MEMOIRS OF THE FACULTY OF SCIENCE, KYOTO UNIVERSITY, SERIES OF GEOL. & MINERAL.,
Vol. XLIV, No. 1, pp. 1-65, 1977

Computer Aided Data Analysis for Geologic Problems

——Data Analysis of Grain-size Distributions of Bottom Sediments——

By

Kaichiro YAMAMOTO

(Received April 27, 1977)

Abstract

Data analysis of grain-size distribution of the recent sediments in Chijiwa Bay, Nagasaki Prefecture, was performed as a case study for showing the usefulness of computer aided data analysis. Multivariate analyses discriminated several sediment types which are confirmed to correspond to distinct depositional environments.

Introduction

Multivariate analysis was recognized to be one of the most effective methods for data analysis in order to interpret the complicated geological phenomena. Its application, however, was not realized until the recent development of the electronic computer, as it involves complicated and enormous computations. Since then, a number of achievements of data analyses have appeared in many fields of science. Furthermore, the computer has promoted the development of many unique applications. However, the applicational techniques of the computer analysis have not yet been settled in geology.

During the last five years I have been developing computer programs at the Data Processing Center, Kyoto University, with collaboration of many colleagues. In this paper, I attempt to historically review data analysis in geology, and describe the outline of the computer supported system which I compiled. Finally I present a case study which shows the usefulness of computer aided data analysis.

Acknowledgement

I wish to express thank to Prof. KAMADA of Nagasaki University for his kind offer of the sedimentological data from Chijiwa Bay, Nagasaki Prefecture, to Dr. Kiyoshi WADATSUMI of Osaka City University for his financial support and encouragement, and to Dr. Shinjiro MIZUTANI of Nagoya University and Dr. Kunihiro ISHIZAKI of Tohoku Univeristy for their suggestion especially on analytical techniques. I also

thank Dr. Yasuo NOGAMI of Primate Research Institute, Kyoto University, for his help with the manuscript.

Special thanks are due to Prof. R. REYMENT of Uppsala University (Sweden), Prof. D. F. MERRIAM of Syracuse University (U. S. A.), Dr. J. E. KLOVAN of University of Calgary (Canada), and Dr. J. M. PARKS of Lehigh University (U. S. A.), for their most fruitful suggestion and constructive criticism.

I am also grateful to Prof. Ichiro MIYAKE of Doshisha University for preparation of the SPSS program in the Data Processing Center, Kyoto University, and introducing me its procedures.

I wish to thank Prof. Keiji NAKAZAWA of our Department, and Prof. Setsuo OKUDA of Disaster Prevention Institute, Kyoto University for their encouragement and fruitful discussion during this work.

The work has been performed by using the facility of Data Processing Center, Kyoto University for the computer analysis. A part of the programs was sponcered by the center for preparation.

Historical Review

The electronic computer, which has much higher speed and more reliable performance than any pre-existed calculator, has made the statistical processing, which involves enormous amount of computations, easier. As the statistical processing becomes more popular by the advent of computers, the geostatistics dealing with mainly the multivariate analysis has been rapidly developed. Although the effectiveness of multivariate analysis was already realized in geology, the analysis was performed by conventional hand-computation (MILLER and KAHN, 1962) and less popular than the univariate analysis, because the number of variables and the amount of data were limited due to a lack of high-speed computers.

Establishment of multivariate analysis for geological use has realized the classification of sediments (IMBRIE and PURDY, 1962) using factor analysis, and numerical taxonomy using cluster analysis (SOKAL and SNEATH, 1963). Many results from these kinds of studies came out in the middle of 60's (MCCAMMON, 1966; PARKS, 1966; WEBB and BRIGGS, 1966).

Although the data analysis using computer enables the composite and objective comprehension of geologic information, the application of multivariate analysis was limited in some divisions of geology because of the descriptive nature of geologic data. In other words, the usage of computer and analytical methods required the data of numeric and qualitative nature. Lately, however, statistical methods dealing with non-parametric data such as original scale and nominal scale have been invented (KRUSCAL, 1964a, 1964b; SAMMON, 1969; CHEETHAM and HAZEL, 1969; HAZEL, 1970), and are applied to the geologic problems (COLLYER and MER-

RIAM, 1973).

In addition to the ordinal calculation and geostatistical use, the computer, which is combined with digital incremental plotters, developed in the late 60's, is utilized for the spatial data handling; for example, contour mapping and triangular diagrams (HARBAUGH and MERIAM, 1968; YAMAMOTO and NISHIWAKI, 1975a). It became common to display β-diagram (ROBINSON, 1963; NOBLE and EBERLY, 1964), π-diagram (WARNER, 1969) and triangular diagram (LUMSDEN, 1973) by computer. Lately the system simulation of sedimentary processes and tectonic movement has been advanced by Harbaugh and Bonham-Carter (1970), Hattori and MIZUTANI (1971), KOMAR (1973), DAVIS and FOX (1972), and YAMAMOTO (1974a).

MIZUTANI (1974) attempted a histrical review of the computer applications to geologic problems in Japan and other countries. Many textbooks and introductory articles are: (1) HARBAUGH and MERRIAM (1968) on computer application; (2) HARBAUGH and BONHAM-CARTER (1970) on computer simulation; (3) VISTELIUS (1967) on geomathematics; (4) BLACKITH and REYMENT (1971) on morphometrics; and (5) DAVIS (1973) on statistics in geology. From 1966 to 1970, Kansas Geological Survey had been publishing the computer contribution series which contain computer programs and their application examples. Many leading articles in the field of mathematical geology appear in the Journal of the International Association for Mathematical Geology that started in 1969, and Computer and Geosciences began in 1975.

According to the research survey conducted by MERRIAM (1969), 91% of scientists who answered the questionaire use the computer in some way, and 39% of them use KGS (Kansas Geological Survey) computer programs for simulation, plotting, correlation analysis, multivariate analysis, and trend surface analysis, and many of them want to obtain the programs concerning the items mentioned above. The report indicates that the computer usage for data analysis and plotting is indespensable in the geological sciences.

Data Analysis in Geology

The problem under study is usually treated through a process shown in Fig. 1.

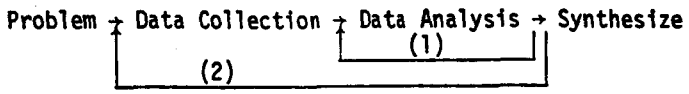


Fig. 1. General process of data analysis.

In the process, the data analysis is used to draw a critical information from the

input data, and to evaluate it. In practice the procedure consists of selection of proper method for analysis, calculation, and evaluation of the result. The procedure is repeated as shown by the arrow in the text-figure, to obtain the final information several analyses. The whole procedure is called computer aided data analysis.

During the procedure, the computer should not only perform the necessary analyses, but also manage the data file by storing the data obtained from each analysis, and by feeding them selectively back to the following analyses. The data analysis requires a well designed package program which is composed of a number of functions to the required procedures and of the function of data management. This demand has promoted the use of SPSS (the Statistical Package for the Social Sciences, NIE *et al.*, 1975), by which the data analysis of geologic use can be carried out (YAMAMOTO and NISHIWAKI, 1975b; NISHIWAKI and YAMAMOTO, 1975).

SPSS is, however, not perfectly applicable to the geological sciences, for the geologic data is somewhat different from that in the social sciences, and most of statistical procedures are not designed for specifically for geological data analysis. The program was modefied to be more applicable for the data analysis of geological use.

Data analysis using computer is divided into three categories:

(1) Recognition of simirality (classification)
ex. factor analysis, cluster analysis, descriminant analysis, variance analysis, t-test, chi-square test, *etc.*

(2) Trend analysis
ex. trend-surface analysis, time-trend analysis, multi-pass filtering, spectral analysis, *etc.*

(3) Plotting and display
ex. contour map, block-diagram, triangular diagram, histogram, scattergram, pi and beta diagram, rose diagram, strike-line map, form line contour, *etc.*

Computer Supported System for Data Analysis of Geologic Problems

In order data analyses, it is necessary for a package program to have both data management functions and sufficient variety of the statistical procedures. In addition, these functions must be simple. I have attempted therefore to compile programs chosen from SPSS, BMD (Biomedical computer Programs Dixon, 1973), BMDP (Biomedical Computer Programs P program, DIXON, 1975) and KGS, in order to achieve more efficient data analysis. This assemblage of programs works as a single system in practical use. The system is called "Computer Supported System for Data Analysis of Geologic Problems" and will be described in the following paragraphs.

The system is designed to fulfil the following requirements:

Sufficient procedure function: The main requirement of data analysis involves the use of statistical analysis. For the general statistical analysis, I have used programs in BMD (DIXON, 1973) and SPSS (NIE, 1973), which are available at the Data Processing Center of Kyoto University. I have also converted three programs, BMDP1M, BMDP2M, and BMDP3M of BMDP (DIXON, 1975), for a FACOM 230–75 computer, in order to perform cluster analysis (Yamamoto, 1974c, 1975a), because no procedure for cluster analysis is available in BMD and SPSS. I have added further programs from the KGS for cluster analysis (PARKS, 1970), cluster analysis of qualitaitive data (BONHAM-CARTER, 1967), and calculation of π-diagram (WARNER, 1969). In addition to the programs mentioned, I have included programs designed by myself in the system.

Data mangement function: This facility is required to store the analyzed data in data files, and to maintain them for retrieval and processing. The result of one analysis will be registered in data file for the following analysis, if necessary. This ensures the data management, selection of proper data, and transforming them. It is indespensable, when we deal with a large amount of data, and operate the reported feed-back procedures. SPSS facilitates this ability.

Program of easy usage: It is important that we can prepare, withought difficulty, the control parameters which inform a program of the contents of procedures. Since data analysis generally involves a considerable number of procedures, this is an essential factor in performing the analysis effectively. I have used the programs of SPSS and BMDP, as they are designed to require no apecific language for the control parameters, and alllow free format data input.

Interchangeability of data among the programs: When we perform a series of calculations, the output of one procedure is often input into the following analysis. This process requires the interface between SPSS and other programs. I have employed a temporary data file to establish a network among the programs, to make them function as a system, and I have modified programs which could only use punch cards as I/O (input/output) medium, to utilize the other modes of I/O medium (i. e., tape or disk).

Throughout the operations, programs from SPSS, BMD, BMDP, and KGS, and the other programs comprise a system. The structure and function of the system is described in Fig. 2.

The system is composed of two main facilities, as illustrated in Fig. 2; the control of data file which consists of a SPSS system file, and the three subsystems which perform the data analysis. The data file will be treated by the data processing subsystem, and the result is displayed by the display subsystem. The data transfor-

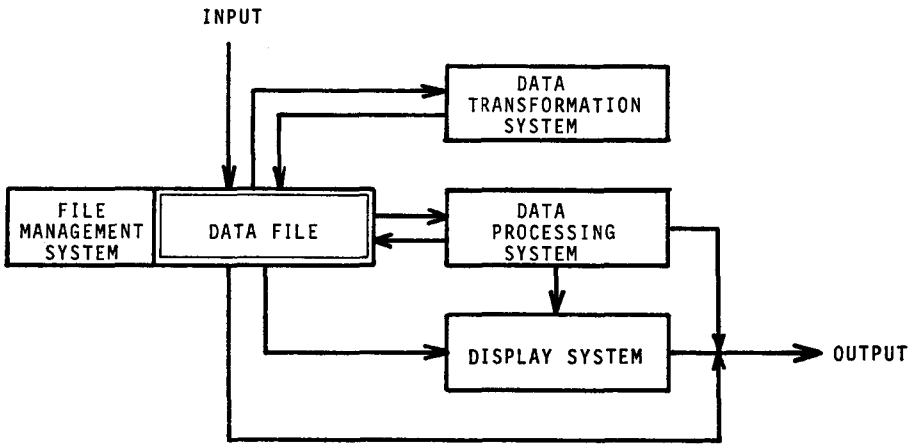


Fig. 2. Schematic illustration of the computer supported system for the data analysis of geologic problems.

mation subsystem transforms the data by interpolation, when necessary. The flow of the data among the subsystem can be seen in Fig. 2.

Since SPSS can only accept the data in matrix form, no program that employs the data in the other forms can be incorporated in the system. In practice, however, this limitation poses no serious difficulty, for the data are usually prepared in matrix form, and most of data can be described in the form.

The contents of each subsystem are follows:

The datafile management subsystem: Functions of the data file management of SPSS include storage, retrieval, selection and transformation of data (NIE, *et al.*, 1973; YAMAMOTO and NISHIWAKI, 1975b: NISHIWAKI and YAMAMOTO, 1975). The SPSS system file utilized as data file storage consists of two components; reference information about the data matrix such as size of the matrix and labels, and raw data in matrix form i. e. variable by sample (called a case). As the subsystem can accept data form and provide data to the other programs, it manages the data file for the whole system (NIE, *et al.*, 1975, p. 80).

The data processing subsystem: This is a subsystem to perform mainly the statistical procedures. It contains the subprograms of statistical procedures of SPSS (NIE, *et al.*, 1975, p. 96), and the statistical package programs of BMD (DIXON,1973) and BMDP (DIXON, 1975). Besides, the other programs are supplied into the subsystem; Q-mode cluster analysis using principal component analysis (PARKS, 1970) and cluster analysis of qualitative data (BONHAM-CARTER, 1967) from KGS, and Q-mode factor analysis for large-scale data (KLOVAN and IMBRIE, 1971), and non-linear map-

ping algorithm (SAMMON, 1969; HOWARTH, 1973). I have added a few analytic programs designed by myself; trend-surface analysis, with polynomial and double Fourier approximations (YAMAMOTO, 1973a), trend-surface analysis with moving averages and maps of summit and river levels (YAMAMOTO and NISHIWAKI, 1975c), time trend analysis of stratigraphic sequences (YAMAMOTO and NAKAGAWA, 1974).

The display subsystem: The subsystem displays the data in files and output from analytic procedures, according to the plotting programs, by X–Y plotter, line printer, typewriter, and other media. The programs in SPSS, BMD, and BMDP, for preparing histograms and scattergrams are included in the subsystem. I have designed the subsystem programs for preparing contour map with an X–Y plotter (YAMAMOTO and NISHIWAKI, 1975a) and with a line printer (YAMAMOTO, 1976), for checking the input data of contour program (YAMAMOTO and NISHIWAKI, 1975c), for perspective plotting in three-dimensional space (YAMAMOTO, 1973a), for preparing triangular contour diagrams, columnar sections, strike-line map, and for estimation of structure contours from strike-dip data and mapping (YAMAMOTO and NISHIWAKI, 1975d).

Case Study

Data analysis of grain-size distribution of the recent marine sediments was performed by statistical and computer techniques, to evaluate the computer supported system in practical use. The analytical procedures and results are described and illustrated as a flow diagram. The results are also compared with those studied previously.

The case study deals with the unconsolidated bottom sediments of the Chijiwa Bay, Nagasaki Prefecture, southwest Japan. All the original data are analyzed by KAMADA (1966).

To show the flow and the results of the analysis as clearly as possible, the procedures adopted in each step of the analysis are not discussed in detail. Those will be published separately. The sedimentary environment is not considered, for it is beyond the scope of the data analysis.

Data Analysis of Grain-size Distribution of the Recent Sediments of Chijiwa Bay

Several papers have been published on the bottom sediments of the Chijiwa (or Tachibana) Bay (see locality map, Fig. 3). KAMADA (1966) classified the bottom sediments and estimated their sedimentary environments. INOUE (1970) discussed the relation-ship of the bottom sediment distribution to the current system in the bay. KAMADA, *et al.* (1973) used the statistical parameters of grain-size distribution

(TRASK, 1932), and classified the bottom sediments according to the standard of INMAN and CHAMBERLEIN (1955) into:

Type 1: composed of highly sorted fine sand; medium phi 2.0 to 3.0.

Type 2: characterized by sandy constitution; median phi 0 to 3, sorting coefficient 1.25 to 3.0 in general; subdivided into 2A and 2B.

Subtype 2A: composed of coarser grains; median phi about 2.0; skewness more than 1.0.

Subtype 2B: composed of finer grains; containing more than 5% muddy materials; skewness less than 1.0.

Both types are gradually merge into each other near the point, where skewness is 1.0.

