WOODS HOLE OCEANOGRAPHIC INSTITUTION

REFERENCE NO. 65-42

A COMPUTER PROGRAM FOR GRAIN-SIZE DATA

WOODS HOLE, MASSACHUSETTS

WOODS HOLE OCEANOGRAPHIC INSTITUTION

Woods Hole, Massachusetts

Reference No. 65-42

TECHNICAL REPORT

Submitted to the United States Geological Survey under Contract USGS-14-08-0001-8358

A Computer Program for Grain-Size Data

by

John Schlee

and

Jacqueline Webster

In citing this report in a bibliography, the references should be followed by the phrase UNPUBLISHED MANUSCRIPT.

August 1965

APPROVED FOR DISTRIBUTION John M. Hunt

John M. Hunt, Chairman
Department of Chemistry and Geology

TABLE OF CONTENTS

	Page No.
Abstract	2
Introduction	3
General Program Description	6
Interpolation	9
Interpolation of Tails	15
Summary	18
References	19
Appendix	
Sediment Analysis	A-1
CPI	A-15
Foplot	A-20

LIST OF ILLUSTRATIONS

		Page No.
Figure 1.	Flow diagram of program for grain-size data	7
Figure 2.	Frequency curve based on data points from an actual analysis and interpolated values obtained by continuous parabolic interpolation	10
Figure 3.	Plot of two decreasing functions, with values obtained from probability function and an exponential function	17

Flow Diagram	A-8,9,10
Program Listing	A-11,12,13,14
Flow Diagram	A-18
Program Listing	A-19
Example of Sample	A-21

ABSTRACT

The computer program presented here seeks to improve estimation of statistical parameters for grain-size data by use of interpolated values. Interpolation is made by fitting a series of overlapping parabolas to the data, and follows the method of Snyder (1961). The values are used in moment formulas to compute standard statistical measures. Skewness and kurtosis are reduced by the interpolation data, and extreme positive values of kurtosis tend to be greatly reduced. The program also picks major modes, the median, and sediment type.

INTRODUCTION

Statistical summation of data from grain-size analysis has had as its purpose the numerical description of the sediment and use of the values to infer an origin of the sediment. Beginning with studies of Udden (1914) more than 50 years ago and continuing with the work of Krumbein (1934, 1938), Otto (1939), Inman (1952), and Folk and Ward (1957), geologists have attempted to describe the grain size of sediment with a number or a series of numbers. In this endeavor they have employed the formulae and methods of statisticians because grain sizes are continuously distributed. Our purpose is to describe a computer program which approximates a continuous distribution and yields parameters by the method of moments.

Statistical summation of size curves has been performed by moment measures (Krumbein and Pettijohn, 1938, p. 239-253) and more commonly with graphic measures which approximate moment measures. Before the advent of modern electronic computers, graphic measures were used because of the tediousness of the calculation for moment measures. Graphic measures (Inman, 1952, p. 130; Folk and Ward, 1957, p. 12-14) make an approximation of the moment measures by use of percentiles picked at selected places on the curve. For most curves, the values obtained by these measures are similar to those computed by moment measures. However, at best, they analyze the central 90 percent of the frequency distribution and ignore the extreme ends of the distribution--the parts most important in an accurate estimate of the kurtosis and skewness.

An inspection of four computer programs in current use (Table 1) shows that most calculate both graphic and moment measures. To increase accuracy, three programs (University of Washington, University of Missouri and the United States Geological Survey) locate the percentiles by non-linear interpolation for computation of the graphic measures; none of them uses interpolation in connection with calculation of moment measures.

The authors wish to acknowledge the advise and assistance of Mr. Walter L. Anderson, and Mr. David S. Handwerker of the United States Geological Survey in calling to our attention the method of continuous parabolic interpolation, and in computing several analyses. We have used a program devised by Dr. Lloyd Breslau of the Office of Naval Research for sediment type, using the classification of Shepard (1954) for mixtures of sand, silt, and clay. Where appreciable amounts of gravel are present, the sediment designation is not valid because the program is not set up to handle gravel percentages.

TABLE 1

Comparison of computer programs for statistical analysis of grain-size

	Florida State Univ.[4/]	Univ. of Washington (Collias & others, 1963)	Univ. of Missouri (Kane & Hubert, 1963)	United States Geological Survey[5/]	This Program
Graphic	No	Yes	Yes	Yes	No
Moment	Yes	Yes	Yes	Yes	Yes
Interpolation for Moments	No	No	No	No	Yes (Optional)
Sediment Description	No	Yes	Yes	No	Yes
Calculation of mode	No	No	No	Yes	Yes
Input Form	Sample Weight	Frequency weight per cent or sample weight	Sample weight	Sample weight, cumulative weight per cent, or percentiles	Frequency weight per cent
Computer language	FORTRAN	FORTRAN	FORTRAN	ALGOL	FORTRAN

4/ H. G. Goodell, Personal Communication, April 8, 1963.

5/ Computing program for 220, United States Geological Survey informal compilation, February 1963.

GENERAL PROGRAM DESCRIPTION

The purpose of the program is to analyze statistically grain-size data of sediments by computing the mode(s), median, arithmetic mean, standard deviation, skewness, and kurtosis (Krumbein and Pettijohn, 1938, p. 239-253). Our program calculates statistical parameters by the method of moments. We seek to improve this method by obtaining intermediate points between the measurements through continuous parabolic interpolation--a procedure designed specifically to eliminate the sharp peaks often produced by conventional interpolation methods (Snyder, 1961). Use of moment measures to summarize size curves can be criticized because of their limitations. As Inman (1952, p. 143) has noted, the estimate of standard deviation is affected by the skewness. Other difficulties have been discussed by McCammon (1962). In part, skewness may reflect the mixing of two lognormal distributions, as has been pointed out by Spencer (1963, p. 180). If such mixing does take place, then all methods presently in use are affected, for all assume normal or near normal distributions, not polymodal ones.

For each sediment sample, the weight per cent for each size class is read into the General Electric 225 computer from a punched card. A flow diagram of subsequent steps is shown in Figure 1. Note that there are two options included in the program, one to allow suppression of the continuous parabolic interpolation, and another to allow print-out of the interpolated data. In the normal operation, interpolation is made, and no print-out of interpolated data occurs. Approximately 35 seconds of General Electric 225 computer time are used to analyze one sediment in the normal manner. Approximately 100 seconds are used if the interpolated data are printed out, and approximately 25 seconds are used if no interpolation is made.

