A STATISTICAL ANALYSIS

OF

SAND GRAIN SIZE

IN

SAN SALVADOR, BAHAMAS

by

Robert B. Gardner

Bahamian Field Station
San Salvador, Bahamas

Occasional Paper 1993, No. 1

A STATISTICAL ANALYSIS OF SAND GRAIN SIZE IN SAN SALVADOR, BAHAMAS

Robert B. Gardner
Department of Mathematics
Louisiana State University in Shreveport
Shreveport, Louisiana 71115

Abstract. We introduce two statistical tests which can be applied to sand grain size data. The first test allows the comparison of mean grain size of two samples. This test is applied to two samples collected at Sandy Point on San Salvador, Bahamas. The second test allows the comparison of average $\Phi 50$ values between two sites where several samples have been collected at each site. The shoreline of San Salvador is broken into four compass directions and the sides of the island are compared for differences in grain size. This test lends statistical significance to certain claims which have already appeared.

INTRODUCTION

We address two topics in this paper. The first provides a statistical test for comparing the means of grain size of two sand samples, given the grain size distribution of each. We run the test on samples taken on San Salvador Island, Bahamas in March of 1989. The second topic is a statistical analysis of data collected by Lee *et al* [1986] on the island. We break the shoreline of the island into four compass directions and perform statistical comparisons to test for significant differences in $\Phi 50$ values.

San Salvador Island (also known as Watling Island), Bahamas is located at 24°00′ N latitude and 74°30′ W longitude. It is one of the outermost of a chain of approximately 700 islands. It is famous for being the first landing site in the New World of Christopher Columbus in 1492.

PROCEDURE

Samples were taken at six locations on the island (primarily on the western side). See Figure 1 for the specific locations. Each sample was taken near the high tide mark, with the exception of one sample taken from Sandy Point at about 75 meters from the shore (the "dune" sample), to be used in comparison with the Sandy Point sample taken from nearer the water (the "shore" sample). Sample size was about 100 to 150 grams in size. They were air dried for five days and ran through a sample splitter three times to obtain a random sample approximately 50 grams in size. The samples were then placed in a Cenco-Meinzer sieve shaker. Each sample required two runs, first with screens of size 0, 0.5, 1.0, 1.5, and 2.0 Φ values, and then that collected in the sieve pan was ran with screens of size 2.5, 3.0, 3.5, and 4.0 Φ. Each run was 15 minutes. The results were weighed on a Mettler PE 3600 electronic scale to 0.01 gram accuracy. Figure 2 gives a histogram for each sample and Table 1 lists the percentage distributions in the nine Φ classes.

ANALYSIS OF DATA AND COMPARISON OF SANDY POINT SAMPLES

We begin with a conventional analysis of the seven samples collected by the author. First we calculate the Inclusive Graphic Skewness (Sk_I) by the following formula [Folk 1974]:

$$Sk_I = \frac{\Phi 16 + \Phi 84 - 2\Phi 50}{2(\Phi 84 - \Phi 16)} + \frac{\Phi 5 + \Phi 95 - 2\Phi 50}{2(\Phi 95 - \Phi 5)}.$$

This measure includes 90 percent of the curve, as oppossed to other measures of skewness which consider less of the curve. As can be seen from Table 2, following Folk's "verbal" guidelines, the Rice Bay, Snow Bay, and Telephone Pole Reef samples are near symmetrical in distribution. The remaining samples are fine-skewed, with the Grotto Beach sample being strongly fine-skewed.

We also compute the Graphic Kurtosis by [Folk 1974]:

$$K_G = \frac{\Phi 95 - \Phi 5}{2.44(\Phi 75 - \Phi 25)}.$$

We see (Table 2) that all but one sample has K_G roughly between $0.90 - 1.11$, indicating (following Folk's guidelines) that the distributions are mesokurtic (meaning not excessively peaked or flattened). The Grotto Beach sample does have $K_G = 1.31$ indicating a leptokurtic (excessively peaked) distribution.