Type 3: muddy sediments; median phi 3.0 to 8.0; subdivided by 4ϕ and 6ϕ median into 3A, 3(s–s), and 3B subtypes. It is a character of the sediments

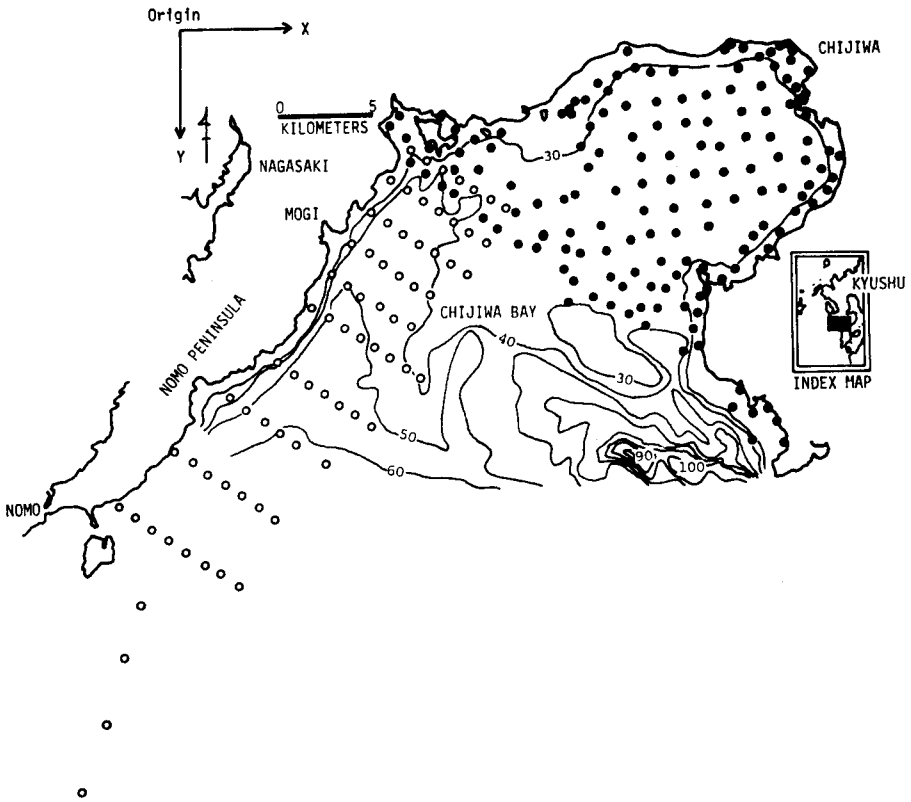


Fig. 3. Location of sampling stations: open circles represent the points of No. 1001 to 1075, and solid circles those of No. 2001 to 2160.

belonging to 3A and 3(s–s) that the subtypes are less than 1.0 in skewness. The sorting coefficient and its deviation increase, correlating to the median phi, but the sorting coefficient concentrate in 3.0 to 4.0 near the boundary between the subtype 3A and type 4.

Type 4: containing more than 50% muddy materials; median phi more than 8.0.

Type 5: composed of pebbly material; median phi less than 0.0 (more than 1.0 mm in diameter); generally containing abundant fragments of organic origin.

In addition to recognizing these types of sediments, KAMADA *et al.*, (1973) confirmed the areal distribution of these types in the bay, as shown in Fig. 4.

All these data were kindly offered by Prof. KAMADA for this case study. Using multivariate analyses and several kinds of display techniques, data analysis of grain-size distributions was performed on Recent bottom sediments in the Chijiwa Bay.

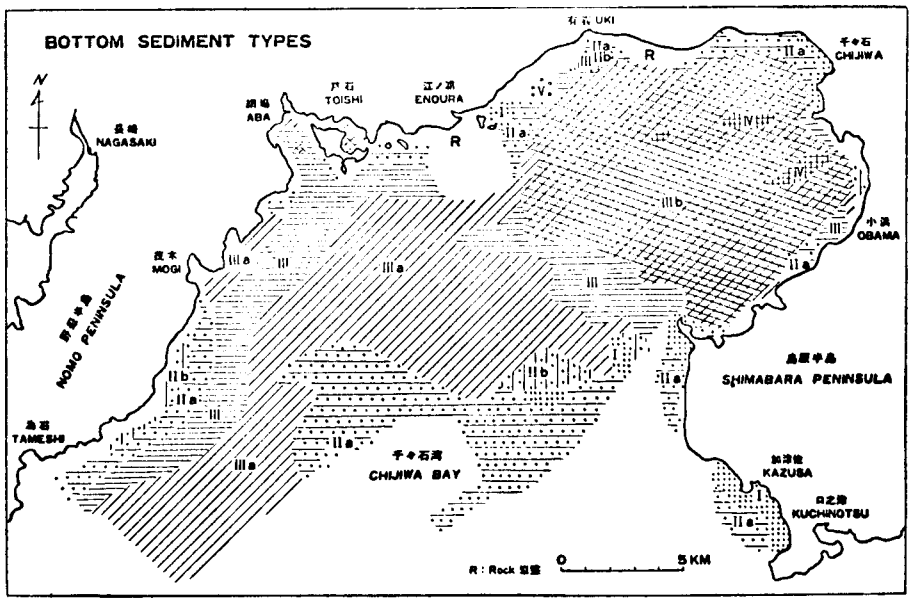


Fig. 4. Areal distribution of the sediment types defined by Kamada et al. (1973).

Procedure

The data analysis on grain-size distribution of the Recent bottom sediments in the Chijiwa Bay was carried out using the computer supported system. The flow of whole procedures and applied methods are shown in Fig. 5. Main procedures adopted at each step are summarised in Table 1.

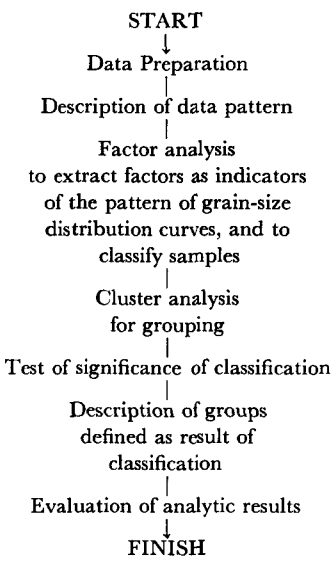


Fig. 5 Flow chart of data analysis.

Table 1. Summarized table of procedures at each step of data analysis, Chijiwa Bay.

Step 1. Preparation of data, original and transformed
* Data transformation

Step 2. Description of data pattern
* Preparation of scatter diagrams between each pair of statistical and compositional parameters
* Preparation of histograms on sediment types defined by Kamada et al. (1973)
* Preparation of triangular contour diagrams of sand-silt-clay system
* Preparation of contour maps of water depth, median phi, gravel and clay contents

Step 3. Factor analysis
* R-mode factor analysis
* Q-mode factor analysis

Step 4. Cluster analysis
* Preparation of dendrogram
* Simplification of dendrogram

Step 5. Test of significance of classification
* Test for multivariate equality of means among groups
* Computation of F-distances among groups
* Discriminant analysis

Step 6. Description of group
* Breakdown
* Preparation of triangle contour diagram for each group

Step 7. Evaluation of analytic results
* Crosstabulation between two classifications, by Kamada et al. (1973) and this analysis

Step 1. Preration of orginal and transformed data

The data to be stored in the SPSS system file should be in "case-variable" matrix form (see Fig. 6). A case is the basic unit of analysis for which measurements have been obtained. In this study, the cases correspond to the samples collected from the bay bottom; No. 1001 to No. 1075 and No. 2001 to No. 2160 (see locality map Fig. 3).

CASE (Sample)	VARIABLE (Measurement) 1	2	3	4	5	6	7
0001	1.5	A	1	12.3	6	5	AA
0002	2.3	A	3	13.5	8	5	AC
0003	1.4	B	6	23.4	3	6	AD
0004	0.8	C	2	18.1	2	4	BA
0005	3.1	B	8	43.6	9	9	BD

Fig. 6. Data form (matrix form).

Each case is composed of values for one or more measurements that have been taken. These measurements are termed variables, and each case within a study will have one value for each of variables. For example, the variables correspond to the median size, sorting coefficient, skewness and others in this study.

To each of variables is given a name which is used as index (see, Table 2).

Table 2. Processing steps and data stored into data file of Chijiwa Bay.

Step	Procedure	Variable Name	Attribute
1	Preparation of original data	STANO	Station number of sampling point
		X	x-coordinate of station
		Y	y-coordinate of station
		DEP	water depth at station
		MDMM	median diameter in mm
		MDPH	median diameter in phi-scale
		SO	sorting coefficient (Trask, 1932)
		SK	skewness (Trask, 1932)
		G	gravel contents, % (SHEPARD, 1954)
		SA	sand contents, % (SHEPARD, 1954)
		SI	silt contents, % (SHEPARD, 1954)
		CL	clay contents, % (SHEPARD, 1954)
		MUD	mud contents, %, SI+CL
		VAR001	weight % of grains coarser than -2ϕ
		VAR002	between -2ϕ and -1ϕ
		VAR003	between -1ϕ and 0ϕ
		VAR004	between 0ϕ and 1ϕ
		VAR005	between 1ϕ and 2ϕ

Table 2. Continued.

Step	Procedure	Variable Name	Attribute
		VAR006	between 2ϕ and 3ϕ
		VAR007	between 3ϕ and 4ϕ
		VAR008	between 4ϕ and 5ϕ
		VAR009	between 5ϕ and 6ϕ
		VAR010	between 6ϕ and 7ϕ
		VAR011	between 7ϕ and 8ϕ
		VAR012	between 8ϕ and 9ϕ
		VAR013	between 9ϕ and 10ϕ or finer than 9ϕ
		VAR014	finer than 10ϕ
		TYPE	sediment type defined by KAMADA et al. (1973)
2	Data transformation	VAR101	phi percentile of 5%
		VAR102	of 10%
		VAR103	of 15%
		VAR104	of 20%
		VAR105	of 25%
		VAR106	of 30%
		VAR107	of 35%
		VAR108	of 40%
		VAR109	of 45%
		VAR110	of 50%
		VAR111	of 55%
		VAR112	of 60%
		VAR113	of 65%
		VAR114	of 70%
3	R-mode factor analysis	F1	first R-mode factor score
		F2	second one
		F3	third one
	Q-mode factor analysis	QFL1	first Q-mode factor loading
		QFL2	second one
		QFL3	third one
		QFL4	fourth one
	Classification	GROUP	code of groups defined on the basis of factor analysis
4	Classification by cluster analysis	CAG3	code of groups defined on the basis of cluster analysis
		CAG3D	code of groups, more detail one of CAG3

In this study, the cases are diveded into two groups termed subfiles, in which the cases of No. 1001 to 1075 and the cases of No. 2001 to 2160 are stored separately.

The data stored into the file are composed of the following variables.

Geographic variables: coordinate value, depth and number of sampling points. The coordinate value is fixed by the coordinate system (Fig. 3).

Statistical parameters (KAMADA *et al.*, 1973): median size (in millimeter and phi scale), sorting coefficient and skewness defined by TRASK (1932). They are not only used as indices to show some characters of the grain-size distribution, but also in this instance, for the sake of comparison between the previous and present methods.

Compositional parameters (KAMADA *et al.*, 1973): gravel, sand, silt, clay (SHEPARD, (1954) and mud (silt+clay) contents. They are used to reveal some characters of the grain composition.

Sediment types (KAMADA *et al.*, 1973): according to INMAN and CHAMBERLAIN (1955).

Frequency scores: percentages of every unit phi, ranging from -2ϕ to 9 or 10ϕ and both tails, by KAMADA. The measurements are analysed in this study.

Phi percentiles: percentiles of every 5%, ranging from 5 to 70%. The values are computed from the cumulative proportion by linear interpolation. The grain-size distribution was observed on particles up to 10ϕ. The percentiles higher than 70% can not be calculated for most samples.

Step 2. Description of data pattern

Several tables and figures are produced to extract some characters from the data to be analyzed. With the aid of the tables and figures produced, the analytic methods employed in the following steps are to be selected. In this case study, the following ones are prepared (see Table 1).

Scatter diagram: a diagram of data points based on two selected variables as shown in Fig. 7. In this study, two kinds of scatter diagrams are produced using the SPSS statistical procedure SCATTERGRAM; one is based on each pair of median, sorting coefficient and skewness parameters and the other on each pair of sand, silt and clay content values.

Histogram: an illustration of the absolute or relative frequencies as shown in Fig. 8. In this case study, three kinds of histograms on the sediment types are obtained using

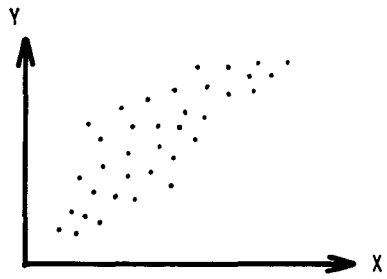


Fig. 7. Scatter diagram.

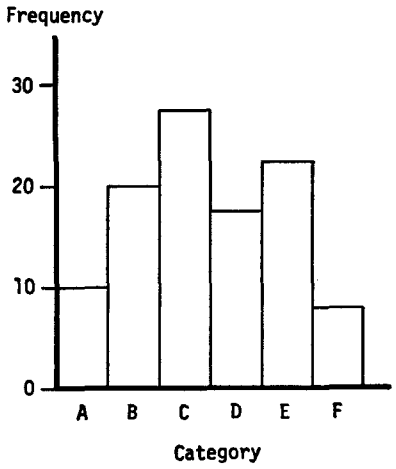


Fig. 8. Histogram.

the SPSS statistical procedure FREQUENCIES; one dealing with all the cases (samples), the second with the cases of station No. 1001 to 1075, and the third with the cases of station No. 2001 to 2160. Further, the histogram on grain-size distribrtion is produced for each case by using a specific program prepared for this study.

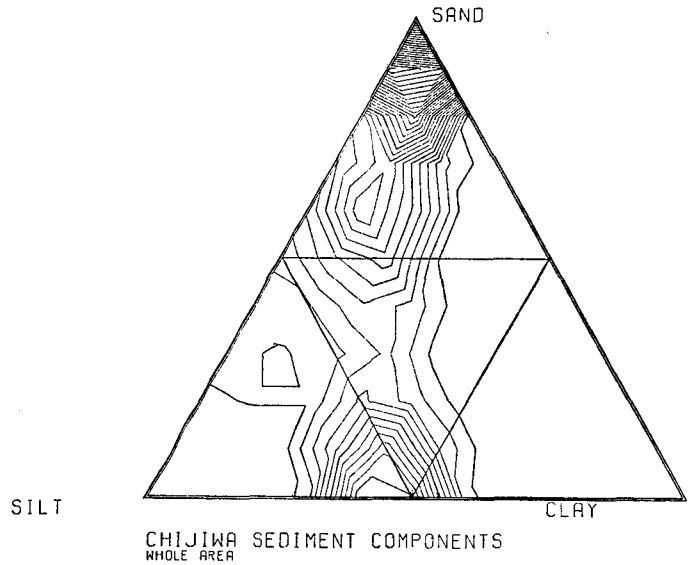


Fig. 9. Triangular contour diagram, showing concentration pattern of sample points based on three variables.

Triangular contour diagram: a contour map showing concentration pattern of data points in a three component system based on three variables, as shown in Fig. 9.

Contour map: In this case study, four contour maps are prepared using the CMPP2 program (YAMAMOTO and NISHIWAKI, 1975c); on the water depth of sampling points, median ϕ, gravel and mud contents.

Step 3. Factor analysis

Factor analysis is characterized by distinctive data reduction capavility. Given an matrix of correlation coefficients for a set of variables or cases, factor analysis techniques show whether some underlying patterns of relationships exist, so that the data may be reduced to a smaller set of factors or components.

If factor analysis is applied to a matrix of correlations between cases, it is called Q-mode factor analysis, while the more common variety based on correlations among variables is known as R-mode factor analysis.

The factor analysis that subsumes a fairly large variety of procedures is carried out in four stages: (1) calculation of the correlation matrix, (2) extraction of the principal factors, (3) rotation to a terminal solution, and (4) estimation of factor scores.

The first stage involves a calculation of appropriate measures of association for a set of relevant variables or cases. The second stage is to compute eigenvalues and eigenvectors of the correlation matrix. The eigenvalue indicates the degree of representation of the extracted factor to the information exhibited in the original data. The eigenvalues are cumulated in order of factor extractions and is listed in values relative to the information content of the original data, termed cumulative proportion of total variance. The eigenvector shows the relation of the extracted factors to each variable or case. Appropriate factors are then selected, considering the eigenvalue and the cumulative proportion of total variance. The eigenvectors of the selected factors are expressed in a "variable/case-factor" matrix form, that is called the principal factor loading matrix.