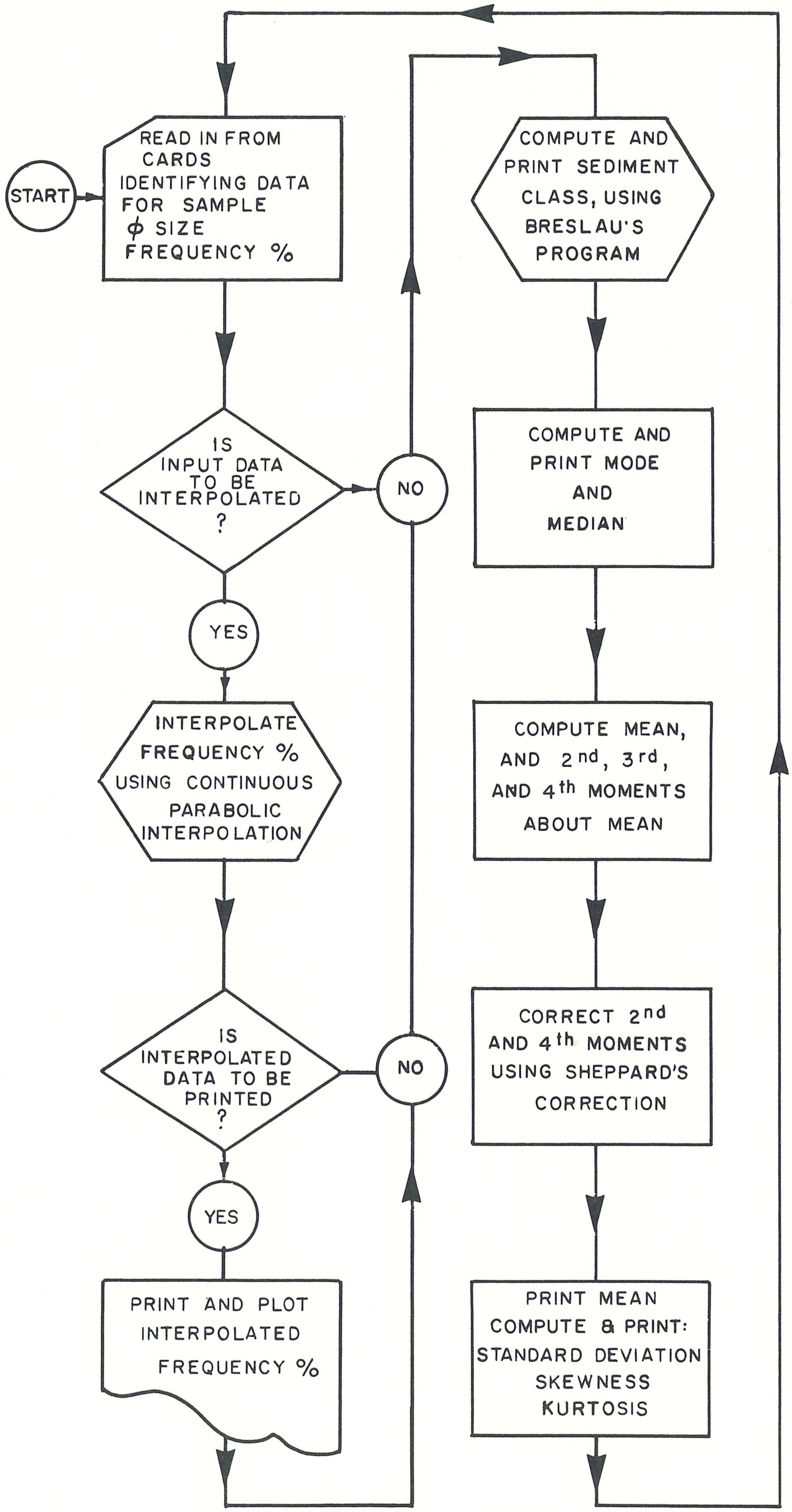


Figure 1. Flow diagram of program for grain-size data

To correct for the error introduced by use of grouped data, Sheppard's correction (Kenny and Keeping, 1954 p. 95-96) is applied in computation of the second and fourth moments. The error, though slight, arises because we assume that each value within a phi class is centered at the mid-point of each class.

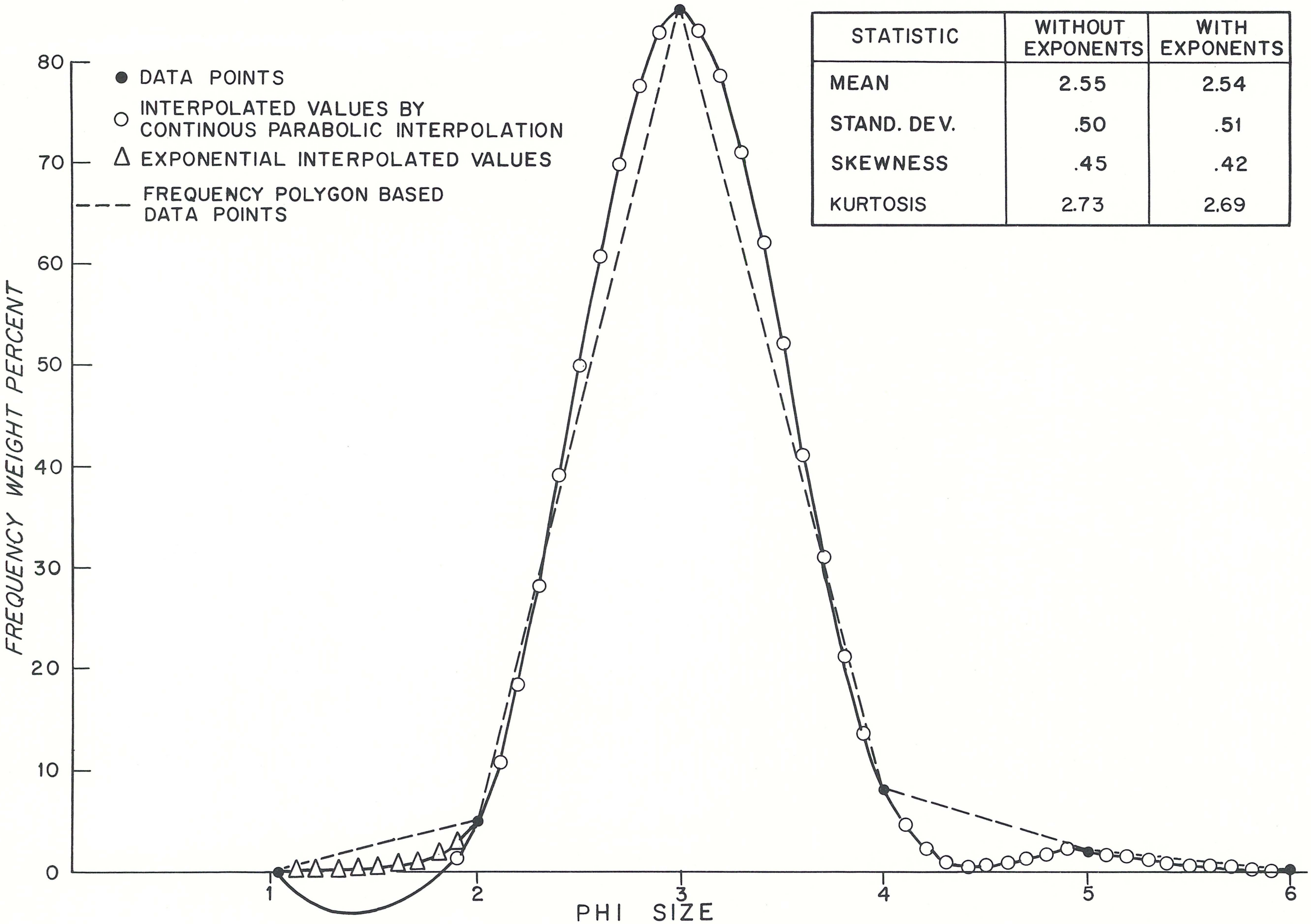
Samples are sized by sieves (gravel) and settling tube (sand, silt, clay). For the few samples where the percentage of sediment less than one micron is greater than five per cent, the size curve is extrapolated into the finer classes on probability paper so that no more than five per cent is in the finest size class.

INTERPOLATION

Interpolation of intermediate points between the data points was made in order to approximate, as closely as possible, the continuous distribution. Without this interpolation, the approximating curve obtained by drawing chords between data points can depart significantly from the size distribution curve, particularly in the tails and on the flanks of the distribution (Figure 2).

Two methods of interpolation were tried, the continuous parabolic (Snyder, 1961) and the three-point Lagrangian (Rattray, 1962). The Lagrangian method consists of taking the average of values from two overlapping parabolas. It was found after computing a few examples, that this method did not yield curves that were as smooth as the ones yielded by continuous parabolic interpolation.

Continuous parabolic interpolation is accomplished by fitting a series of parabolas through the data points and obtaining the interpolated values from the smoothed curve at a designated interpolation interval. The method fits the curve through data points and eliminates sharp changes in slope. In the area where two parabolas overlap, a series of values intermediate between the two curves is obtained from a group of parabolas of intermediate curvature generated between the two original overlapping ones. The intermediate parabolas are obtained by extending the limbs of the two overlapping parabolas to a common ordinate. The three points necessary to describe the intermediate parabolas consist of the two data points common to the overlapping parabolas plus a third one taken at a fixed increment (usually 0.1 phi in our case) on the common ordinate. Mathematically, the intermediate values are obtained by linear transformation of the polynomials which describe one parabola to those which describe the next overlapping parabola.



STATISTIC	WITHOUT EXPONENTS	WITH EXPONENTS
MEAN	2.55	2.54
STAND. DEV.	.50	.51
SKEWNESS	.45	.42
KURTOSIS	2.73	2.69

Figure 2. Frequency curve based on data points from an actual analysis and interpolated values obtained by continuous parabolic interpolation.

Almost any interpolation interval can be employed[6/]; however, for all our analyses, a 0.1 phi interval has been used. As shown for one representative sample, (Table 2) no change takes place in values for mean, standard deviation, and skewness between the 0.1 phi interval and finer ones. Only in estimation of kurtosis is there a change and this is a minor one. Based on this analysis, plus others, we have concluded that little is gained by using a finer interval than 0.1 phi.

We wish to emphasize that the interpolated values are in no way a substitute for additional data. The values are only points of a curve fitted to existing data points. More data would give a more accurate representation of the curve and cause a shift which interpolated values never could cause. The influence of more data points can be seen in Table 3. Notice that the addition of more data causes a marked change, particularly in estimation of skewness and kurtosis. The effect is similar to that of interpolation, in that positive values are lessened; parameters based on more data points, or a closer interpolation interval, converge toward a similar value. Presumably the parameters based on interpolation of the largest number of data points give the most accurate statistical estimates.