We now wish to compare the mean Φ values for the "dune" and "shore" Sandy Point samples. For this, we take a novel approach. First, instead of dealing with the $\Phi 50$ as a mean, we will compute a weighted arithmetic mean (Folk calls this the "phi arithmetic mean") for each sample as follows:

$$\overline{y} = \frac{\sum \Phi_{i+1/2} m_i}{\sum m_i}$$

where m_i is the percent of the sample in the Φ_i class and $\Phi_{i+1/2} = \dfrac{\Phi_i + \Phi_{i+1}}{2}$, that is $\Phi_{i+1/2}$ is the midpoint Φ value for the class. The m_i's can be thought of as masses, and since we are using percents, $\sum m_i = 100$ (this is similar to Folk's "Method of Moments"). Performing this computation we get for the dune sample $\overline{y_1} = 1.00$ and for the shore sample $\overline{y_2} = 0.73$. Now we wish to test the null hypothesis H_0: "There is no significant difference in the means." To run a statistical test, we cannot have only two samples. Since each sample consists of thousands of sand grains, it seems reasonable to us to treat, say, each percent as a single sample and therefore have 100 "samples" from each sight. We shall perform a t test to check H_0. For the conventional t test, there are three assumptions. The first is that the samples are independent. This is no problem. The second is that the distributions are normal. We are dealing with Φ values so this should not be a problem. The third is that variances are the same. We must decide whether or not this is the case. For this, we need to compute the sample variances using [Ott 1988, p. 287]: $s^2 = \dfrac{\sum(y - \overline{y})^2}{n-1}$. Since each sample consists of 9 subsamples (one for each Φ class) we will use the y value from each class and weigh $(y - \overline{y})^2$ according to the percentage of the sample in that class. We get for the dune sample $s_1^2 = 0.2529$ and for the shore sample $s_2^2 = 0.1795$. The test statistic for our test of the null hypothesis of "No significant difference in variances" is [Ott 1988, p.293]: $F = \frac{s_1^2}{s_2^2}$. We get $F = 1.4089$. We have $df_1 = df_2 = 99$. At $\alpha = 0.05$ the F statistic is 1.39. So we reject the null hypothesis; that is, there seems to be a difference in the variances. So we will use the "approximate t test for independent samples with unequal variances" [Ott 1988, p.175]. We have the test statistic

$$t' = \frac{\overline{y_1} - \overline{y_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

where

$$df = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)c^2 + (1 - c)^2(n_1 - 1)}$$

and

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}.$$

For our data, we find $t' = 1.6573$ and $df = 192$. The F statistic (with $\alpha = 0.05$) is then 1.654 and we reject our original null hypothesis that the means of the Φ's are the same. So the dune sample seems to have a lower $\Phi 50$ value and hence a larger average diameter. Interestingly, we can reach this conclusion at the $\alpha \geq 0.05$ level of confidence, but not at a higher level of confidence.

This approach provides a simple way to compare the means of two samples. In fact, by using this particular t test, we need not even assume that the variances are the same.

ANOVA OF GRAIN SIZE SAMPLES

We now turn our attention to the paper of Lee *et al* [1986]. They collected 26 samples from various locations along the shore of the island. See Figure 1 and Table 3. They state that "Beaches of the east side of San Salvador Island can be characterized generally as being fine to smaller medium sized grains that are moderately well sorted... Beaches on the south side generally are larger medium to coarse grained sands that range from very well sorted to poorly sorted... Beaches on the west side of the Island are dominantly coarse grains but range from fine to very coarse in size... Beaches on the north side are medium to coarse grained, only moderatley well to poorly sorted. Thus, each side of San Salvador Island seems to show reasonably distinct and consistent beach sand characteristics." This claim is not backed up by statistical tests, although this seems to be the trend by "eyeing" the data. We will now break the shore of the island into four parts and perform the relevant tests.

