BEACH EROSION BOARD
OFFICE OF THE CHIEF OF ENGINEERS

RELATIVE EFFICIENCY OF BEACH SAMPLING METHODS

TECHNICAL MEMORANDUM NO. 90
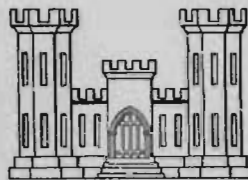
RELATIVE EFFICIENCY OF
BEACH SAMPLING METHODS

TECHNICAL MEMORANDUM NO. 90

BEACH EROSION BOARD

CORPS OF ENGINEERS

SEPTEMBER 1956

FOREWORD

Samples of beach sand for engineering and geological purposes are collected with one or more of three objectives in mind: to estimate the mean particle size on the beach, to estimate the relative variability of the deposits with locality and season, or to estimate systematic changes (gradients) in beach properties from one locality to another. Each of these purposes requires consideration of the sampling method to be used. Because of natural variations in the deposits on backshore and foreshore, as well as in the nearshore submerged areas, some form of stratified sampling is commonly best adapted to estimation of mean particle size. Studies of relative variability can be handled by multilevel designs in which samples are collected at several different spacings to bring out the magnitude of changes between closely spaced and more distantly spaced samples. Gradients are effectively studied with systematic samples laid on a grid.

The present report compares a number of sampling designs to indicate some of the factors involved in beach sampling for different purposes. It is part of a continuing study on beach material characteristics emphasizing some of the principles for planning beach sampling operations.

This report has been prepared by Dr. William C. Krumbein in pursuance of Contract DA-49-055-eng-35 with the Beach Erosion Board, and with the active collaboration of Dr. Howard A. Slack. Both Dr. Krumbein and Dr. Slack are members of the Geology faculty at Northwestern University. The writers are indebted to numerous individuals for suggestions and data used in this report. Dr. E. C. Dapples and a group of graduate students at Northwestern University collaborated in the designed experiment at Illinois Beach near Waukegan, Illinois. Mr. J. M. Caldwell and Charles T. Fray of the Beach Erosion Board staff, with the assistance of other members of the staff, helped collect the samples at Ocean Beach, Maryland. Messrs. R. O. Eaton, J. V. Hall, Jr. and G. M. Watts of the Beach Erosion Board staff have been especially cooperative in discussing beach sampling problems and in placing data on beach deposits at the writers' disposal.

Views and conclusions stated in this report are not necessarily those of the Beach Erosion Board.

TABLE OF CONTENTS

RELATIVE EFFICIENCY OF BEACH SAMPLING METHODS

by

W. C. Krumbein and H. A. Slack, Northwestern University

INTRODUCTION

In an earlier report (published in 1954 as Beach Erosion Board Technical Memorandum No. 50) the statistical implications of several methods for sampling beach materials were discussed in a preliminary way. It was pointed out that some sampling procedures are to be preferred because they produce more reliable results for a given expenditure of effort. The present report is an extension of the earlier one in that it presents the results of several sampling experiments designed to show more explicitly how estimates of certain beach material properties may vary as a result of the sampling plan adopted.

The areas selected for study include a sand and gravel beach along Lake Michigan near Waukegan, Illinois, and a sand beach at Ocean Beach, Maryland. The heterogeneity of the deposits near Waukegan affords an illustration of the problems encountered among beaches composed of materials of different sizes, whereas the relative homogeneity of the beach at Ocean Beach represents a more usual situation along extensive sand beaches.

Purpose. Samples may be collected in numerous ways from selected points on a beach, as at the berm crest, Bascom's mid-tide point (1950), the low water line; along profiles normal to the shore; or over unit sampling areas on the beach. Further, the individual samples may be collected on a random, systematic, or stratified basis, as was pointed out in the earlier report. As a result of these varying ways of collecting samples, there is a fair likelihood that the several estimates of mean values for the same beach could be noticeably different.

The number of samples collected is also a factor in the reliability of the estimates made. Evidently ten samples are better than one, and one hundred samples are better than ten. How many should be collected to obtain an estimate within a given confidence band? Questions such as these can be answered in part by designed experiments, and it is the purpose of this report to examine some of these questions explicitly. On the basis of the experiments some specific sampling procedures can be recommended.

Objectives of Beach Sampling. The data from the sampling experiments are treated in terms of three general objectives of beach sampling, which may be stated as follows:

1. Estimation of mean particle size in the beach area. This information is of value in comparing beaches among themselves, in studying the size changes that may occur on a single beach during the seasons, and for setting up specifications for beach fill where the engineering design calls for sand having the same particle size characteristics as the beach being stabilized.

2. Estimation of the relative variability of the deposits in each of several beach zones (backshore, foreshore, etc.) as a basis for detailed description or study of sediment response to varying wind and hydraulic forces, or for accumulation of a body of knowledge for design of general beach sampling plans.

3. Estimation of gradients in beach particle properties along and across beaches. Such data are of value in studying the rates at which particles may be sorted by waves, currents, or wind, as well as to investigate the relative effects of different agents in distributing materials downbeach from a source area.

In each instance methods of computation and specific examples are cited to show the method used in handling the problem.

REVIEW OF BEACH ZONES AND BEACH POPULATIONS

The most fundamental natural divisions of a beach area are (1) the nearshore bottom zone, (2) the foreshore, and (3) the backshore. Each of these zones may be subdivided further in special studies, but their over-all characteristics are controlled by the following features:

1. The Nearshore Bottom. This zone is by definition practically always submerged, inasmuch as it extends outward from the mean low water line to some arbitrary depth, usually taken as about -30 feet. The material in the shoreward part of this zone is subject to movement by breaking waves and by currents, depending on the energy conditions prevailing at the time.

2. The Foreshore. This zone extends by definition from the mean low water line to the crest of the main berm at or near the high tide line. This zone is alternately covered with water and exposed to the air during each tidal cycle. All parts of it are subjected to the action of breaking waves as the tide rises and falls, and the submerged portions at any tidal stage are subject to current activity.

2

The upper limit of the average wave uprush defines the normal berm crest; this crest is topped only by occasional waves or during storm conditions.

3. The Backshore. This zone extends by definition from the crest of the berm inland to the belt of dunes if dunes are present, or to some other limiting feature such as a lagoon or bank. The backshore is always exposed to the air except during brief periods of unusual wave wash or during storms. It is thus a zone that on the whole is relatively free from hydraulic forces after it is established by outward building of the whole beach. The backshore is subject to wind action in that the dry sand is picked up selectively by onshore winds and blown toward the dune belt. The selective action of the wind represents winnowing in which certain particles are selected for transport according to their size, shape, and density. With negligible exceptions only sand and silt are picked up by the wind, and such pebbles as may be present remain on the backshore. On a stable beach the backshore is a relatively unchanging part of the beach, insofar as daily processes are concerned. From time to time during the year, parts of the backshore may temporarily be converted to foreshore as new berms are built, but there is an average position of the main berm crest that may persist for a number of years. This is controlled by the average amount of shore drift, by the tidal range, and by the incidence of storms of given magnitude along that part of the coast.

3a. The Dune Belt. This zone of shore terrain is not considered as part of the beach proper, but it is closely related in a genetic sense to the beach deposits. The dunes represent accumulation of material swept inland from the beach by wind, and as such they are a partial segregation of the original beach deposits. Most of the sand is probably derived from the backshore, although at times essentially dry portions of the foreshore may contribute significant quantities of sand to the dune. In one sense dunes provide a relatively permanent trap area for beach-derived materials. The dunes are beyond the reach of any but the most severe storms (hurricanes, etc.) and although the dune belt may become fairly wide, in general it remains as an essential part of the shore terrain. One may argue from this point of view that estimates of the initial properties of the beach material as a whole should include the properties of the dune particles.

In a statistical sense the several beach zones may be considered as natural "sampling strata", each of which has its own characteristic particle populations. The variability from point to point within each zone may be different for the several zones; thus, the range of particle sizes on the foreshore may be greater than on the backshore. As a result, samples from one of the zones may show considerable

differences in particle size and other properties when compared to samples from another zone. Such differences affect the estimated mean particle size for the beach as a whole, and they may indicate that a weighted mean for the several strata may be better than a single composite mean computed without regard to the distribution of samples over the zones.

SAMPLING EXPERIMENTS AT ILLINOIS BEACH ON LAKE MICHIGAN

General Remarks. Beaches along the western shore of southern Lake Michigan commonly consist of mixtures of sand and pebbles. At times an individual beach may be made almost wholly of sand; at other times gravel dominates, and most commonly the main part of the beach may be sand with gravel stringers on the backshore and more prominently along the foreshore. The variability of the exposed beach is paralleled by a similar variability of the nearshore submerged zones, as maps prepared by the Illinois Division of Waterways show (Fisher, 1954). Linear subparallel gravel zones occur in places along the shore, which in some cases lie tangent to the shore line or may angle off into deeper water.

Beaches as variable as these pose special problems of sampling, inasmuch as any estimates of mean particle size will depend strongly on where the samples are collected and on how many samples are used. Maps showing the variability of the beach and underwater deposits are also affected, inasmuch as the sample spacing along profiles, and the spacing of the profiles along the shore in part control the accuracy with which contour lines of mean particle size can be drawn between the sampling points.

In most situations a practical compromise between costs and desired degree of reliability has to be made. It would be advantageous if such adjustments could be made on a quantitative basis. To do so requires knowledge about the variability of the deposits in each beach zone. When such data are available, cost functions can be applied at least to the problem of estimating mean particle size or mean heavy mineral content. Evaluation of optimum sample spacing for preparation of maps is a somewhat more difficult problem.

In order to arrive at some basis for decision on the manner of collecting samples, on sample spacing, and on the number of samples needed, the main experiments reported in this paper were designed for the highly variable deposits along some Lake Michigan beaches. It is thought that if the problem can be handled under somewhat difficult conditions, the corresponding solution for more homogeneous beaches would represent mainly a simplification of the complexities encountered.

In the following sections, accordingly, experiments on Lake Michigan beaches are first described, and the results are then compared with those for a more homogeneous sand beach along the ocean coast.

Comparison of Sampling Methods for Estimation of Mean Values

The beach area at Illinois Beach State Park, some 6 miles north of Waukegan, Illinois, is several miles long, and consists of moderately wide foreshore and backshore, and a belt of dunes and glacial lake ridges extending for a half mile or more inland. The site selected for study was a north-south, east-west square area along the shore line, 300 feet on edge, picked at random from the southern half of the beach area. This square was large enough to include the foreshore, backshore, and the first few ridges of low dunes. No attempt was made to extend the sampling underwater, inasmuch as suitable equipment was not available.

Figure 1 shows the sampling square and some of its natural features. The lakeward edge of the square was approximately parallel to the lake shore. The foreshore was well developed with a prominent berm marking its landward edge. The average width of the foreshore was about 60 feet. The backshore was about 80 feet wide, and its landward edge was marked by a transition to wind-rippled clean sand that marked the edge of dominant wind-deposited sand. The portion of the dune belt included in the square contained a low fore-dune ridge and a more prominent ridge about 60 feet farther inland.

Figure 1 shows the variability of the beach deposits. As indicated, several sub-parallel zones of sand with scattered pebbles, areas of clean sand, and patches of gravel were distributed over the backshore. The foreshore had narrow similar bands of variable composition, and the most recent berm, about 15 feet from the water's edge, consisted of a narrow band of clean gravel.

Sampling Designs. In all, seven parallel sampling designs were applied to the beach area. In the first plan 16 samples were distributed completely at random over the square. These were selected by using a table of random numbers to obtain 16 pairs of numbers that defined x- and y- coordinates of the samples. The position of these 16 samples is shown in Figure 2A. They tended to occur more frequently in the south and east-central part of the area. The pattern is typical of randomly spaced points, in that some tendency to clustering occurs, and there may be large bare spots.

The 300-foot sampling square was next divided into 16 smaller squares, each 75 feet on edge. These were considered to be "sampling cells" for part of the experiment. The cells did not necessarily coincide with the natural beach zones, but were thought of as units
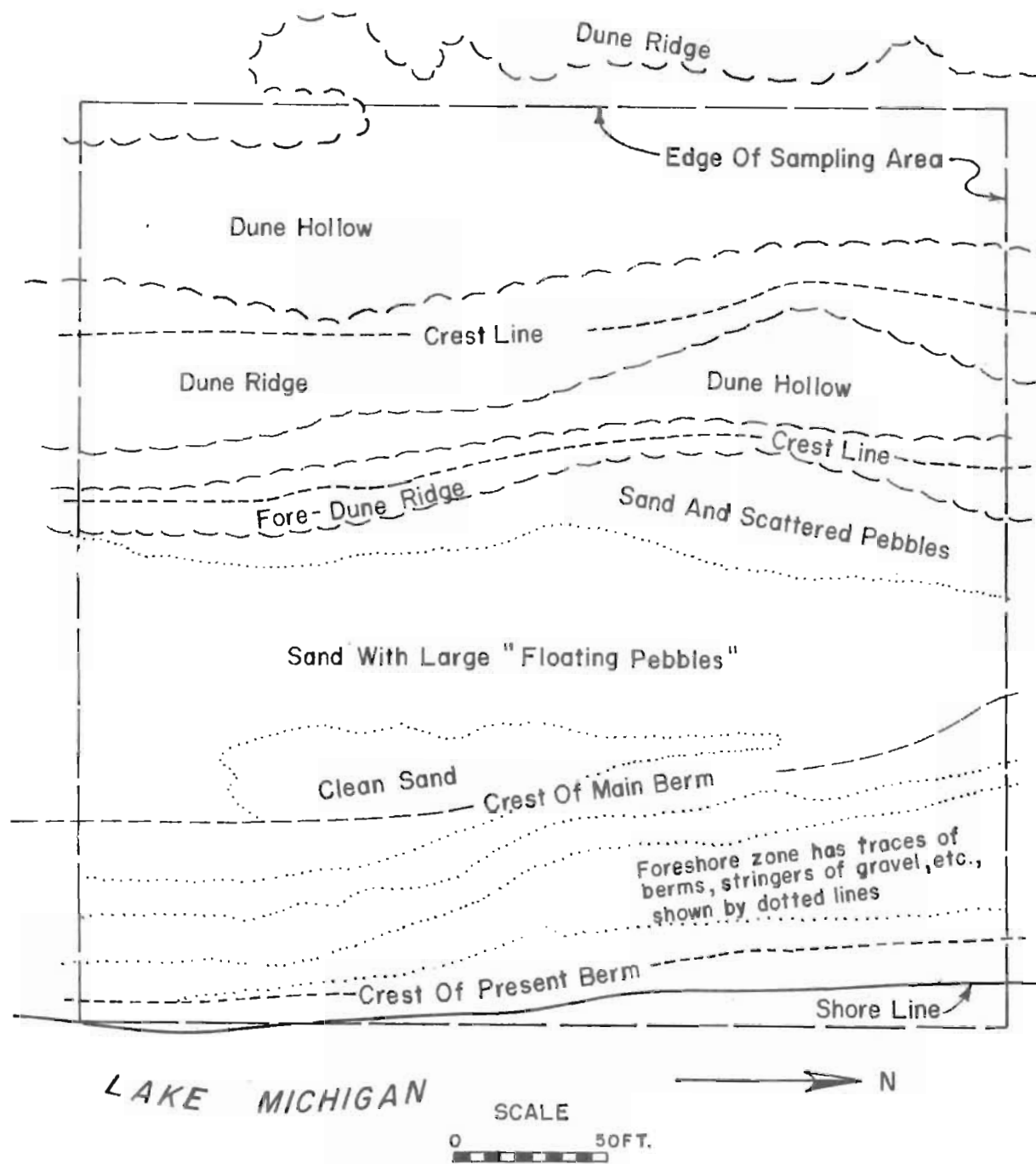
5

FIGURE I · SAMPLING AREA AT ILLINOIS BEACH STATE PARK

within a larger square dropped at random over the beach area. This choice was dictated by pedagogical considerations, inasmuch as the experiment was conducted with a group of graduate students, and part of the problem was the preparation of maps from data distributed in various ways in sampling cells laid over a heterogeneous deposit.

