5 cy

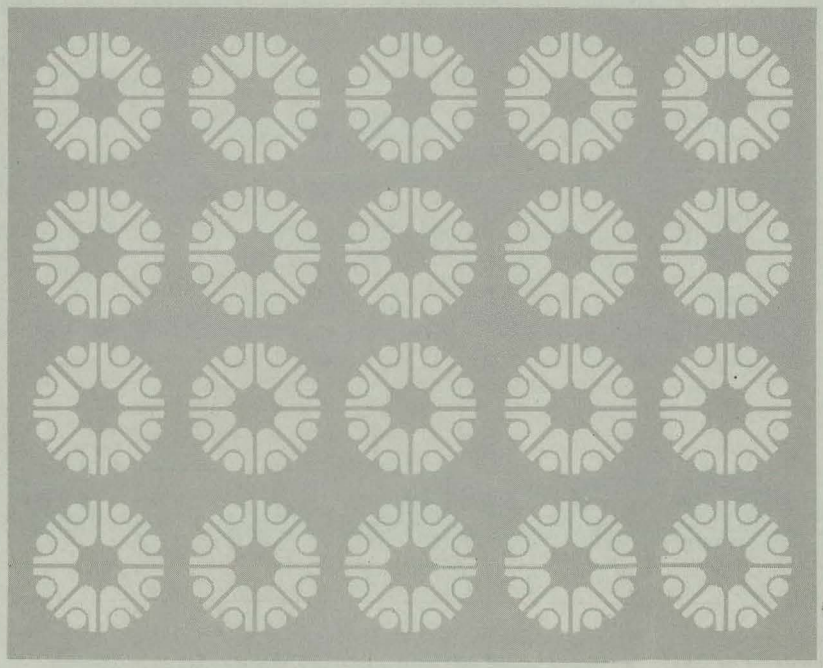
BNWL-B-217

Battelle
Pacific Northwest Laboratories
Richland, Washington 99352

This report is intended primarily for internal use by the sponsoring organization and Battelle.

STATISTICAL ANALYSIS OF SOIL PLUTONIUM STUDIES, NEVADA TEST SITE

SEPTEMBER 1972



MASTER

BNWL-B-217

DISCLAIMER

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
U.S. ATOMIC ENERGY COMMISSION
Under Contract AT(45-1)-1830

BNWL-B-217

STATISTICAL ANALYSIS OF SOIL PLUTONIUM STUDIES, NEVADA TEST SITE

by

L. L. Eberhardt[1] and R. O. Gilbert[2]

September 1972

BATTELLE
PACIFIC NORTHWEST LABORATORIES
RICHLAND, WASHINGTON 99352

1 Research Associate
Ecosystems Department

2 Senior Research Scientist
Mathematics and Physics Research Department

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

1.0 INTRODUCTION AND SUMMARY

This report has been written to summarize the results of our efforts to date in the plutonium environmental studies program of The Nevada Applied Ecology Group. Many of the field studies are currently not completed, but it seems desirable to produce a summary analysis of the statistical aspects at this time. We believe the report will be useful in planning further studies at the Test Site. It should also serve as a basis for discussion of the kinds of data needed for statistical appraisals of particular problems.

Much of our effort thus far has been directed towards the problems of estimating the inventory of plutonium in soil. The only extensive set of data available to us thus far is from the GMX area. That data suggests that it is quite feasible to proceed with an inventory for GMX with the tools at hand. The other kinds of objectives (described under SAMPLING TECHNIQUES) require more data and various decisions as to what is wanted from the studies.

The report provides some preliminary results on the use of the FIDLER instrument in Area 13, but we believe it is not profitable to try to do too much interpretation of such comparisons without more crosschecking FIDLER counts against "wet" chemistry or similar analyses. Another section of the report provides a basis for decisions on the number of replicates for inter-laboratory comparisons. It should be noted that this section serves for planning purposes -- once the study is completed, the results should be analyzed by other procedures (most likely an analysis of variance).

2.0 SAMPLING TECHNIQUES

Statistical methods in sampling mainly serve two purposes: (1) to secure unbiased ("accurate" in statistical jargon) estimates of some item or items of interest, and (2) to do so with a minimum cost in terms of sampling effort (or dollars). Much of the choice of a particular design thus has to do with efficiency -- getting the best results for the available funds, or, determining the minimum cost for an advance specification of chance errors considered acceptable or tolerable.

Selecting a sampling plan also depends very much on the objectives of the study. The NAEG plutonium studies appear to have a wide range of objectives, so it seems useful to sort out and comment on some possibilities here, under the following headings:

(1) Inventory

This is the most straightforward objective, being simply that of determining the total quantity of plutonium on a given area. Various complications have to be considered, but it does seem that an efficient sampling plan can be produced (one is described for the GMX area later in this report).

(2) "Clean-up"

We here suppose a decision is made to reduce the quantity of plutonium in a particular area. Such a decision necessarily hinges on an evaluation of costs and consequences (e.g., erosion subsequent to clean-up operations), and thus requires the results of the wide variety of studies now underway. But given such a decision, and assuming no practicable method exists (or is required) to remove all of the plutonium from an area, the sampling methods then have two objectives. One is to determine the initial distribution of plutonium in space, and the other is to assure that the removal

process meets some pre-assigned standards. Undoubtedly the standards will take the form of some statements as to residual quantities or concentrations of plutonium that may be acceptable. To do the job for minimum cost, one would thus like to seek out an optimum scheme -- one which minimizes the amount of sampling to be done before, during, and after the actual field operation. Such an optimal scheme would seem to require a cost function which specifies thing like soil removal costs, transportation, processing, and storage costs, as well as sampling costs (field and analytical). No doubt the "first pass" in clean-up can be expected to be less expensive than any subsequent operations, so there is a need to balance the advantages of meeting specifications in one pass against the need to minimize the amount of soil to be handled. All in all, sampling for "clean-up" appears rather more complicated than it would be for inventory purposes Actual designs will require some experience and specifications.

(3) Determining hazards

This is the most complicated of the several categories of objectives. Probably two subdivisions can be considered. One is the redistribution of plutonium, due mostly to wind action (resuspension). We have not tried to look into the statistical aspects of resuspension studies in any detail, but it seems clear that there are a variety of questions, at present mostly having to do with the complications of detailing the actual physical processes under field conditions. Once a methodology is decided on, there will be various sampling questions having to do with such things as estimating average rates and directions of movement.

A second category concerning hazards has to do with radioecology and the redistribution of plutonium through animal behavior and food-chains. In this case, sampling schemes are required to deal with a number of plant and animal

species, with space, and with changes in time. Sampling plans are thus likely to become extremely complicated, and probably need to begin with some sort of model representing what is known about the biology and radioecology of plutonium. We have been working on the development of optimum sampling plans for variations over time (e.g., fitting uptake and retention curves) and feel that enough results and data are available to produce a fairly simple initial rationale for such studies.

The above material characterizes the different objectives that need to be considered before starting out to do some sampling. We now proceed to describe the main sampling schemes. The underlying concepts are not at all complicated -- one can obtain a good understanding of the main features from a short book ("Sampling: A Quick Reliable Guide To Practical Statistics" by M. J. Slonim, Simon and Schuster, New York, 1960) that does not contain one formula or equation. Application of such methods does, however, call for both algebra and experience. The following items list some standard schemes:

(1) Systematic sampling

In sampling in two dimensions, this is commonly known as "grid" sampling. One simply lays out a uniformly spaced grid of sample locations and goes to work. There are some pretty substantial advantages in terms of simplicity and ease of operations. The disadvantages are somewhat more subtle. An obvious prospect for trouble is in circumstances where a systematic pattern exists in the distribution of the item being studied, so that sample estimates are either too high or too low depending on how the sampling grid happens to hit the existing pattern. A less obvious disadvantage is that there are uncertainties in determining the variability of the actual distribution of, say, plutonium from data taken systematically. Specification of

standards for inventory or clean-up depends on measuring the variability of plutonium distribution, and one might thus be lead astray by data from a "grid" sample. Nevertheless, practical field operations may need to be based on systematic samples, as for example, in clean-up operations.

(2) Simple random sampling

Randomized sampling constitutes the basis for virtually all of the statistical theory of sampling. It depends on selecting sample units by some chance process in which each sampling unit has a known probability of selection. In simple random sampling the probabilities are all equal. In the two-dimensional situation considered here, one arranges a grid of convenient units covering the area to be sampled and selects those units in the sample by drawing coordinates from a table of random numbers.

If it should somehow be true that the item under study (e.g., plutonium concentration) varies randomly over the area under study, then systematic and simple random sampling give essentially the same results. Since it is seldom true that such variables are truly randomly distributed in space (i.e., both natural and man-made processes rarely operate in a random fashion), only the random sampling method has predictable attributes (predictable in a statistical sense; "on the average"). Thus if "tolerances" or standards are to be met, with some agreed-on risk of error, random sampling methods are necessary.

(3) Stratified sampling methods

In practice, there generally is some advance knowledge about the population of items to be sampled. There is a variety of methods for using such advance information to produce an efficient sampling method. One of the best known such methods is stratification. In essence, one forms several separate strata by grouping units known to be "alike". Then each stratum is sampled at random. In the plutonium study at GMX area, a field survey instrument was used to delineate

three areas having different count-rates. Each such area is thus a convenient stratum (see Figure 3.1, page 3.3).

The complications in stratified sampling mostly have to do with determining how large a sample is required, and how to allocate (distribute) that sample to the several strata. In situations like soil sampling for plutonium, it usually is desirable to determine sampling intensity from data on the size of the stratum and advance estimates of the variability within the stratum. General experience with sampling for radionuclides shows that variability increases with mean concentration, so in effect, the number of samples required in a stratum increase with its size and mean concentration.

(4) Double sampling

This is a special method, of particular interest here because we had hoped to be able to use it to combine an inexpensive ("quick and dirty") method of measuring plutonium concentration with an expensive, but accurate, method (wet chemistry) to cut costs in inventory work. Double sampling basically amounts to calibrating readings from the inexpensive method by doing part of the analyses by both methods. Statistical details have to do with determining efficiency of the method and deciding how many "double" samples to take (usually by minimizing a cost function). There are several minimum requirements for use of the method (i.e., it may not be advisable in a given case). Use of the term "calibration" may be misleading in the sense that double sampling methods require recalibration in each and every application.

(5) Other methods

There are other methods that involve an auxiliary variate measured on each unit in the entire population (ratio and regression methods), methods that use "clusters" (usually groups of contiguous units), and methods for subsampling (or sampling in "stages"). A synopsis of such methods is provided by the previously

referenced book by Slonim, while technical details are available in several texts (we have used W. G. Cochran's "Sampling Techniques", Second Edition, J. Wiley and Sons, New York, 1963).