In the third stage the matrix is rotated to search for simple and interpretable factors. Each variable or case is made exclusively related to a specific factor, during the rotation. The relation between them is shown by a large factor loading. Each factor will be more easily interpreted from the variable(s) or case(s) related to it. The cases can be classified into the groups related to factors.

In the fourth stage each factor attribute is numerically estimated, that is termed factor score. The score is given to each case or variable in R- or Q-mode factor analysis.

In this case study, the following factor analyses are carried out (see Table 1).

R-mode factor analysis on phi percentiles: DAVIS (1970) performed the R-mode factor analysis of frequency scores. The R-mode factor analysis on phi percentiles is chosen to extract factors, which may adequately reflect some characteristics of patterns observed in the cumulative frequency curve of grainsize distribution. The BMD08M program is used for this analysis.

Q-mode factor analysis on frequency scores: KLOVAN (1966) and DRAPEAU (1973) perfomed a classification using the Q-mode factor analysis of frequency scores. In this study the same method is adopted, and the CABFAC program (KLOVAN and IMBRIE, 1971) is used.

Factor scores, factor loadings, and the results of the R- and Q-mode factor analyses are stored in the data file to be used in the following steps (see Table 2).

Step 4. Cluster analysis

Cluster analysis is used to classify the cases (Q-mode) or the variables (R-mode) into groups. Among various methods available procedure was selected in this study, in order to classify the cases (samples).

The distances (i. e. dissimilarity) among the cases are computed first. For example, if two cases (i_1 and i_2) have $X_{i_1 j}$ and $X_{i_2 j}$ variables ($j=1 \sim p$), respectively, the distance between two cases is:

$$d_{i_1 i_2}=\left[\sum_{j=1}^{p}(x_{i_1 j}-x_{i_2 j})^2\right]^{\frac{1}{2}}$$

The distances among the cases are expressed in "case-case" matrix form named the distance matrix. Based on the distance matrix, the similarity among the cases are represented as a hierarchical dendrogram shown in Fig.10. The cases correspond to the branches at the left ends of the tree on the dendrogram, and the branches are united to the root.

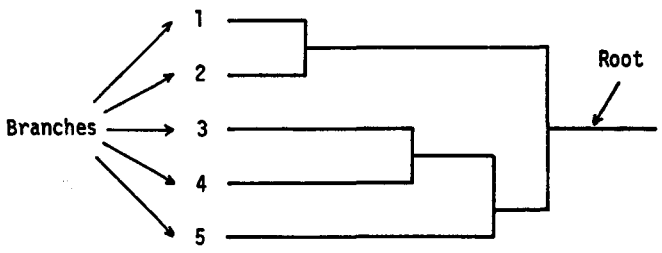


Fig. 10. Dendrogram of cluster analysis. Ends of branches correspond to the samples (cases).

Cluster analysis has been successfully applied not only to paleontology, such as the numerical taxonomy of species (SOKAL and SNEATH, 1963), classification of faunal assemblages (HAYAMI and NAKANO, 1968; HAZEL, 1970; UJIIE, 1973), but also to sedimentary petrology, such as the classification of limestones (PARKS, 1969). It has, however, rarely been applied to the grain-size distributions.

In this case study, the chi-square distance based on frequency score is adopted as a trial, and the computed results are stored into the data file to be used in the following steps. The results of classification are stored in the data file (see Table 2).

Step 5. Test of classification

The results of factor analysis and cluster analysis are compared with the interpretation by KAMADA *et al.* (1973). The classified groups make their own assemblages in multidimensional space, where each variable used for the classification defines an axis, as shown diagramatically in Fig. 11. To test whether the groups are clearly separated in the space, discriminant analysis is applied to the classification.

Equality of means and F-value distances are examined with discriminant analysis.

Test for multivariate equality of means: to test the hypothesis that all the center of groups are same in the multidimensional space mentioned above, using student t-test. If the hypothesis is verified, the classification is, in turn, considered to be invalid statistically.

Calculation of F-value distances among the groups: to compute the distance of each pair of assemblages in the space, and in practice, to search for the distance from the center of an assemblage to that of the other. The larger the distance value is, the more suitable the classification is between the two assemblages.

Discriminant analysis: to re-classify on the basis of the distance from each case to the center of each group. The distance is determined taking the group dispersion

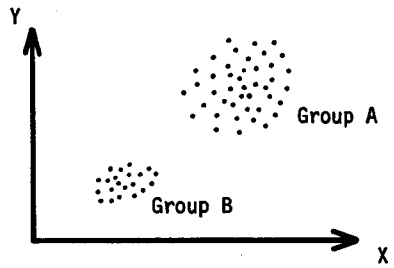


Fig. 11. Groups in two dimensional space.

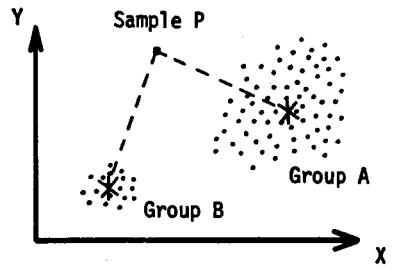


Fig. 12. Descrimination: sample P is classified into the group A, when the group A is larger than the B, even when the distances are same.

into account. For instance, if the distances from a sample to two groups are same in absolute value, the sample should be allied to more widely distributed group, as shown in Fig. 12. According to this standard, each case is discriminated into the most probable group. The fewer incorrectly discriminated cases, the more suitable the classification is. Of course, a group from which few cases are re-classified into another is regarded to be clearly separated.

Step 6. Description of groups

According to the clsssification, which is obtained by the cluster analysis, some groups are defined, and their characteristics are examined. For the purpose are produced some tables and figures, the breakdown, and the triangle contour map of each group.

Breakdown: to compute the means and the deviations for statistical parameters and compositional parameters, and to test the univariate equality of means for F-values, using SPSS statistical procedure BREAKDOWN. The larger the F-values are, the clearer the differences of the parameters among the groups.

Triangle contour diagram: to search for the concentration pattern of cases on the sand-silt-clay triangle diagram for each group, using the TRICON program.

Step 7. Evaluation of analytical results

To compare this classification with that done by KAMADA, *et al.* (1973), the crosstabulation table is produced on both classifications by using the SPSS statistical procedure CROSSTABS. The table indicates the joint frequency of each group for each classification, as shown in Fig. 13. A case, which is belonging to the "A" group by classification B and to the "b" group by classification A, is printed in the crossing area of "A" group row and "b" group column.

		Classification A			
		a	b	c	d
Classification B	A	27	8	2	3
	B	0	11	5	4
	C	3	1	12	21

Fig. 13. Crosstabulation table: joint frequency between two classifications are shown in squares.

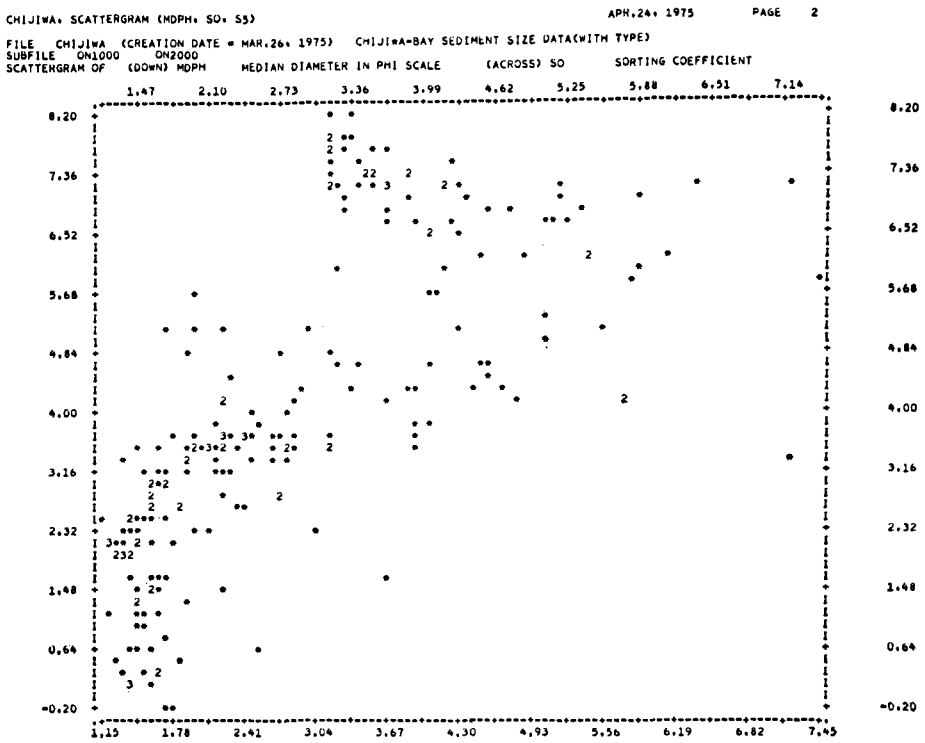
Results of Analysis

Description of data pattern

Scatter diagram: In the scatter diagrams based on each pair of median (MDPH), sorting coefficient (SO) and skewness (SK) (Fig. 14), a positive correlation between

MDPH and SO is confirmed (Fig. 14a), but no definite assemblage is observed.

On the other hand, in the scatter diagrams based on each pair of compositionsl parameters, a negative correlation between SA and CL (Fig. 15b) and a positive one between SI and CL (Fig. 15c) are confirmed. As shown in Figs. 15a and b, the samples contain neither silt nor clay, if their sand contents exceed 90%. They are almost constant in silt quantity, but increase exceedingly in clay, if their sand contents are less than 30%. The SI–CL scatter diagram indicates a tendency that



CHIJIWA, SCATTERGRAM (MDPH, SO, S5) APR,24, 1975 PAGE 3

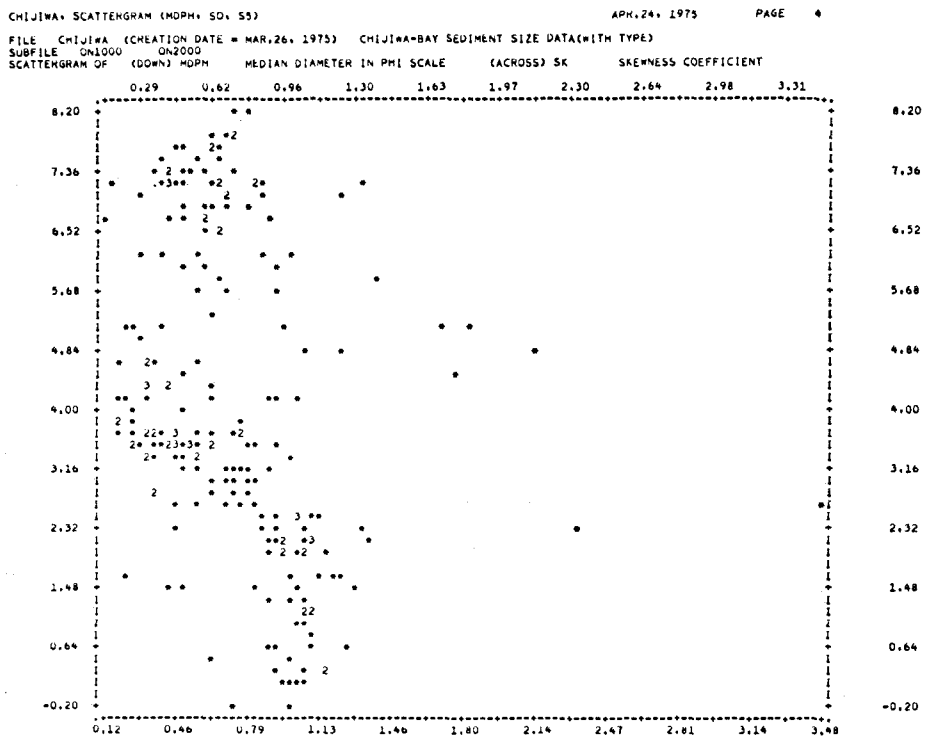
STATISTICS,,

CORRELATION (R)=	0.70355	R SQUARED =	0.49499	SIGNIFICANCE =	0.00001
STD ERR OF EST =	1.55910	INTERCEPT (A) =	0.91374	SLOPE (B) =	1.13520
PLOTTED VALUES =	235	EXCLUDED VALUES=	0	MISSING VALUES =	0

'********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

(a)

Fig. 14. Scatter diagram among statistical parameters: (a) median against sorting(MDPH–SO); (b) median against skewness (MDPH–SK); (c) sorting against skewness (SO–SK).



CHIJIWA, SCATTERGRAM (MDPH, SO, SS) APR.24, 1975 PAGE 5

STATISTICS..

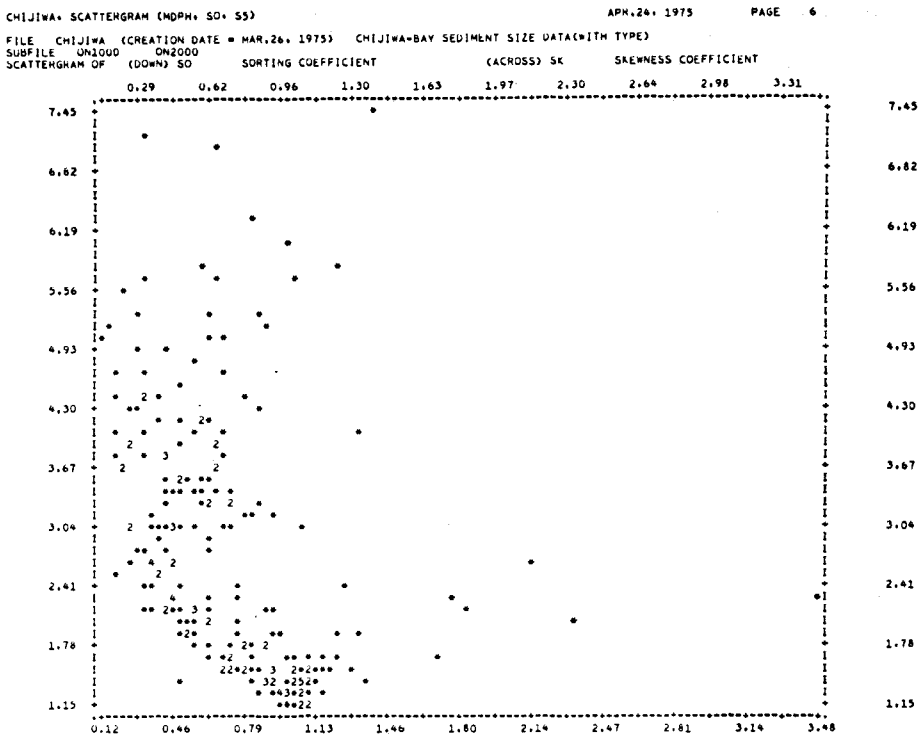
CORRELATION (R)=	-0.31061	R SQUARED =	0.09648	SIGNIFICANCE =	0.00001
STD ERR OF EST =	2.08541	INTERCEPT (A) =	5.38531	SLOPE (B) =	-1.73873
PLOTTED VALUES =	235	EXCLUDED VALUES=	0	MISSING VALUES =	0

'********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Fig. 14. Continued (b)

the samples concentrate on two lines, which cross each other near a point of 35% silt and 15% clay. The silt contents are mainly increasing, so far as the mud contents are limited within the point. On the other hand, only the clay contents remarkably increase, if the mud contents exceed the point. These facts suggest that the bottom sediments are classifiable on the basis of grain-size distribution, and safely grouped into several assemblages corresponding to sedimenary enviroments.

Histogram: In the histogram of frequency distribution of sediment types (Fig. 16),



CHIJIWA, SCATTERGRAM (MDPH, SO, SS) APR.24, 1975 PAGE 7

STATISTICS..