The Eastern Part will include samples 1 – 6, the Southern will include 7 – 13, the Western 14 – 20, and the Northern 21 – 26. Keeping with tradition, we shall deal with the Φ values as opposed to the millimeter measurements. We approach the problem by considering each $\Phi 50$ value as a sample and comparing the average $\Phi 50$ values for each quadrant of the island. We will use the following analysis of variance (ANOVA) formulas from Mendenhall [1983, p. 554]:

$$SST = \frac{n_1 n_2}{n_1 + n_2}(\overline{y_1} - \overline{y_2})^2$$

$$SSE = \sum_{i=1}^{n_1}(y_{1,i} - \overline{y_1})^2 + \sum_{i=1}^{n_2}(y_{2,i} - \overline{y_2})^2$$

$$MSE = \frac{SSE}{n_1 + n_2 - 2}$$

$$MST = \frac{SST}{df} = SST$$

$$F = \frac{MST}{MSE}$$

where the samples from Population 1 are $y_{1,1}, y_{1,2}, \ldots, y_{1,n_1}$ and the samples from Population 2 are $y_{2,1}, y_{2,2}, \ldots, y_{2,n_2}$. At $\alpha = 0.05$ we have an F statistic of 4.96. In Table 4 we present the F value for several tests. In each case, the null hypothesis is that there is no significant differences between the mean values. Calculating the average $\Phi 50$ for each side we get 2.08 for the East, 0.97 for the South, 1.15 for the West, and 0.92 for the North. This seems to indicate that the East side is composed of a finer grain sand. If we check Table 4, we find that that is the trend. Notice that the F value for the East versus West test is high, although not high enough to reject the null hypothesis at $\alpha = 0.05$. In fact, the only tests which yield a rejection of the null hypothesis are those involving the East versus some other region or regions. The strongest rejection comes from the North versus East test. We also get a rejection for the East versus South and West. In addition, if we pool the North, South, and West sides and compare them with the East side, we get a rejection. We can therefore say with a good deal of confidence, that the average grain size is smaller on the eastern (open ocean facing) side of the island. However, as Table 4 indicates, we cannot statistically justify any other such general claims about grain size. Lee *et al* [1986] also make statements concerning the distributions of the grain sizes ("well sorted", "poorly sorted", etc.), but we will not address that issue here.

CONCLUSION

We have presented two tests which give statistical significance to certain claims involving sand grain size data. The tests were applied to data gathered on San Salvador Island, Bahamas. In the first test, we used a slightly modified t test to compare the weighted arithmetic means ("phi arithmetic means" of Folk [1974]) of two sand samples. This test was used to demonstrate the difference in mean grain size of two samples taken at Sandy Point. The "shore" sample was found to have smaller average diameter than the "dune" sample. The second test is an application of analysis of variance to average $\Phi 50$ values of several samples taken at different sites. The island was broken into four sides (North, South, East, and West) and comparisons were made between these. The only statistically significant differences involved comparing the East side of the island with one or more of the other sides. It seems that this side, which is exposed to the open ocean, has a significantly higher $\Phi 50$ value and hence a smaller average grain size.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Ron Lewis of Auburn University for his useful advice during the preparation of this paper.

REFERENCES

FOLK, R. L., 1974. *Petrology of Sedimentary Rocks*, Hemphill Publishing Co., Austin.

LEE, Y. I., LINDSEY, B., MAY, T., AND MANN, C. J., 1986. Grain Size Distribution of Calcareous Beach Sand, San Salvador Island, Bahamas: Occasional Paper 1986, No. 1. CCFL Bahamian Field Station, San Salvador, Bahamas.

MENDENHALL, W., 1983. *Introduction to Probability and Statistics*, Duxbury Press, Boston.

OTT, L., 1988. *An Introduction to Statistical Methods and Data Analysis*, PWS-Kent Publishing Co., Boston.

Table 1. Distributions of grain size in sand samples collected on San Salvador Island in March of 1989. Entries are the percentage by weight of each sample occurring in the given Φ class. The precise sample sites are given in Figure 1.