The second sampling plan consisted in collecting one sample at random in each of the 16 cells, as shown in Figure 2B. In this method the sample in each cell was located by taking a pair of random numbers to locate a coordinate point within the square. Comparison of Figures 2A and 2B shows that the second plan assured more even distribution of samples over the area, although some samples were closer together than others because of their random distribution in the cells.

The third sampling plan involved the collection of a sample at the center of each cell, which also yielded 16 samples. These are equally spaced, as shown in Figure 2C, and they provide equally spaced samples for mapping purposes. The systematic samples have an element of randomization in their collection inasmuch as the position of the major sampling square was itself randomized on the beach.

The fourth sampling plan involved the collection of four clusters of four samples each as shown in Figure 2D. One cluster was confined to each row of cells, but the position within the row was randomized. This plan provides a spotty coverage, although the total number of samples is 16 as in the other designs.

The fifth sampling plan was a variant of the cluster design involving two-stage subsampling. A second position was selected in the rows and corresponding columns of cells that had the clusters, but in this case one pair of the cluster samples was combined with another pair spaced farther apart than in the original cluster. This is a three-level "nested sampling" design, and is shown in Figure 2E.

The sixth sampling plan was based on natural divisions of the shore area. In its original version this sample stratification plan was based on four strata, one consisting of the dunes, another of the foreshore, and two sub-divisions of the backshore. Four samples were collected at random in each stratum, as shown in Figure 2F. Subsequently the samples in the major designs were redistributed according to three basic strata, dunes, backshore and foreshore, to obtain the best estimate from all the independently collected samples. This layout is shown in Figure 2G.

The final sampling plan involved purposive selection of specific points to be sampled. This was done by Dr. E. C. Dapples, who examined
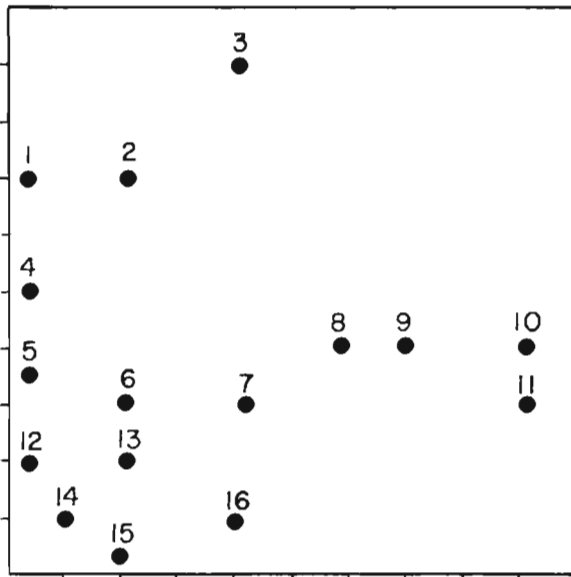
A. Simple Random Samples
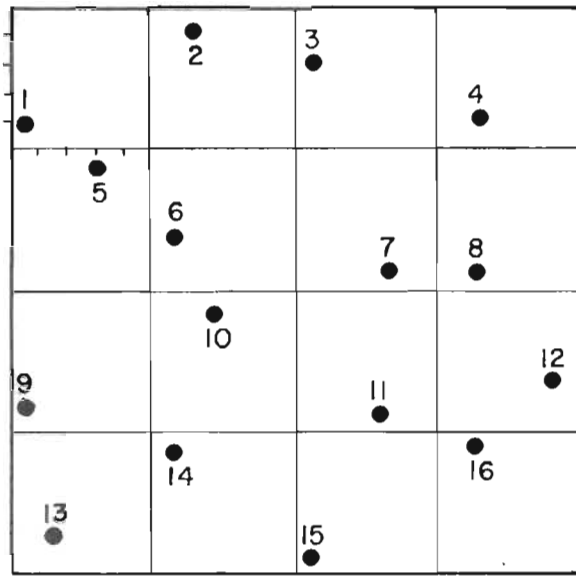
B. Random In Cells
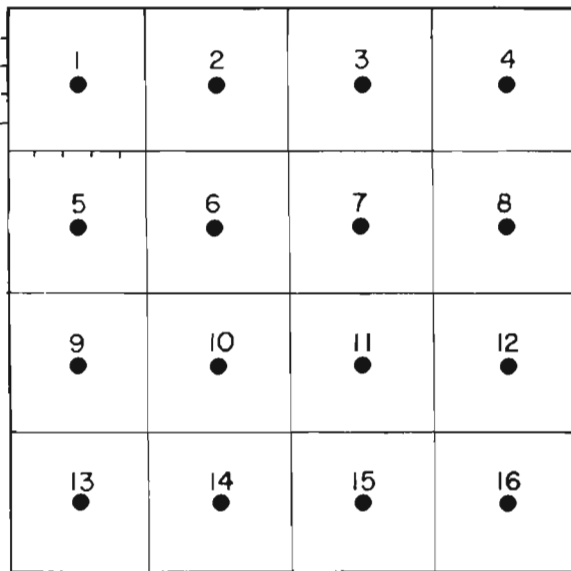
C. Systematic In Cells
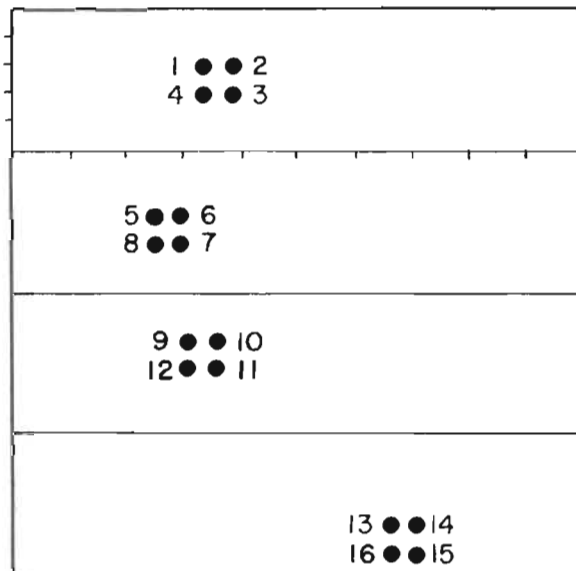
D. Clusters Of Four

FIGURE 2 · SAMPLING DIAGRAMS, ILLINOIS BEACH EXPERIMENT
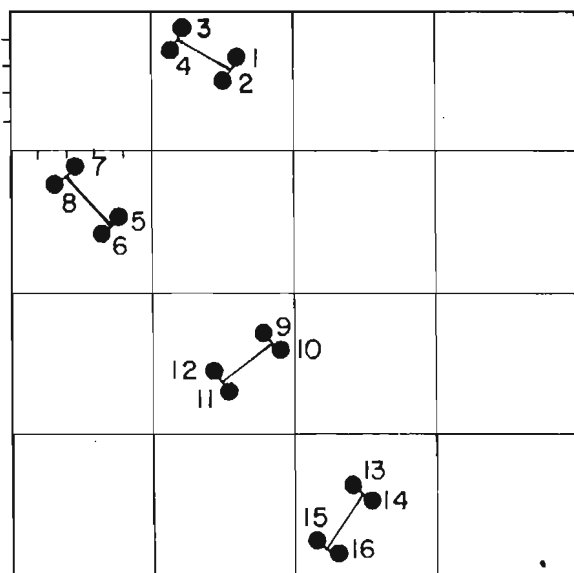
E. Three Level Design
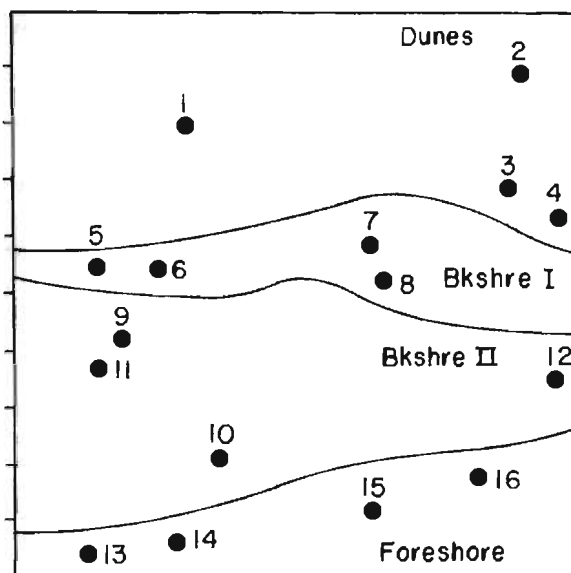
F. Stratified Random
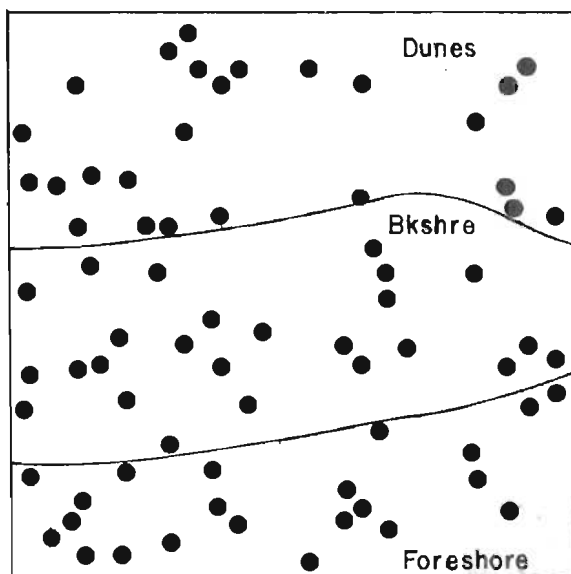
G. Three Strata (Combined Samples)
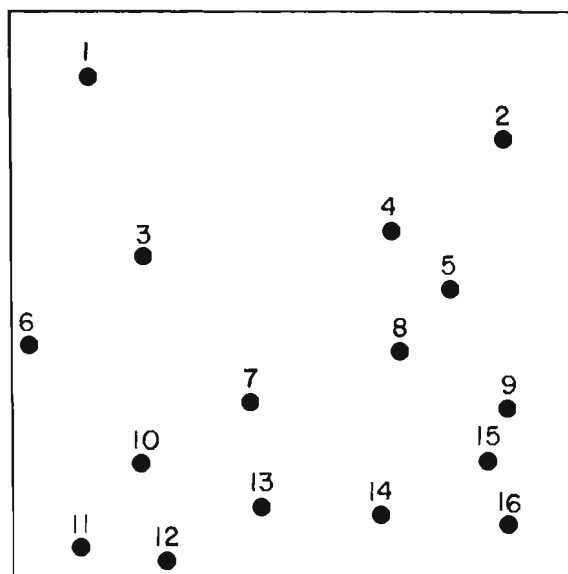
H. Purposive Selection

FIGURE 2 · SAMPLING DIAGRAMS, ILLINOIS BEACH EXPERIMENT

9

the beach in detail, and then designated 16 of the available samples that he believed would provide a satisfactory estimate of the mean particle size. This auxiliary experiment was included to determine whether samples collected on a non-probability basis (subject-matter knowledge) were as good or better than those collected by randomization processes. The samples so selected are shown in Figure 2H.

Each group of samples was analyzed in a standard manner for particle size distribution by sieving, heavy mineral content by magnetic fractionation, content of acid solubles (mainly calcite and dolomite grains), and roundness and sphericity of the particles. The analyses were performed by class members, and in order to avoid operator bias the samples were distributed in a randomized manner among the analysts. This assured that no important systematic operator errors would enter the results, although the non-systematic variations among operators might inflate some of the error terms in the experiment. Previous studies on this point indicated that the additional errors introduced were not great enough to mask essential differences between samples except in some mineral identification studies and in visual measurement of roundness and sphericity.

Particle Size Data. As the samples were collected, it was noted that many of them contained scattered pebbles, and in a few instances the samples were almost entirely gravel. The mixed materials when analyzed sometimes showed the presence of two distinct size distributions. These were in effect mixtures in varying proportions of pebbles with sand in the interstices, although in other instances the sample consisted of uniform sand with one or more "floating pebbles".

One way of handling problems of this sort is to separate the mixture into two strata, sand and pebbles, and to treat each separately in terms of its mean value. Such a solution was hindered by occurrence of some samples in which a single size distribution was present that ranged from pebbles to sand without a break between. It was decided on this basis to use a weighted mean particle diameter, whose value depends on the relative proportion of each size grade in the sample. This resulted in some instances in a mean value that lay between the two peaks on the distribution curve. However, such weighted means do indicate the relative coarseness of the deposit as a whole, and they provide a single mean value for comparative purposes.

Without exception the dune sand was well sorted and without pebbles. Some of the backshore samples also were well-sorted sand, but others displayed irregular size curves. The particle size data are summarized in Table 1, which lists the logarithmic mean (phi mean) of each sample according to its grouping in the sets of samples. Discussion of the phi mean as a measure of average particle size is given in Krumbein and Pettijohn (1938, Chaps. 8 and 9).