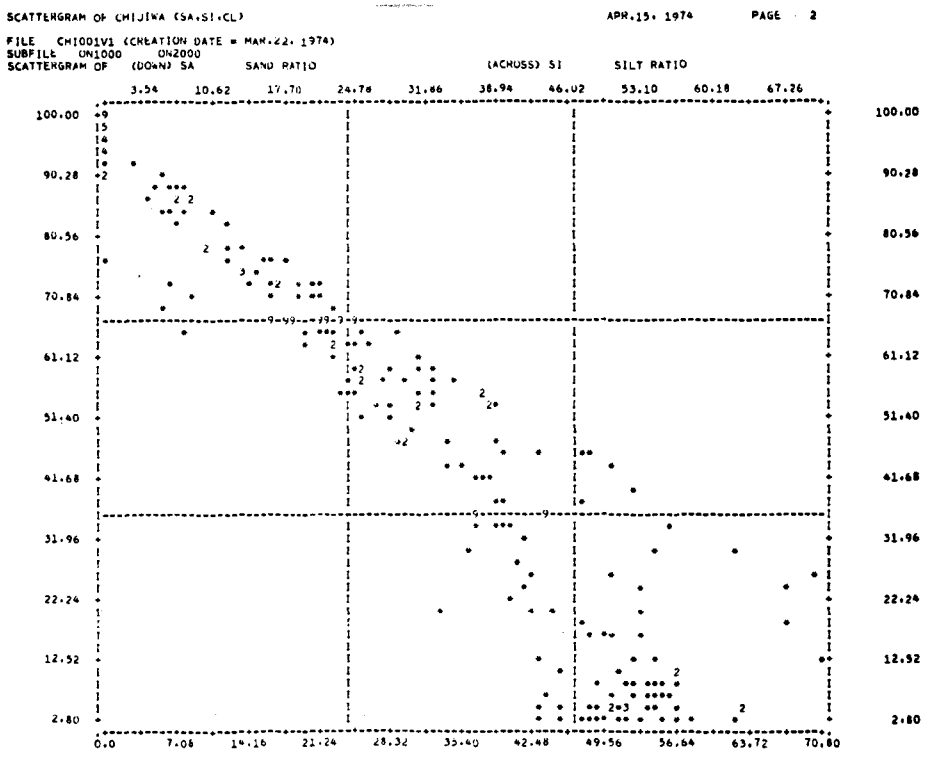
CORRELATION (R)=	-0.37205	R SQUARED =	0.13842	SIGNIFICANCE =	0.00001
STD ERR OF EST =	1.26210	INTERCEPT (A) =	3.76415	SLOPE (B) =	-1.29077
PLOTTED VALUES =	235	EXCLUDED VALUES=	0	MISSING VALUES =	0

'*********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Fig. 14. Continued (c)

Fig. 16c indicates more variety of sediment types than Fig. 16b does. The collecting area of samples No. 1001 to No. 1075 seems to be different somewhat in environment from that of the samples No. 2001 to No. 2160.

The figures on grain-size distribution (Fig. 17) illustrate that the samples may be divided into a few groups on the basis of the curve patterns. Then, the curve patterns are compared with each other by using the multivariate analysis, in order



(a)

Fig. 15. Scatter diagram among compositional parameters: (a) sand vs silt (SA–SI); (b) sand vs clay(SA–CL); (c) silt vs clay (SI–CL).

to classify the samples objectively according to the frequency distribution of grain-size.

Triangle contour diagram: In the sand-silt-clay contour diagram (Fig. 18), some concentrated area are confirmed. Fig. 18b is somewhat different in pattern from Fig. 18c. Namely, the both figures show a common concentrated center, but the other center different; one is situated on the side of less clay content, while the other is on the side of less sand content. It suggests that the collecting area of the samples No. 1001 to No. 1075 is not the same in environment as that of the samples No. 2001 to No. 2160.

Contour map: The bathmetric map constructed by computer contouring (Fig. 19)

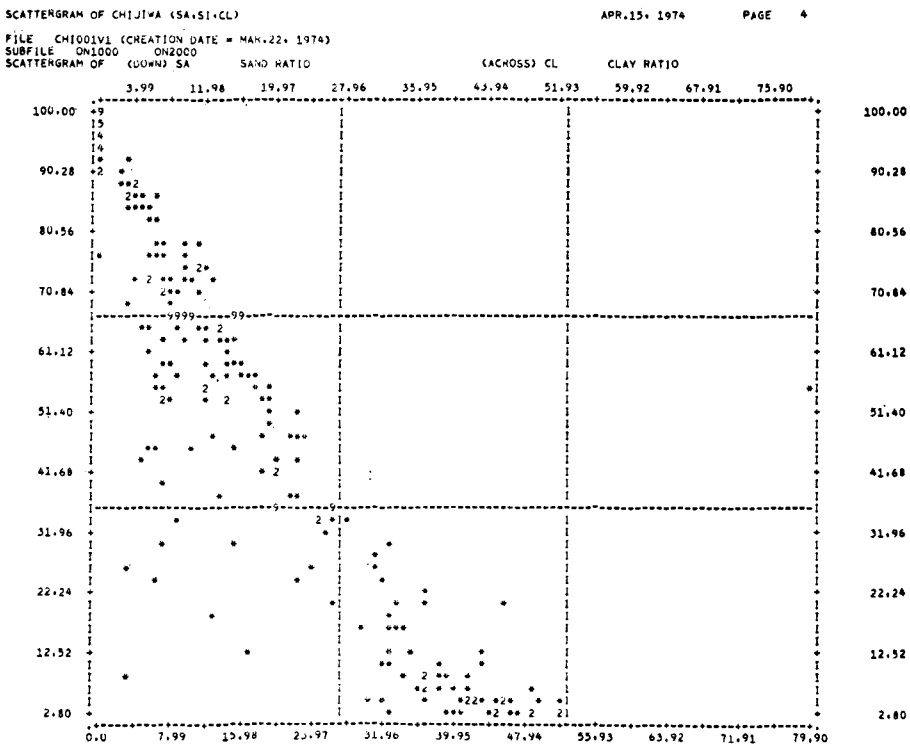


Fig. 15. Continued (b)

shows a ridge structure in the central portion of the bay, surrounded a flat bottom, and a submarine canyons also along the Nomo Peninsula. The samples No. 1001 to No. 1075 are collected from the submarine canyons, and those No. 2001 to 2160 from the central portion of the bay (see Fig. 3). The geomorphologic defferences of the bottom for both sampling areas should have influence on the characteristics of sediments.

The contour maps on the distribution of median phi scale, gravel contents and mud contents (Figs. 20 to 22) show that the mud contents (Fig. 22) are rather high in the nearshore area, and indicate that an spatial distribution pattern is correlatable to that of median phi (Fig. 20). The distribution of gravel contents are pathy in certain limited areas along the seashore line and the Nomo Peninsula (Fig. 21).

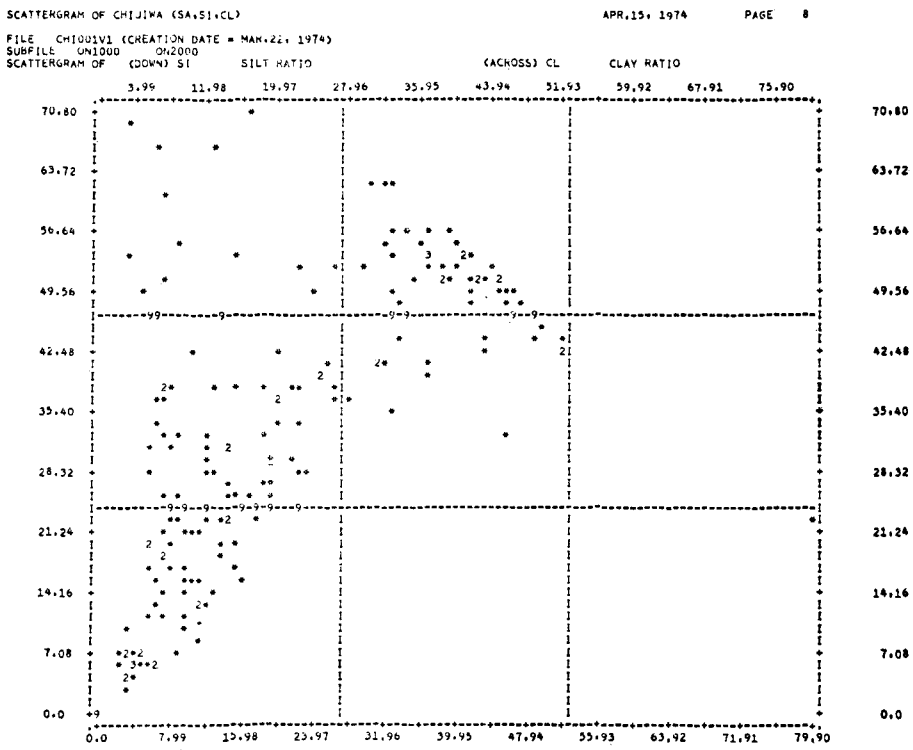
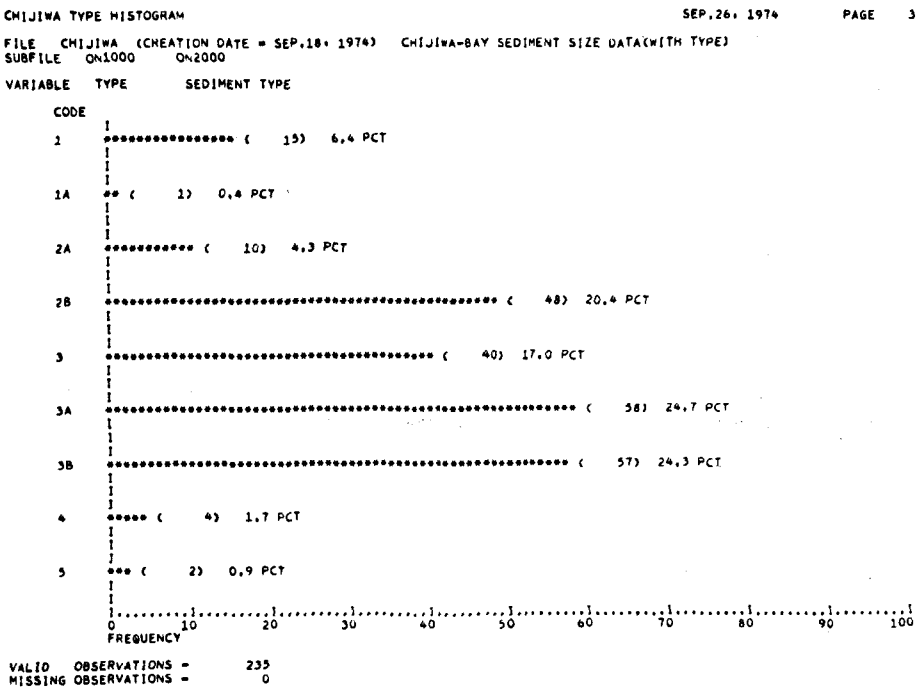


Fig. 15. Continued (c)



(a)

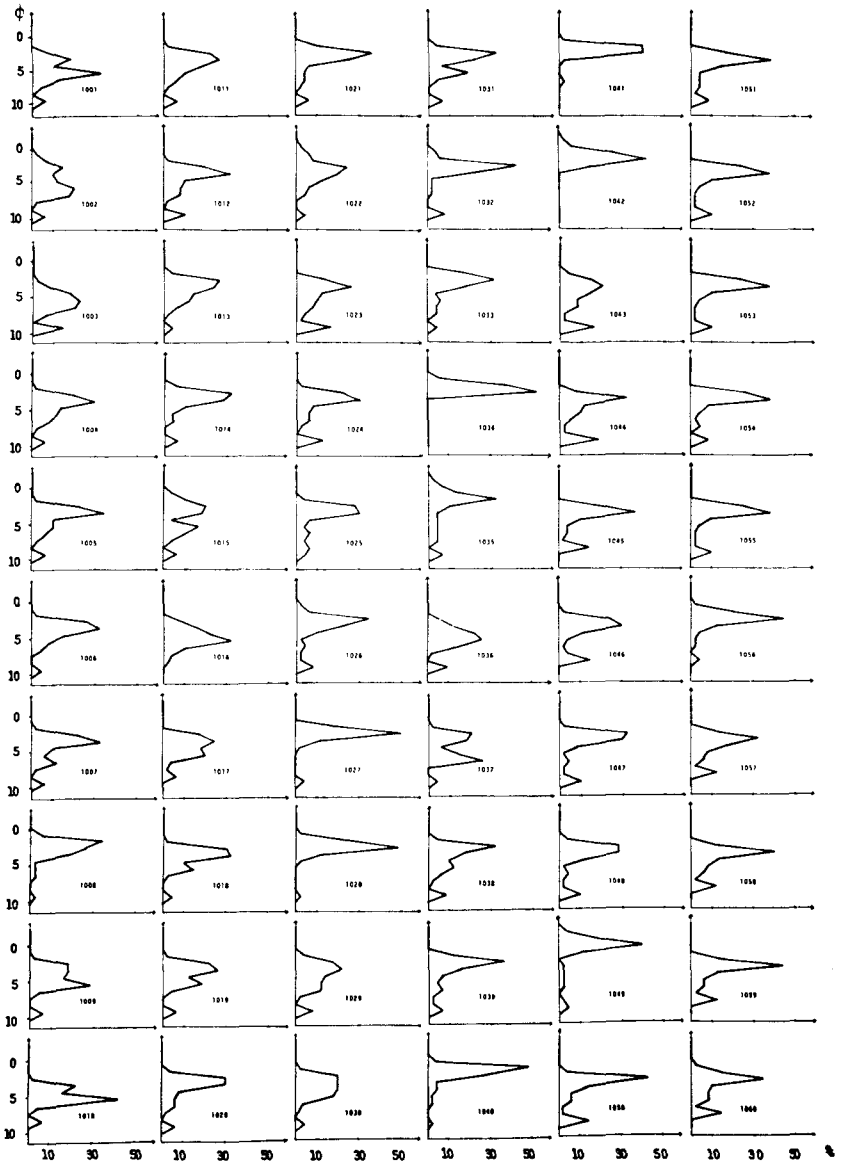
Fig. 16. Frequency of the sediment types defined by Kamada et al. (1973): (a) whole samples; (b) samples No. 1001 to 1075; (c) samples No. 2001 to 2160.

```
CHIJIWA TYPE HISTOGRAM                                                    SEP,26, 1974        PAGE    6

FILE   CHIJIWA  (CREATION DATE = SEP,18, 1974)   CHIJIWA-BAY SEDIMENT SIZE DATA(WITH TYPE)
SUBFILE   ON1000

VARIABLE   TYPE       SEDIMENT TYPE

      CODE
           I
      1    *** (     1)   1,3 PCT
           I
           I
           I
      2A   ********* (     4)   5,3 PCT
           I
           I
           I
      2B   **************************************** (    18)  24,0 PCT
           I
           I
           I
      3    ********************************** (    16)  21,3 PCT
           I
           I
           I
      3A   ********************************************************************************** (    36)  48,0 PCT
           I
           I
           I.........I.........I.........I.........I.........I.........I.........I.........I.........I
           0         5        10        15        20        25        30        35        40        45        50
           FREQUENCY

VALID   OBSERVATIONS =         75
MISSING OBSERVATIONS =          0
```

Fig. 16. Continued (b)

```
CHIJIWA TYPE HISTOGRAM                                                    SEP,26, 1974        PAGE    8

FILE   CHIJIWA  (CREATION DATE = SEP,18, 1974)   CHIJIWA-BAY SEDIMENT SIZE DATA(WITH TYPE)
SUBFILE   ON2000

VARIABLE   TYPE       SEDIMENT TYPE

      CODE
           I
      1    *************** (    14)   8,7 PCT
           I
           I
           I
      1A   ** (     1)   0,6 PCT
           I
           I
           I
      2A   ******* (     6)   3,7 PCT
           I
           I
           I
      2B   ********************************* (    30)  18,7 PCT
           I
           I
           I
      3    ************************* (    24)  15,0 PCT
           I
           I
           I
      3A   ************************ (    22)  13,7 PCT
           I
           I
           I
      3B   ********************************************************** (    57)  35,6 PCT
           I
           I
           I
      4    ***** (     4)   2,5 PCT
           I
           I
           I
      5    *** (     2)   1,2 PCT
           I
           I
           I.........I.........I.........I.........I.........I.........I.........I.........I.........I
           0        10        20        30        40        50        60        70        80        90       100
           FREQUENCY

VALID   OBSERVATIONS =        160
MISSING OBSERVATIONS =          0
```

Fig. 16. Continued (c)



(a)

Fig. 17. Frequency distribution of grainsize, samples from the bottom of the Chijiwa Bay.