The importance of interpolation in calculation of the higher moments was also noted in 16 additional samples analyzed for skewness and kurtosis. Differences of estimates between interpolated and non-interpolated analyses range from 0.02 to 3.89; differences are greatest for high (peaked) positive values of kurtosis. The reason for this is pointed out in Figure 2. The

---

6/ Up to 2000 items of data (including interpolation values) can be stored in the GE 225 computer during analysis of a sample. Thus, where fewer size classes are present, a smaller interpolation interval can be used.

TABLE 2

Changes of statistical parameters
caused by different interpolation intervals

	Interpolation Interval			No Interpolation
Sample W-140	0.1	0.05	0.01	
Mean	2.44	2.44	2.44	2.45
Standard Deviation	0.43	0.43	0.43	0.39
Skewness	0.00	0.00	0.00	0.02
Kurtosis	-0.41	-0.40	-0.40	-0.71

TABLE 3

Comparison of statistical parameters from interpolated and non-interpolated runs. Basic data at interval indicated at head of column.

Sample CC4	INTERPOLATED			NOT INTERPOLATED		
	Whole Phi	Half Phi	Quarter Phi 7/	Whole Phi	Half Phi	Quarter Phi
Mean	0.49	0.50	0.49	0.49	0.50	0.49
Standard Deviation	0.44	0.38	0.35	0.22	0.34	0.34
Skewness	0.06	0.17	0.24	1.16	0.27	0.25
Kurtosis	1.35	3.36	4.55	81.08	5.60	4.89

7/ Interpolation interval at 0.05 phi; all others, at 0.1 phi.

non-interpolated analysis emphasizes the peakedness of the distribution. The fullness of flanks is not included and hence the peak is accentuated and the magnitude of the fourth moment, kurtosis, is exaggerated. Interpolation, when coupled with moment measures, offers a more accurate estimate of higher moments by more closely approximating a continuous distribution.

The preceding example helps to point out what kurtosis really measures, namely the fullness of the peak. Most texts refer to kurtosis as a measure of the peakedness (Arkin and Colton, 1950, p. 9); some refer to it as a measure of the development of the tails (Kenney and Keeping, 1954, p. 27). In a sense both definitions are closely related to the fullness or thinness of the flanks in the main part of the frequency distribution (Snedecor, 1956, p. 201). High values indicate a thinning of the flank and thereby an emphasis of the peak and tails. Negative values indicate very full flanks and hence the peak and tails stand out less clearly.

INTERPOLATION OF TAILS

Both the Lagrangian and continuous interpolations have a drawback, in that restricted sequences of negative interpolated values show up on the tails of a few distribution curves (Figure 2, between 1 phi and 2 phi). These result where there is a large differential between two adjacent data points. The parabolas that are fitted during interpolation between the two data points can be quite sharp and thereby project below the zero per cent line. For a size curve, these values are meaningless, and they adversely affect the moment measures. Initially, to overcome this difficulty, the program was arranged so that a negative value was regarded as a zero by the computer in calculation of the moments, but the effect of this is to truncate the distribution in the tails.

Because of the sensitivity of kurtosis to the position of data points in the tails, we considered that a better solution was needed. A solution to the problem was found by substituting a different function -- one that decreases rapidly from the last data point and asymptotically approaches zero. To approximate this relation we used an exponential function (Krumbein and Pettijohn, 1938, p. 208-211):

$$y = y_0 \, e^{\pm ax}$$

The negative function was used at the right end of the curve for descending values and the positive form was used for ascending values at the left side of the frequency curve. In order to compute a (slope coefficient), values other than zero have to be assumed at the extreme ends of the distribution; we used 0.01 per cent for this value. Once the slope coefficient is known, interpolated values of y can be determined by substituting in the formula for each value of x over the interpolation interval.

The interpolated values obtained by the formula are similar to the values obtained by extrapolation on probability paper (Figure 3). Two examples are shown: one, for decrease of five per cent over a one-half phi interval, and two, for a decrease of eight per cent over a whole phi interval. Slopes of exponential and probability curves are also very similar because both curves describe a similar function. Hence, we decided to use the exponential function to approximate the extreme ends of the curve, wherever negative numbers were encountered.

The effect of this approach on the statistical parameters in comparison to the one where zeros were used is surprisingly small. The parameters are changed very little by the new method; a shift of only 0.04 in the kurtosis takes place in the example shown (Figure 2). Hence, though we have continued to use the exponential method for interpolation at the ends of the curve where negative values are present, the justification for continued use of the method is not strong.

Figure 3. Plot of two decreasing functions, with values obtained from probability function and an exponential function.

SUMMARY

The computer program for size analysis, used by the Woods Hole Oceanographic Institution - United States Geological Survey Atlantic Continental Margin project, fits a continuous distribution to discrete data by the method of continuous parabolic interpolation. In a few samples, negative size values were computed in the tails of the distribution as a result of this interpolation. To eliminate this difficulty, an exponential interpolation is substituted whenever negative values occurred in the tails. From the interpolated size data, moment measures are then calculated to characterize the size curve. When interpolated values as well as original measurements are used, rather than original measurements alone, the moment measures more truly represent the characteristics of the size distribution.

REFERENCES

Arkin, H., and Colton, R. R., 1950, Statistical Methods: New York, Barnes and Nobles, Inc., 224 p.

Collias, E. E., Rona, M. R., McManus, D. A., and Creager, J. S., 1963, Machine processing of geological data: Univ. of Washington Tech. Rept. No. 87, 119 p.

Folk, R. L., and Ward, W. C., 1957, Brazos River bar: a study in the significance of grain size parameters: Jour. Sed. Petrology, v. 27, p. 3-26.

Inman, D. L., 1952, Measures for describing the size distribution of sediments: Jour. Sed. Petrology, v. 22, p. 125-145.

Kane, W. T., and Hubert, J. F., 1963, Fortran program for calculation of grain-size textural parameters on the IBM 1620 computer: Sedimentology, v. 2, p. 87-90.

Kenney, J. F., and Keeping, E. S., 1954, Mathematics of statistics, part one: Princeton, New Jersey, D. Van Nostrand Company, Inc., 348 p.

Krumbein, W. C., 1934, Size frequency distributions of sediments: Jour. Sed. Petrology, v. 4, p. 65-77.

______________, 1938, Size frequency distributions of sediments and the normal phi curve: Jour. Sed. Petrology, v. 8, p. 84-90.

______________, and Pettijohn, F. J., 1938, Manual of sedimentary petrography: New York, Appleton-Century, 549 p.

McCammon, R. B., 1962, Moment measures and the shape of size frequency distributions: Jour. Geology, v. 70, p. 89-92.

Otto, G. H., 1939, A modified logarithmic probability graph for the interpretation of mechanical analyses of sediments: Jour. Sed. Petrology, v. 9, p. 62-76.

Rattray, M., 1962, Interpolation errors and oceanographic sampling: Deep-Sea Research, v. 9, p. 25-37.

Shepard, F. P., 1954, Nomenclature based on sand-silt-clay ratios: Jour. Sed. Petrology, v. 24, p. 151-158.

Snedecor, G. W., 1956, Statistical methods: Ames, Iowa, The Iowa State College Press, 534 p.

Snyder, W. M., 1961, Continuous parabolic interpolation: Hydraulics Div. Jour., Am. Soc. Civil Engineers Proc., v. 87, no. Hy 4, p. 99-111.

Spencer, D. W., 1963, The interpretation of grain-size distribution curves of clastic sediments: Jour. Sed. Petrology, v. 33, p. 180-190.

Udden, J. A., 1914, Mechanical composition of clastic sediments: Geol. Soc. America Bull., v. 25, p. 655-777.

APPENDIX

PROGRAM NAME: Sediment Analysis

PROGRAM TYPE: General Electric 225 FORTRAN Program

DATE: October 1964

PROGRAMMERS: Jacqueline Webster and John Schlee

PURPOSE:

a) To analyze statistically grain-size data from natural sediments, i.e. to compute the mode, median, arithmetic mean, standard deviation, skewness, and kurtosis for a given size distribution.

b) To classify sediments according to Shepard's (1954) classification for sand, silt, and clay.

METHOD:

For each sediment sample, the frequency per cent with its corresponding phi size is read into the computer from punched cards. These input data are printed out. Then, interpolated data may be obtained from the given data, and printed out, if desired. (See write-ups of continuous parabolic interpolation subroutine for details of interpolation method.)

Let frequency per cent be represented by F_i and phi size by $\emptyset_i$, i going from 1 to N, where N is the number of data points in the sample (either interpolated points or only the original data, depending on choice made by the user of the program).

First the sediment class for the given sample is computed, according to Shepard (1954) and printed out.

The modes of the sample are found by examining the first differences of the frequency per cent. When the first difference, $\Delta_i = F_i - F_{i-1}$, changes sign, the center of the phi class corresponding to F_{i-1} is taken as a mode, provided the frequency per cent for that class, F_{i-1}, is greater than 5 times the class interval, $\Delta\emptyset$. This latter provision sets an arbitrary limit to eliminate minor modes within the distribution.