Φ	Sandy Point "dune"	Sandy Point "shore"	Rice Bay	Snow Bay	Telephone Pole Reef	Grotto Beach	Sue Point
0.0	8.2	23.2	0.1	1.1	0.2	24.2	28.1
0.5	45.4	56.9	0.2	6.4	3.6	41.1	43.2
1.0	37.1	19.1	1.0	22.4	35.4	18.3	20.2
1.5	6.8	0.6	4.5	26.8	41.1	6.1	4.6
2.0	1.8	0.02	26.7	29.9	16.4	4.5	2.0
2.5	0.5	0.04	43.4	12.1	3.1	4.5	1.2
3.0	0.1	0.0	21.8	1.0	0.2	1.2	0.5
3.5	0.03	0.0	2.3	0.4	0.0	0.1	0.04
4.0	0.03	0.0	0.0	0.0	0.0	0.0	0.0

Table 2. Cumulative distribution of Φ values. ΦX is the Φ value below which $X\%$ of the sample occurs. These Φ values are used in the calculation of the Inclusive Graphic Skewness (Sk_I) and the Graphic Kurtosis (K_G). The entries marked with a "*" were determined with linear extrapolation.

SITE	$\Phi5$	$\Phi16$	$\Phi25$	$\Phi50$	$\Phi75$	$\Phi84$	$\Phi95$	Sk_I	K_G
Sandy Point "dune"	-.05*	.10	.20	.45	.80	.90	1.30	1.32	.92
Sandy Point "shore"	-.20*	-.05*	0.0	.25	.45	.60	.90	1.00	.90
Rice Bay	1.45	1.70	1.85	2.20	2.50	2.70	2.95	4.4	.95
Snow Bay	.30	.70	.90	1.40	1.80	1.95	2.35	2.14	.93
Telephone Pole Reef	.55	.70	.80	1.25	1.45	1.60	1.95	2.67	.88
Grotto Beach	-.20*	-.10*	0.0	.30	.75	1.00	2.20	.97	1.31
Sue Point	-.25*	-.10*	-.05*	.25	.60	.80	1.40	.86	1.04

Table 3. The data of Lee *et al* [1986]. The entries are the 50^{th} percentile ($\Phi 50$), Inclusive Graphic Skewness (Sk_I), and the Graphic Kurtosis (K_G). The sample sites are given in Figure 1.

SAMPLE	$\Phi 50$	Sk_I	K_G
1	2.47	-.23	1.19
2	2.47	-.03	1.02
3	1.89	-.20	.97
4	1.84	.15	1.15
5	2.12	.02	.86
6	1.69	.09	.83
7	1.36	-.49	1.41
8	1.40	.27	.82
9	1.06	-.56	.62
10	0.51	.50	.91
11	0.29	-.18	1.23
12	0.64	-.04	.64
13	0.56	.18	1.18
14	-0.20	.16	1.62
15	2.25	-.28	1.12
16	2.25	-.13	1.05
17	0.89	-.13	1.02
18	2.74	-.46	1.00
19	0.09	.20	1.35
20	0.53	.24	.66
21	0.03	-.15	1.06
22	0.42	-.0074	.79
23	1.89	-.31	1.06
24	0.47	.24	.98
25	1.32	-.50	.77
26	1.40	-.46	.81

Table 4. Results of several ANOVA tests comparing Φ50 values of sand grain size from different sides of San Salvador Island. The data analyzed is that of Lee *et al* [1986]. See the text for further details.