TABLE 1
ILLINOIS BEACH - MEAN PARTICLE SIZE OP BEACH SAMPLES
(Expressed as Phi Means and Arranged According to the Sampling Plans Used)

Sample Number	Simple Random (Fig.2A)	Random in Cells (Fig.2B)	Systematic in Cells (Fig.2C)	Clusters of Four (Fig.2D)	Three-Level Design (Fig.2E)	Stratified Random (Fig.2F)	Purposive Selection (Fig. 2H)
1	1.96	2.08	1.92	1.91	1.91	1.85	1.92
2	1.90	1.88	1.95	1.96	1.98	1.75	2.05
3	1.84	2.00	2.02	1.98	1.44	2.05	1.22
4	1.36	1.80	2.09	1.96	2.06	1.52	1.23
5	1.90	1.99	1.90	1.63	1.63	1.75	1.16
6	0.47	1.60	1.85	1.55	1.68	1.22	1.90
7	2.00	1.00	1.65	1.68	2.06	1.23	2.00
8	0.90	1.16	1.85	1.72	2.00	1.22	0.55
9	0.55	1.50	0.31	1.20	1.20	1.14	-1.84
10	1.07	1.51	1.50	1.18	0.80	2.08	1.44
11	-1.84	1.65	-1.43	0.80	1.00	1.40	0.01
12	1.87	0.14	1.26	0.95	1.05	0.26	-0.10
13	1.44	1.21	-2.24	-1.03	-1.03	0.01	1.92
14	-0.22	1.72	0.23	-0.45	-0.98	0.48	-1.03
15	-0.10	-1.75	-0.44	-0.98	0.64	0.69	1.83
16	1.92	1.83	-2.75	1.01	0.74	1.83	-2.75

TABLE 2
ILLINOIS BEACH-SUMMARY OF
DATA ON MEAN PARTICLE SIZE

	(1) No. of Samples	(2) "Grand Mean" of Each Set $\bar{X}_p$	(3) Variance of the Grand Mean $V(\bar{X}_p)$	(4) Standard Error of the Grand Mean $S(\bar{X}_p)$	(5) Relative Efficiency	(6) 95% Confidence Band
A Simple Random	16	1.06	0.0724	0.27	1.00	0.49 to 1.63
B Random in Cells	16	1.33	0.0569	0.24	1.13	0.82 to 1.84
C Systematic in Cells	16	0.73	0.1931	0.44	0.61	-0.21 to 1.67
D Clusters of Four	16	1.07	0.2636	0.51	0.53	-0.02 to 2.16
E Three-Level Design	16	1.14	0.1684	0.41	0.66	0.27 to 2.01
F Stratified Random=	16	1.40	0.0226	0.15	1.80	1.09 to 1.71
G Three Strata (Combined Data)	72	1.21	0.0083	0.09	2.97	1.03 to 1.39
H Purposive Selection	16	1.21	0.0729	0.27	1.00	0.64 to 1.78

In sieving the samples a sufficiently large portion (which always included all the pebbles present) was analyzed, so that for all practical purposes the mean particle size of the sample was determined with only a very small analytical error. The mean particle sizes may thus be considered as known parameters associated with each carton of sand.

In setting up statistical measures for comparing relative sampling efficiency, it is appropriate to define the word "sample" more carefully. In common engineering and geological practice a unit volume of sand (as a half-pint ice cream carton) collected at a given point on the beach is a sample of sand. However, each such sample is itself only one of a very large number of similar samples that could be collected on the beach. In fact, the number of 3-inch circles in a 300-foot square is of the order of 1.5 million. Hence, each carton of sand is an individual in this "super-population" of all possible cartons that could be collected.

On this basis, the mean particle size obtained from a carton of sand represents a single observation in some 1.5 million such observations that could be made on the beach. The 16 cartons of sand in each sampling design thus represent a "super-sample" of 16 from this larger "super-population". There is a more formal way of stating this relation (Miller, 1954), but the essential point is that if the 16 mean values from each sampling design are themselves averaged, the resulting grand mean from the "super-sample" is an estimate of the population mean particle size in the entire 300-foot square.

Table 2 lists the grand means of the particle size distributions in the group of samples, expressed as the arithmetic means of the phi means. Inasmuch as each group of sand samples was independently collected (with a few exceptions to be noted), the grand means represent independent estimates of the population mean, each based on 16 observations, except for the combined design of 72 samples. The grand means vary from 0.73 to 1.40, representing a range in the geometric mean diameter from 0.60 to 0.38 millimeter. This is an illustration of how widely the estimates of a total population mean may vary when only 16 samples are collected in a set. The best estimate of the population mean particle size for the 300-foot square, based on 72 samples, is 1.21 in phi units, corresponding to a geometric mean diameter of 0.43 millimeter.

In addition to computing the grand mean for each group of sand samples, it is possible to compute the standard deviation of the individual means in each group. It was stated in the earlier report (Beach Erosion Board, Technical Memorandum No. 50) that different sampling designs may require different manners of computing the variability in the group. Methods of computation for each kind of sample are included in an Appendix to this report. The computed values of the variance of the means are shown in the third column of Table 2. These variances

show an even greater relative range of values than do the grand means. The several groups of 16 samples show a range from 0.0226 to 0.2636, which is about tenfold. A wide range is characteristic for sets of relatively small samples. The value is reduced to 0.0083 for the complete set of 72 independently collected samples. The large reduction is partly associated with the greater number of samples involved.

The relative values of the variances obtained from the several sampling plans can be used to compare the relative efficiency of the sampling plans. The completely random samples provide a "norm" for the comparison, inasmuch as basic sampling theory relates to completely randomized samples. In making this comparison the standard error of the mean is computed for each sampling plan, and ratios are taken of each standard error to the standard error of the random samples. In this way the relative "efficiency" of each sampling plan with respect to the norm may be expressed. An alternative is to use the ratio of the variances directly, in which case the relative precisions are equivalent to the square of the relative efficiencies in Table 2.

The standard errors of the grand means are shown in column 4 of Table 2. They were obtained by taking the square root of the variances in the preceding column. The ratio of the standard error of each group to that of the random samples is shown in column 5 headed "Relative Efficiency". Inasmuch as the random sampling provides the norm, its own ratio is 1.00. The column shows that the relative efficiency of the several sampling plans ranges from 0.53 to 1.80 for the sets of 16 samples. The lowest value is associated with the cluster samples, only 53% as efficient as the random samples, to the stratified samples, which are 80% more efficient than the completely random samples. The entire set of 72 independent samples, computed as a stratified sample representing the three main shore zones, is nearly three times as efficient as the set of 16 random samples. This improvement arises from the combination of stratification plus a greater total number of samples.

An interesting sidelight in this experiment is the relative efficiency of the "purposive selection" group of samples. It found a value of the grand mean almost identical to that from the entire set of 72 samples, but its relative efficiency is about the same as that of the random samples. This result is borne out by experience, which appears to indicate that expert knowledge provides a sound basis for estimating a mean value, but that the variability of which the estimate is subject is very difficult to control.

Figure 3 shows the preceding relations graphically. The horizontal bars on the left opposite each sampling plan represent the 95 percent confidence band about the grand mean of the 16 samples in the set. A

Grand Means (Phi Units)

A
B
C
D
E
F
G
H

A. Simple Random
B. Random in Cells
C. Systematic in Cells
D. Clusters of Four
E. Three-Level Design
F. Stratified Random
G. Three Strata (Combined Data)
H. "Purposive Selection"
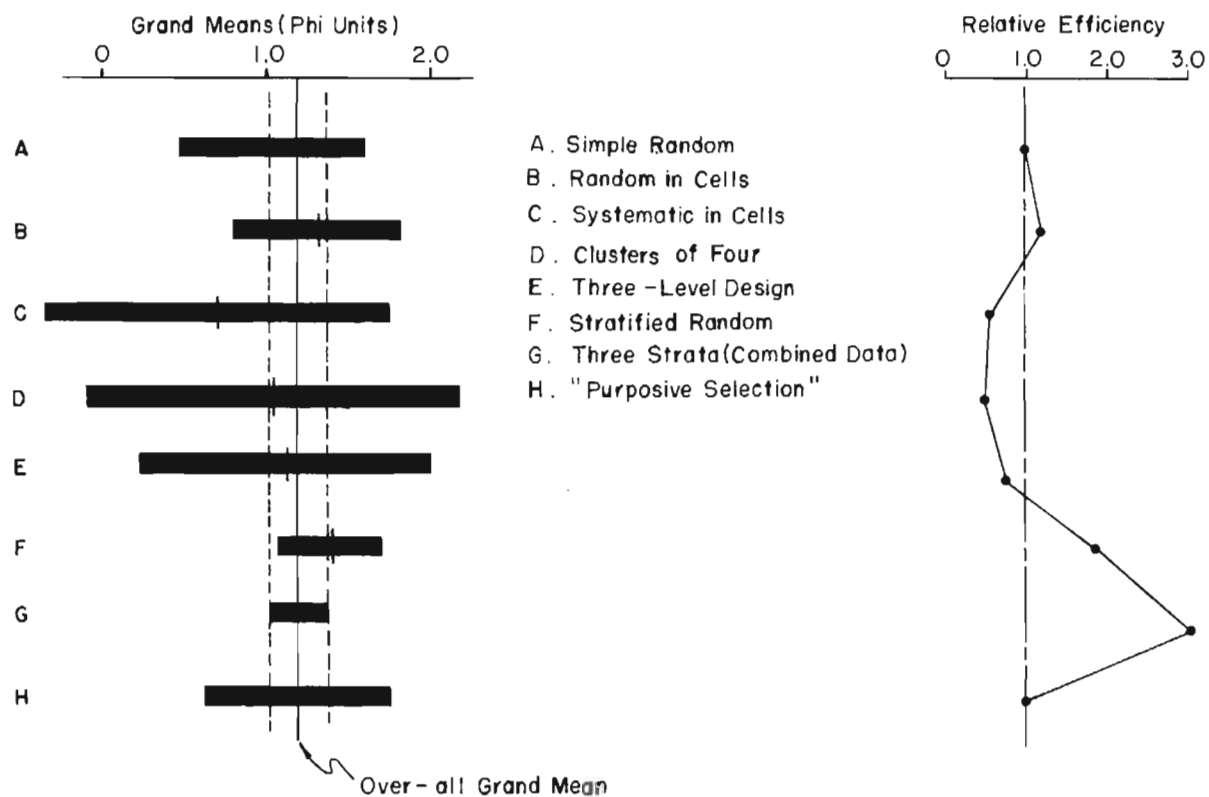
Over-all Grand Mean

Relative Efficiency

FIGURE 3 · CONFIDENCE BANDS AND RELATIVE EFFICIENCY
OF SAMPLING DESIGNS FOR PHI MEAN PARTICLE DIAMETERS

14

confidence interval or band about a mean is that range of values that has a given probability of including the population mean. These intervals are commonly set at the 95 percent confidence level. The limits are given in the last column of Table 2. The grand mean for each set is indicated by the short line segment at the center of each bar. The lateral displacement of the bars is a graphic indication of the differences in the estimates of the overall population mean provided by relatively small sets of samples.

The heavy vertical line through the bars is the average phi mean for all 72 independently collected samples, and the dashed lines on either side of it represent the 95 percent confidence band about this overall mean. As may be seen, this confidence band includes most of the grand means of the samples sets. The systematic-in-cells samples yielded the smallest grand mean, although this is not to be considered as a generalization for that sampling method.

The right hand part of Figure 3 is a graphic representation of the relative efficiency of each sampling plan. The graphed line fluctuates about the random sampling plan value of 1.0, moving to a peak of greatest efficiency opposite the stratified sample plan.

A practical implication of the relative efficiencies of the sampling plans can be stated as follows: for a given level of precision it would take twice as many cluster samples as random samples to get the same reliability in the estimate of the population mean, whereas it would take only slightly more than half as many stratified as random samples to get equal reliability in the estimate of the population mean. Although the specific values obtained in this experiment cannot be applied to other beaches without qualifications, the experiment as a whole strongly supports the generalization that beach samples should be stratified if the main objective of the sampling is to estimate the population mean particle size.

It may be mentioned here that sampling experiments of the type described are most satisfactorily conducted with populations that may be completely studied by census. In geological and engineering studies on beaches it is not possible to take a census of all the sand on the beach. As a result there is no absolute standard of comparison for the experiment, and the values obtained for the relative efficiencies of the sampling plans depend in part on the particular experiment being run.

Acid Solubles Data. Among the other properties measured on the sand samples were several that seemed to show no significant differences between dunes, backshore, or foreshore. That is, there appeared to be no natural stratification of the sort that was observed in the coarser and finer zones represented by particle sizes. It was accordingly decided to test the relative efficiency of the several sampling plans in this case also.

15

The acid solubles represent mainly: the calcite and dolomite grains in the deposits. They are determined by taking a weighted sample, heating it in dilute hydrochloric acid, washing, drying, and reweighing. The percent of acid solubles was computed from these data, and they are listed for each sample in Table 3. As may be noted, there is relatively little difference either among the samples in a given set, or among the sets. Analysis of variance (Krumbein and Miller, 1953) was used to test for significant row and column changes along and across the beach, and the tests failed to show that any marked effect was present. That is, the data do not contradict the inference that the population of acid soluble materials is homogeneous in the sense that samples collected from one part of the 300-foot square are statistically similar to samples collected from other parts of the square.

Table 4 lists the summarized statistical data for the acid solubles, similar to Table 2 for particle size. It will be noted that the grand means of the "super-samples" of 16 cartons vary only slightly, showing a range from 5.18 percent by weight to 5.65 percent by weight. The mean of all 72 independently collected samples is 5.35 percent. The variances of the sets of 16 samples also show a lesser range than did those for particle size. As before, the standard errors of the grand means are used to estimate the relative efficiency of the several sampling plans. These are indicated in the appropriate column, and they show that the random plan is about as good as any. The cluster samples are still the poorest, and interestingly enough the purposive selection set has about a 30 percent increase in efficiency. The 72 independent samples were treated as a random set to obtain the relative efficiency of 1.92, which is entirely a function of the larger number of samples included.

Figure 4 shows the results of the acid solubles graphically, and indicates again that the grand means of the sets of 16 samples vary from set to set, but that the confidence band about the total estimate based on the 72 samples includes all or most of the individual grand means.

The graph of relative efficiency on the right side of the figure also shows the lack of any startling departure from the unit value line. The generalization that arises from this experiment is that when the population is homogeneous, simple random samples are about as effective as any other.