3.0 THE GMX STUDY

GMX is a site in Area 5 where several "safety tests" scattered significant quantities of plutonium, without fission products. It has thus been a good area for testing the utility of a hand-held gamma spectrometer--the "FIDLER" instrument, which measures the 60 KEV gamma rays emitted from Americium 241. Since Americium abundance is correlated with that of Plutonium-239 and 240, it is possible to assess the abundance of plutonium by gamma ray measurements. The "age" of the plutonium and stability of the ^{241}Am to $^{239,240}Pu$ ratios (which depend on the history of the particular bit of Pu being counted) clearly have a great deal to do with the efficiency and reliability of the method.

This section deals with some data obtained in an effort to cross-check FIDLER measurements with "wet chemistry" analyses of corresponding soil samples. Since the time of the survey, various efforts have been made to improve performance of the instrument in the field, so the results given here may no longer apply. However, they do serve to illustrate the statistical methodology required for future studies. Field work in the study was conducted by staff of Reynolds Electric and Engineering Company ("Reeco"), so that details of FIDLER operation and of the field surveys will be found in the appropriate Reeco reports.

FIDLER and "Wet Chemistry" Comparison

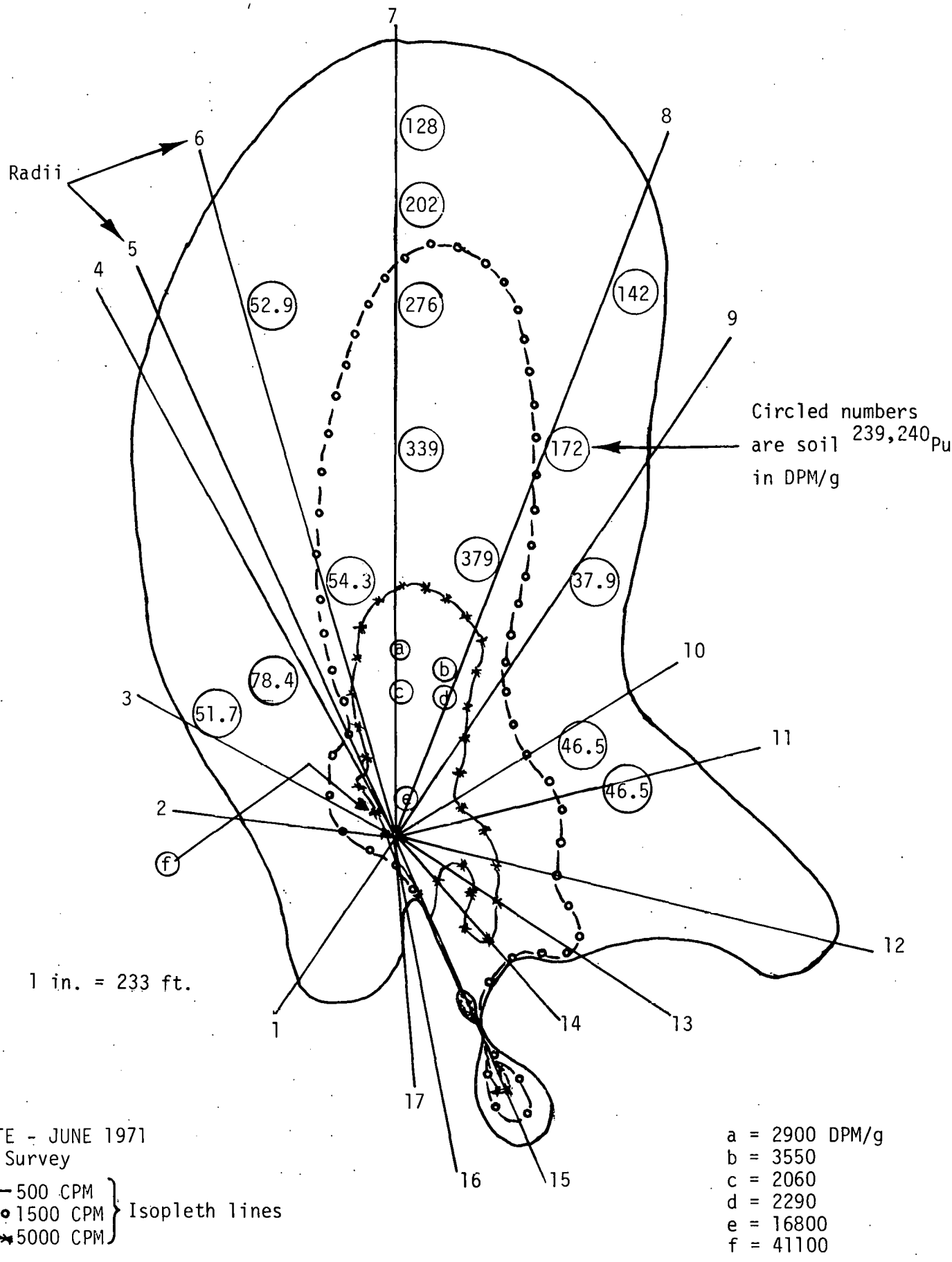
An extensive FIDLER survey of GMX area was conducted in June of 1971. The area was divided into three regions demarcated by isopleth lines drawn at points corresponding approximately to 500, 1500, and 5000 counts per minute (CPM) on the instrument. The higher counts were registered in a relatively small area around a bunker that served as "ground zero" for the tests. This area was bounded by a larger region (1500-5000 CPM), which in turn was

surrounded by a still larger area in which the lowest counts were obtained (500-1500 CPM). These areas (Figure 3.1) are conveniently referred to as "high", "medium" and "low", but have also been referenced as "innermost", "middle", and "outermost", as well as by the count-rate designations 5000+, 1500-5000 and 500-1500.

Seventeen radii were laid out from the bunker, and staked at 50 foot intervals. The angles between the radii were not all equal, so the stakes are not uniformly spaced in circular patterns. In particular, the radius line that goes about due north is the only representative of a large sector of the area. The number of stakes in each isopleth area was determined: 45 in the 5000+ CPM area, 41 in the 1500-5000 CPM area, and 108 in the 500-1500 CPM region. However, 12 of these 108 stakes were then arbitrarily removed from the list because two of the radii (numbers 4 and 5) were very close together, leaving 96 locations in this area as a population of sampling points.

Numbers were then assigned to the stakes and samples selected in each isopleth region by means of a table of random numbers. Initially 20 locations were selected in each region, but a later decision resulted in the use of the first 10 of each set that fell to the north of the bunker. There are thus 3 sets of 10 sample locations, one for each isopleth area (an additional set of 10 samples was later obtained outside the 500 CPM contour, but is ignored here since no analyses were done on those samples). Some further locations were selected for sampling to contrast Pu levels under bushes with those on desert "pavement", since some tests with the FIDLER showed appreciably higher readings under bushes than recorded on nearby pavement areas. However, wet chemistry results are not available for these samples. A more detailed description of the selection procedure (written by R. Lease of Reeco) is appended to this report (Appendix A).

Figure 3.1 Wet Chemistry Determinations of $^{239,240}Pu$ in DPM/g on GMX



The procedure employed at each sampling location was as follows (as given in a letter from R. Lease, October 7, 1971):

1. A standard uncollimated FIDLER reading was taken one foot over the selected samping location. (A standard uncollimated FIDLER reading is taken at one foot above the ground and integrates the activity present in a one meter diameter circle.)

2. Three steel rings, five inches in diameter are then pushed into the soil to a depth of three centimeters. The rings are of steel and designed to hold the collimator in place. Three rings are placed inside the one meter area defined by the uncollimated FIDLER. They are placed in the center of the area in a triangular pattern at 0° (N), 120° and 240° in as close proximity as possible without disturbing one another.

3. A stainless steel lined, lead collimator, five inches inside diameter, is placed on the ring. The FIDLER is inserted into the collimator and a reading recorded (The collimator is built to hold the FIDLER probe one foot above the surface of the ground.)

4. The soil is then completely removed from inside the lower ring to a depth of three centimeters, leaving a flat bottom inside the ring.

5. The soil from each ring is then labeled with its applicable station number and compass designation, i.e., 9-100, 0°. The sample is double bagged and transmitted to the sample preparation laboratory for processing and/or storage.

6. A portion of each sample is retained in a sample library at Mercury.

7. Sample preparation is dependent upon the request of the investigator desiring the sample. Preparation may include sieving to specified size, crushing and/or grinding to 200 mesh or a combination of other preparations.

8. The sample is then bagged and canned in commercial can for shipment or storage.

A further selection was exercised in that not all of the 30 samples were analyzed by wet chemistry methods. Results are available for the following numbers of samples:

500-1500 CPM	10
1500-5000 CPM	4
5000+ CPM	6
	20

The wet chemistry results used here were done on 30 gram aliquots (smaller for the 5000+ CPM area), taken from the top 3 cm of soil in the 0° location at each sampling point. Data used here has been extracted from a letter written by Dr. G. Hamada (then of Teledyne Isotopes, Palo Alta, California) to J. J. Davis, September 16, 1971.

The various sampling and selection procedures described above are such that it seems unwise to regard the data as coming from random sampling of the three isopleth areas. In particular, samples were not taken under bushes, so the data pertain only to the area not covered by bushes. However, the FIDLER instrument is not normally used in the vicinity of bushes (in view of the hazard to the detector), and the comparison of interest here (FIDLER vs. wet chemistry) very likely is not much affected by departures from truly random selection of sample plots.

The original expectation was that the FIDLER readings might serve to reduce the cost of an area inventory by reducing the number of wet chemistry determinations required. The basic idea is that of "double sampling" or "two-phase" sampling, wherein a correlation between an expensive but accurate determination (here, wet chemistry) is used to permit "calibration" of a large number of inexpensive samples to yield a rather precise overall estimate.

FIDLER readings were taken in two modes, collimated (with a lead collar) and uncollimated. The essential determination is that of counts per minute in the 60 KEV channel, which corresponds to Americium-241 (which as mentioned above is in turn associated with the item of real interest, $^{239,240}Pu$). Readings were also taken in the 122 KEV channel as a means of producing a background correction factor.

The collimated FIDLER readings were evidently unsuitable, giving lower values in the 60 KEV channel for the middle isopleth than were obtained in the outer isopleth (lowest Pu levels), and negative readings (middle isopleth) when "background-corrected." Hence the collimated readings are not considered further here.

Turning to the uncollimated FIDLER readings, its quite clear that neither the 60 KEV readings or the background-corrected readings are sufficiently well-correlated with the wet chemistry determinations for double sampling purposes. The background corrections consisted of taking readings in the 60 and 122 KEV channels in areas away from Pu contamination (which can be checked by examining plots of gamma spectra), and taking the ratio 60 KEV CPM/122 KEV CPM as an adjustment to use in Pu contaminated areas. The ratio thus used for the uncollimated mode of operation was 0.6. Data obtained for the outermost (500-1500 CPM) isopleth area are as follows:

	CPM (in thousands)				
REECO Code	60 KEV	122 KEV	Adjusted ("Diff.")	Ratio 60 KEV/122 KEV	$^{239,240}Pu$ (DPM/g)
4-350 0°	4.4	6.5	.5	.68	78.4
8-1000 0°	4.8	7.2	.5	.67	142.
10-350 0°	5.5	8.0	.7	.69	46.5
7-1050 0°	5.0	7.5	.5	.67	202.
11-450 0°	5.4	8.0	.6	.68	46.5
6-800 0°	5.0	7.0	.8	.71	52.9
8-700 0°	5.0	7.5	.5	.67	172.
3-400 0°	5.0	7.5	.5	.67	51.7
9-600 0°	4.5	6.5	.6	.69	37.9
7-1150 0°	4.8	7.0	.6	.69	128.