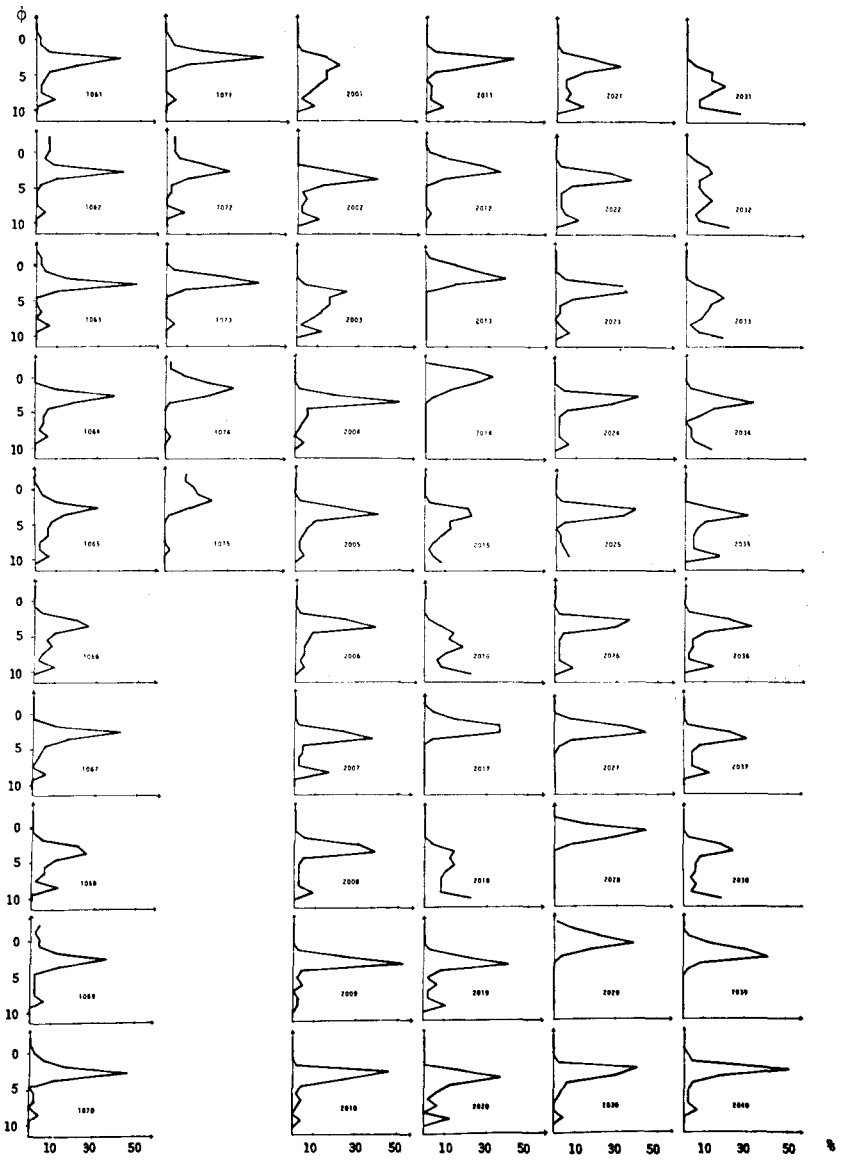


Fig. 17. Continued

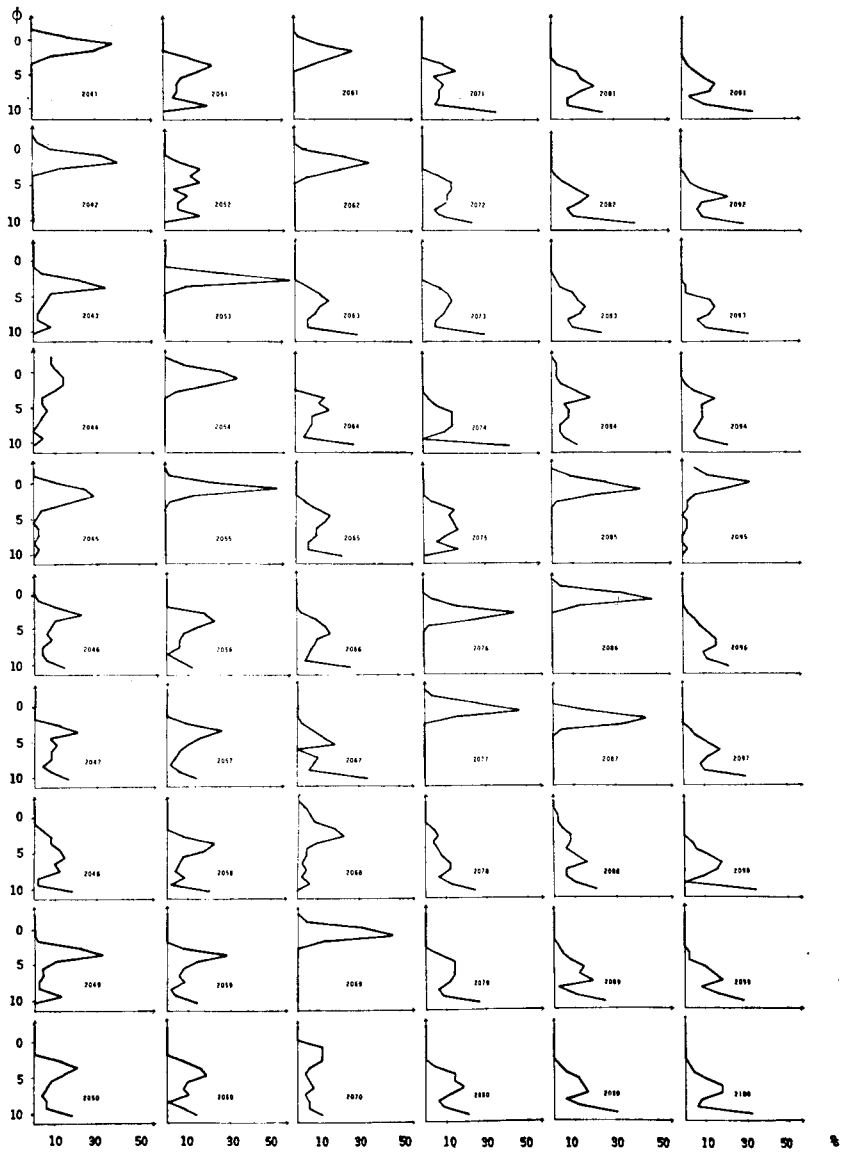


Fig. 17. Continued

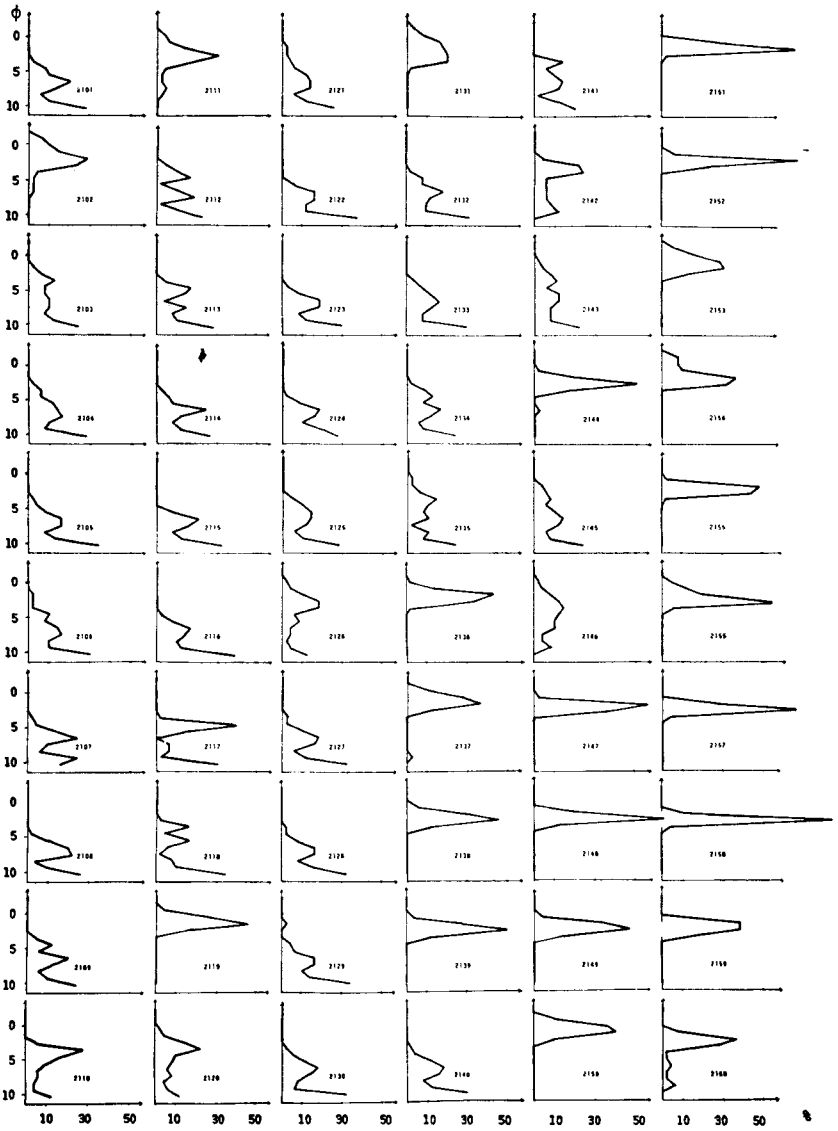
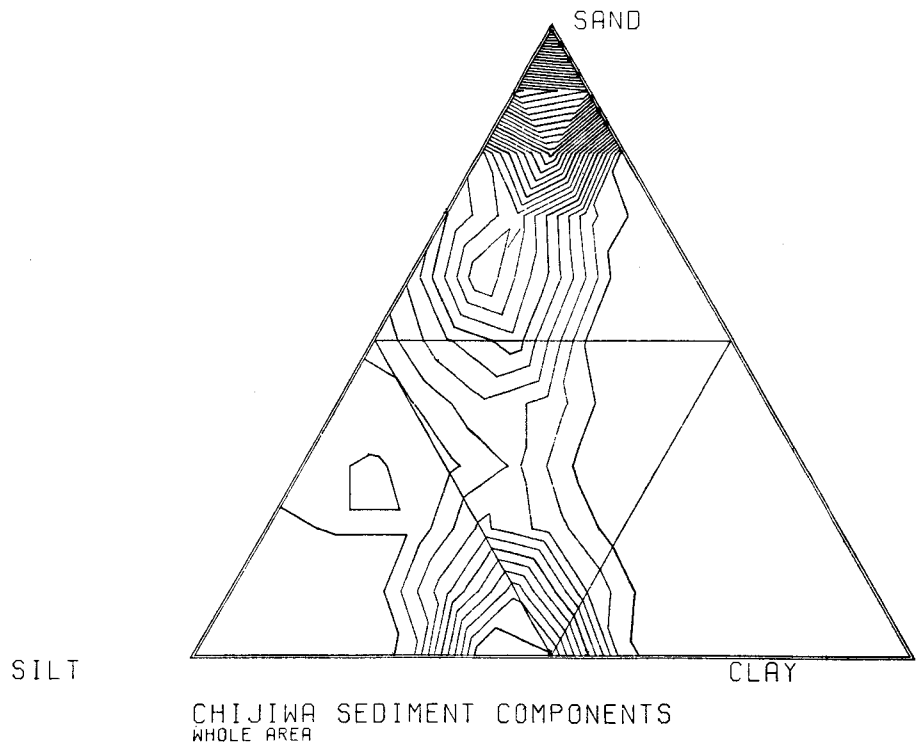


Fig. 17. Continnued



(a)

Fig. 18. Triangular contour diagrams, showing concentration pattern of samples on triangular diagram of sandsilt-clay three components system. (a) all samples; (b) samples No. 1001 to 1075; (c) samples No. 2001 to 2160.

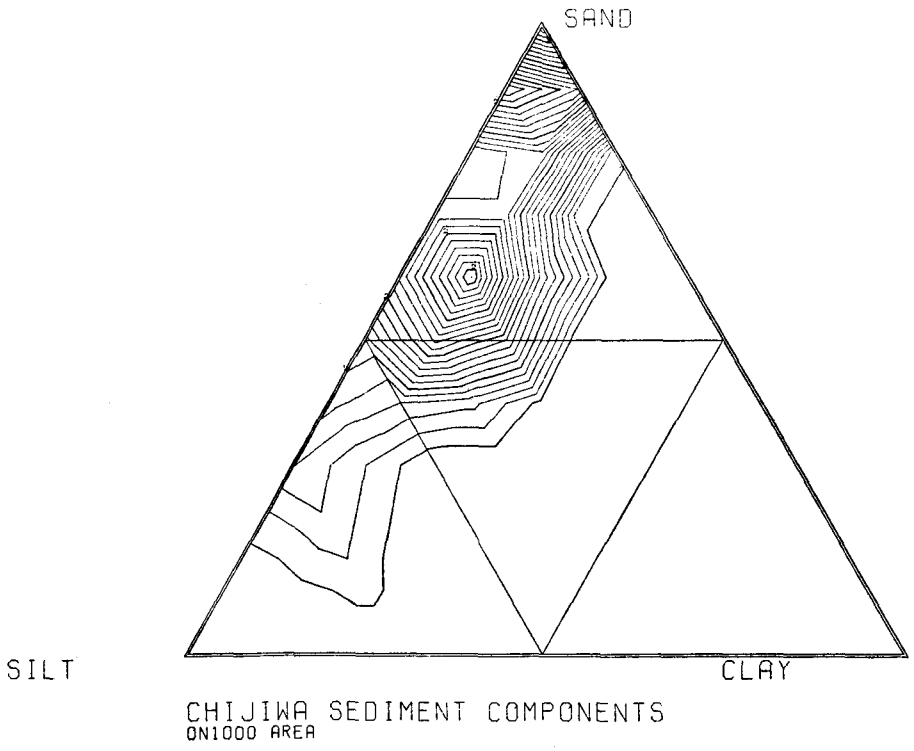


Fig. 18. Continued (b)

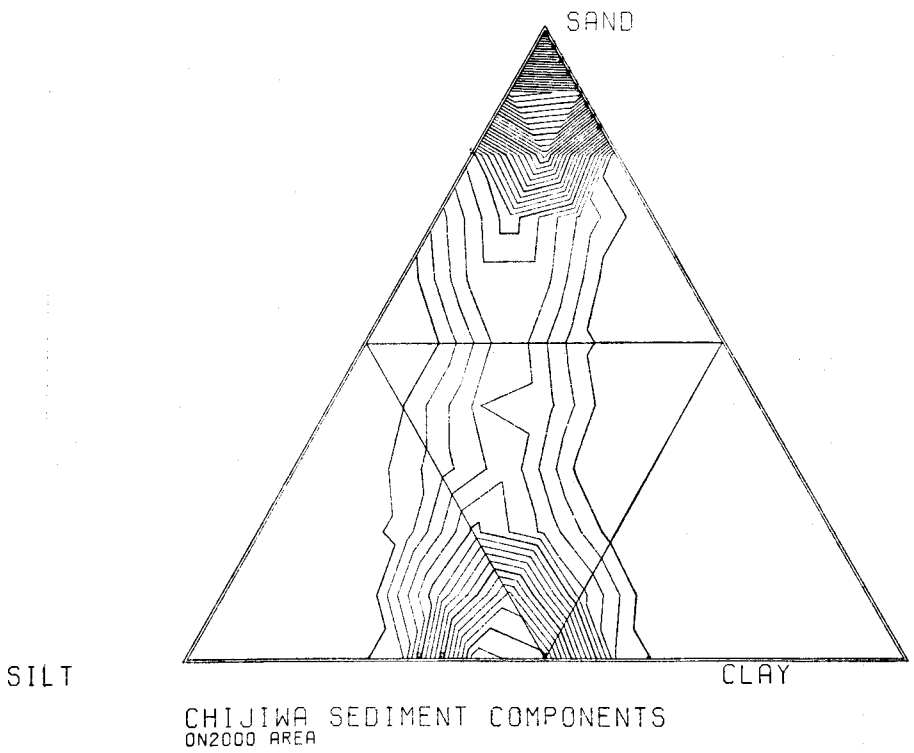


Fig. 18. Continued (c)

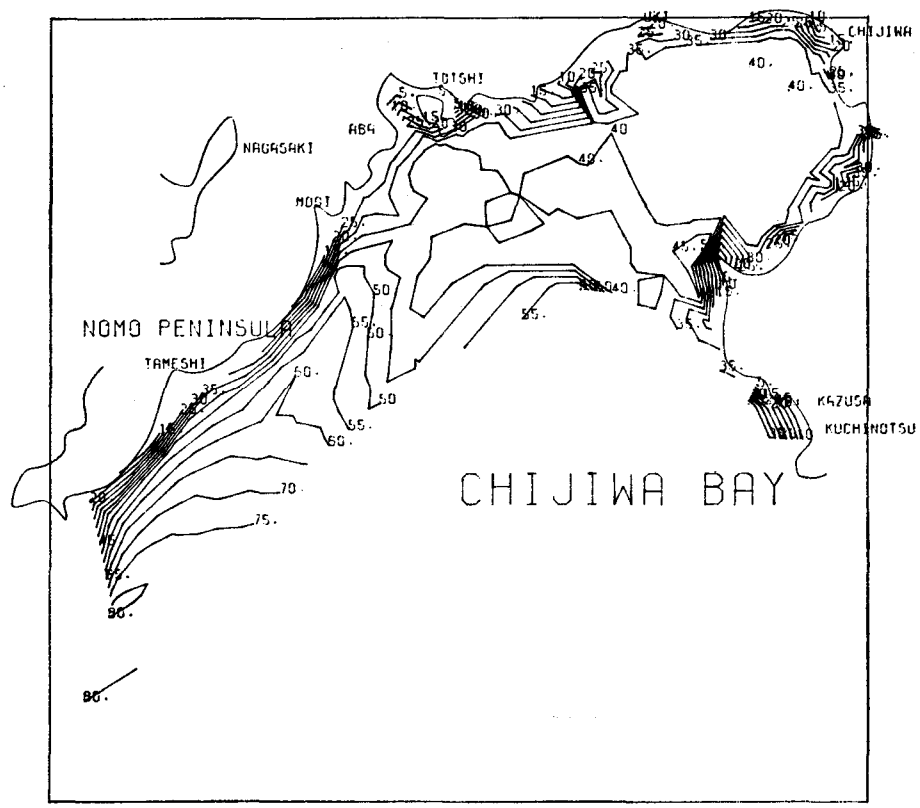


Fig. 19. Contour map of water depth (DEP) in meter.