The median of the sample is found by calculating cumulative frequency per cent, and then interpolating linearly to find the $\emptyset$ value corresponding to a cumulative frequency per cent of 50.

Let C_i be the center of the $\emptyset$ class corresponding to F_i. Then moment measures are calculated as follows:

$$S = \sum_{i=1}^{N} F_i$$

$$n_1 = \sum_{i=1}^{N} F_i \, C_i \, /S$$

$$n_2 = \sum_{i=1}^{N} F_i \, C_i^2 \, /S$$

$$n_3 = \sum_{i=1}^{N} F_i \, C_i^3 \, /S$$

$$n_4 = \sum_{i=1}^{N} F_i \, C_i^4 \, /S$$

The uncorrected moments about the mean are found

$$m_2 = n_2 - n_1^2$$

$$m_3 = n_3 - 3n_2n_1 + 2n_1^3$$

$$m_4 = n_4 + n_1(-4n_3 + 6n_1n_2 - 3n_1^3)$$

Sheppard's correction (Kenney and Keeping, 1954, p. 95-96) is applied to the fourth and second moments about the mean:

$$(m_4)_{corr} = m_4 - \frac{m_2}{2}(\Delta\emptyset)^2 + \frac{7}{240}(\Delta\emptyset)^4$$

$$(m_2)_{corr} = m_2 - (\Delta\emptyset)^2 \, /12$$

where $\Delta\emptyset$ is the $\emptyset$ class interval for the sample.

The standard deviation, skewness, and kurtosis are computed as follows:

$$\sigma = \sqrt{(m_2)_{corr}}$$

$$Sk = m_3/2\sigma(m_2)_{corr}$$

$$K = \frac{(m_4)_{corr}}{[(m_2)_{corr}]^2} - 3$$

For each sample the modes, median, arithmetic mean, standard deviation, skewness and kurtosis are printed out.

USAGE:

a) Deck make-up. To run a set of data, the following card deck must be made up:

1) FORTRAN control cards (*FORTRAN, *XEQ, *BINARY).

2) Binary deck for sediment analysis program.

3) Binary deck for FOPLOT subroutine, which is used to plot the data on the printer.

4) Binary deck for continuous parabolic interpolation subroutine.

5) CALL DATA card (a card with a 1 and a 2 punched in column 1).

6) An identifying card for the sample punched as follows:

columns 1-12	Sample identification, consisting of any combination of alphanumeric characters.
columns 16-17	Degrees portion of the latitude.
columns 19-23	Minutes portion of the latitude to tenths of minutes (the decimal point should be punched).

columns 24-26 Degrees portion of the longitude.

columns 28-31 Minutes portion of the longitude to tenths of minutes (the decimal point should be punched).

columns 33-36 Depth of the sample, in meters.

columns 38-39 Number of data points in the sample.

columns 41-45 Ø size interpolation interval. Can be given to thousandths. The decimal point must be punched.

7) A set of data cards for the frequency per cent and corresponding Ø size punched as follows:

columns 1-5 $Ø_1$

columns 6-10 F_1

columns 11-15 $Ø_2$

columns 16-20 F_2

. .

. .

. .

columns 71-75 $Ø_8$

columns 76-80 F_8

As many cards of the above format as are necessary may be used. The last card of a set may be partially blank.

8) By repeating 6) and 7) as many times as desired, several samples may be analyzed in one machine run.

9) After the last sample to be processed, there must be an end card, which consists of a card with 100 punched in columns 37-39.

b) <u>Options</u>.

1) To use interpolated data in the calculation of the statistical measures, sense switch 17 must be off. To use just the raw input data for this calculation, sense switch 17 must be on.

2) To plot the interpolated frequency per cent on the printer, sense switch 16 must be on.

3) Sense switch 18 can be put on to get a print-out of the intermediate calculations that are done in computing the moments. This option was used in checking the program, and would not be used except in case of some trouble.

c) <u>Limitations</u>.

1) The frequency per cent data must be given at equal Ø intervals. The program checks to see if this is true, and if it is not, skips to the next sample.

2) The maximum number of frequency per cent input values per sample is 30.

3) The maximum number of interpolated points that can be generated is 2000. Thus the Ø interpolation interval must not be chosen so small that this number will be exceeded.

d) <u>Timing</u>.

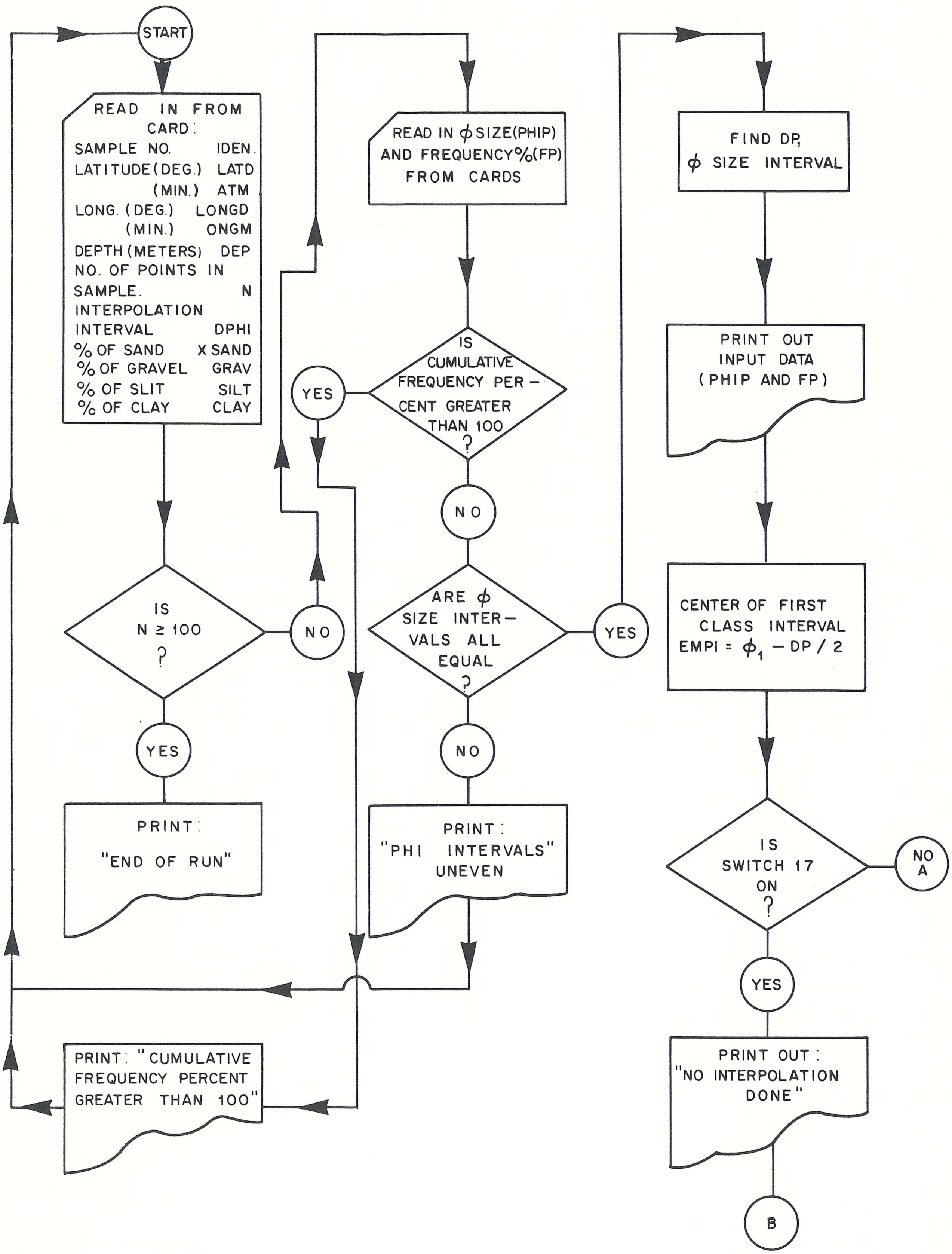
To load the program into the computer takes 25 seconds. Each sediment takes approximately 25 seconds to run if no interpolation is used, 100 seconds if interpolated data are used, and printed out, 35 seconds if interpolated data are used but not printed out.