The 60 KEV readings adjusted for background (column 4 in the above table) are computed as follows: 60 KEV CPM - 0.6 (122 KEV CPM). Plots of 60 KEV or adjusted values against Pu values show little evidence of a relationship, the FIDLER readings being essentially constant, clustered around 5.0 (or .5). The sample correlation coefficient between the adjusted 60 KEV and wet chemistry readings is negative and small ($r = -.540$), so there is little or no evidence of a useful linear relationship at these levels of contamination. The ratio of readings in the 60 KEV and 122 KEV channels (see table above) is so nearly constant as to hint that the main contribution to 60 KEV is some kind of background. Whatever the actual reasons, it seems unlikely that these FIDLERS would be useful for double sampling purposes at the lower concentrations of plutonium.

Assessing the higher levels is complicated by fewer samples. The 4 samples at the intermediate level do suggest a positive correlation between Pu level and FIDLER reading, but are not sufficient to calculate a useful correlation coefficient. It is interesting that both 60 KEV and 122 KEV readings seem below those given above, but the background correction gives higher presumed Pu levels, in accord with the wet chemistry. Data are as follows:

	CPM (in Thousands)				
REECO Code	60 KEV	122 KEV	Adjusted ("Diff.")	Ratio 60 KEV/122 KEV	$^{239,240}Pu$ (DPM/g)
7-750 0°	4.2	5.2	1.1	.81	339.
7-900 0°	4.6	6.0	1.0	.77	276.
8-500 0°	4.8	5.5	1.5	.87	379.
6-400 0°	3.6	5.0	.6	.72	54.3

On November 1, 1971 additional FIDLER readings were taken at the same locations as in the June 7, 1971 survey for the 1,500-5,000 CPM isopleth. These readings were taken in an effort to duplicate the FIDLER results recorded on the initial survey on June 7, 1971. The data are:

REECO Code	CPM (in Thousands) 60 KEV	122 KEV	Adjusted ("Diff.")	Ratio 60 KEV/122 KEV	$^{239,240}Pu$ (DPM/g)
7-750 0°	5.0	5.0	2.0	1.0	339.
7-900 0°	5.0	6.0	1.4	0.83	276.
8-500 0°	5.0	6.0	1.4	0.83	379.
6-400 0°	4.0	5.0	1.0	0.80	54.3

The new adjusted and unadjusted 60 KEV readings are somewhat larger than obtained in the earlier survey. Again, we do not have sufficient data to compute a useful correlation coefficient.

Turning to the highest levels of Pu contamination (innermost isopleth area), it appears that there is a clear correlation between FIDLER and Pu level. The data are:

REECO Code	CPM (in Thousands) 60 KEV	122 KEV	Adjusted ("Diff.")	Ratio 60 KEV/122 KEV	$^{239,240}Pu$ (DPM/g)
7-200 0°	17.0	8.0	12.2	2.12	2060.
7-250 0°	9.	13.	1.2	.69	2900.
8-200 0°	12.	7.	7.8	1.71	2290.
8-50 0°	65.	8.	60.2	8.12	16800.
8-250 0°	13.	6.	9.4	2.17	3550.
5-50 0°	90.	10.	84.	9.00	41100.

The sample size (6) is again small for a useful estimate of the correlation coefficient ($r = .950$, a value which would be obtained with probability about .0025 for 6 observations and true correlation of zero).

A warning needs to be inserted here against the obvious temptation to combine all 3 isopleths and calculate a sample correlation coefficient. Such a procedure will give an evidently "good" relationship, but the correlation coefficient is not reliable when high and low values are deliberately selected for inclusion in the data.

As a sidelight on behavior of the FIDLER instrument, a multiple regression analysis of Pu determination against 60 and 122 KEV readings was calculated for the entire set of data, with the relationship studied being:

$$y = b_o + b_1 X_1 + b_2 X_2 ,$$

where $y = {}^{239,240}Pu$ by wet chemistry,

X_1 = 60 KEV channel readings,

and X_2 = 122 KEV channel readings.

The idea is that the regression coefficients b_1 and b_2 might give some notion of an appropriate adjusting relationship between the two channels. The results appear as Table 3.1, but mostly seem to show a discrepancy in pattern for the 1500-5000 CPM data. Dropping the 500-1500 CPM data gives Table 3.2, but there is little change in the regression coefficients. One would expect b_2 to be negative in terms of the adjustment previously used (subtracting .6 of the 122 KEV channel reading from the 60 KEV channel data). Notice (Table 3.2) that the 1500-5000 CPM data clearly differ from the regression relationship which is largely imposed by the 5000+ CPM data due to its wide range of values.

Gamma Counts and Wet Chemistry

An alternative to use of a field instrument, such as the FIDLER, is to transport soil to the laboratory and measure the gamma ray emissions there. Such a procedure is obviously more time-consuming and expensive than the field operation, but has the advantage that more sensitive devices (e.g., larger crystal) and better geometry can be employed. Further, the detection device is exposed only to a known sample, whereas the field instrument may be influenced by a "hot particle" lying outside the soil actually collected as a sample.

Table 3.1

Multivariate Regression of $^{239,240}Pu$ Wet Chemistry Determinations on 60 & 122 KEV Channel FIDLER Readings

Regression Equation:

$\hat{y} = -3535.49 + 406.78\ X_1 + 208.52\ X_2$

Standard Deviation of Residuals ($s_{y \cdot x}$) = 2338.76

Index of Determination (R-SQ) = 0.938

Variable	Mean	Standard Deviation
y	3535.31	9360.67
X_1	13.63	21.92
X_2	7.32	1.73

Observed Versus Calculated

	Observed	Calculated	Difference	Percent Difference
500 - 1500 CPM	78.4	- 390.255	- 468.655	120.
	142	- 81.5769	- 223.577	274.
	46.5	369.988	323.488	87.4
	202	62.3362	- 139.664	-224.
	46.5	329.31	282.81	85.8
	52.9	- 41.9258	- 94.8257	226.1
	172	62.3362	109.664	-175.9
	51.7	62.3362	10.6362	17.
	37.9	- 349.578	- 387.478	110.8
	128	- 123.282	- 251.282	203.8
1500 - 5000 CPM	339	- 742.692	-1081.69	145.6
	276	- 413.161	- 689.161	166.8
	379	- 436.067	- 815.067	186.9
	54.3	- 1028.46	-1082.76	105.2
5000 + CPM	2060	5047.95	2987.95	59.1
	2900	2836.34	- 63.664	- 2.2
	2290	2805.53	515.532	18.3
	16800	24573.4	7773.38	31.6
	3550	3003.79	- 546.212	- 18.1
	41100	35159.9	-5940.08	- 16.8

Table 3.2

Multivariate Regression of $^{239,240}Pu$ Wet Chemistry Determinations on 60 and 122 KEV Channel FIDLER Readings (Excluding the 500 - 1500 CPM Area)

Regression Equation:

$\hat{y} = -3776.54 + 408.71\ X_1 + 221.03\ X_2$

Standard Deviation of Residuals ($s_{y \cdot x}$) = 3294.28

Index of Determination (R-SQ) = 0.928

Variable	Mean	Standard Deviation
y	6974.83	12311.8
X_1	22.32	28.46
X_2	7.37	2.39

Observed Versus Calculated

	Observed	Calculated	Difference	Percent Difference
1500 - 5000 CPM	339	- 910.6	-1249.6	137.2
	276	- 570.292	- 846.292	148.3
	379	- 599.066	- 978.066	163.2
	54.3	- 1200.03	-1254.33	104.5
5000 + CPM	2060	4939.75	2879.75	58.2
	2900	2775.25	- 124.754	- 4.4
	2290	2675.18	385.18	14.3
	16800	24557.7	7757.74	31.5
	3550	2862.86	- 687.143	- 24.
	41100	35217.5	-5882.49	- 16.7

The study by G. Hamada (referenced above; data in letter to J. J. Davis, 16 September, 1971) provides an illustration. The plutonium data are the same as used above, and the only change is that we now replace a FIDLER count with one made on a laboratory gamma-detection device (sodium iodide crystal). Results for the three isopleth areas follow:

REECO Code	y $^{239,240}Pu$ DPM/g	x ^{241}Am DPM/g	Correlation and Regression Results
500-1500 CPM Area			
4-350 0°	78.4	12.8	
8-1000 0°	142.	14.8	
10-350 0°	46.5	5.3	r = .9718
7-1050 0°	202	23.2	
11-450 0°	46.5	6.4	y = -13.18 + 9.29 x
6-800 0°	52.9	8.3	
8-700 0°	172	18.2	Ratio $\frac{\Sigma y}{\Sigma x}$ = 8.17
3-400 0°	51.7	5.9	
9-600 0°	37.9	5.9	
7-1150 0°	128.	16.5	
SUM	957.9	117.3	
1500-5000 CPM Area			
7-750 0°	339	38	r = .939
7-900 0°	276	42	
8-500 0°	379	41	y = -16.37 + 8.58 x
6-400 0°	54.3	8.8	
SUM	1048.3	129.8	Ratio = 8.08
5000+ CPM Area			
7-200 0°	2060 (30 g)	246	
7-250 0°	2900 (30 g)	259	
8-200 0°	2290 (50 g)	188	
8-50 0°	16800 (50 g)	1870	r = .9995
8-250 0°	3550 (50 g)	304	
5-50 0°	41100 (50 g)	4980	y = 794.7 + 8.15 x
SUM	68700	7847	Ratio = 8.75

As is evident from the correlations, and Figures 3.2 and 3.3, the gamma counts will apparently do very well in predicting concentrations of $^{239,240}Pu$ found by wet chemistry. There is consequently the prospect that combining laboratory gamma counts and wet chemistry determinations may well serve to reduce the costs of doing an inventory. A rough idea of possible reductions in cost is available from the following data (derived from variance formulae given in Cochran (1963, p. 337):

Total cost reduced 25% if:

Ratio of costs	17	7	4	2.7
True correlation	.72	.83	.91	.95

Total cost reduced 50% if:

Ratio of costs	50	15	5.7	3.5
True correlation	.6	.87	.95	.98

The "ratio of costs" entry is the ratio of cost of doing the expensive but accurate determination (wet chemistry) to that of the inexpensive but less accurate method (gamma count). Since these calculations are based on an assumed knowledge of the true correlation, which we can in practice only estimate from sampling data, they will usually tend to overemphasize the savings. In the present instance, a rough guess at the costs involved gives a cost ratio of 4 (G. Hamada, personal communication) and the correlation seems at least .95, so it appears the combined (double sampling) survey might be done for about half the cost of one that used wet chemistry alone.