Other Particle Properties. As pointed out earlier, the beach samples were analyzed for their magnetic mineral content (approximately equivalent to the heavy mineral content), particle roundness, particle sphericity, and content of selected mineral species. Some of these properties showed stratification according to beach zones

TABLE 3

ILLINOIS BEACH-PERCENT OF ACID SOLUBLES IN BEACH SAMPLES

Sample Number	Simple Random (Fig.2A)	Random in Cells (Fig.2B)	Systematic in Cells (Fig.2C)	Clusters of Four (Fig.2D)	Three-Level Design (Fig.2E)	Stratified Random (Fig.2F)	Purposive Selection (Fig.2H)
1	5.6	6.6	4.6	5.7	5.7	6.5	4.6
2	5.8	5.2	5.0	4.2	5.3	5.2	4.7
3	5.8	5.8	5.4	5.5	4.8	4.7	6.0
4	5.3	6.2	5.6	5.5	5.2	4.6	5.0
5	7.0	5.3	5.1	4.5	4.5	5.4	4.6
6	6.1	5.1	5.1	5.3	5.0	6.0	7.0
7	6.1	4.6	4.7	5.0	5.3	5.0	6.1
8	5.1	4.6	4.4	5.3	5.8	5.3	5.6
9	5.6	5.5	4.6	4.7	4.7	5.3	5.7
10	5.8	4.9	5.3	5.6	6.4	6.0	5.5
11	5.7	5.4	5.5	6.4	5.4	5.2	5.7
12	4.8	5.4	5.3	5.6	4.6	4.9	4.9
13	5.5	5.6	5.8	6.6	6.6	5.7	6.3
14	5.0	4.4	4.7	6.4	6.0	5.0	6.6
15	4.9	4.3	6.1	6.0	4.5	6.1	5.8
16	6.3	5.8	5.7	6.0	5.2	4.9	5.7

TABLE 4

ILLINOIS BEACH-SUMMARY OF DATA ON ACID SOLUBLES

Sample Set	No. of Samples	"Grand Mean" of Each Set $\bar{X}_p$	Variance of the Grand Mean $V(\bar{X}_p)$	Standard Error of the Grand Mean $S(\bar{X}_p)$	Relative Efficiency	95% Confidence Band
A Simple Random	16	5.65	0.0203	0.14	1.00	5.35 to 5.95
B Random in Cells	16	5.29	0.0256	0.16	0.88	4.95 to 5.63
C Systematic in Cells	16	5.18	0.0300	0.17	0.82	4.82 to 5.54
D Clusters of Four	16	5.52	0.0729	0.27	0.52	4.95 to 6.09
E Three-Level Design	16	5.32	0.0398	0.20	0.70	4.90 to 5.74
F Stratified Random	16	5.36	0.0440	0.21	0.67	4.87 to 5.77
G Three Strata* (Combined Data)	72	5.35	0.0049	0.07	2.00	5.21 to 5.49
H Purposive Selection	16	5.33	0.0121	0.11	1.27	5.10 to 5.56

* Treated as 72 random samples

17

Weight % of Acid Solubles

5.00   6.00

A

B

C

D

E

F

G

H

A . Simple Random
B . Random in Cells
C . Systematic in Cells
D . Clusters of Four
E . Three – Level Design
F . Stratified Random
* G . Three Strata (Combined Data)
H . "Purposive Selection"

* Treated as 72 random samples

Over – all Grand Mean
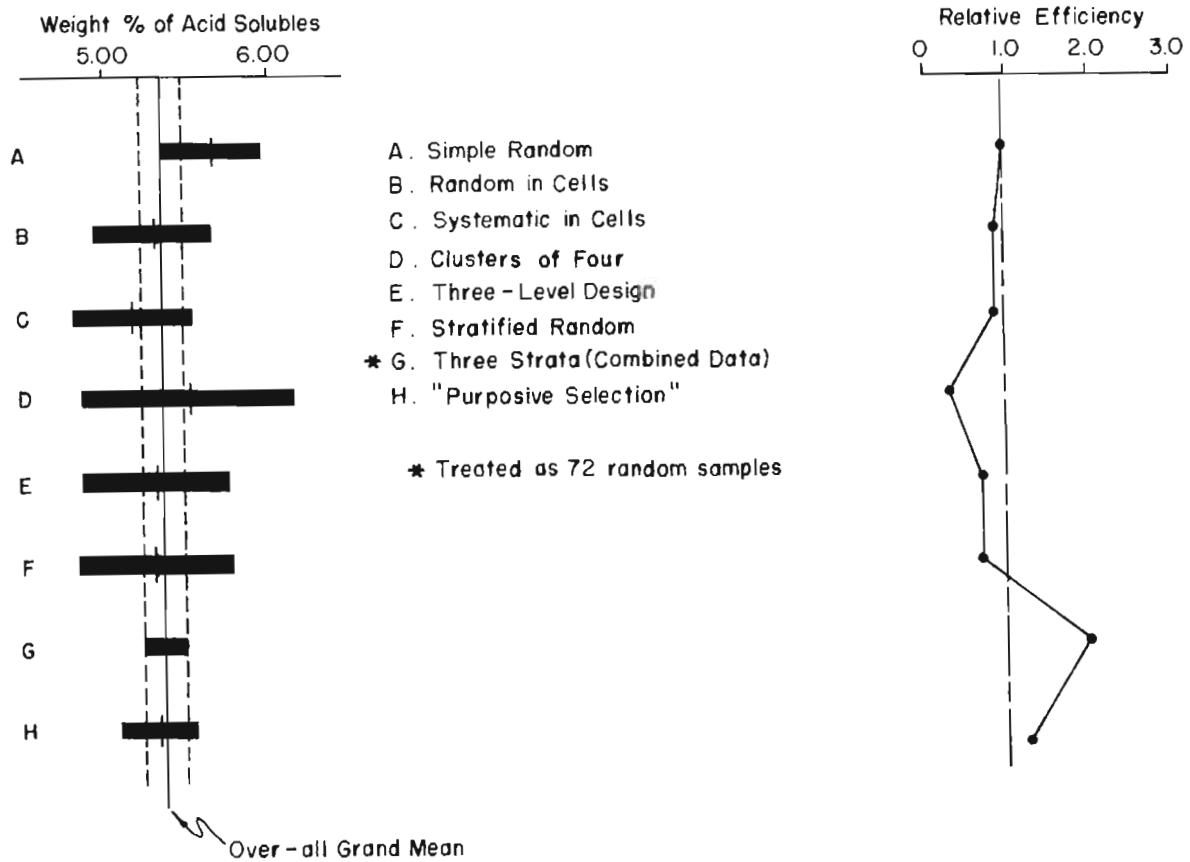
Relative Efficiency

0   1.0   2.0   3.0

FIGURE 4 · CONFIDENCE BANDS AND RELATIVE EFFICIENCY
OF SAMPLING DESIGNS FOR ACID SOLUBLES IN BEACH SAND

18

in a manner similar to particle size, and others appeared to be homogeneous over the beach area. Detailed analyses of these additional properties are not included here, but in general these properties behaved as described earlier elsewhere (Krumbein and Miller, 1953).

As a general summary, it may be pointed out that when a multiple-purpose sample is collected, as one for estimating the population mean of particle size, heavy mineral content, and other properties, it is preferable to use stratified sampling, inasmuch as in homogeneous populations there are no particular disadvantages to stratified sampling, whereas for properties that vary across the beach the stratified samples yield better estimates of mean values.

Sampling for Gradients

The designed experiments reported above were analyzed for the purpose of comparing sampling methods used in the estimation of population means. This, as was mentioned in the Introduction, is an important phase of beach sampling for some purposes. Another purpose of sampling is to estimate changes in mean particle size or other properties along and across beaches. This can be done by collecting samples in such manner that they provide data for detection of systematic changes from one part of the beach to another. In one sense, the occurrence of natural strata or zones on the beach is an indication that the several parts may be significantly different in some of their properties, but there are more formal ways of expressing these relations.

One of the most direct ways of studying beach gradients is to collect sets of samples along profiles normal to the shore line, or along sections laid parallel to the shore line. The analytical data are then plotted on a graph to see whether any trends are discernible. Many such studies have been reported in the literature and they need not be reviewed here. These studies amply demonstrate the occurrence of gradients along and across beaches in such varied properties as particle size, particle shape (sphericity), particle roundness, content of heavy minerals, moisture content, firmness of the beach, beach slope, and others. Normally the gradients across a beach are more pronounced than along the beach. For many properties the along-beach changes are exponential, although some are dominantly linear.

Where the gradients are pronounced the graphs clearly indicate their presence. However, in some instances there may be considerable variability in the data, so that the scatter of individual points on the graph may tend to obscure any gradient that is present. In such cases statistical methods based on regression analysis are helpful in testing the data for significant linear, quadratic, or higher degree changes in the graph along the beach. Within recent years methods have become available for extending the analysis to areas as well as traverses. Contour-type maps of beach attributes can be analyzed

in this manner to show on the one hand the smooth underlying gradient surface and on the other the random variations that are superimposed on the smooth surface. Space limitations prevent illustration of these methods here, but examples of geological maps analyzed by these methods are given by Krumbein (1956).

When samples are collected for qualitative estimation of gradients they may be collected at arbitrary intervals along the traverse, but for formal analysis it is advantageous to have the samples equally spaced along the line or over the beach area. Equal spacing permits use of greatly simplified methods of computation based on orthogonal polynomials (DeLury, 1950; Bennett and Franklin, 1954, p. 255), and furnishes data on linear, quadratic, and higher components in the gradient with a minimum of mathematical effort.

In terms of the sampling designs described here, the most effective for formal regression analysis is the set of systematic samples shown in Figure 2C. This provides equally spaced samples for the formal analysis, and at the same time provides an optimum distribution of points for mapping the beach attributes. If the present experiment had been designed specifically for regression analysis, the cells would have been made about one-fourth as large, yielding 64 samples over the area. This closer spacing would give a better insight into the components that make up the gradients.

An apparent contradiction arises when beach sampling is designed simultaneously for several purposes, such as the estimation of an over-all population mean and for a study of gradients that may be present. For estimating mean particle size, for example, stratified sampling is preferable, whereas for studying particle size gradients, systematic samples are to be preferred. Various combination designs are possible for resolving the contradiction. One of them is to take one sample at each cell center (say) and supplement these with additional samples distributed over the natural beach zones. All of the samples could then be combined in terms of their occurrence within the natural beach zones for estimating the mean, whereas only the systematic samples need be used for studying the gradients. Other types of combined plans can be designed for special studies.

Sampling for Variability Estimates

The third purpose of sampling beaches as mentioned in the Introduction involves estimation of the relative variability of deposits in each of the several beach zones. This kind of information is of value in study-ing the relative homogeneity or heterogeneity of beach deposits in the backshore, foreshore, or nearshore bottom, as a reflection of the physical processes that take place. The variability within any one zone is commonly a function of the sample spacing, so that there are levels of variability associated with samples collected a few feet apart, a few tens of feet apart, and so on up the scale. In most sand beaches it is

probably true that there is less difference between two samples collected a foot apart than there is between the averages of two pairs of samples collected say 100 feet apart. That is, neighboring sites commonly furnish sand more nearly similar than more distant sites. On the other hand, examples can be cited in which closely spaced samples differ markedly, yet the averages of pairs of samples farther removed have an over-all similarity.

Variability of Beach Zones. Knowledge of the levels of maximum variability in beach deposits is helpful in selecting a sample spacing designed to give maximum information per unit of effort or cost. The statistical method involved is one that provides estimates of variance components, and the design can be set up to estimate the components at a number of different sample spacings simultaneously. In fact, for this purpose the cluster samples and multilevel samples (designs 2D and 2E) that appeared relatively inefficient for estimating the mean, provide an effective way of getting at the problem. The cluster sample, for instance, represents a two-stage sampling process, in that the clusters were spaced on the average about 100 feet apart, whereas the samples within a cluster were only a few feet apart. Hence, estimates can be made of the variability at these two levels of sampling.

The nested samples (design 2E) provide an example of three-stage multilevel sampling, in that the main centers were 100 feet apart, the two pairs at each main center were 40 feet apart, and the two samples making up the pairs were two feet apart. Thus this design permits estimation of variance components at three levels of sample spacing.

TABLE 5

ILLINOIS BEACH — VARIANCE COMPONENTS FOR CLUSTER SAMPLES AND THREE-LEVEL DESIGN

(Designs D and E, Figure 2)

Sampling Level	Relative Sample Spacing	Cluster Design (2 levels)	Multilevel Design (3 levels)
Between main cluster centers	100 feet	0.9950	0.7072
Between locations at main cluster centers	40 feet	-	0.3638
Between samples	2 feet	0.2372	0.0356

Table 5 lists the variance components found in the cluster and three-level sampling designs. These sets of samples were not collected wholly independently, and the estimates are therefore somewhat interrelated.

The main cluster centers were common to both designs, and for the three-level experiments one pair of each cluster of four was used in conjunction with another pair collected separately for the three-level design. This overlapping was done to save time in the original study, inasmuch as the principles illustrated are not affected. However, the overlapping samples were not included in the estimates of the means given in an earlier section.

As Table 5 shows, the greatest variability occurs at the top sampling level, and the smallest variability occurs at the smallest sample spacing. This result was expected inasmuch as the whole sampling plan was laid over a variable deposit that included foreshore, backshore, and dunes. On the other hand, it is sometimes found that the maximum variability occurs at an intermediate or lower level, especially when there are abrupt changes from sand to gravel on the beach. In that case some of the pairs at the lowest level might pick up one sample of each.

In fact, the apparently large difference between the variance components at the lowest level in Table 5 was occasioned by the fact that one of the clusters happened to include samples with a wide range of phi means. This greatly inflated the variance contribution from this cluster sample. It provides an illustration of one of the risks involved in using a limited number of samples from a highly variable deposit.

Sample Allocation and Cost Functions. Some of the uses that can be made of the variance components have to do with questions of optimum sample spacing on a grid, optimum number of samples to be collected at each level for maximum reliability in estimating population means, and with questions of optimum distribution of samples at the several levels for maximum returns per unit of cost. For example, inspection of the relative values of the variance components for the three-level example in Table 5 shows that the variability at the lowest stage is only about 1/10 that at the intermediate stage, and only about 1/20 that at the top stage. This suggests qualitatively that it is hardly worth while collecting very many samples only a few feet apart. It would intuitively seem better to collect more samples on the levels where the variability is greatest.

Formal methods are available for making decisions on this point, and they may not always agree with the intuitive reasoning. These methods were illustrated with beach penetrability data in the earlier report (Beach Erosion Board Technical Memorandum No. 50, p. 24). It was shown that in general it is most advantageous to increase the number of largest sampling units rather than to increase the number of samples at a lower level within the largest units, even when the variability is greatest at the lower levels. In terms of the present experiment, for example, it is formally and intuitively more desirable to increase the

22

number of cluster points rather than to increase the number of samples in a cluster. If the present three-level design had been arranged to take eight cluster locations, with two points about each cluster, and only one sample from each point, the efficiency of the design with respect to the random sampling design would have increased from 0.66 to 0.87. This is still not as good as random sampling for reliability of the population mean estimate, but it is better than the sample allocation that was used.

The use of cost functions in distributing sampling efforts is taken up in the Appendix of this report. It is shown there, for example, that in the clusters-of-four design a more optimum allocation would be to collect only one sample instead of four in each cluster. In fact, if one sample were collected at each level for sixteen positions, this would in effect give a random set of samples with a relative efficiency of 1.0.

This analysis indicates that cluster sampling or multilevel sampling in general appears to have its greatest value in furnishing information on levels of variability, unless the number of sampling units at the top level is large. If multilevel sampling is used for estimating the mean, it seems desirable to have at least ten major sampling units. Moreover, multilevel sampling increases in effectiveness as the over-all population becomes more homogeneous, and it may save costs by permitting sampling effort to be concentrated at certain accessible points rather than requiring a wide distribution of individual sampling localities. Cost function analysis, based on some knowledge of the variabilities involved, is a necessity in designing effective multilevel designs.

