New range	0.10–3.1	0.15–2.9	0.046–3.0	0.21–2.3	0.011–0.30	0.050–0.20	0.060–3.0	0.090–2.5
Mean µg/g	1.4	1.7	0.71	1.2	0.11	0.13	0.62	0.75
Median µg/g	1.7	2.0	0.54	1.0	0.10	0.10	0.50	0.60
*Users guide to CLP, µg/L	10		5		0.2		5	
*SW-846 µg/L	7		4		0.2		2	

*Values taken from page D-2 (Table D-3) of ER 1110-1-263 (USACE 1990).

The concentrations shown are for guidance; they will vary with different samples and different matrices. Direct comparison of these values with the QC and QA is not possible because the latter concentrations are µg/g in soil while the reference values are µg/L for an extract.

Table B4. Characteristics of QC/QA ratios for archived VOC results for soils.

	Ethyl-benzene	*Toluene*	*Xylenes (total)*	*Various**	*Summary*
Number of pairs of real values	25	35	36	38	134
Number and % ratios outside the limits of 0.25–4.00	11 (44%)	14 (40%)	18 (50%)	13 (34%)	56 (42%)
Number and % ratios outside the limits of 0.20–5.00	11 (44%)	12 (34%)	12 (33%)	11 (29%)	46 (34%)
Number and % ratios outside the limits of 0.10–10.0	9 (36%)	6 (17%)	9 (25%)	8 (21%)	32 (24%)

* Consists of eight benzene, one chloroform, five 1,2-dichloroethene (total), two methyl ethyl ketone, eleven tetrachloroethene, and eleven trichloroethene ratios.

Table B5. Characteristics of archived total petroleum hydrocarbon (TPH) results from NPD and MRD.

	NPD results		*MRD results*
	$\frac{QC_1}{QC_2}$	$\frac{QC}{QA}$	$\frac{QC}{QA}$
Number of pairs of real values	83	95	61
Number and % ratios outside the limits of 0.25–4.00	5 (6.0%)	13 (13.7%)	28 (45.9%)
Geometric mean of ratios within the limits of 0.25–4.00	0.95	0.98	*
95% confidence limits on mean	0.84–1.07	0.84–1.15	*
99% tolerance limits around mean	0.22–4.07	0.16–6.15	*

* Not calculated due to extremely large % of values outside the range of 0.25–4.00.

TABLE B6. Characteristics of QC/QA ratios for archived explosives results for soils.

	RDX	*1,3,5-TNB*	*1,3-DNB*	*2,4,6-TNT*	*2-Am-DNT*	*2,4-DNT*	*Summary*
Number of pairs of real values	3	13	3	21	14	11	65
Number and % ratios outside the limits of 0.25–4.00	1 (33%)	2 (15%)	0 (0%)	3 (14%)	0 (0%)	1 (9%)	7 (10.8%)
Number and % ratios numerically < 1.00	1 (33%)	9 (69%)	3 (100%)	15 (71%)	8 (57%)	5 (45%)	41 (63.0%)

Table B7. Summary of QC/QA concentration ratios for nine VOCs in groundwater.

Compound	*Total no. of ratios QC/QA*	*No. of ratios excluded (outside 0.30–3.00)*	*Percent exclusions*	*Geometric mean of retained ratios*	*95% confidence limits on means*	*99% tolerance limits for individuals*	*No. previously accepted ratios outside tolerances*
Benzene	46	2	4.3	1.03	0.93–1.14	0.42–2.51	1
Chloroform	26	2	7.7	1.07	0.97–1.17	0.57–2.00	0
Ethylbenzene	37	0	0.0	1.11	0.98–1.25	0.41–3.01	0
Tetrachloroethene	19	0	0.0	0.97	0.84–1.12	0.41–2.30	1
Toluene	36	3	8.3	1.02	0.87–1.18	0.34–3.09	0
1,2-Dichloroethene	44	0	0.0	0.91	0.82–1.00	0.39–2.09	0
Trichloroethene	63	2	3.2	0.94	0.89–1.00	0.52–1.70	1
Vinyl Chloride	9	0	0.0	0.95	0.67–1.34	0.21–4.27	0
Xylene	41	2	4.9	1.09	0.95–1.26	0.34–3.50	0
	Total = 321	Total = 11	Avg = 3.2%	Avg = 1.01			Total = 3

Table B8. Distribution of VOC samples according to the number of analytes present and the number of QC/QA ratios outside 0.40–2.50.

No. analytes above reporting limits for both QC and QA per sample	*Distribution of samples according to number of analytes above reporting limits per sample*	*No. of samples with ratios outside 0.40–2.50*	
		One outlier	*Two outliers*
0	13	0	0
1	50	0	0
2	32	3	2
3	32	5	0
4	21	3	1
5	5	1	0
>5	0	0	0
	Total = 153	Total = 12	Total = 3

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestion for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
May 1996

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Comparison Criteria for Environmental Chemical Analyses of Split Samples Sent to Different Laboratories: Corps of Engineers Archived Data

5. FUNDING NUMBERS
USAED-MRD
MIPR GP95-11

6. AUTHORS
Clarence L. Grant, Thomas F. Jenkins and Anand R. Mudambi

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Cold Regions Research and Engineering Laboratory
72 Lyme Road
Hanover, New Hampshire 03755-1290

8. PERFORMING ORGANIZATION REPORT NUMBER
Special Report 96-9

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of the Chief of Engineers
Washington, D.C. 20314-1000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

Available from NTIS, Springfield, Virginia 22161

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)
Data comparison criteria were developed for quality control (QC) and quality assurance (QA) chemical analyses obtained during environmental studies directed by the U.S. Army Corps of Engineers (USACE). Archived results for 124 sets of eight metals in soils, 69 sets of fourteen volatile organic compounds (VOCs) in soils, 163 sets of total petroleum hydrocarbons (TPH) in soils, 79 sets of six explosives in soils, and 153 sets of fourteen VOCs in groundwater were analyzed statistically. Concentration ratios (QC_1/QC_2 and QC/QA) were lognormally distributed and this was the model used for comparisons. For both metals in soils and VOCs in groundwater, duplicate QC results should yield ratios between 0.50–2.00; less than 4% of the archived results were outside these limits. For QC/QA ratios, the limits suggested are 0.40–2.50; only 10.2% of metals ratios in soils and 5.6% of the VOCs in groundwater exceeded these limits. Considering that both methods are multi-analyte, we find only 4.0% of the metals samples and 2.0% of the VOC samples had more than one offending ratio per sample. Application of these limits to recent analyses produced very similar results. For VOCs, TPH, and explosives in soils, temporary limits of 0.25–4.00 are suggested with the understanding that improvements leading to tightening should be energetically pursued. Even with these wide limits, approximately 42% of VOCs, 14% of TPH, and 11% of explosives QC/QA ratios must be considered outliers. Here, too, recent analyses yielded very comparable percentages. The situation for VOCs in soils requires immediate attention to improve sampling and preparation procedures, but all methods can and should be capable of producing improved agreement between laboratories. Numerous recommendations are offered with this goal in mind.

14. SUBJECT TERMS
Explosives
Groundwater
Metals
Quality assurance
Soils
Split sample analysis
TPH
VOCs

15. NUMBER OF PAGES
40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
UL