A Stratified Sampling Plan for GMX to Estimate Pu Inventory

Since the essential information is at hand, it seems worthwhile to describe a stratified sampling plan to estimate Pu inventory for the GMX area. Readers are reminded that the calculations are for purposes of illustration and planning only. There are several reasons why the data should not be

Figure 3.2 Laboratory Determinations of ^{241}Am vs Wet Chemistry Determinations of $^{239,240}Pu$ in DPM/g on GMX (Outer and Middle Isopleths)

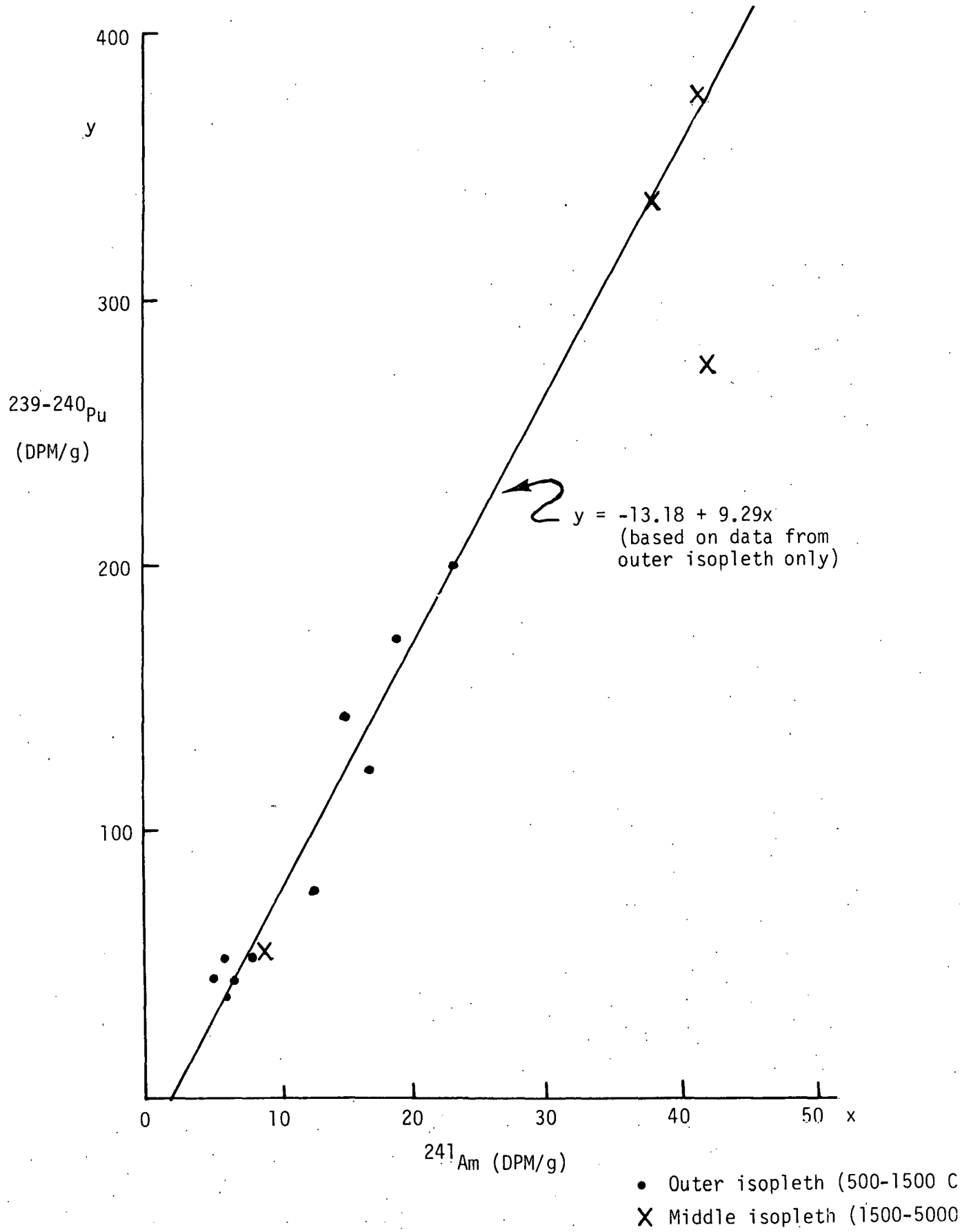
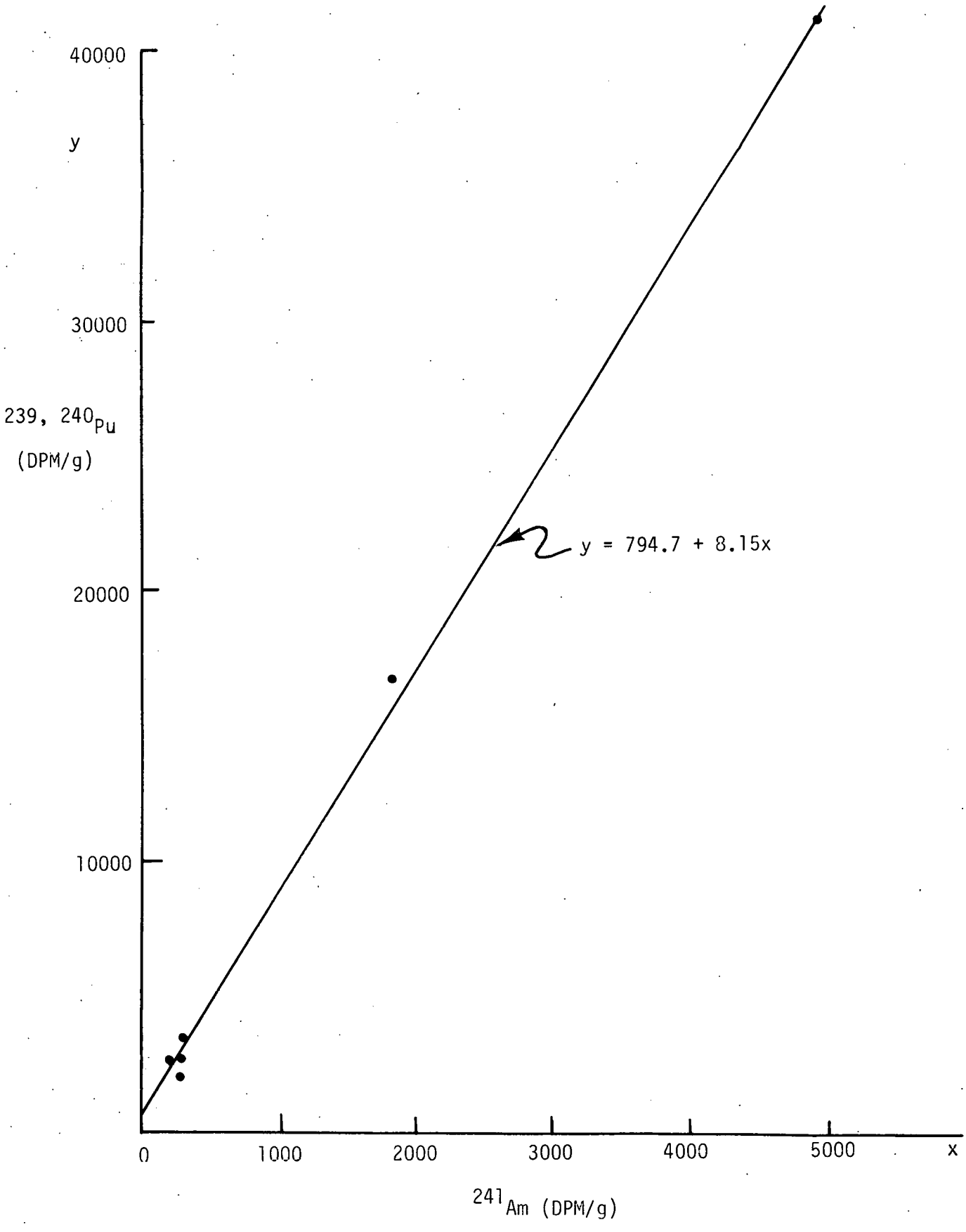


Figure 3.3 Laboratory Determinations of ^{241}Am vs Wet Chemistry Determinations of $^{239,240}Pu$ in DPM/g on GMX (Inner Isopleth)



considered as suitable for estimating the quantity of plutonium (in the surface 3 cm) in the GMX area:

(1) Areas under creosote bushes were not included in the main sampling study.

(2) No samples were taken south of the bunker.

(3) There were various other departures from random sampling, including the fact that 10 of the selected samples were not analyzed.

A fourth, and perhaps more compelling reason is that the study was designed to explore accuracy of the FIDLER instrument, and not for estimating total plutonium on the area. Hence, the available estimate is not suitable due to the large variance associated with it. However, the data do make it possible to lay out an efficient plan for surveying GMX. At the present, it seems likely that the survey might be carried out by using the laboratory gamma scans described above, and cross-checked with a small sample of wet chemistry analyses (say 20 or 30 such analyses). For illustration here, though, we use the available wet chemistry determinations.

The procedure is simple, straightforward and described, e.g. in Cochran (1963). The area has already been stratified - by construction of the isopleth lines. The statistical procedure consists mainly of working out means and variances of Pu concentration in each stratum, and combining them for an overall estimate. The basic data are those given in the above sections (^{239,240}Pu in DPM/g by isopleth area) and an estimate of the relative areas included within the three isopleth areas (obtained by rough planimetering of a plot of the June, 1971 FIDLER survey). We then combine the data in the following table:

Stratum	Area (sq. ft.)	W_h (Proportion)	n_h (Number of samples)	$\bar{y}_h$ (mean Pu conc.)	s_h^2 (variance)	Coefficient of Variation
1 (500-1500 CPM)	366,600	.4892	10	95.8	3618.	.63
2 (1500-5000 CPM)	291,100	.3885	4	262.1	20,980.	.55
3 (5000+ CPM)	91,600	.1223	6	11,450.	243.1×10^6	1.36
	349,300	1.0000	20			

One interesting feature of the above table is that the coefficients of variation (standard deviation divided by mean) for strata 1 and 2 are roughly equal to values observed for concentrations of worldwide fallout radionuclides in soils in various places. One might thus argue this as grounds for doubting that the stratification could be improved for that area. However, a look at the pattern (Figure 3.1) of the observations suggests that a few additional samples might serve to improve the stratification somewhat in that it appears that the "downwind" pattern was more pronounced than that recorded by the instrumental (FIDLER) survey.