Summary Remarks on Illinois Beach Experiment

The expository treatment of sampling in this section was arranged to show that the purpose of the sampling has a strong effect on selection of a sampling design. It was seen that methods that appear to be relatively unsatisfactory for some purposes are highly desirable for others. The four most important designs out of the seven that were used are (1) simple random sampling when the objective is to obtain the best mean value for a homogeneous population, (2) stratified sampling when the objective is to obtain the best mean of a heterogeneous but zoned population, (3) systematic sampling when the objective is to study gradients or to prepare maps, and (4) multilevel sampling when the objective is to study relative degrees of variability within the deposits or zones.

The relative heterogeneity of the Illinois Beach deposits is such that some of the differences between the several sampling methods were made more prominent than they would be on beaches composed entirely of sand. In the latter cases the absolute and relative variabilities are generally smaller, although it is believed that the four generalizations

23

about optimum sampling methods still hold. In the following sections
interest is focused on a sand beach along the Atlantic Coast, and
although data are not available on a variety of sampling plans, an
attempt is made to show how sample allocation within a major design
affects the reliability of the estimated mean particle size.

SAMPLING EXPERIMENTS AT OCEAN BEACH, MARYLAND

Introduction

Ocean Beach is located along the barrier beach south of the inlet
at Ocean City, Maryland. At the time the samples were collected the
beach area was relatively undeveloped, so that the beach deposits were
not subject to disturbances by public bathing or other activities. A
sand ridge about 5 feet high had been bulldozed along the backshore
about 300 feet from the low tide line, but it was felt that this
feature did not seriously disturb the sampling design. Some reworking
of the ridge surface by winds had developed patches of dune sand along
it, but similar small patches occurred on the natural backshore and
would be picked up in occasional samples even in the absence of the
ridge.

Sampling Designs

The entire beach was composed of sand, and a well developed berm
was present at about the limit of the swash marks from the previous
high tide. The sampling design was laid out as six profiles set normal
to a main base line that paralleled the shore as shown in Figure 5.
The base line was 1,600 feet long and was laid out just inland of the
main berm. At each end of the base line a profile about 400 feet long
was laid normal to the shore line extending from the low tide line to
a point on the backshore about 100 feet inland of the sand ridge. Four
additional profiles of the same length were laid out in the central
part of the beach area. These were 100 feet apart, the most southerly
one, designated as number 1, being half way between the two end profiles.

As laid out, the design provided three main profiles, shown as S,
1 and N, each 800 feet apart. These could be used to test for differences
in sand properties at three fairly widely spaced profiles. In addition
the design provided four closely spaced profiles that could be used to
test for differences within a smaller beach area. As originally laid
out, three of the four central profiles were extended 2,000 feet ocean-
ward to depths of approximately -30 feet.

Each profile was divided into 50-foot segments with stakes, and two
samples of sand were collected at each stake by randomizing two positions
in a two-foot circle with the stake at center. The pairs of samples
were thus two feet apart at each stake, the stakes were 50 feet apart
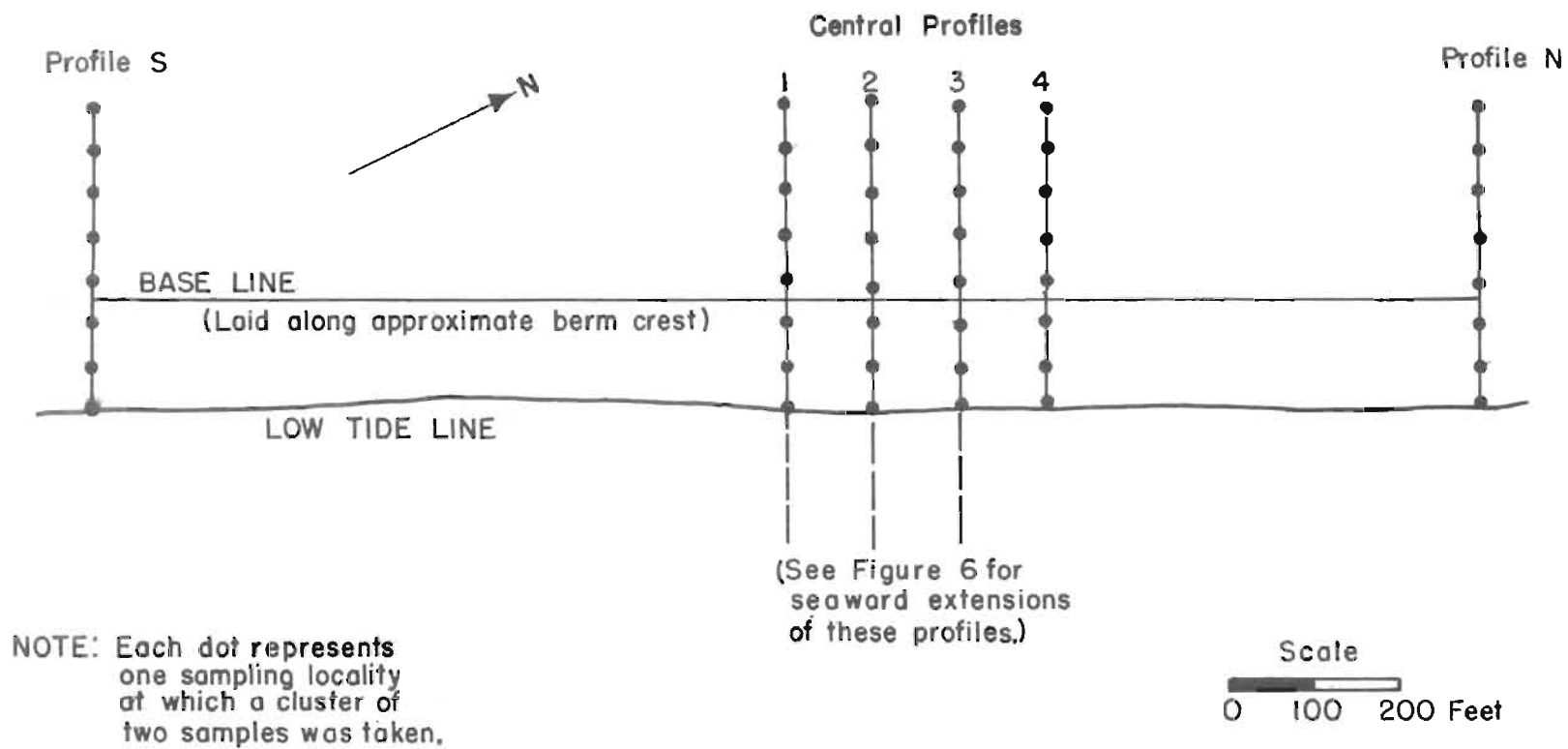along the profile, the central profiles were 100 feet apart along the

FIGURE 5· BEACH SAMPLING DESIGN AT OCEAN BEACH, MARYLAND

beach, and the three main profiles were 800 feet apart along the shore. The four central profiles provided 64 samples and the two end profiles provided 16 each. In the formal design, accordingly, there were eight systematic sampling points along each profile, each systematic point being represented by a cluster of two sand samples. All the samples were probability samples in the sense that the end of the base line was randomized on the beach, and hence all samples collected represented an extension of this same randomization process.

The entire design provided data for answering a number of questions about the mean particle size on the beach, as well as questions on the occurrence of significant changes from one end of the 1,600-foot beach segment to another, as well as within the smaller 300-foot segment represented by the four central profiles. For example, with respect to the total beach segment of 1,600 feet, the following questions can be raised:

1) What is the best estimate of mean particle size that can be obtained from all the pairs of samples from the three main profiles (S, 1, N)?

2) How much less reliable is the estimate of mean particle size obtained from all the pairs of samples along a single profile in the 1,600-foot segment?

3) How much less reliable is the estimate of mean particle size obtained from a single profile with only one sample at each sampling point? In other words, does it pay to take the samples in pairs?

4) How much less reliable is the estimate of mean particle size obtained from only one sample on a single profile? In other words, how good is one sand sample for estimating the mean particle size for the whole 1,600-foot beach segment?

Similar questions can be raised for the central 300-foot beach segment represented by the four closely spaced profiles. As mentioned, the data also supply information on changes in mean particle size along a profile (across the beach), and between profiles (along the beach), for both the shorter 300-foot segment and the longer 1,600-foot segment. These questions are illustrated by examples in the following sections of this report.

Estimates of Mean Particle Size on Exposed Beach

The arrangement of the sand samples in the Ocean Beach experiment is such that the data can be analyzed in at least two general ways for estimates of mean particle size. Inasmuch as there are several different sample spacings in the design, the data may be treated as a multilevel sampling design, which permits estimation of the variability

associated with each sampling level. This type of analysis can provide approximate answers to the kinds of questions listed in the preceding section. A second way that the samples can be analyzed is in terms of the natural beach strata in which the samples happened to fall. Each profile had five backshore sampling points and three foreshore sampling points. By redistributing the samples according to their location in one or the other stratum, the samples can be handled as a stratified set.

In fact, both of these approaches can be treated as one comprehensive design by separating the total variability among the samples into four components: the variability between profiles, the variability between strata within a profile, the variability between positions (sampling points) within a stratum, and the variability between the two samples at each sampling point. This analysis will be conducted in two phases. In the first, the four central profiles are used for various estimates of the mean particle size for the central 300-foot segment of the beach. In the second, the first profile in the central group is used with the two end profiles to estimate the mean particle size for the 1,600-foot beach segment. In this way a number of questions regarding the reliability of one or more samples in estimating the mean may be examined.

The weighting factor used in these computations is the relative area of each stratum. A volumetric weighting factor which takes into account the thickness of deposit in each stratum would be preferable, but such data are not available. The present usage illustrates the process, however, and is uniform in the comparisons made.

Table 6 lists the phi median particle diameters of the sand samples in the total design. The first column under each profile heading represents the first sample collected at the stake, and the second column represents the second sample of the pair. As may be noted, the samples in each set are also classified according to their position in the backshore or foreshore.

The method of analysis of the data is the same as for a multilevel design. Essentially it consists in preparing a series of condensed tables, in the first of which the two samples of each pair are combined. Then the pairs are combined within their strata, then the strata are combined, and finally the profiles for each group are combined. The values in each successive table are squared, and the sums of the squares are used to break out the variability associated with each level. This variability is expressed first as a series of mean squares that can be used to test for significant differences in the variability for each succeeding level of the design. The mean squares may also be used to estimate the component of variance associated with each level, and from these components various other sampling combinations can be evaluated.

27

TABLE 6

OCEAN BEACH-PHI MEDIAN PARTICLE DIAMETERS OF PROFILE SAMPLES

(See Table 11 for Underwater Samples)

STRATA					PROFILES						—
	S		1		2		3		4		N
Backshore	1.54	1.71	1.80	1.80	1.80	1.80	1.85	1.85	1.75	1.75	1.92	1.84
	1.45	1.31	1.95	2.00	1.60	1.60	1.80	1.85	1.75	1.80	1.41	1.45
	1.82	1.38	2.00	1.75	1.65	2.10	1.90	1.70	1.80	1.95	1.86	1.86
	1.94	1.94	1.85	1.85	1.75	1.85	1.60	1.70	1.65	1.50	1.80	1.84
	0.95	1.05	1.75	1.70	1.80	1.75	1.65	1.75	1.70	1.70	1.60	1.76
Foreshore	0.96	1.02	1.65	0.90	1.30	1.45	0.70	0.80	1.60	1.50	1.49	0.93
	1.18	1.23	1.95	0.90	2.05	1.75	1.65	1.55	1.25	1.05	1.18	0.52
	2.30	2.25*	0.80	1.65	1.85	1.60	2.00	1.80	2.15	2.10	2.20	2.06

* Approximate

TABLE 7

OCEAN BEACH-ANALYSIS OF VARIANCE OF PHI MEDIAN FOR
FOUR CENTRAL PROFILES

Source	Sum of Squares	d.f.	Mean Squares	F
Between profiles	0.0951	3	0.0317	<1 NS
Between strata in profiles	1.6353	4	0.4088	3.25*
Between sampling points in strata	3.0208	24	0.1259	2.63**
Between samples in pairs	1.5338	32	0.0479	
Total	6.2850	63		

Variance Components

Level	Differences	"Samples"	Component
Profile	−0.2771	16	0.0000
Strata	0.2829	8	0.0354
Sampling points	0.0780	4	0.0195
Samples	0.0479	1	0.0479

Variability of the Beach Deposits. Table 7 shows the analysis of variance and the variance components for the four profiles in the central set. The F-tests in the upper part of the table indicate first that there is no significantly greater variability between the profiles than there is between strata within a profile. This is indicated by the NS (not significant) symbol after the first line. The star on the second line indicates that there is a significantly greater variability between strata than between sampling positions within a stratum, at the 5% level of significance. This means that the two strata differ more between themselves than do sampling positions within each stratum. Finally, the two stars on the third line indicate that the greater variability between sampling positions within a stratum is highly significant as compared to the variability between the two samples of each pair collected at the stake.

The F-tests are made by dividing each mean square by the one below, so that contrasts are made of each successive pair of levels in the design. The observed F value is compared with tabulated values (Dixon and Massey, 1951, p. 310) at the 5% and 1% levels. The two stars in Table 7 indicate that the last test mentioned shows significance at the 1% level.

The implications of the F-tests are first that there is no significant difference between the means obtained from the four profiles, as compared with the strata means, but there is a significant difference between the two beach strata as compared with positions in a stratum. Lastly, the tests indicate that there is much more variability between the sampling positions than there is between the samples in each pair. The data can also be used for evaluating different sampling arrangements by separating the variance components as shown in the lower part of Table 7. This part of the table indicates that there is no real contribution to the variability made by the four profiles. The contribution of the strata is 0.0354, which is larger than the component due to sampling points in the strata. The largest single component, 0.0479, is associated with the samples in the pairs. This would seem to contradict the F tests, but it does not, inasmuch as the F tests take into account combinations of the variance components associated with each level of the sampling design. As will be shown, the component due to samples is greatly reduced when the relative contribution at each sampling level is taken into account.

For a completely randomized multilevel design the variability of the grand mean (i.e., the mean obtained by using all the samples in the design) can be expressed as the sum of the variance components associated with each sampling level, each component being divided by the number of elements that are involved in the corresponding level.

Effect of Different Sample Allocations. The present design is not a completely randomized one, inasmuch as the samples are systematic

29

within the two beach strata. However, the design is essentially self-weighting in that the number of samples in each stratum is closely proportional to the width of the stratum along the profile. In fact, the grand mean obtained by the present multilevel analysis is the same as the mean computed by the conventional method of obtaining the stratified mean, despite the limitations of the systematic samples as against completely random samples in each stratum. Accordingly, it is believed that the following analysis gives a relative order of errors adequate for rough comparative purposes. As will be seen, the errors are probably overestimated rather than underestimated.