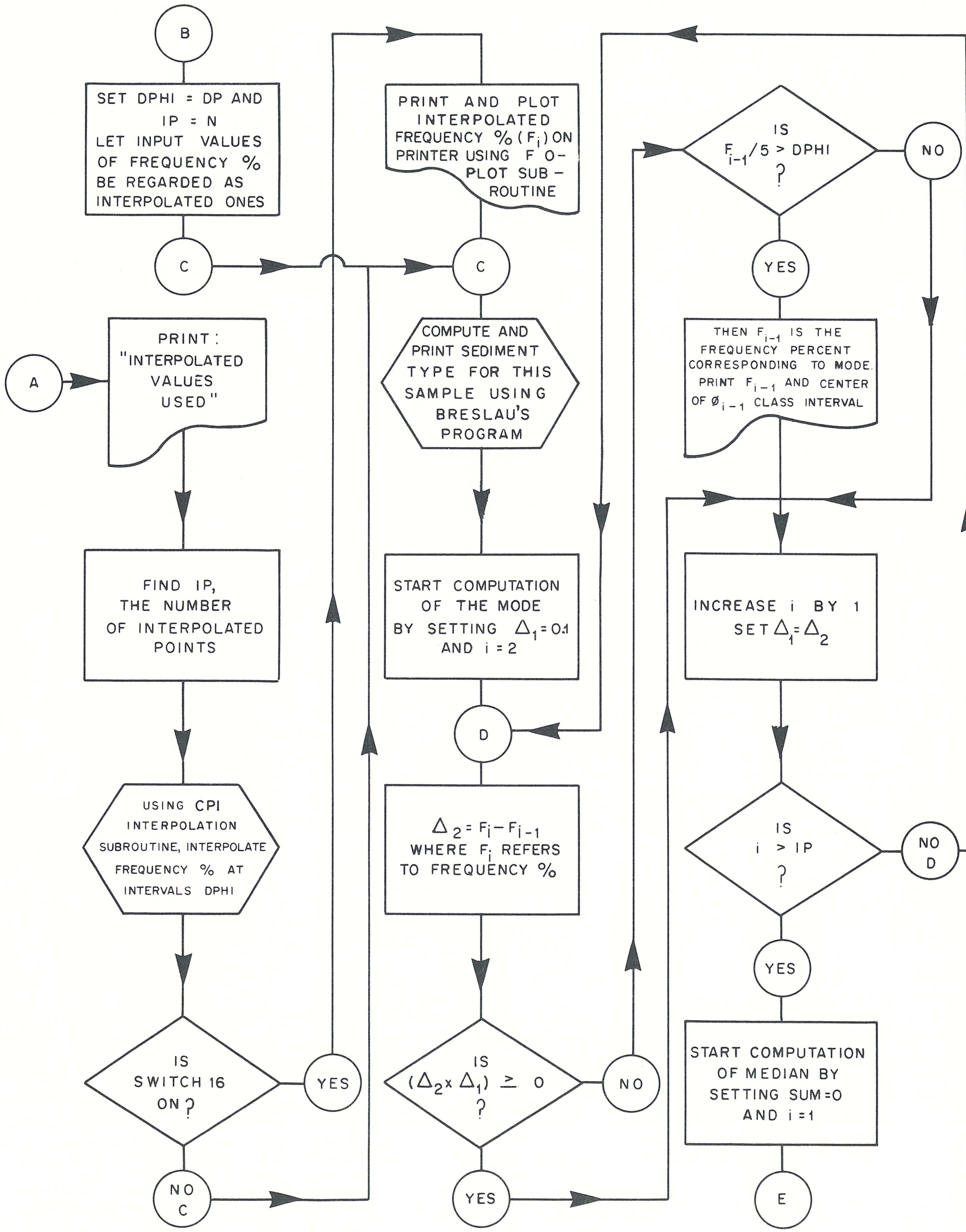
e) <u>Subroutine required</u>.

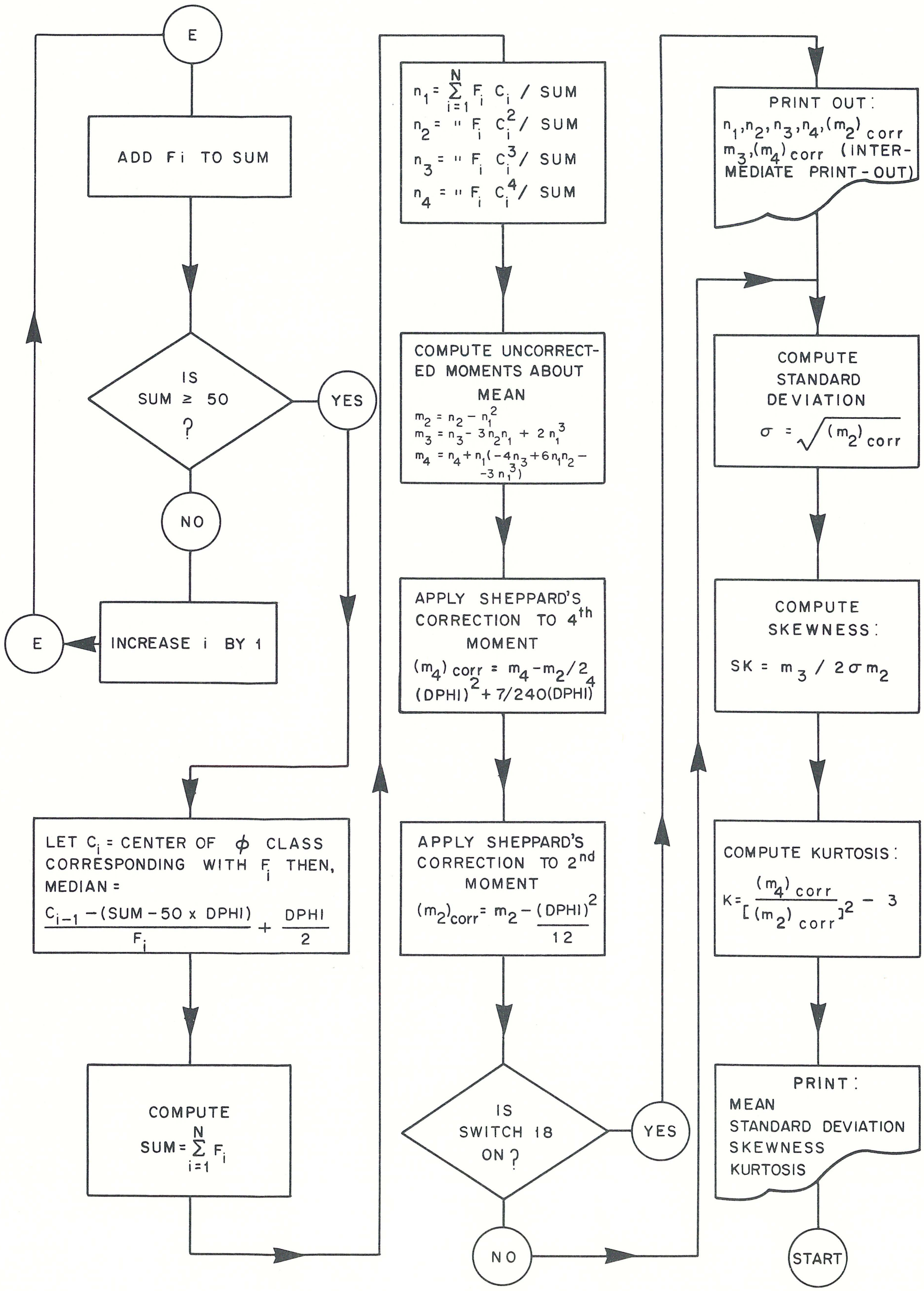
1) CPI - continuous parabolic interpolation.

2) FOPLOT - Plot, using printer.

SEDIMENT ANALYSIS FLOW DIAGRAM




B
SET DPHI = DP AND
IP = N
LET INPUT VALUES
OF FREQUENCY %
BE REGARDED AS
INTERPOLATED ONES
C
A
PRINT:
"INTERPOLATED
VALUES
USED"
FIND IP,
THE NUMBER
OF INTERPOLATED
POINTS
USING CPI
INTERPOLATION
SUBROUTINE, INTERPOLATE
FREQUENCY % AT
INTERVALS DPHI
IS
SWITCH 16
ON?
YES
NO
C
PRINT AND PLOT
INTERPOLATED
FREQUENCY % (Fi) ON
PRINTER USING F O-
PLOT SUB-
ROUTINE
C
COMPUTE AND
PRINT SEDIMENT
TYPE FOR THIS
SAMPLE USING
BRESLAU'S
PROGRAM
START COMPUTATION
OF THE MODE
BY SETTING Δ1 = 0.1
AND i = 2
D
Δ2 = Fi − Fi-1
WHERE Fi REFERS
TO FREQUENCY %
IS
(Δ2 x Δ1) ≥ 0
?
NO
YES
IS
Fi-1/5 > DPHI
?
NO
YES
THEN Fi-1 IS THE
FREQUENCY PERCENT
CORRESPONDING TO MODE.
PRINT Fi-1 AND CENTER
OF Ø i-1 CLASS INTERVAL
INCREASE i BY 1
SET Δ1 = Δ2
IS
i > IP
?
NO
D
YES
START COMPUTATION
OF MEDIAN BY
SETTING SUM=0
AND i=1
E