Combining the data in the above table proceeds simply by using the W_h as weights. We then get the overall estimate of Pu concentration as:

$$(1) \qquad \bar{y}_{st} = \sum_{h=1}^{3} W_h \bar{y}_h = 1549$$

and a variance estimate as:

$$(2) \qquad V(\bar{y}_{st}) = \sum_{h=1}^{3} \frac{W_h^2 s_h^2}{n_h} = 909{,}910 \quad .$$

The square root of this variance estimate usually serves to provide an approximate set of confidence limits for the estimated mean, i.e.,

$$\bar{y}_{st} \pm 2 \sqrt{V(y_{st})}$$

provides limits that can be described as being expected to contain the true, but unknown, mean concentration in 95 percent of such surveys.

In the present case, the standard error (square root of $V(y_{st})$) is too large (954.) to let one place much confidence in the estimate. Clearly, more samples are needed. When variances differ substantially between strata, one obvious thing to do is to increase sample size in the more variable strata. However, the size of the stratum also influences the overall variance. Hence, as a glance at equation (2) above will show, the logical thing is to increase n_h wherever W_h and s_h are large. It turns out that "optimal allocation" (minimum overall variance for fixed cost) distributes samples according to the equation

$$n_h = n \frac{W_h s_h}{\sum W_h s_h} \quad , \tag{3}$$

where n is the overall sample to be distributed (allocated) to the several strata. Calculations with the present data give:

Stratum	$W_h s_h$	Proportion (n_h/n)
1	29.4	.0147
2	56.3	.0282
3	1906.9	.9571
	1992.6	1.0000

It now becomes obvious that the problem lies in the 5000+ CPM stratum, since this apparently calls for over 90 percent of the sampling.

A look at Figure 3.1 and the above allocation suggests that a little more preliminary sampling might save a considerable amount of effort in the final survey. Specifically, the two very high values observed in stratum 3 (5000+ CPM) are quite close to the bunker. If we neglect these two observations, and calculate $W_3 s_3$ from the remaining observations, we get

$$W_3 s_3 = 81.7$$

which is roughly "in line" with strata 1 and 2. There is thus rather compelling evidence that one ought to divide the 5000+ stratum into two new strata. With only two observations near the bunker, it doesn't seem very practical to calculate a variance and estimate sample sizes required. A much more practical route seems to be to analyze say 8 or 10 more samples in the vicinity of the bunker, and then design the final plan.

Insofar as the overall sampling plan is concerned, the need is to reduce the very high stratum to as small an area as possible, so that the W_h^2 term in equation (2) will operate to minimize the overall variance. From a practical standpoint, more samples in the bunker area might conceivably lead to a decision to remove or immobilize the plutonium in that immediate area, thus removing it from consideration in the overall sampling study.

If we assume no change in stratification is made, then a rough estimate of the sample size needed can be obtained by supposing we want to estimate the mean concentration within $\pm$ 15%, so that

$$V(\bar{y}_{st}) = \left[.15\ \bar{y}_{st}\right]^2 = \left[.15(1550)\right]^2 = 54{,}050.$$

Cochran (1963:104) gives an equation for sample size as

$$n = \frac{1}{V(\bar{y}_{st})} \sum \frac{W_h\ s_h^2}{w_h}$$

where w_h = proportion allocated to stratum h. Using this equation and data from the tables above, we get a rough estimate of n = 73.

The preceding material describes the essentials for a survey to estimate the plutonium inventory at GMX. Decisions would need to be arrived at concerning several points before actually beginning the survey. These include: (1) Is the precision of $\pm$ 15 percent adequate? (2) Methods for sampling under bushes need to be arranged. (3) An additional stratum near the bunker should be

delineated (by additional sampling). (4) There may be a need for another stratum beyond the 500 CPM isopleth. (5) Some sampling to depths greater than 3 cm is required.

4.0 AREA 13 STUDY

While this study is incomplete at the time this report is being written, it may nonetheless be useful to summarize the statistical analysis of the initial data. The only results available to us at present are from a series of FIDLER readings taken on Microplot number 1 to explore several questions. The initial results are given below, with details in Appendix B. Four microplots were chosen in January 1972 in a radiation zone of medium intensity to study the distribution of plutonium with respect to vegetation, soil depth, topography, etc. To date, data have been collected only in Microplot 1. The exact location of Microplot 1 was shifted several feet from that originally chosen so that the plot included within its boundaries a variety of shrubs as well as bare ground. The intent of this shift was to obtain a microplot more suitable for determining whether radioactivity under shrubs was greater than that between shrubs (a phenomenon observed at the GMX site). Microplots 2 and 3 were staked and their boundaries delineated by florescent tape while Microplot 4 was only staked. Microplot 1 was divided into approximately 70 half-meter square units laid out in a rectangle with 10 rows numbered from 1 to 10 and 7 columns labeled from A through G.

Since the data on Microplot 1 have been collected, FIDLER readings have been made in Area 13, but these data are not complete at this time. Hence, we discuss below and in Appendix B only the microplot data.

Effect of shrub removal

A single FIDLER instrument was used to take readings one foot above ground level at the centers of 70 half-meter square units before and after shrubs were removed. A paired t-test computed on the 35 units containing shrubs indicates there may be an effect on FIDLER readings due to the presence of bushes. It may be worthwhile to examine the question more thoroughly by obtaining more data.

Differences between machines

Two FIDLER instruments were used to take readings on all 70 units. Tests on two days (January 20 and 21, 1971) both indicated statistically significant differences between instruments.

Analysis with replicate readings

Two readings were taken with each machine in the morning and afternoon of January 21, 1972. An inspection of the data suggested little or no change with time, so the replicate readings were used to examine behavior of the instruments in an analysis of variance. It turned out (see Appendix B) that there was a significant "machines x plots" interaction, suggesting some inconsistency in behavior of the instruments over the plot. One prospective explanation lies in how different operators handle the machines. In this case, we do not have data to attempt to separate an "operator effect" from a "machine effect", but presumably such an experiment could be conducted if it appears useful.

In addition to the above analysis a chi-square test was applied to 5 replicate readings obtained in the 60 KEV and 122 KEV channels (see Appendix B) in rows 5 and 6 which suggests that the 122 KEV readings are more precise (reproduceable) than the 60 KEV readings.

Future studies of the FIDLER instrument

The GMX study yielded relatively poor correlations between FIDLER counts and "wet chemistry" assays for plutonium. The overall indications from that study are that the instrument may well serve usefully for stratification and general survey work, but it did not appear useful for inventory estimates. Various improvements in operating procedures and in the instrument have since

been instituted, so that better performance can be expected. However, it seems to us that further comparisons of the type described here should also include a statistically designed test against wet chemistry analyses or against the laboratory gamma scans discussed under the section above on GMX. Unless there is better evidence of precision in measurement of plutonium concentrations by the FIDLER, there is not much point in very detailed assessment of differences between individual instruments, operators, etc. None of the above should be taken as suggesting that the instrument is not useful; what is needed is a better definition of the circumstances in which it is best used.

5.0 INTERLABORATORY COMPARISONS

In the course of the NAEG study, we were asked about sample sizes for use in interlaboratory comparisons of accuracy of plutonium analysis by "wet" chemistry. We assume the main question to be one of determining how many replicate samples of the same material should be assayed by each laboratory. An answer depends on a number of factors, so that one really needs to make several initial choices and consult some tables for the corresponding number of replicates.

The factors involved include the following:

(1) The risk (α) one is willing to take of deciding that there is a real difference among laboratories when in fact none actually exists. Many investigators routinely set this risk equal to .05, so we have also settled on that level although it is not always advisable to make such an arbitrary choice.

(2) The chances of detecting a real difference among laboratories (the "power" of a test). This amounts to a measure of the sensitivity of the interlaboratory comparison, and is the key issue here. Sensitivity of the comparison depends on the number of replicates run in each lab, which is what we are trying to decide here. If the test is not sufficiently sensitive, there may be no point in making the comparison.

(3) Judging how sensitive the comparison will be requires some way of expressing differences between laboratories. In the present instance, it is most convenient to express this as a ratio (R). Roughly speaking, R represents the minimum error we might expect to detect; i.e., if some lab produces an erroneous analysis that is R times the true value, we might expect to detect that error with probability equal to the stated "power"

(sensitivity) of the comparison. R is influenced by the number of replicate analyses done by each lab.

(4) Variability among replicate analyses done by each laboratory. In practice, duplicate analyses like that for plutonium in soil seldom come out with identical results, due to various differences in amount of plutonium between aliquots, variations in analytical technique, and so on. This variability is the reason for the present study -- if it were zero, there would be no question as to the influence of chance on interpretation of the end result. In the present study, we measure such variability by the coefficient of variation (standard deviation divided by the mean).

(5) Number of laboratories involved in the study. This has a fairly minor influence on the outcome of the study in the present case.

A choice of values for the 5 factors above permits calculating a set of tables wherein one can look up values of R and thus have an indication of the kind of sensitivity resulting from a particular set of circumstances. As noted in (1) above, we adopt $\alpha = .05$ for all calculations here. Also, the more detailed tables in Appendix C (Tables C.4, C.5 and C.6) show relatively small differences as numbers of laboratories change, so we will here illustrate results for 4 laboratories. The main question in planning a study has to do with the variability likely to be encountered, and some experience (described in the appendix) suggests that the coefficient of variation may be as small as .05 (5 percent). However, this is not established with any certainty, so one might (in the absence of any actual data under circumstances of the proposed study) wisely also contemplate what happens if the coefficient of variation is as large as .20 (20 percent). The two tables below give some representative data, and the more detailed tables in Appendix C can be consulted for further results.

Values of R for Coefficient of Variation = .05

Number of Replicates	β = .95	.90	.80	.70
2	1.43	1.38	1.32	1.28
4	1.19	1.17	1.15	1.13
6	1.14	1.13	1.11	1.10
10	1.10	1.09	1.08	1.07

Values of R for Coefficient of Variation = .20

2	4.11	3.55	2.97	2.62
4	1.99	1.86	1.72	1.63
6	1.69	1.61	1.52	1.45
10	1.47	1.42	1.36	1.32

To use the tables, one first selects a value of β (the "power" of the test), where 1-β is the risk we are willing to take of failing to detect a difference when one really exists. As a general rule, β has to be set fairly high, or there simply is no point in doing the comparison. Having selected β, one then consults the tables to see what effect the number of replicates has on sensitivity. For a coefficient of variation of .05 when β=.95 it can be seen that the difference between 2 and 10 replicates amounts to roughly the difference between detecting an error of about 40 percent against finding an error of about 10 percent.

If the coefficient of variation is as high as 20 percent, the situation becomes much worse, as evidenced in the second table above. Here we see that the interlaboratory comparison may not be very satisfactory in any case. If we want to have probability of 95 percent of detecting a difference of R units or better (i.e., β=.95), then 2 replicates can be expected to detect errors of roughly 4 times true concentration, whereas 10 replicates are "good for" detecting an error of roughly 50 percent (R=1.47). Settling for lower power (smaller β) gives better apparent sensitivity but increases the odds on completely missing a "bad" lab.