As with the Illinois Beach experiment, the standard error of the mean for Ocean Beach may be used to compare the relative reliability of different allocations of samples over the profiles and strata. For the four central profiles, using all 64 samples, the grand mean particle size, expressed as the mean phi median, is 1.67, corresponding to an average particle diameter of 0.314 millimeter. The computed standard error for this grand mean is 0.08. In relative terms, obtained by dividing this error by the mean and multiplying by 100, the standard error is 4.5 percent. This magnitude may be used as an approximate yardstick to compare other combinations of samples among the four profiles of the central 300-foot beach segment.

TABLE 8

OCEAN BEACH
APPROXIMATE ERRORS IN ESTIMATING MEAN PARTICLE SIZE AS FUNCTION OF SAMPLE ALLOCATION FOUR CENTRAL PROFILES

Sample Allocation	Approximate Error	Approx. Relative Error
All samples in all four profiles	1.67 ± 0.08	4.8%
All four profiles, but only one sample at each sampling point	1.67 ± 0.09	5.4%
All four profiles, but only one sample from each stratum	1.67 ± 0.11	6.6%
All samples from only one profile	1.67 ± 0.16	9.6%
Only one sample from only one profile	1.67 ± 0.32	19.2%
A closely spaced pair of samples from only one profile	1.67 ± 0.28	16.8%

Table 8 provides some examples to show how the relative error increases as the number of samples is reduced. The questions that were raised are these: how good an estimate of the population mean is obtained if all four of the profiles are used, but if only one sample is taken at each location instead of a pair of samples? Within the limits of rounding in Table 8, the result shows that the gain is only about 0.6 percent. This indicates that on a cost basis the taking of duplicate samples is not worth the added effort. In contrast to this, however, is the situation when only one sample is to be taken along the profile as an estimate of the mean particle size for the 300-foot beach segment. In this case the single sample is subject to an average error of about 19.2 percent, whereas if two closely spaced samples are collected, the error drops to 16.8 percent, a gain of nearly 3 percent.

The other examples in Table 8 are also interesting. They show that if only one profile with 16 samples is used instead of the four profiles, the relative error is only twice as large as for all profiles (9.6 as against 4.8%). This reduction of error proportionally to the square root of the number of sampled units is in accord with general sampling theory. Table 8 also shows that if only one sample is taken from each stratum (foreshore and backshore) on each profile, as against taking one sample per location, the relative error increases from 5.4 to 6.6 percent.

Extension to Larger Beach Segment. A similar analysis of the three main profiles representing the 1,600-foot beach segment provide the analysis of variance and variance components shown in Table 9. The F tests indicate that there is no significantly greater variability between profiles than there is between strata within a profile; and for the longer beach segment there is no longer a significantly greater variability between strata than between sampling points in a stratum. As in the shorter beach segment, however, there is a highly more significant variability among sampling points in a stratum than there is between the paired samples at each sampling point.

The variance components shown in the lower part of Table 9 again indicate no significant contribution by the profiles, but the contributions of strata, sampling points, and samples increase down the design. This is an example in which the variance "payoff" is greatest at the lowest level, although as before, the actual contribution of this lowest sampling level in any given set of samples depends upon how many times each level is represented in the total design. Table 10, which shows some comparative data in the same manner as Table 8, indicates that all the relative errors are somewhat higher than they were for the 300-foot beach segment, which is to be expected, considering that a beach is usually less homogeneous over a 1,600-foot stretch than in a given

TABLE 9

OCEAN BEACH-ANALYSIS OF VARIANCE OF PHI
MEDIANS FOR THREE MAIN PROFILES (S, I, N)

Source	Sum of Squares	d.f	Mean Square	F
Between profiles	0.1739	2	0.0870	<1 NS
Between strata within profiles	1.5081	3	0.5027	1.81 NS
Between sampling points in strata	4.9959	18	0.2775	3.79 **
Between samples in pairs	1.7599	24	0.0733	
Total	8.4378	47		

Variance Components

Level	Difference	"Samples"	Component
Profiles	-0.4157	16	0.0000
Strata	0.2252	3	0.0281
Sampling points	0.2042	4	0.0510
Samples	0.0733	1	0.0733

TABLE 10

OCEAN BEACH
APPROXIMATE ERRORS IN ESTIMATING MEAN PARTICLE SIZE AS
FUNCTION OF SAMPLE ALLOCATION-THREE MAIN PROFILES

Sample Allocation	Approximate Error	Approximate Relative Error
All samples in all three profiles	1.58 ± 0.09	5.7%
All three profiles, but only one sample at each sampling point	1.58 ± 0.10	6.3%
All three profiles, but only one sample from each stratum	1.58 ± 0.15	9.5%
All samples from only one profile	1.58 ± 0.16	10.1%
Only one sample from only one profile	1.58 ± 0.39	24.7%
A closely spaced pair of samples from only one profile	1.58 ± 0.34	21.5%

300-foot stretch. As before, it is seen that the duplicate samples in the entire design improved the reliability of the grand mean by less than 1 percent, and the error involved in using one profile instead of three is only about 1.8 times as great (10.1 as against 5.7) as in using all three. This factor is nearly the square root of 3, which would be in accord with general sampling theory.

The design at Ocean Beach was laid out with engineering practice in mind. Because of the mixture of systematic and stratified samples, rigorous statistical analysis would be much more complicated than the simplified form of analysis used. Despite these limitations, it is believed that some insight was gained into the effects of using one or more profiles and either single or duplicate samples for sampling a given beach segment. The design was deliberately oversampled by use of sample pairs to provide data on these points.

It is to be emphasized again that the relative errors discussed above are probably underestimated, and that for engineering purposes a more accurate estimate of the errors may show them to be larger than is desirable. If each small carton of sand is considered as an individual in a super-population of all such possible cartons that can be collected in the 300-foot segment, it is seen that even a total of 64 samples is not very large, and the eight samples that might be collected on any one profile is very small.

If some speculation on this point is permitted, one may use the results of the present experiment to "feel out" the error that could occur if the design were laid out as follows: the 300-foot segment is divided into backshore and foreshore, and four samples are collected at random in each stratum. Using the values at hand, the stratified mean would be 1.67 and the variance of this mean, based on the observed variances of the two strata, would be 0.0431. The standard error of the mean is the square root of this, or 0.208. The relative error is 0.208/1.67 = 12.4 percent. If the 95 percent confidence limits are used as is conventional, the relative error for eight samples is 2.36 times this large, or about 29 percent. That is, the 95 percent confidence limits about the mean would be $1.67 \pm (2.36 \times 0.208) = 1.67 \pm 0.49 = 1.18$ to 2.16. In terms of median diameters, then, one may say that there is a probability of about 0.95 that the interval 0.223 to 0.441 millimeter includes the population mean for the 300-foot beach segment. These appear to be rather broad limits for decision on such questions as specifying beach fill, for example. Eight samples are certainly too few.

These remarks apply to samples collected on the exposed parts of beaches. Usual engineering practice includes the collection of a series of underwater samples along the projected profiles. If samples are collected at depths of 0, 6, 12, 18, 24 and 30 feet with respect to low water, each profile would provide six more samples,

which would greatly improve the estimate of the population mean if half a dozen profiles are involved in a study. Some information on this point is provided by the underwater samples collected at Ocean Beach.

Underwater Samples

As stated, engineering practice almost always includes collection of underwater samples, which are required if the nearshore bottom is to be included in the plan for estimating mean particle size on the whole beach. The principles of sampling do not change with this enlargement of the sampling area, inasmuch as the main difference is the inclusion of additional sampling strata in the design. Common practice is to collect underwater samples in terms of depth below low tide rather than in terms of distance from shore. Such samples are normally not equally spaced from the shore line, but from an engineering viewpoint depth control may be more desirable than distance-from-shore control.

Underwater Sampling Strata. Very little data are available from formal studies of natural zones in the nearshore submerged portions of beaches. There is probably a fairly pronounced change toward finer sediments from the foreshore to the shallow submerged zone, followed by less pronounced changes farther out. Knowledge of local conditions is probably the best guide in setting up sampling strata for the near-shore bottom at the present state of knowledge. Either two or three sampling strata seem to be indicated in most instances. An alternative way of setting up the strata is to assign each depth zone arbitrarily to a separate stratum. This introduces a larger number of strata and may require some additional computation, but in some instances it may improve the estimate of the mean.

The three underwater profiles at Ocean Beach were sampled systematically at intervals of 250 feet from the low tide line. This permitted the use of two similarly spaced sampling positions on the exposed part of the profiles, in order to include both the backshore and the foreshore in the estimate. The sampling layout is shown in Figure 6. Duplicate underwater samples were collected with a double-tube sampler having a fixed spacing of about 18 inches between the tubes. The phi medians of the 60 samples involved in the design are listed in Table 11. This table includes one backshore and one foreshore position from each profile at the same 250-foot spacing. The entire design is thus a systematic plan involving pairs of samples a few feet apart at each sampling point, sampling points 250 feet apart along the profiles, and profiles 100 feet apart along the shore.

The nearshore zone had tentatively been divided into three sampling strata, but one pair of the samples, at a distance of 1,250 feet from shore, was very poor, and would normally be discarded. Apparently the bottom deposits there were fine silt with a thin veneer of very coarse sand. One tube apparently picked up a skim of the coarser bed. It
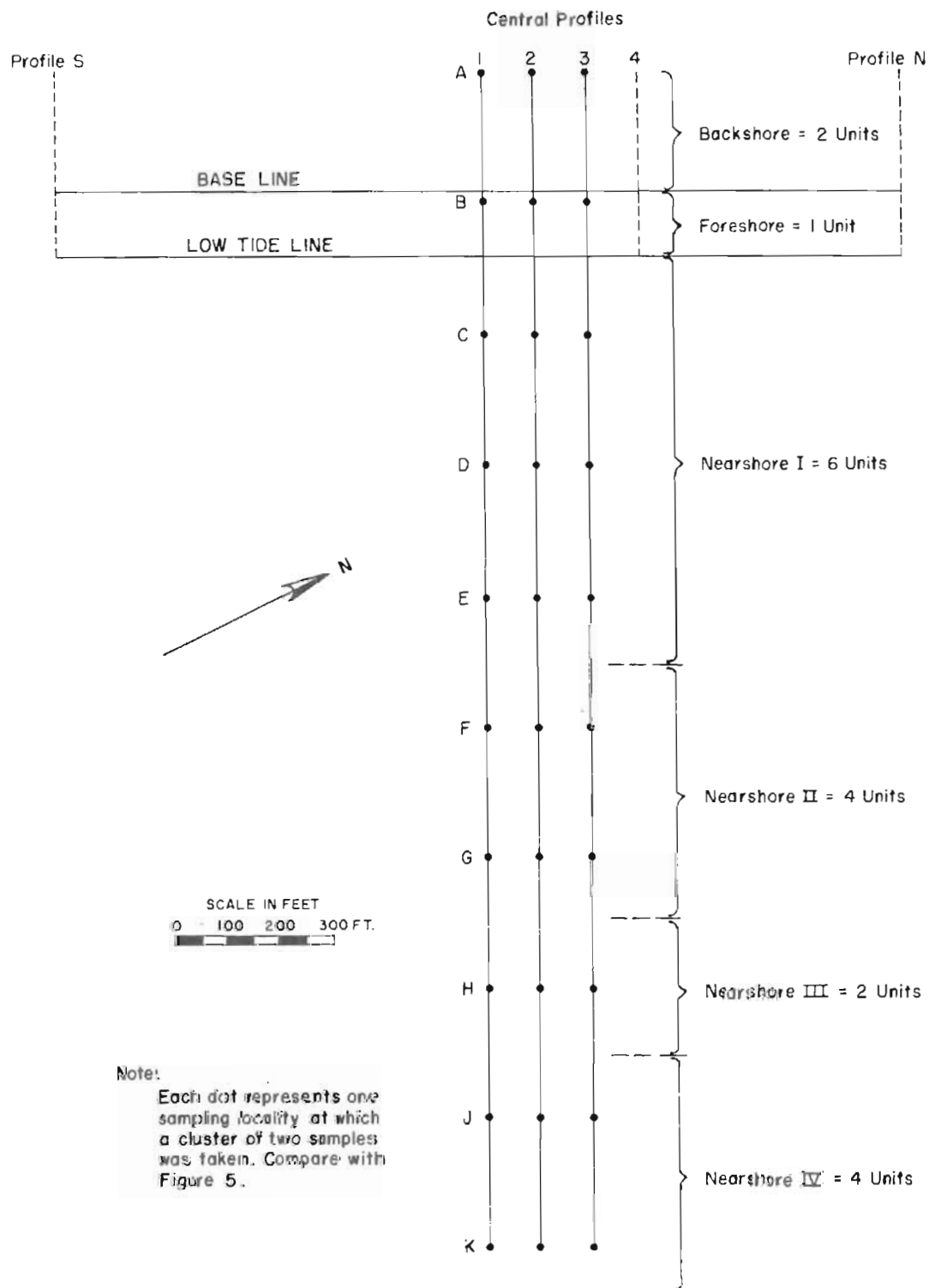
FIGURE 6· UNDERWATER SAMPLING DESIGN AT OCEAN BEACH, MARYLAND

was decided to include the data despite their obvious effect in inflating the variance of the mean. In this way the influence of the poor samples on the grand mean can be shown.

Comparison of Stratified Means. Inasmuch as the sampling design included two closely spaced samples at each sampling locality, it is possible to compare the two mean values obtained by considering the samples in each pair separately. That is, the grand mean can be computed with the first sample of each pair and then with the second sample, to see whether there is any significant difference between them. This was done both with and without the poor samples (starred in Table 11), to determine their effect on the means.

TABLE 11

OCEAN BEACH
PHI MEDIAN DIAMETERS OF UNDERWATER SAMPLES

Position on Profile	Stratum	Relative Stratum Area	Profile 1		Profile 2		Profile 3	
A	Backshore	2	1.80	1.80	1.80	1.80	1.85	1.85
B	Foreshore	1	1.65	1.10	1.30	1.35	0.70	0.80
C	Nearshore I	6	2.70	2.75	2.65	2.60	2.75	2.50
D	"		2.58	2.51	2.69	2.70	2.80	2.70
E	"		2.51	2.65	2.76	2.72	2.77	2.77
F	Nearshore II	4	3.08	3.08	3.08	2.98	3.09	2.94
G	"		3.24	3.09	3.14	3.18	3.14	3.05
H	Nearshore III	2	6.40*	0.55*	3.41	3.41	3.50	3.46
J	Nearshore IV	4	2.39	2.57	2.40	2.39	3.52	3.83
K	"		2.73	2.98	2.43	2.47	2.43	2.33

* Poor sample recovery

In computing the stratified means, each stratum of Table 11 was handled separately for each of the paired sets to obtain the stratum means over the three profiles. These stratum means were then weighted according to the stratum size. Similarly, in computing the standard error of the stratified mean, the variance in each stratum was computed and weighted according to the square of the stratum size. The equations used are given in the Appendix.