E
ADD Fi TO SUM
IS SUM ≥ 50 ?
YES
NO
INCREASE i BY 1
E
LET C_i = CENTER OF φ CLASS CORRESPONDING WITH F_i THEN, MEDIAN =
$\frac{C_{i-1} - (SUM - 50 \times DPHI)}{F_i} + \frac{DPHI}{2}$
COMPUTE $SUM = \sum_{i=1}^{N} F_i$
$n_1 = \sum_{i=1}^{N} F_i C_i / SUM$
$n_2 = " F_i C_i^2 / SUM$
$n_3 = " F_i C_i^3 / SUM$
$n_4 = " F_i C_i^4 / SUM$
COMPUTE UNCORRECTED MOMENTS ABOUT MEAN
$m_2 = n_2 - n_1^2$
$m_3 = n_3 - 3n_2n_1 + 2n_1^3$
$m_4 = n_4 + n_1(-4n_3 + 6n_1n_2 - 3n_1^3)$
APPLY SHEPPARD'S CORRECTION TO 4th MOMENT
$(m_4)corr = m_4 - m_2/2 (DPHI)^2 + 7/240(DPHI)^4$
APPLY SHEPPARD'S CORRECTION TO 2nd MOMENT
$(m_2)corr = m_2 - \frac{(DPHI)^2}{12}$
IS SWITCH 18 ON ?
YES
NO
PRINT OUT: $n_1, n_2, n_3, n_4, (m_2)corr$ $m_3, (m_4)corr$ (INTERMEDIATE PRINT-OUT)
COMPUTE STANDARD DEVIATION
$\sigma = \sqrt{(m_2)corr}$
COMPUTE SKEWNESS:
$SK = m_3 / 2\sigma m_2$
COMPUTE KURTOSIS:
$K = \frac{(m_4)corr}{[(m_2)corr]^2} - 3$
PRINT: MEAN STANDARD DEVIATION SKEWNESS KURTOSIS
START

```
C       SCHLEE PROJ 8358 SEDIMENT SIZE ANALYSIS
        COMMON PHIP[30],FP[30],IDEN[4],F [2000]
C       READ IN SAMPLE NO.,LONGITUDE,LATITUDE,DEPTH,NO. OF SAMPLE PTS.,AND
C       PHI SIZE INTERVAL FOR INTERPOLATION
V        NLT
 15     READ 500,IDEN,LATD,ATM,LONGD,ONGM,DEP,N,DPHI,XSAND,GRAV,SILT,CLAY
  500   FORMAT[4A3,3X,I2,F5.1,I4,F5.1,F5.0,I3,F6.3,4F5.0]
        NP1 =N + 1
C       LAST CARD HAS 100 IN COLS. 37-39
        IF[N-100]2,1,1
 1      PRINT 612
 612    FORMAT [12H1 END OF RUN   /1H1]
        CALL EXIT
C       READ IN PHI SIZE AND FREQUENCY PERCENT
      2 READ 501,[PHIP[J],FP[J],J=1,N]
 501    FORMAT[10F5.2,6F5.2]
C       TEST IF CUMULATIVE FREQUENCY PERCENT IS =100
        CFP=0.
        DO 3 J=1,N
 3      CFP=CFP+FP[J]
        IF[ABSF[CFP-100.]-2.] 16,6,6
 6      PRINT 613,IDEN
 613    FORMAT[10X,4A3,62HCUMULATIVE FREQUENCY PER CENT GREATER THAN 101 O
       XR LESS THAN 99   / 1H1 ]
        GO TO 15
 16     CONTINUE
C       TEST IF PHI INTERVALS ARE EVEN
        DP=PHIP[2]-PHIP[1]
        DO 55 J=3,N
        TEST2=PHIP[J]-PHIP[J-1]
        TEST3=ABSF[DP-TEST2]
        IF [TEST3-.001] 55,55,56
 55     DP=TEST2
        GO TO 62
C       PHI INTERVALS NOT EVEN
 56     PRINT 611,IDEN
 611    FORMAT [10X,4A3,20HUNEVEN PHI INTERVALS    / 1H1]
        GO TO 15
C       PRINT ORIGINAL DATA
 62     PRINT 600
  600   FORMAT[33H1USGS-WHOI SEDIMENT SIZE ANALYSIS///10X,14HIDENTIFICATIO
       XN,6X, 9HLONGITUDE,11X,8HLATITUDE,12X,9HDEPTH [M],10X,12HPHI INTERV
       1AL ///]
        PRINT 601,IDEN,LONGD,ONGM,LATD,ATM,DEP,DPHI
  601   FORMAT[10X,4A3,8X,I3,F5.1,12X,I2,F5.1,14X,F5.0,15X,F6.3///]
        PRINT 602
 602    FORMAT[20X,23HPHI SIZE  FREQ. PERCENT//]
        PRINT 603,[PHIP[J],FP[J],J=1,N]
  603   FORMAT[21X,F5.2,8X,F5.2]
        EMP1 = PHIP[1] -0.5*DP
```

```
C       P IS NO. OF INTERPOLATED POINTS                                         189
        IP = [PHIP[N] - PHIP[1]]/DPHI + 1.5
        IF [SENSE SWITCH 17] 51,52                                              201
 51     IP=N                                                                    202
        PRINT 620
 620    FORMAT [//25H NO INTERPOLATION DONE           //]
        DPHI=DP                                                                 203
        DO 50 I=1,N                                                             204
 50     F[I]=FP[I]                                                              205
        GO TO 54                                                                206
 52     PRINT 621
 621    FORMAT [//25H INTERPOLATED VALUES USED           //]
        DO 100 I=1,N
        J = N-I
        FP[J+2] = FP[J+1]
 100    PHIP[J+2] = PHIP[J+1]
        CALL CPI [PHIP,FP,N,DP,F,IP,DPHI ]
        DPI=DPHI/DP                                                             216
        DO 64 I=1,IP                                                            217
        IF [F[I]] 63,64,64
 63     F[I] =0.
 64     F[I]=F[I]*DPI                                                           218
 54     IF[SENSE SWITCH 16]4,5                                                  220
 4      PRINT 604                                                               230
 604    FORMAT[///25X,40H          INTERPOLATED  FREQUENCY PLOT   //]           234
        CALL FOPLOT [0.,20.,F,IP]                                               240
        PRINT 610                                                               245
 610    FORMAT[1H1]                                                             246
 5      PRINT 609
 609    FORMAT[///]                                                             256
C       THIS IS LLOYD BRESLAUS SEDIMENT CLASS PROGRAM
        IF [XSAND] 878,880,878
 880    XSAND =.001
 878    IF [SILT]883,882,883
 882    SILT=.001
 883    IF [CLAY] 885,884,885
 884    CLAY=.001
 885    SAND=XSAND+GRAV
        IF [SAND-75.] 879,850,850
 879    IF [SILT-75.] 886,851,851
 886    IF [CLAY-75.] 899,852,852
 899    SANSIL = SAND/SILT
        CLYSND=CLAY/SAND
        SILCLY=SILT/CLAY
        IF [SAND-20.] 887,887,890
 887    IF [SANSIL-1.]888,888,853
 888    IF [SILCLY-1.]854,889,889
 889    IF [CLYSND-1.]856,856,855
 890    IF[CLAY-20.]891,891,895
 891    IF [SANSIL-1.] 856,892,892
```

```
 892  IF [SILCLY-1.]858,858,857
 895  IF[SILT-20.] 896,896,859
 896  IF [CLYSND-1.]858,858,853
 850  PRINT 820
 820  FORMAT[13H SAND          ///]
      GO TO 881
 851  PRINT 822
 822  FORMAT[13H SILT          ///]
      GO TO 881
 852  PRINT 824
 824  FORMAT [13H CLAY         ///]
      GO TO 881
 853  PRINT 826
 826  FORMAT[13H SANDY CLAY    ///]
      GO TO 881
 854  PRINT 828
 828  FORMAT [13H SILTY CLAY     ///]
      GO TO 881
 855  PRINT 830
 830  FORMAT[13H CLAYEY SILT   ///]
      GO TO 881
 856  PRINT 832
 832  FORMAT[13H SANDY SILT    ///]
      GO TO 881
 857  PRINT 834
 834  FORMAT[13H SILTY SAND    ///]
      GO TO 881
 858  PRINT 836
 836  FORMAT[13H CLAYEY SAND   ///]
      GO TO 881
 859  PRINT 838
 838  FORMAT[13H SAN SIL CLY   ///]
 881  CONTINUE
C     END BRESLAUS PROGRAM
C     MODE
      DEL1=0.1
      DO 7 I=2,IP
      DEL2=F[I]-F[I-1]
      IF[DEL2*DEL1] 8,7,7
 8    IF [DEL2] 80,7,7
 80   IF [F[I-1]-5.*DPHI]7,7,9
 9    XI=I-2
      XEM = EMP1 + XI*DPHI
      XF=F[I-1]
      I=I+1
      PRINT 605,XEM,XF
  605 FORMAT[   20X,5HMODE ,F5.2,3X,F5.2//]
 7    DEL1=DEL2
C     MEDIAN
      SUM=0.
```

```
      DO 10 I=1,IP
      SUM=F[I]+SUM
      IF[SUM-50.]10,11,11
   10 CONTINUE
 11   XI=I-1
      CT = EMP1 + XI*DPHI
      CTMED=CT-[SUM-50.]*DPHI/F[I]    +0.5*DPHI
      PRINT 606,CTMED
  606 FORMAT[20X,7HMEDIAN=,F5.2//]
C     MOMENTS ABOUT THE MEAN
      SUM=0.
      SUM1=0.
      SUM2=0.
      SUM3=0.
      SUM4=0.
      DO 12 I=1,IP
      XI=I-1
      CT = EMP1 + XI*DPHI
      SUM=SUM+F[I]
      SUM1=SUM1+F[I]*CT
      SUM2=SUM2+F[I]*CT**2
      SUM3=SUM3+F[I]*CT**3
 12   SUM4=SUM4+F[I]*CT**4
      EN1=SUM1/SUM
      EN2=SUM2/SUM
      EN3=SUM3/SUM
      EN4=SUM4/SUM
      ZM2=EN2-EN1**2
      ZM3=EN3-3.*EN2*EN1+2.*EN1**3
      ZM4=EN4+EN1*[-4.*EN3+6.*EN1*EN2-3.*EN1**3]
      DPHI2 = DPHI*DPHI
      ZM4 =ZM4 -0.5*DPHI2*ZM2 + 0.02916667*DPHI2*DPHI2
      ZM2 = ZM2 -DPHI2/12.
      IF[SENSE SWITCH 18]13,14
   13 PRINT 607,EN1,EN2,EN3,EN4,ZM2,ZM3,ZM4
  607 FORMAT[7E16.8//]
C     STANDARD DEVIATION
   14 SIGMA=SQRTF[ZM2]
C     SKEWNESS
      SKEW=0.5*ZM3/[SIGMA*ZM2]
C     KURTOSIS
      ZKURT=ZM4/ZM2**2 -3.
      PRINT 608,EN1,SIGMA,SKEW,ZKURT
  608 FORMAT[20X,5HMEAN=,F5.2//20X,19HSTANDARD DEVIATION=,F5.2//20X,9HSK
     XEWNESS=,F5.2//20X,9HKURTOSIS=F5,2]
C     READ IN ANOTHER SET OF DATA
      GO TO 15
V      LST
      END
```