The main outcome of the above exercise seems to be that one must have a good notion of how well the labs can repeat their analyses (coefficient of variation). If the replicate results are quite variable, then increasing the number of replicates permits detecting important errors, where as a small number of replicates seems so insensitive that is likely best not to do the study.

6.0 REFERENCES

Aitchison, J. and Brown, J. A. C., "The Lognormal Distribution", Cambridge University Press, 1969, 176 pages.

Bowman, K. O., "Sample Size Requirement: Continuation", Oak Ridge National Laboratory, ORNL - 4712, UC-32-Mathematics and Computers, August 1971.

Cochran, W. G., "Sampling Techniques", John Wiley and Sons, Inc., 2nd Edition, 1963, 413 pages.

Kastenbaum, M. A., Hoel, D. G., and Bowman, K. O., "Sample Size Requirements: Tests of Equality of Several Gaussean Means", Oak Ridge National Laboratory, ORNL-4468, UC-32-Mathematics and Computers, Nov. 1969.

Pearson, E. S. and Hartley, H. O., "Charts of the Power Function for Analyses of Variance Tests, Derived from the Non-Central F-Distribution", Biometrika, Vol. 38, 1951, p. 112.

Scheffe, H., "The Analysis of Variance", John Wiley and Sons, Inc., New York, 1963, 475 pages.

Slonim, M. J., "Sampling: A Quick Reliable Guide to Practical Statistics", Simon and Schuster, New York, 1960.

Snedecor, G. W. and Cochran, W. G., "Statistical Methods", The Iowa State University Press, Ames, Iowa, 1967, 593 pages.

APPENDIX A

PROCEDURE USED TO SELECT GMX SOIL SAMPLING LOCATIONS

I. Desert Pavement-Downwind (North)

A. Staked Locations

1. There are 17 unevenly spaced radii from the GMX bunker. The radii are numbered 1 through 17 in a clockwise direction with number 1 pointing south-southwest, and number 7 pointing almost true north.

2. The radii are staked at 50-foot intervals away from the bunker. Additional stakes are on the radii within 300 feet of the bunker, but these were not considered in the soil sampling location selection process.

B. Isopleth Areas

1. From FIDLER monitoring data for each staked location, isopleths were drawn which represented 500 c/m, 1500 c/m, and 5000 c/m. Thus, three isopleth areas were established, and these were 500 to 1500 c/m, 1500 to 5000 c/m, and greater than 5000 c/m.

2. There were 45 stakes available in the greater than 5000 c/m area, and 41 stakes available in the 1500 to 5000 c/m area. Initially, there were 108 stakes available in the 500 to 1500 c/m area. However, 12 of these locations on radius number 5 were not considered because they were very close to radius number 4. Thus, 96 staked locations were available for selection purposes in this isopleth area.

C. Staked Location Selection Process

1. All stakes in an isopleth area were numbered in a systematic manner. Thus, stakes 1 through 45 were in the greater than 5000 c/m area, stakes 1 through 41 were in the 1500 to 5000 c/m area, and stakes 1 through 96 were in the 500 to 1500 c/m area.
2. Starting with radius number 7, pointing about true north, stakes were numbered sequentially starting with the stake nearest the bunker in each area, and taking each radius in clockwise order.
3. Random numbers were taken from a table in the NBS Handbook of Mathematical Functions (The numbers had five digits, but only the first two were used). The random numbers were taken in the order they appeared in the table except that numbers greater than the number of stakes in an area were disregarded as were repetitions of numbers already selected for the particular area. When 20 stakes were selected in this manner for the greater than 5000 c/m area, the selection of random numbers from the table continued in the same manner until 20 stakes were selected in the 1500 to 5000 c/m area, and then 20 stakes in the 500 to 1500 c/m area.

D. Revised Sampling Procedure

1. After the above sampling locations were selected to meet existing requirements the numbers of samples to be taken were changed, and additional selection of locations was necessary.
2. Accordingly, the first 10 randomly selected locations which fell in the north one-half of the GMX Site were selected. It happened that more than 10 of the initial 20 locations per isopleth area were in the north portion of each area.

3. An additional 10 locations were selected outside and a maximum of 500 feet away from the 500 c/m isopleth. These locations were on radii extensions to allow documentation of sampling sites, and were selected similarly to the original locations except that 100-foot intervals on 9 radii were used to provide 45 locations for possible selection.

II. Undershrub

A. Sampling Location Requirements

1. Three samples shall be collected from each of the isopleth areas.
2. Locations selected should be near a desert pavement sampling location.
3. Each selected location should be under a single bush (creosote, if possible).

B. Method of Selection

1. The previously selected sampling locations were examined in the random order in which they were selected.
2. Those locations which met the above criteria of being under a creosote bush and near a previous sampling location were selected until there were three locations per area as required.

III. Desert Pavement-Upwind

A. Sampling Location Requirements

1. Three samples will be collected in each isopleth area of the south one-half of the GMX Site.
2. The locations will be selected in the same manner as for downwind sample locations.

B. Method of Selection

1. It happened that at least three locations of the original 20 locations selected per area were available in the south portion of each isopleth area.
2. The first three listed by the method using the table of random numbers were selected for each of the three isopleth areas.
3. Three locations outside the 500 c/m isopleth were selected in the same manner as for these sample location types in the downwind area.

APPENDIX B

ANALYSES OF FIDLER READINGS ON MICROPLOT 1 IN AREA 13

Effect of Shrub Removal

The same FIDLER instrument was used to obtain 60 KEV and 122 KEV readings (uncollimated) in the center of each square a distance of one foot from the surface both before and after shrubs were removed. The differences in the background corrected 60 KEV readings ("after" minus "before") X 10^{-3} were obtained and are given in Table B.1. Those squares which contained bushes are indicated by circling the difference reading for that square. The average of the circled numbers is 465.7 CPM of ^{241}Am which was significantly greater than zero at the $\alpha = .05$ level (2-tailed) using a paired t test[1] (Snedecor and Cochran (1967), page 93):

$$t_{34} = \frac{\bar{x}_D \, n^{\frac{1}{2}}}{s_D} = \frac{(465.7)\ (35)^{\frac{1}{2}}}{1342.3} = 2.05^{*}$$

where s_D = standard deviation of the differences. The 95% confidence interval about the true mean difference is $12 \leq \mu_D \leq 919$ CPM computed as $\bar{x}_D \pm 2s_D/n^{\frac{1}{2}}$. (We are assuming here that the differences are normally distributed). These results suggest that FIDLER readings may tend to be higher after shrub removal.

Differences Between Machines

On both Jan. 20 and 21, FIDLER readings were taken on all 70 plots using 2 different machines. The average difference in net (adjusted) CPM of ^{241}Am in 60 KEV channel in these two machines was determined by finding

---

[1] α is the risk one is willing to take of deciding that the "after" reading is greater than the "before" reading when in fact this is not the case.

* indicates statistical significance at the $\alpha = .05$ level of risk.

Table B.1 Adjusted 60 KEV "After" Minus "Before" Shrub Readings† of ^{241}Am Using One FIDLER Instrument on Microplot 1 of Area 13

Row \ Column	A	B	C	D	E	F	G
1	.6	.4	-.9	-1.2	-1.2	-.9	.1
2	.1	-.6	-1.2	-.9	.1	-1.2	-1.2
3	.4	-.9	.1	.1	-1.2	.1	.4
4	1.4	1.1	-1.2	-.2	-1.2	1.1	1.1
5	1.4	1.4	-.6	1.1	.4	.4	.1
6	.1	2.8	.1	1.4	1.4	1.1	.4
7	-2.2	.1	1.1	.4	.4	.8	.4
8	1.1	2.1	1.1	.4	.4	1.1	.1
9	.1	.4	.4	4.1	3.4	.1	-.9
10	.4	.1	-.6	1.4	.4	- .9	.4

†Table entries are ("after" minus "before") CPM X 10^{-3}

the difference between the readings of machine 1 and machine 2 in each plot then averaging these differences over the 70 plots. The usual two-tailed paried t test was computed on the sample of differences. For the data collected on Jan. 20, the average difference $\bar{x}_D$ = 1143 CMP was significantly greater than zero (α = .01), the 95% confidence interval being $954 \leq \mu_D \leq 1331$ CPM. For the Jan. 21 data the average difference was 380 CPM (significantly greater than zero at the α = .05 level) with a 95% confidence interval of $141 \leq \mu_D \leq 619$ CPM.

The t tests and confidence intervals were computed as follows:

January 20, 1972:

$$t_{69} = \frac{\bar{x}_D(n)^{\frac{1}{2}}}{s_D} = \frac{(1142.9)(70)^{\frac{1}{2}}}{788.4} = 12.13^{**}$$

$$\bar{x}_D \pm 2s_D/(n)^{\frac{1}{2}} = 1142.9 \pm 2(788.4)/(70)^{\frac{1}{2}} = 954.4 \text{ and } 1331.4$$

January 21, 1972:

$$t_{69} = \frac{(380)(70)^{\frac{1}{2}}}{1000.37} = 3.18^{**}$$

$$\bar{x}_D \pm 2s_D/(n)^{\frac{1}{2}} = 380 \pm 2(1000.37)/(70)^{\frac{1}{2}} = 140.9 \text{ and } 619.1$$

These tests suggest that different FIDLER machines may not give identical readings.

** indicates statistical significance at α = .01 level of risk.

Analysis with Replicate Readings

On rows 5 and 6 of Microplot 1, two readings were obtained from each of two FIDLERS on each plot in both the morning and afternoon of January 21. An inspection of these data indicated the duplicate readings of each machine did not change from morning to afternoon indicating that a time effect did not occur. Hence, the data were analysed as a 2-way mixed analysis of variance with machines and plots being the "fixed" and "random" main effects, respectively (Scheffé (1959), page 269). (Machines are considered to be a "fixed" effect since we are concerned only with the performance of the particular machines used. Plots are "random" effects since they can be considered as a random sample from a large population of potential plots).