By using the first sample of each pair the stratified mean was found to be 2.79, corresponding to an average median diameter of 0.145 millimeter. The second sample of each pair yielded a stratified mean of 2.57, corresponding to 0.168 millimeter. This difference was considerably reduced when the poor samples were excluded. In this case the values

36

were 2.74 and 2.67, corresponding to average median diameters of 0.150 and 0.157 millimeter respectively. This illustrates the effect of a single poor sample in a limited number of observations.

The standard error of the stratified mean for the first samples of each pair was found to be 0.24, corresponding to a relative standard error of 8.6%. For the second set the standard error was 0.25, corresponding to 9.7%. These estimates included the poor samples. Their exclusion reduced the relative error by several percent. By using the standard errors with the poor samples included, a conservative value is obtained which indicates that a set of 30 samples (one from each pair), arranged as 10 samples per profile over six strata, provided an estimate of the grand mean with a relative standard error of less than 10%. Inasmuch as the estimates included samples from the exposed beach, these estimates apply to the entire beach zone from within the backshore to water depths of the order of 30 feet.

Even with inclusion of the poor samples, the estimates based on one sample from each pair were not significantly different, inasmuch as the two-thirds confidence bands about the two means overlap. For the first set the confidence limits were found to be 2.79 ± 0.24 = 2.55 to 3.03; and from the second set they were 2.57 ± 0.25 = 2.32 to 2.82. The range from 2.55 to 2.82 is in common, and includes both computed means.

It is judged from this experiment that collecting and analyzing duplicate samples at each sampling locality is not worth the added effort. The magnitude of the relative error (which here included the additional hazard of picking up occasional poor underwater samples) suggests that additional profiles be included in a beach study in order to keep the relative error safely under 10%. The practical implication is that for any given segment of beach being studied, a number of profiles commensurate with the scale of the study be employed, with a minimum set at four or preferably six. This point is mentioned again in the following section.

SUMMARY REMARKS ON ESTIMATION OF MEAN PARTICLE SIZE

The sampling experiments described here lead to the general conclusion that stratified sampling, with computation of a weighted mean, yields a more reliable estimate of the population mean particle size than does an unweighted mean. As far as generalizations may be drawn from the experiments, they seem to point in the following direction:

1) Estimates of mean particle size of beaches should include contributions from samples taken from each of the natural zones on the beach. This implies that the profile should extend from well within the backshore (if not completely across it) to a point offshore at some fixed depth, as

30 feet. For some purposes it may be desirable to extend the landward end of the profile into the dune belt.

The boundaries of the natural beach zones should be indicated on the profiles. On the exposed beach these are indicated by the seaward edge of the foredune, by the bevel at the major berm near the high tide line, and by the low water datum line. The nearshore bottom can normally be divided into from two to four natural zones, depending upon changes in bottom slope, presence or absence of bars, and perhaps major changes in texture. The width of each zone can be indicated on the profile as a weighting factor for later use.

2) Allocation of samples along the profile can in general be handled in three ways. In the first, the number of samples per stratum is proportional to the stratum width. This was the method followed in the Ocean Beach design, where the uniform spacing of 50 feet between sampling points gave a number of samples roughly proportional to stratum width. More formally, the number of samples per stratum can be made directly proportional to width, starting with one or more samples in the narrowest stratum. This is a form of proportional sampling, which assure that each stratum makes a contribution to the weighted mean proportional to its relative magnitude.

A second way of allocating samples is to have an equal number of samples from each stratum, regardless of relative width. Thus, in its simplest form, one sample could be taken from the center of each stratum. This assures that each stratum is included in the final estimate, but it takes no account of the relative stratum weighting in the sampling plan. In one of the Illinois Beach experiments four random samples were taken from each beach stratum, which is an example of this method.

The third method of allocation distributes the samples over the strata in proportion to the relative variability in each stratum. If the foreshore is four times as variable as the backshore, for example, this would assign four times as many samples to the foreshore, regardless of stratum width. This method requires knowledge of the variability in each stratum, which is usually not available before samples are taken, unless data from a previous survey can be used. As illustrations of this sort of allocation, a sampling of the backshore and foreshore at Illinois Beach (omitting the dune belt) would assign 4.5 samples in the foreshore for each backshore sample. The equations for such allocations are given in the Appendix. Experience thus far available suggests that the foreshore variability is commonly from four to ten times as great as the backshore variability, as measured by the stratum variances.

3) The present experiments suggest that more than one profile should be sampled for any given stretch of beach under study, and that each profile should be sampled at least once in each natural beach zone. This would supply the minimum number of samples deemed adequate for even approximate estimates of mean particle size. If six profiles are used, each with say five sampling strata, the minimum sampling plan would embrace 30 samples. It is probably fair to say that such a set would provide an estimate of the population mean with a relative error of about 10 percent at the 95% confidence level in most cases. To cut this error in half would require four samples from each stratum on each profile, although application of the principles of optimum allocation may permit nearly this amount of improvement with something less than four times the same number of samples.

4) Computation of the weighted mean particle size is probably preferably done with either the log median (phi median) or the log mean (phi mean) inasmuch as they are both normally distributed. There may be some advantages to using the phi mean in subsequent operations, inasmuch as the final result is an estimate of the phi mean of the population, whereas the final result in using the phi median is the arithmetic mean of the phi medians. Many beach particle populations are lognormally distributed, so that the differences in final results are commonly slight.

When several profiles are used in the sampling, the mean particle size can be computed for all samples in one operation by combining the samples from each profile according to their position in the several strata. If desired, separate means may be computed for each stratum over all the profiles, or for each profile over its strata. Such supplementary data are useful in evaluating differences between strata across the beach, or between profiles along the beach.

5) It seems desirable to express the estimated mean particle size in terms of some standard confidence limits. The 95 percent confidence limits are conventionally used, although for some purposes the 67 percent confidence limits, based directly on the standard error, may be suitable. Use of the probable error, which represents the 50 percent confidence limits, is usually regarded as being less discriminatory than seems desirable. The probable error is computed by taking 0.6745 times the standard error.

Use of confidence limits is helpful in two ways. In the first place the limits indicate how much reliability may be placed in the estimate. This information is of value in indicating how narrowly the mean has been pinned down for

later use in design specification. In addition, the habit
of computing confidence limits helps increase awareness
of the general framework of precision within which conven-
tional beach sampling is conducted. Confidence limits are
discussed in the Appendix.

6) The tentative proposals for beach sampling presented in Beach
   Erosion Board Technical Memorandum No. 50 are in general
   supported by the added information from the present experi-
   ments. The remarks regarding sample size and depth, the
   need for randomization processes in sampling, the use of
   more than one profile in beach sampling, and the spacing of
   profiles at some multiple of the profile length, seem all
   to be verified by the present study.

7) The question of the specific number of samples required for
   any given beach study is one that still needs considerable
   qualification. As the present experiments show, the variance
   of the mean, shown as $V(\bar{X}_p)$ in Tables 2 and 4, depends in
   part on the sampling design (simple random, systematic,
   stratified, etc.) as well as on the total number of samples
   collected. Beach sampling almost always has an element of
   systematic sampling in it, inasmuch as the basic sampling
   reference line is the profile. Profiles as a rule are
   spaced a fixed distance apart along the beach, and this
   distance may vary for different beaches. It was also seen
   to be desirable that the several natural beach   zones be
   explicitly represented in the sampling. From these two
   features it would appear that beach sampling in the future
   may tend to emphasize some form of combined systematic-
   stratified sampling design. In such designs the systematic
   element of profile spacing may introduce some component of
   variability, and other components will be introduced by each
   sampling stratum represented by the natural beach zones.
   Although the analysis of the Ocean Beach profiles suggested
   no between-profile variance component at either the 100-foot
   or 800-foot spacings (Tables 7 and 9), it is probably not
   safe to conclude that this is a general rule.

   The interplay of these components of variability, as re-
flected by the allocation of samples in the design, results
in different levels of reliability of the means when different
numbers of samples are distributed in different ways over
the beach zones in a set of profiles. Until sufficient data
are assembled and analyzed to provide information on the
average values of these components on a variety of beaches,
sample planning cannot be undertaken on a completely
quantitative basis.

The writers are of the opinion that there are at present enough beach data in the literature and among beach engineers and geologists to provide a basis for estimating the orders of magnitude of the variances associated with particle sizes in dunes, backshores, foreshores, and one or more nearshore bottom zones. As this knowledge is organized it will become increasingly possible to set up engineering specifications in terms of the expected reliability of estimated means of beach properties. By use of cost functions it should also be possible to state specifically what sort of sample allocation within a proposed design will give satisfactory results for minimum cost.

8) In the light of the foregoing remarks it would appear that the recommendation mentioned earlier (item 3 in this list) with respect to beach sampling provides a pragmatic solution to beach sampling in the present state of knowledge. This recommendation may result in some over-sampling, but it appears to err on the conservative side:

Six profiles extending from within the backshore to a depth of 30 feet, with five sampling strata (say the backshore, the foreshore, and three nearshore bottom zones), and with one sample per stratum per profile, would provide a set of 30 samples. It is believed that such a set would estimate the population mean particle size with a relative error of about 10 percent at the 95 percent confidence level.

This pragmatic suggestion is based on the findings at Illinois Beach which indicated that the set of 16 stratified samples (design 2F) gave a relative error of $0.15/1.40 = 10.7$ percent. Considering the heterogeneity of the deposits, it would seem that in more homogeneous instances a set of 30 samples taken as suggested above, should generally give results within the 10 percent restriction.

CONCLUDING REMARKS

This report emphasizes the estimation of mean particle size on beaches because of the increasing importance of that aspect of beach studies in the design and specification of beach fill. The material presented here will be drawn upon in a report being prepared for such specifications. The other two features of beach sampling, related to studies of gradients in beach populations, and to studies of variability within the several beach zones, were touched upon in the Illinois Beach design. These features are important enough to deserve specific treatment, although space limitations prevent such additional development here.

As the senior author's experience with beach problems grows, it becomes increasingly evident that there is no unique answer to the problem of the number of samples needed for beach analysis. Nevertheless, it is apparent that certain underlying principles of beach sampling have emerged, and the most important of these has been recognition of the need for formal planning of sampling designs. Such planning will be facilitated as better knowledge of the natural variations in beach zones becomes available. With such information it will be possible to enlist the aid of mathematical statisticians in the development of general sampling plans that should have wide applicability when estimates of the expected variations are used in the designs. Moreover, more rigorous methods for computing the variances in mixed sampling designs can be made available to engineers and geologists by such active collaboration of statisticians. The resulting improved quantitative estimates of beach characteristics will in turn help sharpen specifications needed in some branches of beach engineering design.

It may be remarked that all the examples given in this report are based on average particle diameters, and nothing was said about the sorting coefficients of the sand. The degree of sorting varies widely from place to place on some beaches, and it is possible to apply statistical methods to the estimation of an average degree of sorting. These methods do not differ in principle from those used with the median or log mean diameters, except that transformations of the observed values may be required to facilitate statistical analysis. This aspect of beach studies is being examined and will be treated explicitly in subsequent reports.

REFERENCES

Anderson, R. L., and Bancroft, T. A., Statistical Theory in Research: McGraw-Hill Book Co., New York, 1952.

Bascom, W. N., The relationship between sand size and beach-face slope: Trans. Am. Geophys. Union, Vol. 32, pp. 866-874, 1951.

Bennett, C. A., and Franklin, N. L., Statistical analysis in chemistry and the chemical industry: John Wiley & Sons, New York, 1954.

Cochran, W. G., Sampling Techniques: John Wiley & Sons, New York, 1953.

Davies, O. L., Design and analysis of industrial experiments: New York, Hafner Pub. Co., 1954.

DeLury, D. B., Values and integrals of the orthogonal polynomials up to n = 26: Univ. Toronto Press, Toronto, 1950.

Dixon, W. J., and Massey, F. J., Introduction to statistical analysis: McGraw-Hill Book Co., New York, 1951.

Fisher, R. O., Some characteristics of bottom sediments along the Illinois shore line of Lake Michigan: Proc. Fourth Conf. on Coastal Engineering, Berkeley, 1954, pp. 107-118.

Krumbein, W. C., and Pettijohn, F. J., Manual of Sedimentary Petrography, New York, Appleton-Century Book Co., 1938.

Krumbein, W. C. and Miller, R. L., Design of experiments for statistical analysis of geological data: Jour. Geology, Vol. 61, pp. 510-532, 1953.

Krumbein, W. C., Regional and local components in facies maps: Bull. Amer. Assoc. Petroleum Geologists, Vol. 40, September 1956.

Miller, R. L., A model for the analysis of environments of sedimentation: Jour. Geology, Vol. 62, pp. 108-113, 1954.

Potter, P. E. and Siever, R., A comparative study of Upper Chester and Lower Pennsylvania stratigraphic variability: Jour. Geol., Vol. 63, pp. 429-451, 1955.

APPENDIX

INTRODUCTION

The preceding text omitted technical details of the statistical methods used in the sampling designs, in order that the main theme of the report could be emphasized. This appendix is added to make more explicit the methods of computation used with various kinds of sampling designs. The treatment is expository rather than theoretical, and it is assumed that the reader interested in further detail will refer to the authors cited.

Three topics are emphasized in this Appendix. The first is concerned with the computation of means and variances for each type of sampling design. The second is concerned with computing confidence limits for the population mean, and the third topic concerns the application of cost functions to the optimum allocation of samples within a given design.

COMPUTATION OF MEANS AND VARIANCES

Simple Random Sampling

The writers follow Cochran (1953) in most of their definitions and usages of sampling terms. Basic sampling theory is related to purely random samples, and for such a design the sample mean ($\bar{X}$) and sample variance (S^2) are defined as follows (Cochran 1953, chapter 2; Dixon and Massey 1951, p. 19):

$$\bar{X} = \sum X_i / N \qquad \dots \dots \dots \qquad (1)$$

$$S^2 = \sum (X_i - \bar{X})^2 / (N-1) \quad \dots \dots \qquad (2)$$

where X_i is an individual random observation and N is the total number of observations.

Computation of (2) is most conveniently performed by using an algebraic equivalent as follows:

$$S^2 = \frac{\sum X_i^2 - (\sum X_i)^2 / N}{N - 1} \qquad \dots \dots \qquad (3)$$

The variance expressed in this manner is the sample variance, and it may be converted to the variance of the mean, $V(\bar{X}_p)$, by dividing by the number of samples. Thus,

$$V(\bar{X}_p) = S^2 / N \quad \dots \dots \qquad (4)$$

The square root of this is the standard error of the mean, $S(\bar{X}_p)$. The subscript p has been used here to indicate that the grand mean is intended, as it is used in text Tables 2 and 4. This helps distinguish it from other means as used below, although normally no subscript is needed.