Write-up for Continuous Parabolic Interpolation Subroutine

PROGRAM NAME: CPI

PROGRAM TYPE: General Electric 225 FORTRAN subroutine

DATE: October 1964

PROGRAMMERS: Jacqueline Webster and John Schlee

PURPOSE:

To obtain interpolated data points between given data points such that a smooth curve is produced through the data points.

METHOD:

The method used is the one described by Snyder (1961). However, on the tails of the curve, that is, between the first and second original data points on the curve, and between the next-to-last and last points, an exponential function is fitted to the curve, if it is found that the continuous parabolic interpolation give negative values of the dependent variable.

Let

h_j = the independent variable

q_j = the dependent variable to be interpolated

Δh = interval at which h is given

N = number of original data points given for q and h (i.e. j runs from 1 to N)

To interpolate q at a point h between h_j and h_{j+1} the following calculations are made:

$$Y = 3q_j - 3q_{j+1} + q_{j+2}$$

$$k = (h - h_j)/\Delta h$$

$$x = k + 1$$

$$A = q_{j-1} + k(Y - q_{j-1})$$

Then the interpolated value of q at h is given by

$$q = A + x\,[q_j - A + 0.5k(q_{j+1} - 2q_j + A)]$$

If the value of q given by the above equation is negative, and if j = 1 or j = N-1, then the following exponential interpolation method is used:

$$S = \log\,(q_j/q_{j+1})/\Delta h$$

$$L = S(h - h_j)$$

$$q = q_j/e^L$$

USAGE: CALL CPI (H, Q, N, DELH, F, M, DELHC)

Where H, Q, and F are floating point dimensioned variables representing, respectively: the original values of the independent variable, the original values of the dependent variable, and the interpolated values of the independent variable.

N and M are fixed point variables representing, respectively, the number of original data points, and the number of interpolated data points that are to be found.

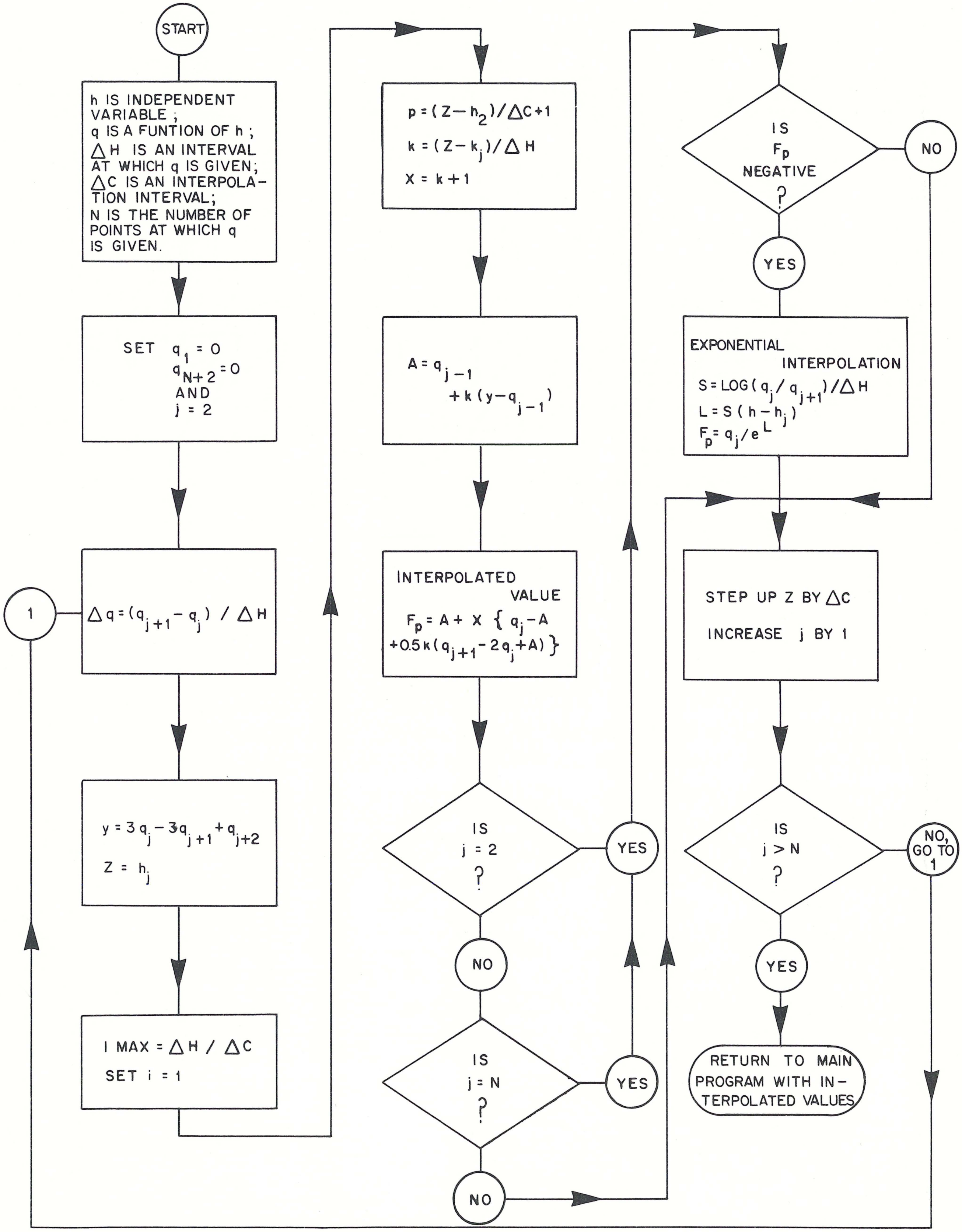
DELH and DELHC are floating point variables representing, respectively, the original spacing interval of the independent variable, and the spacing interval of the independent variable at which interpolation is to be made.