The data are given in Table B.2 and the analysis of variance table is as follows:

2-Way Mixed Analysis of Variance

Source	d.f.	SS	MS	EMS	F test
Machines (fixed)	1	5.232	5.232	$\sigma_e^2+4\sigma_{AB}^2+56\sigma_A^2$	2.69
Plots (random)	13	68.345	5.257	$\sigma_e^2+8\sigma_B^2$	29.17**
M x P interactions	13	25.253	1.942	$\sigma_e^2+4\sigma_{AB}^2$	10.78**
Error	84	15.141	0.180	σ_e^2	
Total	111				

We note first of all that the interaction F test is statistically significant which results from the fact (see Table B.2) that on most plots Machine 1 gave higher readings than Machine 2, while just the opposite occurred on the remaining plots. This may be due to an "operator effect" since different operators may have handled the same machine at different

Table B.2 Replicate Adjusted 60 KEV Readings† of ^{241}Am Using Two FIDLER Machines On Microplot 1 of Area 13

		Machine 1				Machine 2			
Row	Column								
5	A	8.4	8.4	8.4	8.4	7.0	7.0	7.0	7.0
5	B	8.4	8.4	8.4	8.4	9.0	7.0	7.0	7.0
5	C	9.4	9.4	8.4	8.4	8.0	8.0	8.0	8.0
5	D	8.4	8.4	8.4	8.4	8.0	8.0	9.0	8.0
5	E	10.4	10.4	10.4	10.4	9.0	9.0	9.0	9.0
5	F	8.4	8.4	8.4	8.4	7.0	7.0	7.0	7.0
5	G	9.4	8.4	9.4	8.4	8.0	8.0	7.0	8.0
6	A	7.4	7.4	6.8	6.8	7.0	7.0	7.0	7.0
6	B	7.8	8.4	7.8	7.8	7.0	7.0	9.0	7.0
6	C	7.8	7.8	7.8	7.8	7.0	7.0	8.0	7.0
6	D	8.8	8.8	8.8	8.8	10.0	10.0	10.0	10.0
6	E	8.8	8.8	8.8	8.8	10.0	10.0	10.0	10.0
6	F	8.8	8.8	7.8	7.8	10.0	8.0	10.0	10.0
6	G	8.4	8.4	8.4	7.8	7.0	7.0	7.0	7.0

† Table entries are in CPM x 10^{-3}.

times and plot locations. We have not been supplied information concerning when operators changed during the day so as to be able to separate out an "operator effect" from the "machine effect". The F test for "plot effects" was statistically significant indicating differences between plots (a not unexpected result), whereas the F test for "machine effect" is meaningless due to the significant "interaction effect".

We note from the analysis of variance table that an estimate of the variance (s^2) of FIDLER readings is given by the error MS estimate .180, which, when multiplied by 1000 (the original counts were divided by 1000) gives $s^2 = 180$ or $s = 13.42$ CPM, where s is the standard deviation.

Replicate Readings in 60 and 122 KEV Channels

In rows 5 and 6 on Jan. 21, a total of 5 replicate 60 KEV and 122 KEV readings were taken on each of 2 machines in each plot. Of the 28 sets of 5 replicate readings in the 60 KEV channel, 12 sets (43%) showed no variation between replicate readings. In the 122 KEV channel the percentage was 80% (25 sets). The data can be arranged in a 2x2 table and a chi-square test of homogeneity computed to test whether 60 KEV readings are more precise (reproducible) than the 122 KEV readings. The necessary data are:

Observed Frequency

	Change	No Change	Total
60 KEV	$O_1 = 16$	$O_3 = 12$	28
122 KEV	$O_2 = 3$	$O_4 = 25$	28
	19	37	56

Expected Frequency if 60 KEV and 122 KEV Readings are Equally Precise

Change	No Change	Total
$E_1 = 9.5$	$E_3 = 18.5$	28
$E_2 = 9.5$	$E_4 = 18.5$	28
19	37	56

Hence

$$\chi_1^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = 13.46^{**}$$

This chi-square test is highly significant suggesting that the 122 KEV replicate readings are more likely to be identical for a given plot and machine than are replicate readings from the 60 KEV channel. This does not imply, however, that the 122 KEV readings are necessarily any more unbiased (accurate) than the 60 KEV readings.

APPENDIX C

POWER CALCULATIONS FOR INTERLABORATORY COMPARISONS

Following Scheffé (The Analysis of Variance, H. Scheffé, J. Wiley, 1959) p. 62), we refer to power calculation tables with:

I = Number of laboratories being compared ("treatments"),

J = Number of replications per laboratory (to be calculated)

and

$$\phi = \left(\frac{J}{2I}\right)^{\frac{1}{2}} (\Delta/\sigma),$$

where

$\nu_1 = I - 1$ = degrees of freedom for treatments,

$\nu_2 = I(J - 1)$ = degrees of freedom for the experiment,

σ = Variation among replicate analyses done by each laboratory

and

Δ = real difference between two labs for which we want a power of at least β of detecting.

We assume the observations to be lognormally distributed and operate on the basis of a logarithmic transformation of the data. Hence, $\Delta = \log R$, where R is the ratio of "high" to "low" laboratory, or, more appropriately, the ratio of a labs biased value to the true value.

Under the assumption of lognormality, we have (The Lognormal Distribution J. Aitchison and J. A. C. Brown, Cambridge University Press, 1969, p. 8):

$$f(x) = \frac{\exp\left[-\frac{1}{2\sigma^2}(\log_e x - \mu)^2\right]}{\sigma x (2\Pi)^{\frac{1}{2}}} \quad ,$$

and

$$E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$V(x) = \left[\exp(2\mu + \sigma^2)\right]\left[\exp(\sigma^2) - 1\right]$$

so that

$$C^2 = \frac{V(x)}{\left[E(x)\right]^2} = \exp(\sigma^2) - 1 \quad ,$$

where C = coefficient of variation. Hence, in log-transformed data we can specify C and obtain σ from the relation just above that $\sigma = \left[\log_e (C^2 + 1)\right]^{\frac{1}{2}}$. Note that C will be nearly equal to σ when σ is small; i.e., $e^x \doteq 1 + x$ for x small

Thus, we can conveniently specify the problem in terms of the coefficient of variation and a bias (R) in terms of a ratio. As an example, consider C = .1, R = 2, and I = 8:

$\sigma^2 = \log(1.01) = .00995$ so that $\sigma = .1$ is adequate.

$\Delta = \log_e 2 = .693.$

$\phi = (\frac{1}{2} J/8)^{\frac{1}{2}} .693/.1 = 1.73\sqrt{J}$

If we refer to Pearson-Hartley charts (E.S. Pearson and H.O. Hartley, Biometrika, Vol. 38, pp. 115-122 (1951)) with $\nu_1 = 8\text{-}1 = 7$ and $\alpha = .05$, we find for J = 2, $\phi = 2.44$ and $\nu_2 = 8(2\text{-}1) = 8$, that β is about .93, so that if we adopt a rule of $\beta = .90$, two replicates is sufficient. If, however,

a more realistic comparison is indicated, let R = 1.25, then

$$\phi = (\tfrac{1}{2}\, J/8)^{\frac{1}{2}}\ .223/.1 = .56\ \sqrt{J} \quad .$$

Trying J = 9, we get $\phi = 3(.56) = 1.68$ and $\nu_2 = 8(8) = 64$. From the charts, β is somewhat higher than .90, and J = 8 seems a closer choice.

More detailed tables have been prepared by Kastenbaum, Hoel, and Bowman (ORNL-4468) and by Bowman (ORNL-4712). In these tables K is the number of treatments (I above), N the number of replicates (J above) and the tabulation is in terms of the "standardized range", τ, which is here Δ/σ. Also β is given as $1-\beta$. Hence, the above example gives

$$K = 8 \text{ and } \tau = (\log_e 1.25)/.1 = 2.23$$

and the nearest value of τ tabulated is 2.274, so that the power is a little lower than the required .90.

In practice, it doesn't seem likely that there will be much interest in interlaboratory comparisons if sizable numbers of replicates are involved. Thus, it appears most useful to convert the ORNL tables to read off the ratio R for a given coefficient of variation and number of replicates likely to be used in practice. The conversion is:

$$R = \exp(\tau\sigma) = \exp\left\{\tau\left[\log_e (C^2 + 1)\right]^{\frac{1}{2}}\right\}$$

which lets us take a column of tau values under a given "BETA" in the ORNL tables and convert it to entries in a new table having a fixed coefficient of variation (C) and reading in terms of R. The resulting tables can then be used to evaluate the sensitivity of an interlaboratory study for various numbers of replicates and several levels of power.

The attached tables (Tables C.4, C.5 and C.6) are all calculated for $\alpha = .05$ and show entries of R for from 2 to 10 replicates and $\beta = .95$, .90, .80 and .70. Tables for K = 4, 8 and 15 (number of laboratories in the study) are given for a range of values of the coefficient of variation.

The tables are not very useful unless one has a reasonable idea of the coefficient of variation to be expected in a study. Partial data on one such study were made available to us by the Western Environmental Research Laboratory (Las Vegas) of the U.S. Environmental Protection Agency in connection with the NAEG study. These data are analyzed below, in Tables C.1, C.2 and C.3, using one-way analyses of variance (AOV) to obtain estimates of the coefficient of variation for 3 separate samples. The data are assumed to be lognormally distributed. Hence, the data were log transformed before the AOV was computed. These results indicate that C may be as small as .04 and at least as large as 0.13 for samples of known concentration of ^{239}Pu. Hence, the discussion in the text (p. 5.3) when C is as large as .2 may be relevant. We saw there that for 4 laboratories, one needed 10 replicates to detect a 50% error in a lab's determinations with probability .95 when C = .2. If no more than 5 labs could be utilized then a 56% error would be detected with probability .95 if C = .15. In sample #2 (Table C.2) below for which the concentration of ^{239}Pu was unknown the estimated value of C was .35, a considerably larger value than obtained for the known concentration samples. An examination of R values in Table C.4 on page C.9 indicates poor sensitivity for C = .30 or larger, even for 10 replications. These data suggest that we may have less sensitivity to detect laboratory errors in samples of unknown concentration than those of known concentration.

Table C.1 Interlaboratory Comparisons: Sample #1
December 1970, Known Concentration of ^{239}Pu

(Data Transformed To $\log_e$)

		Replicates		
		1	2	3
	A	2.688	2.728	2.747
	B	2.697	2.738	2.698
LAB	C	2.76	2.766	2.747
	D	2.791	2.76	2.646
	E	2.711	2.749	2.732
	G	2.702	2.69	2.723

One Way Analyses of Variance (AOV)

Source	d.f.	SS	MS	F
Lab	5	.005266	.00105	.785 N. S.†
Error	12	.01604	.001337 = s^2	
TOTAL	17			

Therefore, $C = \left[\exp(s^2) - 1\right]^{\frac{1}{2}} = \left[\exp(.001337) - 1\right]^{\frac{1}{2}} = .037.$

(† N.S. = not significant)

Table C.2 Interlaboratory Comparisons: Sample #2
January 1972, Unknown Concentration of ^{239}Pu

(Data Transformed To $\log_e$)

Replicates

	1	2	3	4
A	-3.912	-3.912	-3.507	--
C	-3.474	-3.442	-3.507	--
D	-3.016	-3.689	-4.017	--
E	-3.689	-3.963	-3.689	--
G	-3.381	-3.381	-3.442	--
H	-3.507	-3.474	-3.507	--
I	-3.442	-2.313	-3.772	-3.612
J	-3.101	-3.037	--	--
K	-3.411	-3.650	-3.772	--

One Way AOV

Source	d.f.	SS	MS	F
Lab	8	1.106	0.138	1.2 N. S.
Error	18	2.069	0.115 = s^2	
TOTAL	26			

Therefore, $C = \left[\exp(0.115) - 1\right]^{\frac{1}{2}} = .35$

Table C.3 Interlaboratory Comparisons Sample #3, February 1971, Known Concentration = .8065 pCi/g ^{239}Pu

(Data Transformed To $\log_e$)

	Replicates		
	1	2	3
A	.788	.788	.788
C	.888	.940	.837
D	.751	.425	1.058
E	.833	.788	.833
G	.732	.770	.742
J	.728	.802	.811

One Way AOV

Source	d.f.	SS	MS	F
Lab	5	.04274	.01068	.605 N. S.
Error	12	.21196	.01766 = s^2	
TOTAL	17			

Therefore, $C = \left[\exp(.01766) - 1\right]^{\frac{1}{2}} = 0.13$.