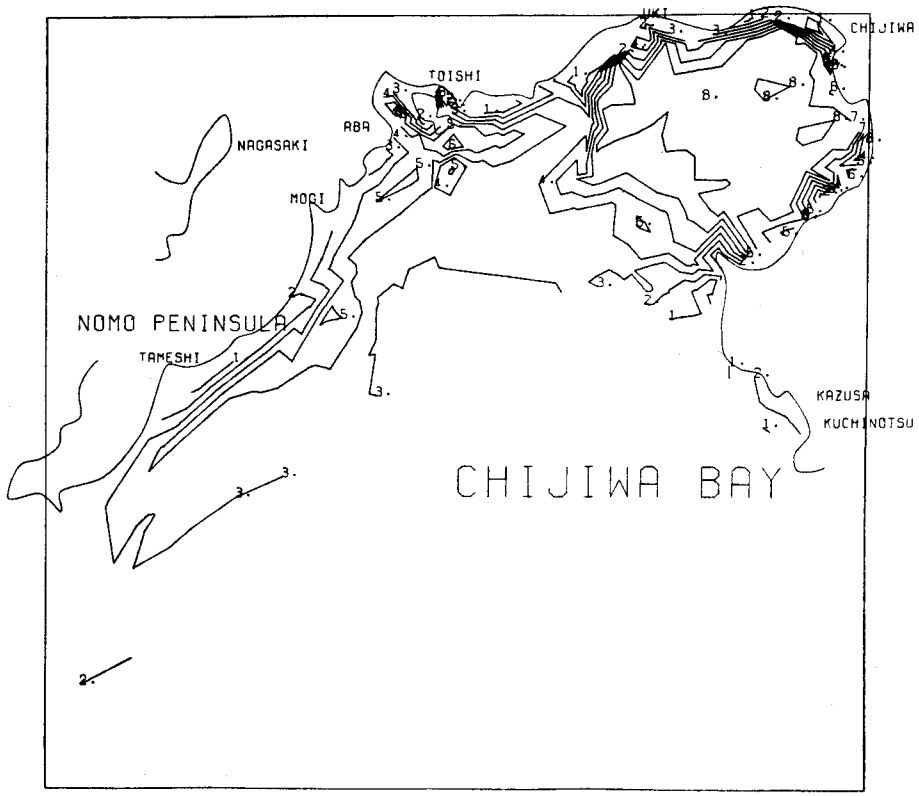


Fig. 20. Distribution of median ϕ (MDPH).

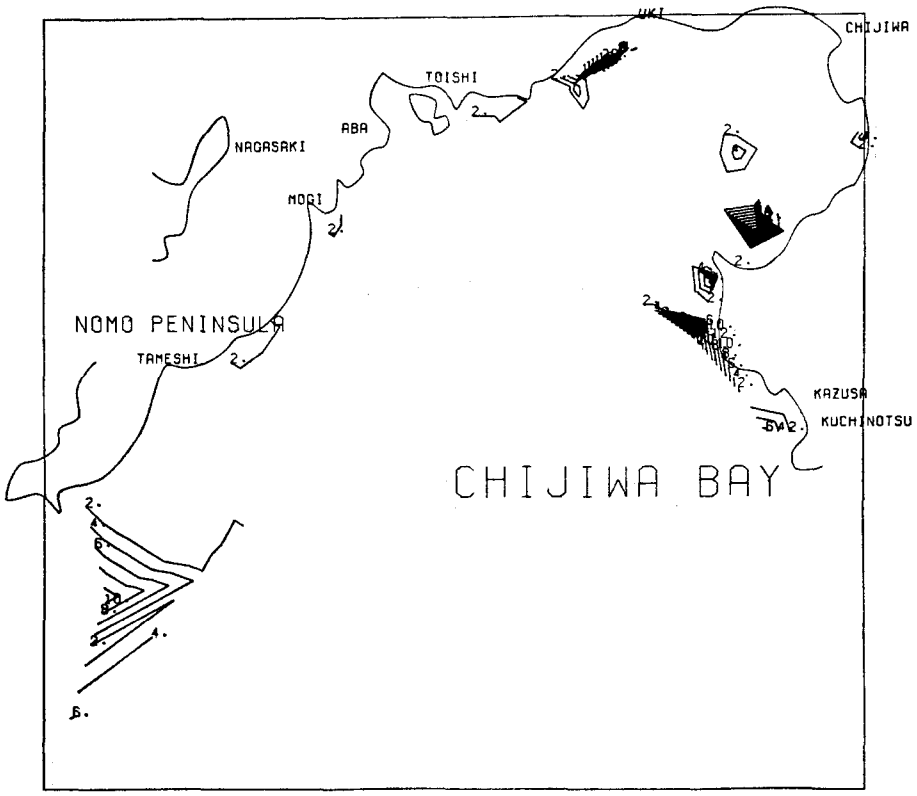


Fig. 21. Contour map of gravel contents (G) in weight percent.

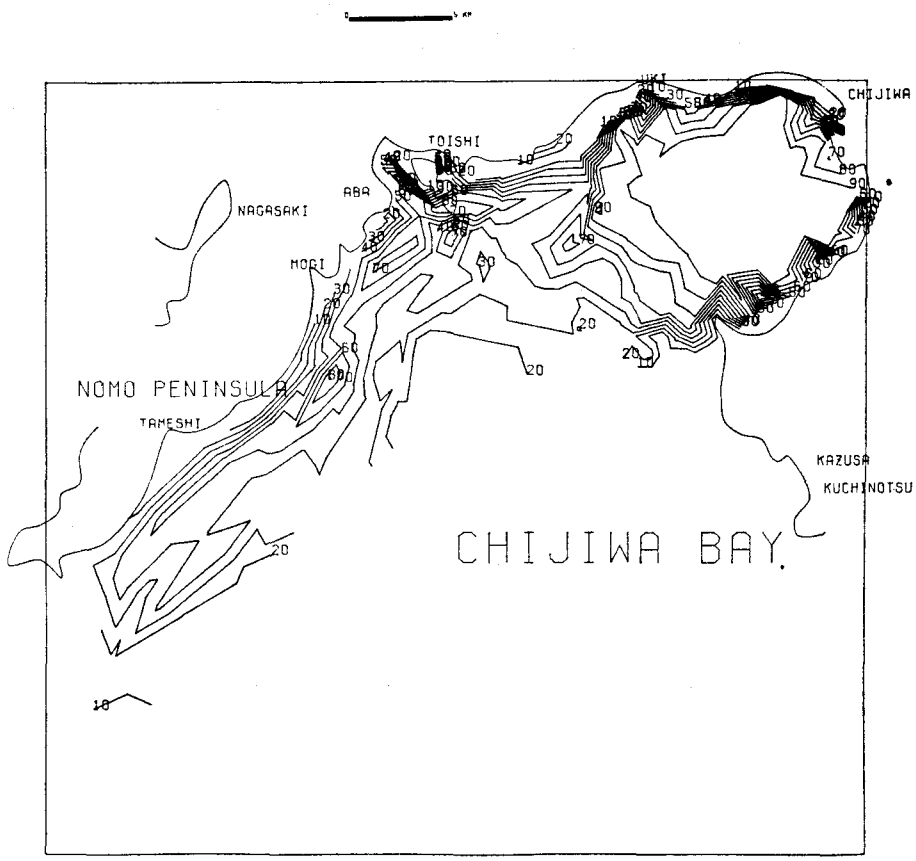


Fig. 22. Contour map of mud contents (MUD) in weight percent.

Factor analysis

R-mode factor analysis: Three factors are extracted, judging from the eigenvalue and cumulative proportion of total variance. To obtain the final factor loading matrix, three principal components are rotated (Table 3). The values are diagrammatically shown in Fig. 23. As the result suggests, the factors seem to have the following attributes on the cumulative curve pattern of grain-size distribution.

F_1...Size factor. Position of curve, for the factor loadings are positive and almost same.

F_2...Shape factor. Range of the curve.

F_3...Shape factor. Distortion of the curve. The negative score means that the curve is distorted on the larger side of the proportion (%). The positive means the contrary case, though the distortion will be smaller than that of the negative.

The correlation of factor scores to each statistical parameters (MDPH, SO and SK) are shown in Fig. 24a–c. Fig. 24a indicates strongly high correlation between F_1 and MDPH, but Figs. 24b and 24c do not illustrate any one.

Factor scores are plotted to investigate the relations among newly extracted factors, F_1, F_2, and F_3, and among the samples based on the factor scores (Fig. 25a–c). On the basis of the scattered diagram of F_1 and F_2 factor scores, the samples are tentatively classified into four groups named A, B, C, and D, which seem to take their positions parallel with each other on the diagram. The cumulative curves of several representative samples of the four groups are shown in Fig. 26a, and their positions on the scattered diagram of F_1 and F_2 in Fig. 26b.

The cumulative curves are reconstructed using the model obtained as the result of the factor analysis. Although those of the samples which show unimodality of frequencies of grain-size distribution are well reconstructed, those of showing multimodality are not well reconstructed.

Q-mode factor analysis: The computed eigenvalues and cumulative proportion of total variance are shown in Table 4a. Four factors are extracted, and they explain 94.7% of total variance. The eigenvectors (the principal factor loading matrix

Table 3. Results of R-mode factor analysis using correlation matrix.

Variable	Initial			Quartimax rotated		
	F_1	F_2	F_3	F_1	F_2	F_3
VAR101 (ϕ 5)	0.892	0.409	0.104	0.885	−0.435	0.020
VAR102 (ϕ10)	0.950	0.313	0.059	0.945	−0.332	−0.003
VAR103 (ϕ15)	0.974	0.219	−0.001	0.971	−0.229	−0.044
VAR104 (ϕ20)	0.987	0.146	−0.040	0.985	−0.149	−0.067
VAR105 (ϕ25)	0.993	0.087	−0.067	0.991	−0.086	−0.082
VAR106 (ϕ30)	0.995	0.028	−0.086	0.995	−0.025	−0.088
VAR107 (ϕ35)	0.996	−0.023	−0.087	0.996	0.026	−0.080
VAR108 (ϕ40)	0.994	−0.068	−0.074	0.995	0.067	−0.057
VAR109 (ϕ45)	0.992	−0.107	−0.022	0.993	0.101	−0.029
VAR110 (ϕ50)	0.988	−0.141	−0.022	0.990	0.128	0.009
VAR111 (ϕ55)	0.983	−0.173	0.011	0.986	0.153	0.048
VAR112 (ϕ60)	0.978	−0.202	0.055	0.980	0.172	0.096
VAR113 (ϕ65)	0.970	−0.219	0.103	0.973	0.179	0.146
VAR114 (ϕ70)	0.962	−0.226	0.119	0.964	0.183	0.164
EIG*	13.32	0.55	0.07	(13.32	0.53	0.09)
CUM·PCT**	95.2	99.1	99.6	95.1	98.9	99.6

* Eigenvalue. Sum of squares of factor loadings in the parentheses

** Cumulative proportion of total variance (%)

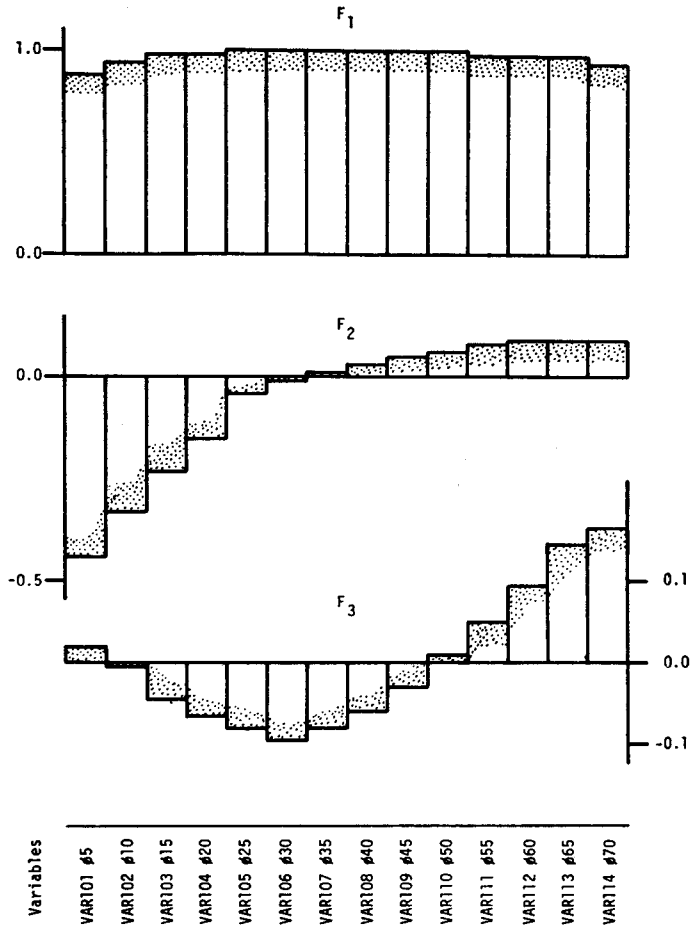
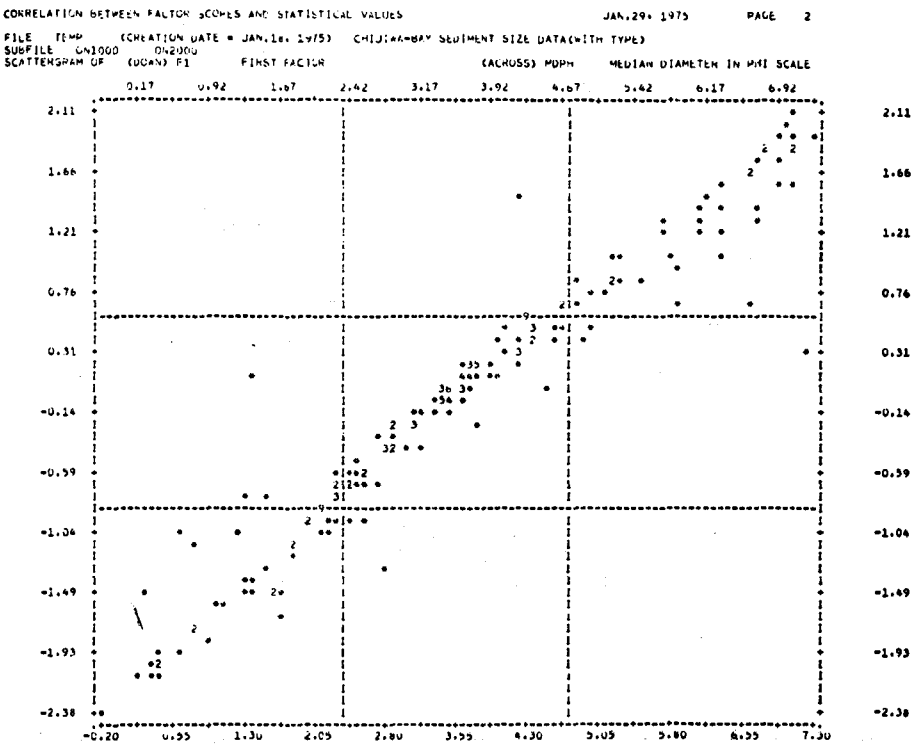


Fig. 23. Final factor loadings of R-mode factor analysis.

of these factors) are rotated by the varimax method. Then, the samples are plotted in a coordinate, where each pair of vectors define X- and Y-axes (Fig. 27). Fig. 27 indicates that most samples are concentrated near the factor axes.