For example, if we had 10 values of H and Q given at H = 1, 2 ... 10, and we wished to obtain values of Q for H = 1, 1.1, 1.2 ... 9.8, 9.9, 10, the calling sequence for the subroutine would be

```
CALL CPI (H, Q, 10, 1.0, F, 100, 0.1)
```

where H and Q must have a dimension of at least 10, and F must have a dimension of at least 100.

CONTINUOUS PARABOLIC INTERPOLATION FLOW DIAGRAM



```
C      CONTINUOUS PARABOLIC INTERPOLATION
       SUBROUTINE CPI[H,Q,N,DELH,F,M,DELHC ]
       DIMENSION Q[1],H[1],F[1]
       Q[1] = 0.
       Q[N+2] =0.
       DO 1 J =2,N
       DELQ = [Q[J+1] -Q[J]]/DELH
       Y = 3. *Q[J] -3. *Q[J+1] + Q[J+2]
       Z = H[J]
       IMAX = DELH / DELHC
       DO 2 I=1,IMAX
       JP = [Z-H[2]]/DELHC + 1.5
       XK = [Z-H[J]]/DELH
       X = XK + 1.
       A = Q[J-1] +XK*[Y-Q[J-1]]
       FC     = A + X*[Q[J] - A + 0.5*[Q[J+1] -2.*Q[J] + A]*XK]
       IF [J-2] 13,12,13
    13 IF [J-N] 5,12,5
    12 IF [FC] 10,5,5
    10 Z = H[J] + DELHC
       IF[J-2] 4,3,4
    3  Q[J] = 0.01
       PRINT 8
    8  FORMAT [38HEXPONENTIAL INTERPOLATION AT BEGINNING  ///]
       GO TO 7
    4  IF [J-N] 5,6,5
    6  Q[J+1] = 0.01
       PRINT 9
    9  FORMAT[32HEXPONENTIAL INTERPOLATION AT END  ///]
    7  SLPE = LOGF[Q[J]/Q[J+1]]/DELH
       DO 11 II=2,IMAX
       JP = [Z-H[2]]/DELHC + 1.5
       ANTI = SLPE*[Z-H[J]]
       F[JP] = Q[J] / EXPF[ANTI]
    11 Z = Z+DELHC
       GO TO 1
    5  F[JP] = FC
    2  Z = Z    + DELHC
    1  CONTINUE
       F[JP+1] = 0.
       RETURN
       END
```

PROGRAM NAME: FOPLOT

PROGRAM TYPE: FORTRAN Subroutine Subprogram

DATE: August 1963

PROGRAMMER: W. Hosken

PURPOSE:

To plot floating point numbers (ordinate) against an equally spaced abssica on the GE 225 ON-LINE PRINTER

USE: CALL FOPLOT (TMIN, TMAX, X, N)

Where, TMIN is the minimum value to plot

TMAX is the maximum value to plot

X is the NAME (subscripted variable name) of the series to be plotted

N is the length (number of values) to plot

NOTE:

The range (TMIN, TMAX) is divided into 100 intervals and a value anywhere inside an interval i will be plotted using print wheel i. When a value falls outside the range (TMIN, TMAX) "OUT OF RANGE" will be printed along with the actual value.

Paper space control is entirely in the main program and no headings are printed.

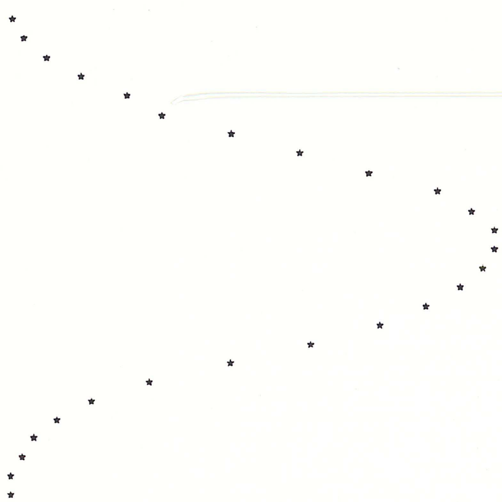
USGS-WHOI SEDIMENT SIZE ANALYSIS

IDENTIFICATION	LONGITUDE	LATITUDE	DEPTH (M)	PHI INTERVAL
W 140	65 58.0	41 49.0	91.	0.100

PHI SIZE	FREQ. PERCENT
1.50	0.
2.00	14.00
2.50	41.00
3.00	37.00
3.50	8.00
4.00	0.

INTERPOLATED VALUES USED

INTERPOLATED FREQUENCY PLOT



0.
0.33920000E-00
0.79359999E 00
0.13584000E 01
0.20288000E 01
0.28000000E 01
0.38128000E 01
0.50703999E 01
0.63615999E 01
0.74751999E 01
0.82000000E 01
0.85168000E 01
0.85664000E 01
0.83776000E 01
0.79792000E 01
0.74000000E 01
0.64928000E 01
0.52384000E 01
0.38576001E 01
0.25712001E 01
0.16000000E 01
0.98560001E 00
0.58080001E 00
0.32320001E-00
0.15040001E-00
0.

SAND

MODE 2.45 8.57

MEDIAN= 2.44

MEAN= 2.44

STANDARD DEVIATION= 0.43

SKEWNESS=-0.00

KURTOSIS=-0.41

Woods Hole Oceanographic Institution
Reference No. 65-42

A COMPUTER PROGRAM FOR GRAIN-SIZE DATA, by John Schlee, Jacqueline Webster. 40 p. August 1965. Contract No. USGS-14-08-0001-8358.

The computer program presented here seeks to improve estimation of statistical parameters for grain-size data by use of interpolated values. Interpolation is made by fitting a series of overlapping parabolas to the data, and follows the method of Snyder (1961). The values are used in moment formulas to compute standard statistical measures. Skewness and kurtosis are reduced by the interpolation procedure, when compared to results from non-interpolated data, and extreme positive values of kurtosis tend to be greatly reduced. The program also picks major modes, the median, and sediment type.

1. Computer Program
2. Grain Size
3. Statistical Method

I. John Schlee
II. Jacqueline Webster
III. Contract No. USGS-14-08-0001-8358

This card is UNCLASSIFIED

Woods Hole Oceanographic Institution
Reference No. 65-42

A COMPUTER PROGRAM FOR GRAIN-SIZE DATA, by John Schlee, Jacqueline Webster. 40 p. August 1965. Contract No. USGS-14-08-0001-8358.

The computer program presented here seeks to improve estimation of statistical parameters for grain-size data by use of interpolated values. Interpolation is made by fitting a series of overlapping parabolas to the data, and follows the method of Snyder (1961). The values are used in moment formulas to compute standard statistical measures. Skewness and kurtosis are reduced by the interpolation procedure, when compared to results from non-interpolated data, and extreme positive values of kurtosis tend to be greatly reduced. The program also picks major modes, the median, and sediment type.

1. Computer Program
2. Grain Size
3. Statistical Method

I. John Schlee
II. Jacqueline Webster
III. Contract No. USGS-14-08-0001-8358

This card is UNCLASSIFIED

Woods Hole Oceanographic Institution
Reference No. 65-42

A COMPUTER PROGRAM FOR GRAIN-SIZE DATA, by John Schlee, Jacqueline Webster. 40 p. August 1965. Contract No. USGS-14-08-0001-8358.

The computer program presented here seeks to improve estimation of statistical parameters for grain-size data by use of interpolated values. Interpolation is made by fitting a series of overlapping parabolas to the data, and follows the method of Snyder (1961). The values are used in moment formulas to compute standard statistical measures. Skewness and kurtosis are reduced by the interpolation procedure, when compared to results from non-interpolated data, and extreme positive values of kurtosis tend to be greatly reduced. The program also picks major modes, the median, and sediment type.

1. Computer Program
2. Grain Size
3. Statistical Method

I. John Schlee
II. Jacqueline Webster
III. Contract No. USGS-14-08-0001-8358

This card is UNCLASSIFIED

Woods Hole Oceanographic Institution
Reference No. 65-42

A COMPUTER PROGRAM FOR GRAIN-SIZE DATA, by John Schlee, Jacqueline Webster. 40 p. August 1965. Contract No. USGS-14-08-0001-8358.

The computer program presented here seeks to improve estimation of statistical parameters for grain-size data by use of interpolated values. Interpolation is made by fitting a series of overlapping parabolas to the data, and follows the method of Snyder (1961). The values are used in moment formulas to compute standard statistical measures. Skewness and kurtosis are reduced by the interpolation procedure, when compared to results from non-interpolated data, and extreme positive values of kurtosis tend to be greatly reduced. The program also picks major modes, the median, and sediment type.

1. Computer Program
2. Grain Size
3. Statistical Method

I. John Schlee
II. Jacqueline Webster
III. Contract No. USGS-14-08-0001-8358

This card is UNCLASSIFIED