TEST	F
E vs W	4.62
N vs S	.02
N vs E	12.91
N vs W	.17
S vs E	3.95
S vs W	3.77
E vs S & W	7.90
E & S vs W & N	1.02
S & W vs N & E	2.00
E vs N & S & W	9.89

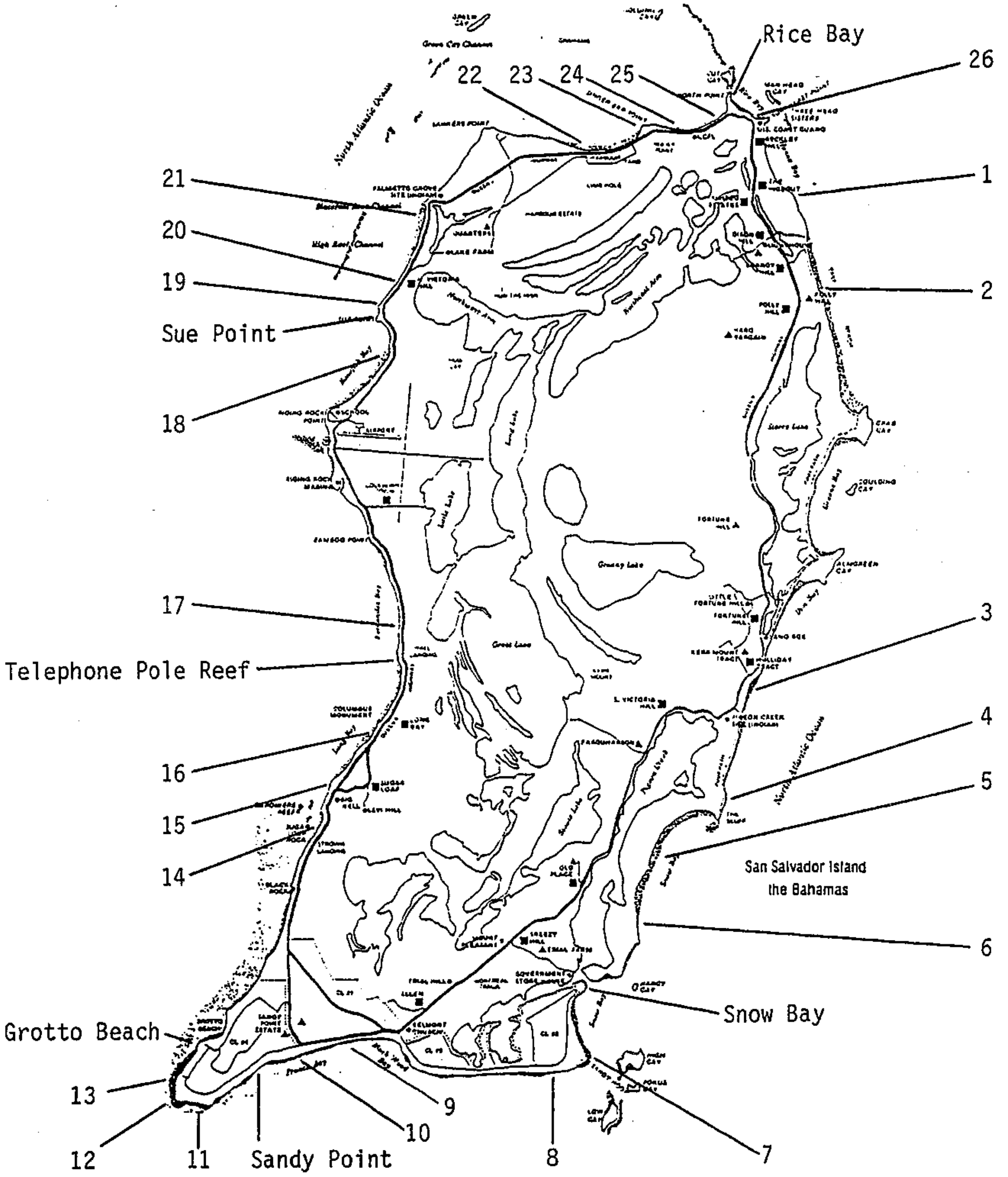

Figure 1. Sample sites on San Salvador Island, Bahamas. Samples 1 through 26 were gathered by Lee *et al* [1986] in March of 1983 and March of 1984. The remaining samples were collected by the author in March of 1989. All samples were collected between mean sea level and the high tide mark with the exception of the Sandy Point "dune" sample which was gathered at about 75 meters from the shoreline. North is at top.