KLOVAN (1966) employed the Q-mode factor analysis for grain-size frequency distribution data and classified recent sediments into four groups on the basis of grain-size characters. According to his procedure most of the samples are well classified into any of four groups, named 1, 2, 3, and 4, each of which is related to one of four factors. There are, however, still remaining several samples which have not any significant relation to each factor.

By factor scores of Q-mode analysis, variable(s) is/are selected, which is/are closely related to each group, as shown in Table 4b. The group 1 is especially related to the 7th variable (frequency of 3 to 4ϕ), the group 2 the 13th variable (frequency of 9ϕ +), the group 3 the 4th variable (frequency of 0 to 1ϕ), and the group 4 the 6th variable (frequency of 2 to 3ϕ). Namely, the groups 1 to 4 are considered to be constituted from the silt, clay, gravel and sand rich samples, respectively.



(a)

Fig. 24. Scatter diagrams between each pair of factor scores of R-mode analysis and statistical parameters: (a) first factor score vs median (F_1–MDPH); (b) second vs sorting coefficient (F_2–SO); (c) third vs skewness (F_3–SK).

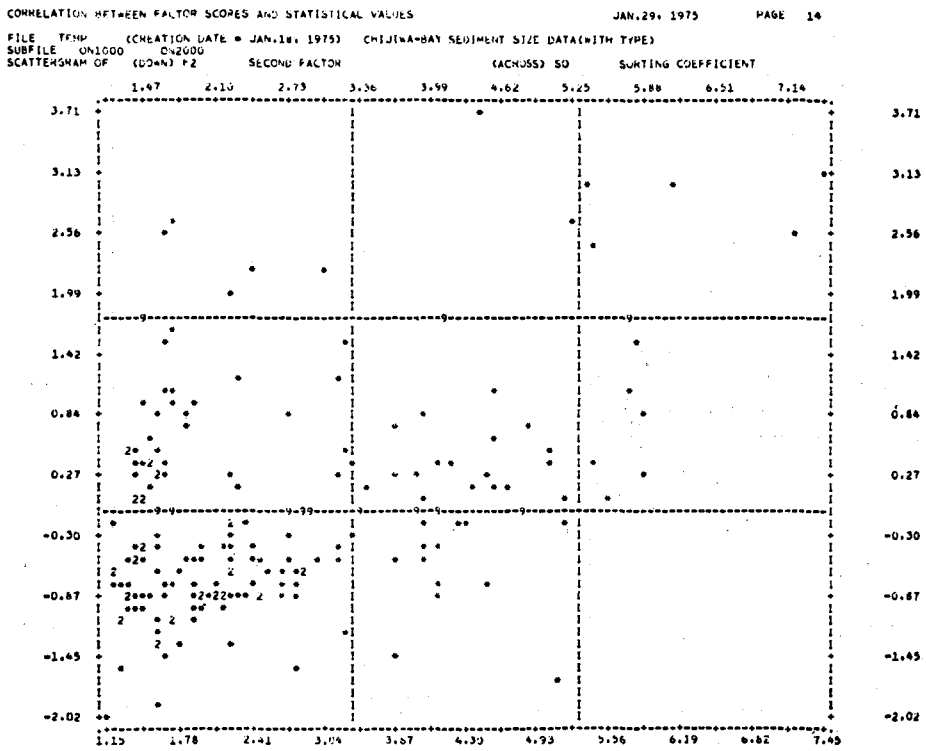


Fig. 24. Continued (b)

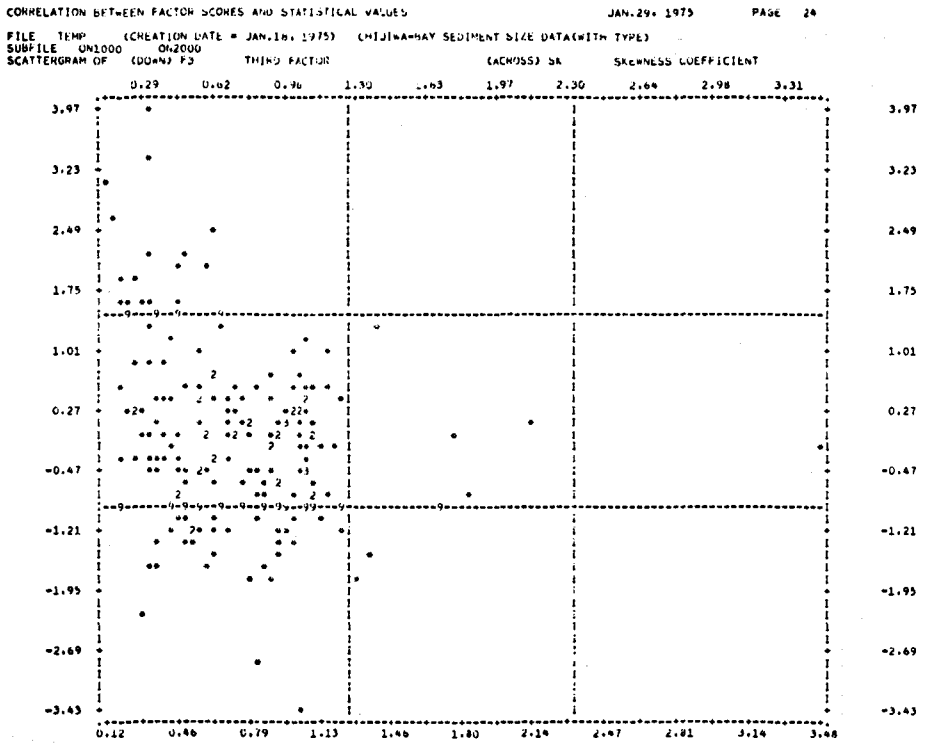
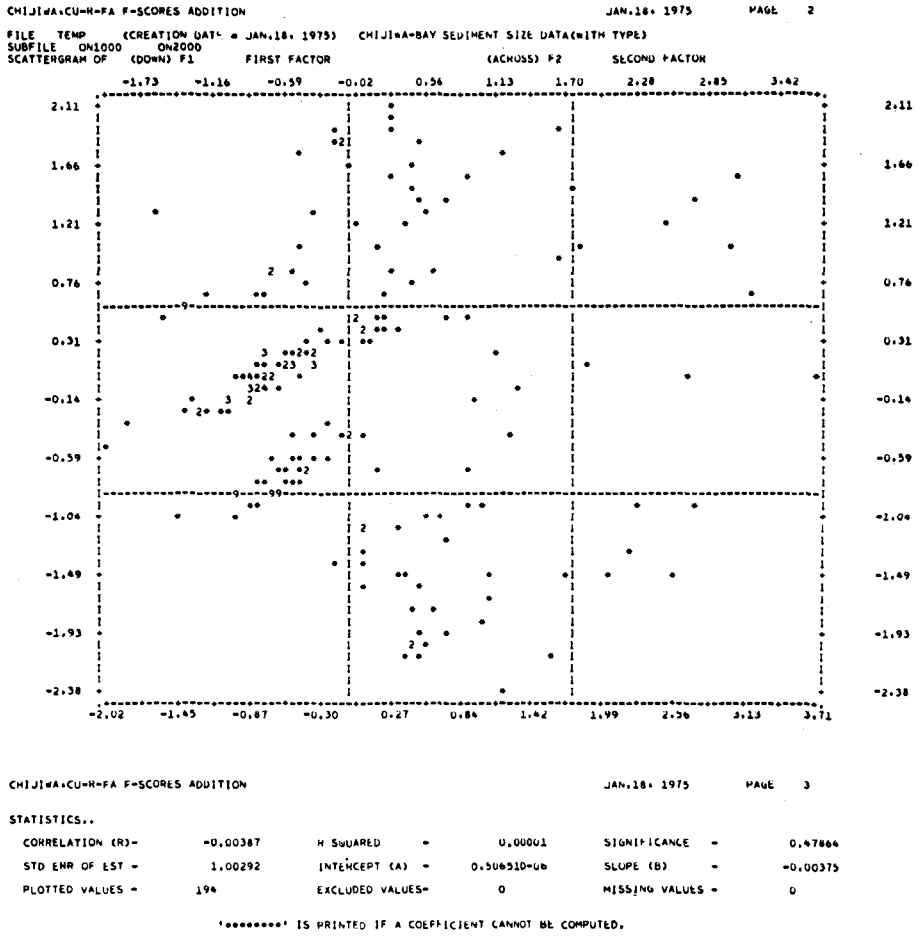


Fig. 24. Continued (c)



CHIJIWA:CU=H=FA F-SCORES ADDITION JAN.18, 1975 PAGE 3

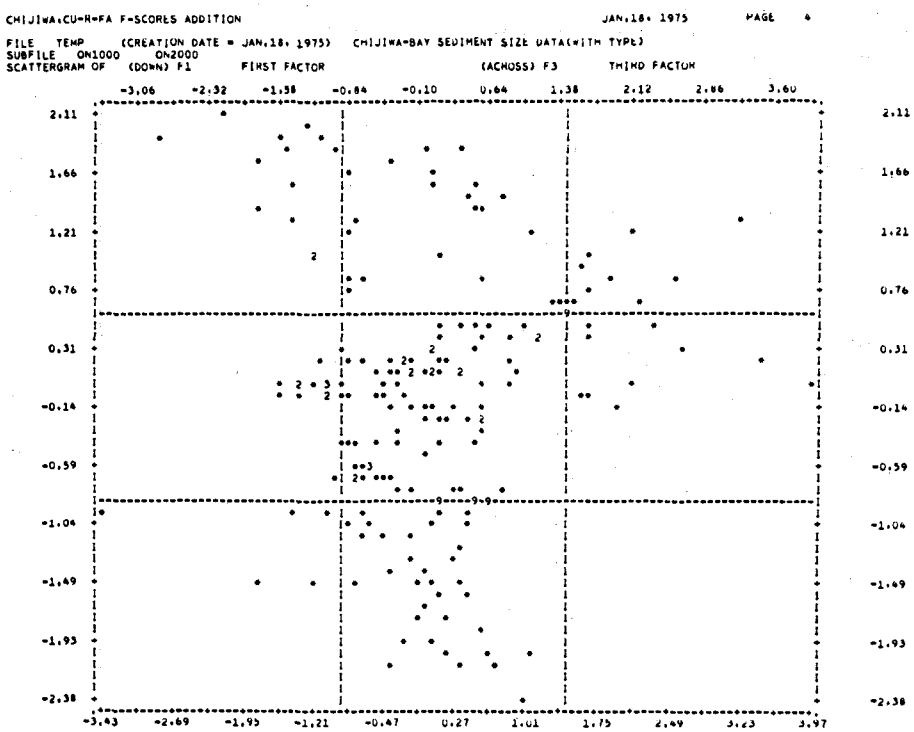
STATISTICS..

CORRELATION (R)=	-0.00387	R SQUARED =	0.00001	SIGNIFICANCE =	0.47864
STD ERR OF EST =	1.00292	INTERCEPT (A) =	0.506510D-06	SLOPE (B) =	-0.00375
PLOTTED VALUES =	194	EXCLUDED VALUES=	0	MISSING VALUES =	0

'********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

(a)

Fig. 25. Scatter diagram among factor scores of R-mode analysis: (a) first vs second factor score (F_1–F_2); (b) first vs third (F_1–F_3); (c) second vs third (F_2–F_3).



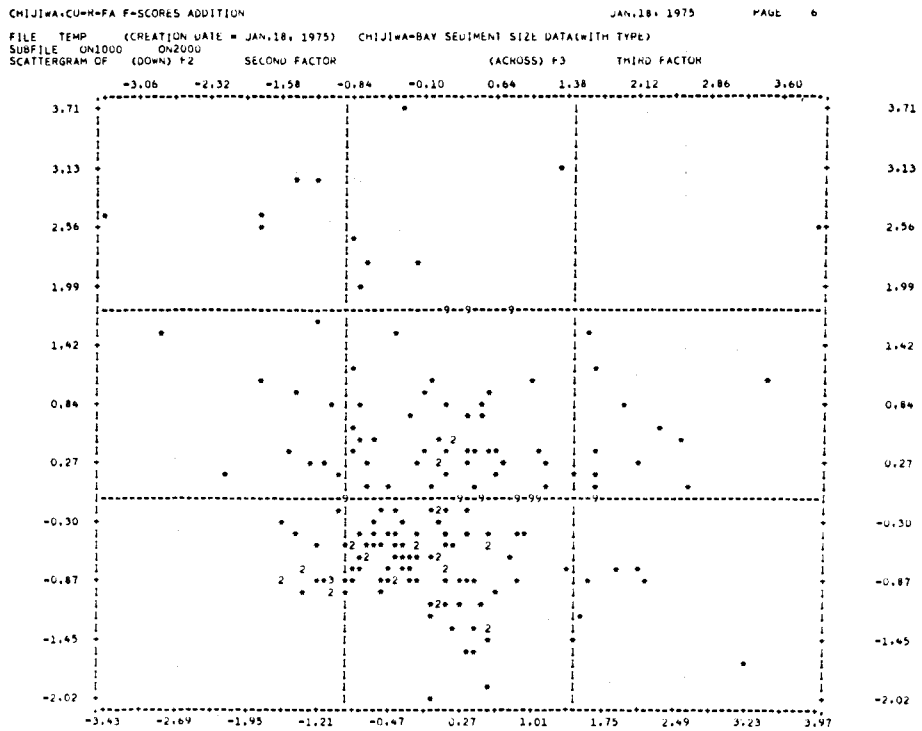
CHIJIWA:CU=R=FA F=SCORES ADDITION JAN.18, 1975 PAGE 5

STATISTICS..

CORRELATION (R)=	0.00786	R SQUARED =	0.00006	SIGNIFICANCE =	0.45671
STD ERR OF EST =	1.00290	INTERCEPT (A) =	0.493620-06	SLOPE (B) =	0.00730
PLOTTED VALUES =	194	EXCLUDED VALUES=	0	MISSING VALUES =	0

'********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Fig. 25. Continued (b)



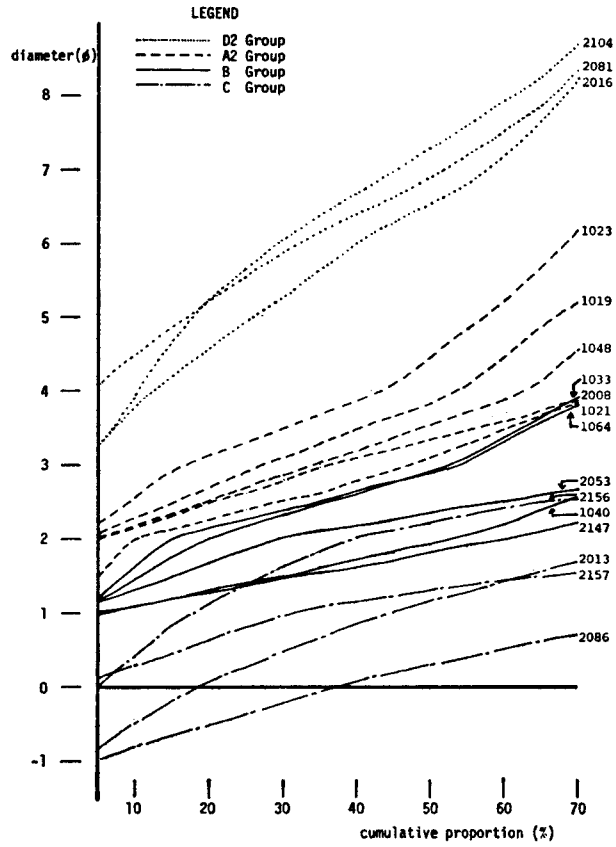
CHIJIWA,CU=H=FA F=SCORES ADDITION JAN.18, 1975 PAGE 7

STATISTICS..