Table C.4 Tables Of R For Intercomparing Four Laboratories

Number Of Replicates	β = .95	.90	.80	.70
		C.V. = .01		
2	1.074	1.066	1.056	1.049
3	1.044	1.040	1.035	1.031
4	1.035	1.031	1.027	1.024
5	1.030	1.027	1.023	1.021
6	1.026	1.024	1.021	1.019
7	1.024	1.022	1.019	1.017
8	1.022	1.020	1.017	1.016
9	1.021	1.019	1.016	1.014
10	1.019	1.017	1.015	1.014
		C.V. = .02		
2	1.153	1.136	1.116	1.102
3	1.091	1.082	1.071	1.063
4	1.071	1.064	1.056	1.050
5	1.061	1.055	1.048	1.043
6	1.054	1.049	1.042	1.038
7	1.049	1.044	1.039	1.034
8	1.045	1.041	1.036	1.032
9	1.042	1.038	1.033	1.030
10	1.039	1.036	1.031	1.028
		C.V. = .05		
2	1.428	1.376	1.316	1.275
3	1.245	1.219	1.188	1.167
4	1.189	1.170	1.147	1.131
5	1.160	1.144	1.125	1.111
6	1.141	1.127	1.110	1.098
7	1.127	1.115	1.100	1.089
8	1.117	1.106	1.092	1.082
9	1.109	1.098	1.086	1.077
10	1.102	1.092	1.081	1.072
		C.V. = .10		
2	2.039	1.892	1.731	1.625
3	1.548	1.485	1.412	1.361
4	1.414	1.368	1.316	1.279
5	1.345	1.308	1.265	1.235
6	1.301	1.270	1.233	1.207
7	1.271	1.243	1.210	1.187
8	1.248	1.223	1.193	1.171
9	1.230	1.207	1.179	1.159
10	1.215	1.194	1.168	1.149

Table C.4 Continued

Number Of Replicates	β = .95	.90	.80	.70
		C.V. = .15		
2	2.902	2.595	2.272	2.068
3	1.923	1.807	1.675	1.586
4	1.678	1.599	1.508	1.445
5	1.558	1.495	1.422	1.371
6	1.483	1.430	1.368	1.325
7	1.431	1.385	1.330	1.292
8	1.393	1.351	1.302	1.267
9	1.363	1.325	1.280	1.248
10	1.339	1.303	1.261	1.232
		C.V. = .20		
2	4.114	3.548	2.974	2.624
3	2.383	2.193	1.984	1.845
4	1.989	1.865	1.725	1.630
5	1.801	1.706	1.596	1.521
6	1.688	1.608	1.516	1.453
7	1.610	1.541	1.461	1.405
8	1.553	1.491	1.420	1.370
9	1.509	1.453	1.387	1.342
10	1.473	1.422	1.361	1.319
		C.V. = .30		
2	8.140	6.535	5.031	4.179
3	3.623	3.204	2.761	2.480
4	2.771	2.519	2.244	2.063
5	2.393	2.207	2.000	1.862
6	2.173	2.023	1.854	1.740
7	2.027	1.899	1.754	1.656
8	1.921	1.809	1.681	1.594
9	1.840	1.740	1.625	1.546
10	1.776	1.685	1.580	1.508

Table C.5 Tables Of R For Intercomparing Eight Laboratories

Number Of Replicates	β = .95	.90	.80	.70
		C.V. = .01		
2	1.072	1.065	1.057	1.051
3	1.048	1.043	1.038	1.034
4	1.038	1.035	1.031	1.028
5	1.033	1.030	1.026	1.024
6	1.029	1.027	1.024	1.021
7	1.027	1.024	1.021	1.019
8	1.025	1.023	1.020	1.018
9	1.023	1.021	1.019	1.017
10	1.022	1.020	1.017	1.016
		C.V. = .02		
2	1.150	1.135	1.117	1.105
3	1.098	1.089	1.078	1.070
4	1.079	1.072	1.063	1.057
5	1.067	1.062	1.054	1.049
6	1.060	1.055	1.048	1.044
7	1.055	1.050	1.044	1.040
8	1.050	1.046	1.041	1.037
9	1.047	1.043	1.038	1.034
10	1.044	1.040	1.036	1.032
		C.V. = .05		
2	1.418	1.373	1.321	1.284
3	1.264	1.239	1.208	1.186
4	1.209	1.189	1.166	1.149
5	1.178	1.162	1.142	1.127
6	1.158	1.143	1.126	1.113
7	1.143	1.130	1.114	1.103
8	1.132	1.120	1.105	1.095
9	1.123	1.112	1.098	1.088
10	1.115	1.105	1.092	1.083
		C.V. = .10		
2	2.009	1.884	1.743	1.648
3	1.598	1.534	1.459	1.407
4	1.461	1.414	1.359	1.320
5	1.388	1.349	1.304	1.271
6	1.340	1.307	1.268	1.239
7	1.307	1.277	1.242	1.216
8	1.281	1.254	1.222	1.199
9	1.261	1.236	1.206	1.185
10	1.244	1.221	1.193	1.174

Table C.5 Continued

Number Of Replicates	β = .95	.90	.80	.70
		C.V. = .15		
2	2.838	2.578	2.295	2.111
3	2.016	1.897	1.760	1.667
4	1.764	1.680	1.582	1.514
5	1.633	1.566	1.487	1.432
6	1.550	1.493	1.426	1.379
7	1.492	1.442	1.383	1.341
8	1.448	1.403	1.350	1.312
9	1.414	1.373	1.324	1.289
10	1.386	1.348	1.303	1.271
		C.V. = .20		
2	3.995	3.516	3.014	2.696
3	2.538	2.340	2.118	1.970
4	2.125	1.991	1.839	1.735
5	1.918	1.814	1.694	1.611
6	1.790	1.703	1.602	1.532
7	1.701	1.626	1.538	1.476
8	1.636	1.568	1.490	1.434
9	1.585	1.524	1.452	1.401
10	1.543	1.487	1.421	1.375
		C.V. = .30		
2	7.792	6.450	5.133	4.351
3	3.977	3.526	3.043	2.733
4	3.057	2.776	2.468	2.264
5	2.626	2.418	2.184	2.028
6	2.370	2.202	2.011	1.882
7	2.198	2.055	1.893	1.781
8	2.074	1.949	1.806	1.707
9	1.979	1.867	1.738	1.650
10	1.903	1.801	1.684	1.603

Table C.6 Tables Of R For Intercomparing 15 Laboratories

Number Of Replicates	β = .95	.90	.80	.70
		C.V. = .01		
2	1.075	1.068	1.060	1.054
3	1.052	1.047	1.042	1.038
4	1.042	1.039	1.034	1.031
5	1.036	1.033	1.030	1.027
6	1.033	1.030	1.027	1.024
7	1.030	1.027	1.024	1.022
8	1.027	1.025	1.022	1.020
9	1.026	1.024	1.021	1.019
10	1.024	1.022	1.020	1.018
		C.V. = .10		
2	2.057	1.936	1.797	1.702
3	1.660	1.593	1.514	1.458
4	1.515	1.465	1.405	1.363
5	1.435	1.393	1.344	1.309
6	1.382	1.346	1.304	1.273
7	1.344	1.313	1.275	1.247
8	1.316	1.287	1.252	1.228
9	1.293	1.266	1.235	1.212
10	1.274	1.250	1.220	1.199
		C.V. = .30		
2	8.363	6.992	5.617	4.786
3	4.449	3.939	3.392	3.037
4	3.398	3.078	2.725	2.490
5	2.894	2.657	2.391	2.212
6	2.593	2.402	2.185	2.038
7	2.391	2.230	2.045	1.918
8	2.244	2.103	1.941	1.830
9	2.132	2.006	1.861	1.761
10	2.043	1.929	1.797	1.706

DISTRIBUTION

OFFSITE

AEC Chicago Patent Group

1 G. H. Lee

AEC-NV National Environmental Research Center, Las Vegas NV

1 L. M. Dunn

AEC-NV Office of Effects Evaluation, Las Vegas NV

1 E. M. Douthett
1 H. J. Kayuha

AEC-NV Plutonium Ad Hoc Committee of NAEG, Las Vegas NV

1 W. J. Bair
1 F. D. Cluff
50 P. B. Dunaway
1 J. W. Healy
1 J. L. Olsen
1 O. Raabe
1 C. R. Richmond
1 E. M. Romney
1 J. N. Stannard
1 R. C. Thompson

AEC-NV Program Element Managers, NAEG, Las Vegas NV

1 L. L. Eberhardt
1 E. B. Fowler
5 D. W. Hendricks
1 W. E. Martin
1 P. L. Phelps
1 E. M. Romney
1 R. E. Stanley

AEC-NV Radiological Operations Division, Las Vegas NV

5 O. D. T. Lynch, Jr.

E G & G, Santa Barbara CA

1 W. A. Rhoades

DOW Chemical, Golden CO

5 J. F. Willging

DISTRIBUTION (Cont.)

OFFSITE

USAEC-Hq. Division of Biomedical & Environmental Research, Wash. DC

1	C. L. Osterberg
5	J. Swinebroad
1	J. Totter

USAEC-Hq. Division of Regulatory Standards, Wash. DC

1	J. J. Davis
1	G. Hamada

University of Washington, Seattle WA

3	A. H. Seymour

Washington State University, Pullman WA

1	V. Schultz

AEC Technical Information Center

5	Special Distribution in UC-48

ONSITE

AEC Chicago Patent Group

1	R. W. Poteat

AEC Richland Operations

2	W. E. Lotz

Atlantic Richfield Hanford Company

1	R. E. Isaacson
1	B. J. McMurray

Battelle Northwest

1	E. L. Alpen
1	T. M. Beetle
1	L. L. Eberhardt
1	R. Emory
1	T. P. O'Farrell
1	R. F. Foster
50	R. D. Gilbert
1	E. Klepper
1	C. W. Lindenmeir
1	K. R. Price
1	W. H. Rickard

DISTRIBUTION (Cont.)

ONSITE

Battelle Northwest

1	W. L. Templeton
1	J. M. Thomas
1	B. E. Vaughn
3	Technical Information
2	Technical Publications