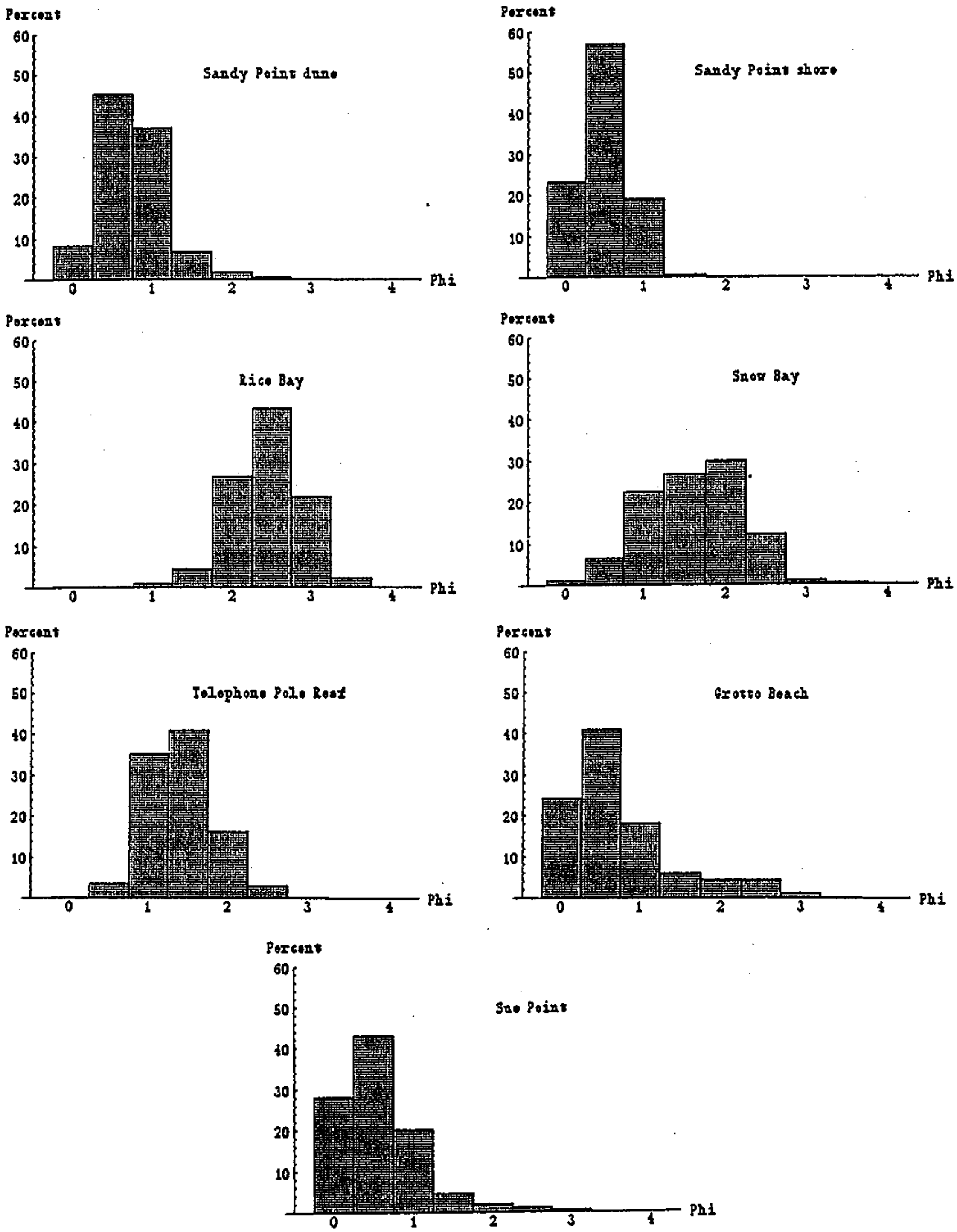

Figure 2. Histograms of samples gathered by the author in March 1989. The precise sample locations are given in Figure 1.