Use of the above definitions for the sample observations of the phi mean particle size as given in column 1 of Table 1 for the set of 16 simple random samples collected at the points shown in Figure 2A, yields the following estimates for the beach population:

$$\bar{X}_p = 1.06 \qquad s^2 = 1.1584 \qquad V(\bar{X}_p) = 0.0724 \qquad S(\bar{X}_p) = 0.27$$

These values are shown in line 1 of Table 2. Similar computations for the observations on percent of acid solubles given in column 1 of Table 3 yield the estimates shown in line 1 of Table 4.

Stratified Random Sampling

In this case there are k strata with n samples each, so that N = kn. The grand mean is computed as a weighted mean involving the individual strata means each weighted according to the area of the stratum (Cochran, Chapter 5, p. 67 ff.):

$$\bar{X}_p = (1/A) \ (a_1\bar{X}_1 \ + \ a_2\bar{X}_2 \ + \ \ldots \ + a_k\bar{X}_k) \qquad (5)$$

where a_1, a_2, ... a_k are the areas of the individual strata, A is the total sample area, and is equal to the sum of the a's. The individual stratum means are indicated as $\bar{X}_1$, $\bar{X}_2$, ... $\bar{X}_k$.

The variance of the grand mean is also weighted but the weighting is proportional to the square of the stratum areas;

$$V(\bar{X}_p) = (1/A^2) \left[a_1^2 \ V(\bar{X}_1) \ + \ a_2^2 \ V(\bar{X}_2) + \ldots a_k^2 V(\bar{X}_k) \right] \ldots (6)$$

In this case $V(\bar{X}_1)$, $V(\bar{X}_2)$, ... $V(\bar{X}_k)$ are the variances of the individual stratum means, computed in accordance with equations (3) and (4). It is to be noted that (6) is the variance of the stratified mean computed in accordance with (5).

In the Illinois Beach stratified sampling plan there were four strata each with four samples, so that k = 4 and n = 4. The relative areas of the strata shown in Figure 2F were expressed as the number of unit sampling cells in each stratum, so that A = 16, and

a_1 (dunes) = 6.89 units

a_2 (backshore I) = 1.15 units

a_3 (backshore II) = 5.51 units

a_4 (foreshore) = 2.45 units

The values of the sample observations in the stratified set, shown in Figure 2F, are given in column 6 of Tables 1 and 3, and the computed population parameters are tabulated on line 6 of Tables 2 and 4.

The Illinois Beach samples were collected at random in each stratum independently, as is called for by theory, whereas the Ocean Beach underwater stratified samples were taken systematically. Equation (6) was used for the latter samples as though they had been collected at random.

Cluster Sampling

In this design there are k clusters of n samples each. The grand mean, $\bar{\bar{X}}_p$, is computed as in the case for simple random samples (equation 1). The double bar above the X indicates that there are really two levels of sampling in this design. The total variance can be separated into two parts. The first is contained in the between-cluster mean square, S_b^2, and the second is estimated by the within-cluster mean square S_w^2. The between-cluster mean square is (Cochran, p. 219):

$$S_b^2 = n \sum_i^k (\bar{X}_c - \bar{\bar{X}}_p)^2 / (k - 1) \ldots \ldots \qquad (7)$$

where $\bar{X}_c$ is the individual cluster mean. The within-cluster mean square is:

$$S_w^2 = \sum_i^k \sum_i^n (X - \bar{X}_c)^2 / k(n - 1) \ldots \ldots \qquad (8)$$

where X is a single observation. Equation (8) is equivalent to the average of the k within-cluster variances.

The variance of the grand mean can be had directly from the relation:

$$V(\bar{\bar{X}}_p) = S_b^2 / kn \ldots \ldots \qquad (9)$$

as shown in Cochran, p. 225. It can also be shown that the between-cluster mean square estimates the sum of the between-cluster variance, S_u^2, and the within-cluster variance, S_w^2, (Cochran, p. 219). Thus, for estimation of variance components, S_u^2 is found from the relation:

$$S_u^2 = (S_b^2 - S_w^2)/n \tag{10}$$

so that the variance of the grand mean in equation (9) can also be shown as :

$$V(\bar{\bar{X}}_p) = (S_u^2/k) + (S_w^2/kn) \tag{11}$$

This form is useful when the effect of other combinations of k and n are to be studied, as mentioned in the text. It was illustrated for three sampling levels in Beach Erosion Board Technical Memorandum No. 50, p. 24.

The Illinois Beach cluster samples in Figure 2D had k = 4 and n = 4. The between-clusters mean square was 4.2172, from which the variance of the grand mean was found to be 4.2172/16 = 0.2636 as shown on line 4 of Table 2. The variance components were estimated as $S_u^2 = 0.9950$ and $S_w^2 = 0.2373$.

Multilevel (Nested) Sampling

This is a more general case that includes cluster sampling as a two-level design. A three-level design is illustrated in the Illinois Beach experiment in Figure 2E. In this case k major groups were chosen at the top level, m subgroups were taken within each major group, and n samples were taken within each subgroup. There is a mean for each level, $\bar{X}$, $\bar{\bar{X}}$, and $\bar{\bar{X}}_p$, the last being the grand mean. The total variability can be separated into three parts, each represented by a mean square. The between-major-units mean square is (Cochran, p. 230):

$$S_b^2 = mn \sum_i^k (\bar{\bar{X}} - \bar{\bar{X}}_p)^2/(k - 1) \tag{12}$$

The between-subgroups mean square is:

$$S_w^2 = n \sum_i^k \sum_i^m (\bar{X} - \bar{\bar{X}})^2/k(n-1) \tag{13}$$

and the between subsubunits (samples) is:

$$S_{ww}^2 = \sum_i^k \sum_i^m \sum_i^n (X - \bar{X})^2/km(n - 1) \tag{14}$$

The grand mean is computed as in equation (1), where N = kmn, and the variance of the grand mean is (Cochran, p. 230):

$$V(\bar{\bar{X}}_p) = S_b^2/kmn \tag{15}$$

The three mean squares provide estimates of the variance components at each sampling level:

$$S_u^2 = (S_b^2 - S_w^2)/mn$$

$$S_s{}^2 = (S_w{}^2 - S_{ww}{}^2)/n \qquad\qquad (16)$$

$$S_{ww}{}^2 = S_{ww}{}^2$$

where $S_u{}^2$ is an estimate of the variance component at the top level, $S_s{}^2$ is an estimate of the variance component at the subgroup level, and $S_{ww}{}^2$ is itself an estimate of the variance component at the lowest level. The variance of the grand mean may also be expressed in terms of these components as (Cochran, p. 229):

$$V(\bar{\bar{X}}_p) = (S_u{}^2/k) + (S_s{}^2/km) + (S_{ww}{}^2/kmn) \quad \ldots \qquad (17)$$

For the three-level design of Figure 2E, $S_b{}^2$ was found to be 3.5918, which would yield a variance of the grand mean of 0.2245 by equation (15). However, when the number of samples selected at any level in the design exceeds 1/20 of all such elements in the population, a finite population correction is applied at that level (Cochran, pp. 17; 220). In the present instance all four of the beach strips were sampled so that a correction factor was applied at this level by use of equation (17). The effect of this correction was to change the unadjusted variance of 0.2245 to 0.1684. The latter value is shown in Table 2, line 5.

Systematic Sampling

For this type of sampling the mean is computed as shown in equation (1), but there seems to be no generally satisfactory method for computing the variance of a single systematic sample. Cochran (1953, chapter 8) discusses systematic sampling in detail and points out that a systematic sample may be considered as a particular case of cluster sampling in which the systematic set is one cluster. Several equations are provided for estimating the variance of the mean, based on k systematic samples of n observations each. The methods in part express the variance of the systematic sample in terms of the variance of a single random sample.

The Illinois Beach design had 16 samples representing one systematic set, so that a compromise method of computation was used, by considering the set as four systematic samples of four items each. The computed variance differed in value depending on the sample combination selected, but tended to lie between that for cluster sampling and for simple random sampling. The value entered in lines 3 of Tables 2 and 4 is believed to represent at least the proper order of magnitude. If the population mean M is known, the variance of the systematic mean $V(\bar{X}_{sy})$ is given by Cochran (p. 163) as:

$$V(\bar{X}_{sy}) = (1/k) \sum_i^k (\bar{X} - M)^2 \quad \ldots \qquad (18)$$

where $\bar{X}$ is the mean of the individual systematic samples.

Systematic sampling is important in most beach studies, and Cochran cites the need for additional research in systematic sampling (p. 168)

Random in Cells

Like the systematic samples the use of sampling cells with a sample randomized in each presents a problem of computing the variance. The mean, however, is computed in the usual way (equation 1). Inasmuch as random in cells allows some tendencies toward clustering, it was felt appropriate for illustrative purposes to consider the samples as a random set. Hence equations (3) and (4) were used directly to compute the estimates of the population parameters shown in line 2 of Tables 2 and 4 from the data tabulated in column 2 of Tables 1 and 3.

Purposive Selection

Cochran points out (p. 7) that sampling theory does not apply to purposive selection of the units inasmuch as it contains no element of random selection. However, inasmuch as the samples selected by Dr. Dapples for this purpose are part of one or another of the other designs, they were treated as simple random samples in order to obtain some estimates of the population parameters for illustrative purposes. Consequently, equations (1-4) were used to obtain the values given on line 8 of Tables 2 and 4 using the observations tabulated in column 7 of Tables 1 and 3.

Mixed Designs

Some of the examples cited in the text, especially those at Ocean Beach, are not standard sampling designs inasmuch as they represent mixtures of several kinds of sampling. Theory for such mixed designs can presumably be developed by mathematical statisticians, but in order to avoid complexities in exposition, compromise methods of computation were used to gain at least qualitative estimates of the variability present. For example, the four closely spaced profiles at Ocean Beach represent basically a systematic set of samples distributed over natural beach strata, and involving clusters of two individual samples at each sampling point. Obviously the rigorous computation of variances in such a design requires the guidance of a mathematical statistician. For present purposes, however, it was felt that the methods used, which involved treatment by multilevel techniques, (see Tables 7 and 9), including use of average degrees of freedom at some levels, gave some insight into the problem of sample allocation, inasmuch as all such sample allocations assumed use of the same basic design.

CONFIDENCE LIMITS ON MEANS

Confidence limits for the true population mean are:

$$\bar{X}_p - t_{\frac{1}{2}\alpha} S(\bar{X}_p) \quad \text{and} \quad \bar{X}_p + t_{\frac{1}{2}\alpha} S(\bar{X}_p) \dots \dots \quad (19)$$

where $t_{\frac{1}{2}\alpha}$ is the value of the normal deviate corresponding to the desired confidence probability α. Its value may be found in tables of t for any specified α and for any given total number of samples (see Table 5, Dixon and Massey, 1951, p. 307).

The 95% confidence band for the true population mean is determined for the case of the simple random set (Figure 2A) in the following way:

$t_{\frac{1}{2}\alpha}$ for α = 95% and for N-1 = 15 degrees of freedom = 2.13

Confidence limits are therefore

1.06 - (2.13)(0.27) to 1.06 + (2.13)(0.27) = 0.49 to 1.63 as shown on line 1 of Table 2.

APPLICATION OF COST FUNCTION FOR OPTIMUM ALLOCATION OF SAMPLES

One advantage of multilevel sampling is that it provides data for optimum allocation of the number of samples to be taken at each level of the design. This may be especially important in redesign of a preliminary sampling plan for subsequent more detailed analysis of a sampling unit. The optimum sampling and subsampling fractions are found by applications of cost functions to combinations of the several variance components.

Cochran (1953, p. 225) describes the application of a cost function in a two-level nested design and Potter and Olson (1954) discuss its application to additional sampling levels. In applying the cost function, the cost (or time) required to make an observation is so distributed over the levels of the design that the variance of the grand mean is a minimum.

Application of a cost function to two levels can be handled by setting up the cost function as follows:

$$C = C_c k + C_s kn \qquad (20)$$

where C, the total cost, is made up of a component C_c that is the cost (or time) required to select and locate the top level sampling positions which in this case are the cluster centers; and a component C_s which is the time required to collect and analyze an individual sample in the cluster.

Cochran (1953, p. 226) shows the process by which the variance is minimized and derives the solution for the optimum number, n', of samples per cluster:

$$n' = \sqrt{(S_w^2/S_u^2)(C_c/C_s)} \qquad (21)$$

A-7

Assume that in a particular case it costs twice as much to arrive at and stake out the cluster positions as to analyze a single sample. Using the values for S_u^2 and S_w^2 for the Illinois Beach cluster samples would then give:

$$n' = \sqrt{(0.2372/0.9950)\,(2/1)} = 0.69$$

The nearest whole number to this is 1, which would imply that if 16 samples are to be collected it is more efficient to collect one at each of 16 random locations than to collect four clusters of four. It is to be emphasized that this is not a general result but that the optimum allocation can differ depending upon the relative variabilities as well as upon the relative costs at the sampling levels. In applied work it is common practice to add a constant cost in the original function which relates to overhead, perhaps transportation to and from the beach, and similar items.

SAMPLE ALLOCATION OVER STRATA

It was mentioned in the text that one way of allocating samples to the strata could be based on the relative variability within each stratum. By this type of allocation it is possible to collect such numbers of samples from each stratum that the contribution made by the stratum variability to the variance of the grand mean is the same for all strata.

For two strata whose variances are S_a^2 and S_b^2 respectively, the number of samples (n') from the more variable stratum per unit sample in the less variable stratum is proportional to the ratio of the variances. Thus if $S_a^2 > S_b^2$,

$$n' = S_a^2 / S_b^2 \tag{22}$$

As an example the foreshore variance at Illinois Beach was estimated as 2.1706 and the estimate of the backshore variance was 0.4782. Hence, for each backshore sample it would be necessary to collect (2.1706)/(0.4782) - 4.5 foreshore samples. Potter and Siever (1955) discuss the theory and apply it to more than two levels of sampling. In contrast to this type of allocation is one that relates the sample size in a stratum to the product of the stratum size and the stratum standard deviation. This is discussed by Cochran on p. 74.

CLOSING REMARKS

The material presented in this Appendix is useful mainly as a guide to further reading in standard statistics reference books. It is likely that some terms were used relatively loosely, and that some applications of variance equations to particular sampling designs need more rigorous

examination. It was emphasized earlier that the collaboration of mathematical statisticians will be needed in many specific instances, especially for mixed sampling designs.

The experiments reported here are all based on limited numbers of samples, and hence the means and variances that were used to test relative efficiencies of the designs are themselves subject to sampling fluctuations. It is believed, however, that the generalizations regarding the value of stratified sampling for estimating population means are valid, inasmuch as the experiments support theory in the sampling of populations with gradients. It is hoped that this Appendix will provide at least a start toward expressing beach sampling problems in the framework of statistical design.

9622 047