CORRELATION (R)=	-0.06384	R SQUARED =	0.00408	SIGNIFICANCE =	0.18826
STD ERR OF EST =	1.03300	INTERCEPT (A) =	-0.23889D-05	SLOPE (B) =	-0.06118
PLOTTED VALUES =	194	EXCLUDED VALUES=	0	MISSING VALUES =	0

'********' IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.

Fig. 25. Contuned (c)



(a)

Fig. 26. (a) Cumulative curves of representative samples of groups defined by factor analysis, and (b) their points plotted on F_1–F_2 diagram of R-mode factor scores, where arrows show ascending order of cumulative curves in (a).

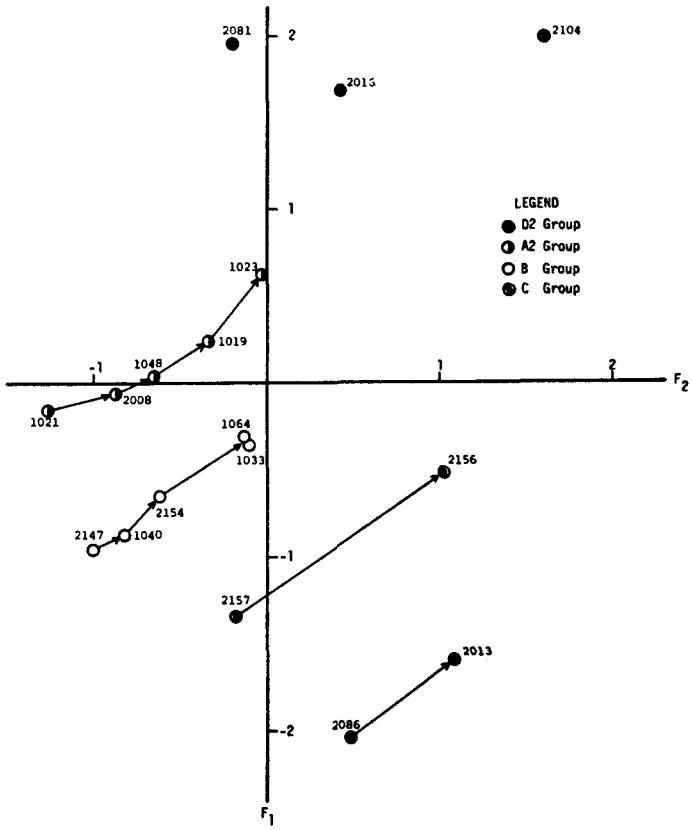


Fig. 26. Continued (b)

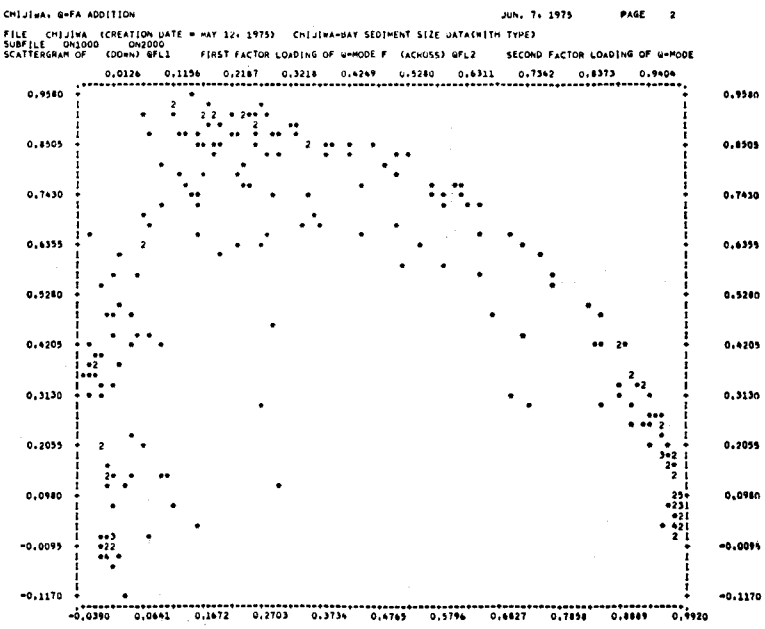
Table 4. Results of Q-mode factor analysis: (a) Eigenvalue and cumulative proportion of total variance; (b) varimax factor scores.

(a)

Factor	Eigenvalue	Cumulative Proportion
1	133.12	56.65
2	53.76	79.53
3	25.65	90.44
4	10.05	94.72
5	4.46	96.62
6	3.17	97.97
7	1.45	98.58

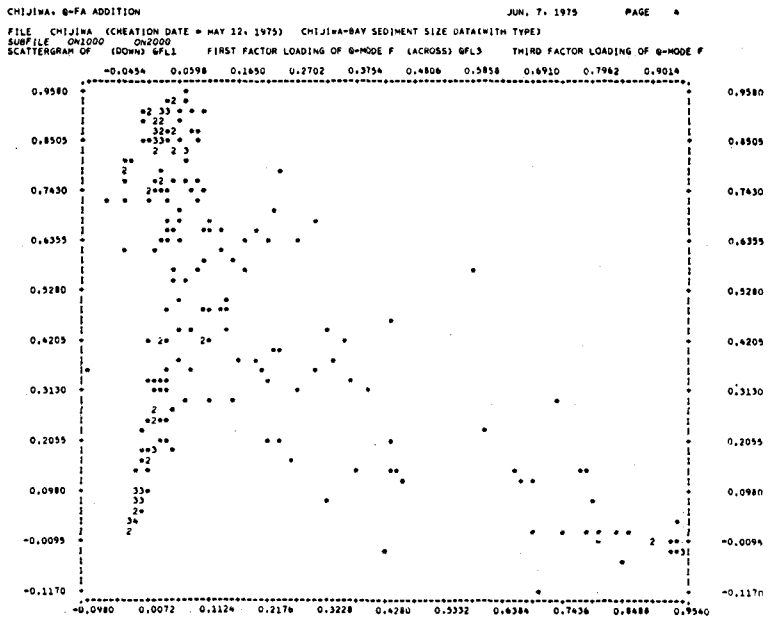
Table 4. Continued (b)

Variable	Factor 1	2	3	4
VAR001	0.001	−0.001	0.031	−0.002
VAR002	0.007	−0.005	0.140	0.019
VAR003	0.019	−0.013	0.433	0.080
VAR004	−0.012	−0.006	0.741	0.036
VAR005	−0.139	0.044	0.439	−0.525
VAR006	0.344	−0.042	−0.158	−0.808
VAR007	0.840	−0.039	0.073	0.150
VAR008	0.318	0.155	0.096	0.154
VAR009	0.208	0.233	0.087	0.118
VAR010	0.053	0.327	0.012	0.002
VAR011	−0.013	0.287	−0.011	−0.023
VAR012	0.053	0.146	−0.001	−0.011
VAR013	−0.077	0.840	−0.057	−0.056

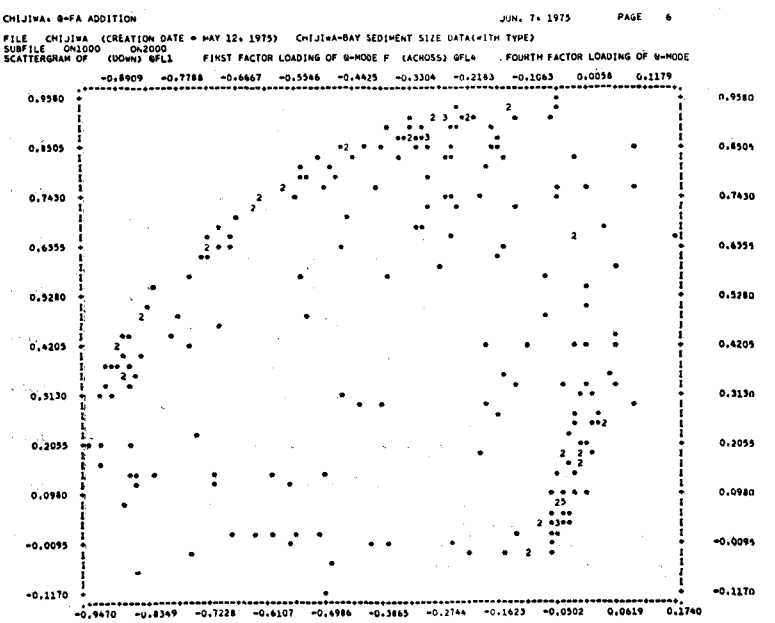


(a)

Fig. 27. Scatter diagrams among factor loadings of Q-mode analysis: (a) first vs second factor loading (QFL1–QFL2); (b) first vs third (QFL1–QFL3); (c) first vs fourth (QFL1–QFL4); (d) second vs third (QFL2–QFL3); (e) second vs fourth (QFL2–QFL4); (f) third vs fourth (QFL3–QFL4).

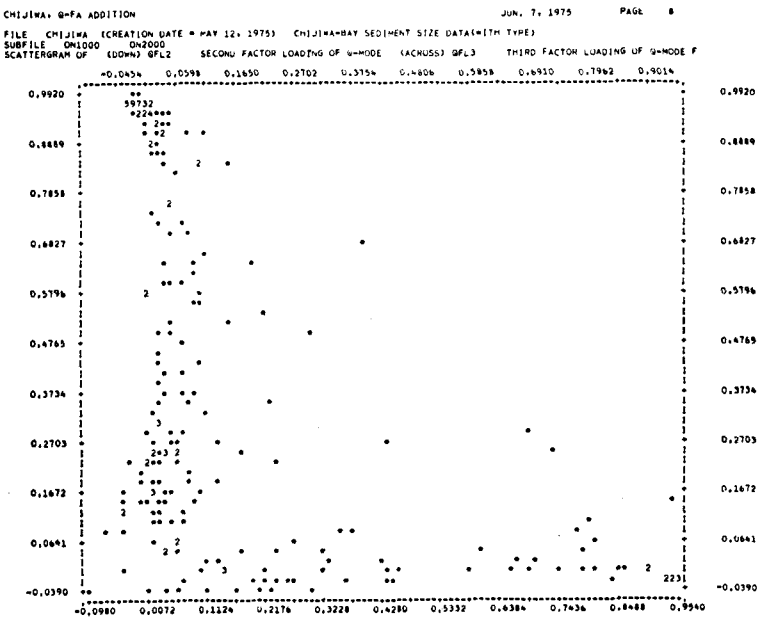


(b)

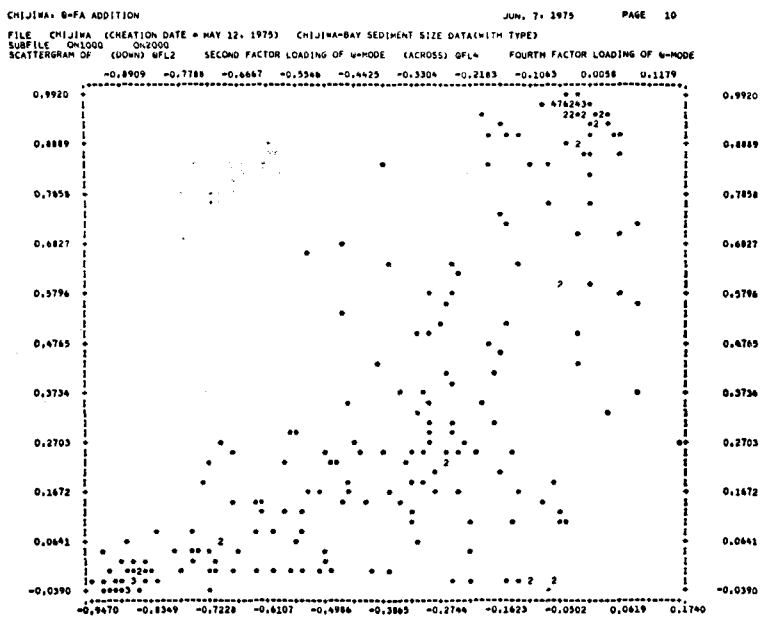


(c)

Fig. 27. Continued (b), (c)



(d)



(e)

Fig. 27. Continued (d), (e)

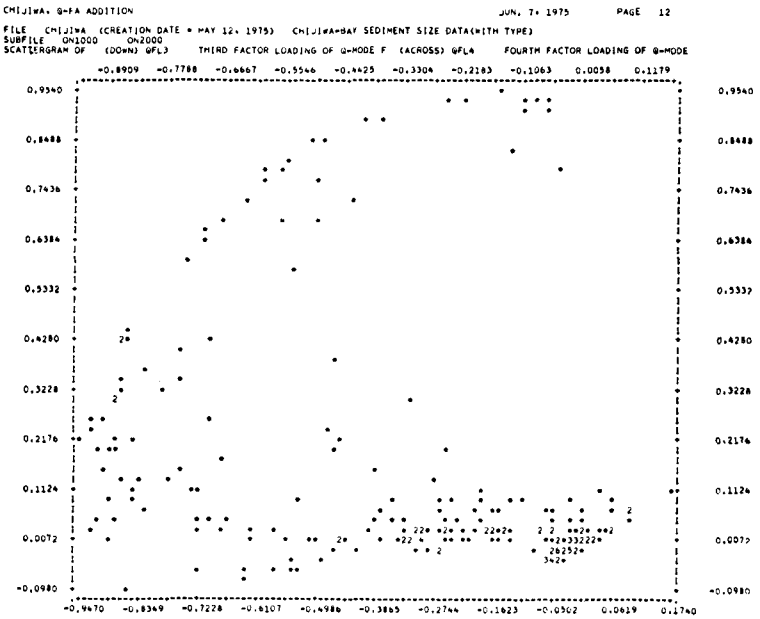


Fig. 27. Continued (f)

Cluster analysis

Several methods of the cluster analysis were carried out, and it was clarified that the cluster analysis based on chi-square distance of frequency scores is the most suitable in this case. The analyzed results are discussed below.

Four major branches are observed in the simplified figure (Fig. 28). The upper branch is further subdivided into three subbranches. As the samples belonging to the lowest branch are rather limited, the branch is omitted in this case study. The second and fourth major branches are subdivided into two subbranches respectively. Thereby, seven clusters are observed in the figure, and named as the A_1, A_2, D_2, D_1, C, B_2, and B_1 group in descending order. The group names correspond to the classification by the factor analysis. By using the dendrogram of the original outputs (Fig. 29), the more detailed classification is attained, as shown in the figure.

The convergence of each cluster is very good, and the separation of one cluster from the other is also good. It is rather easy to recognize the cluster structure.

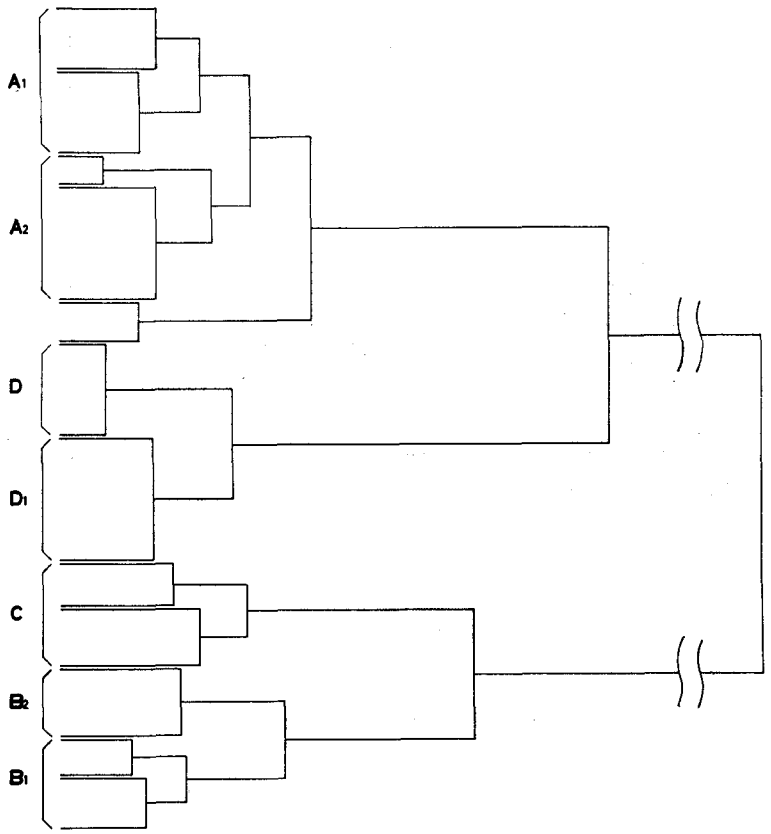


Fig. 28. Simplified dendrogram of Q-mode cluster analysis, compiled from original output shown in Fig. 29.

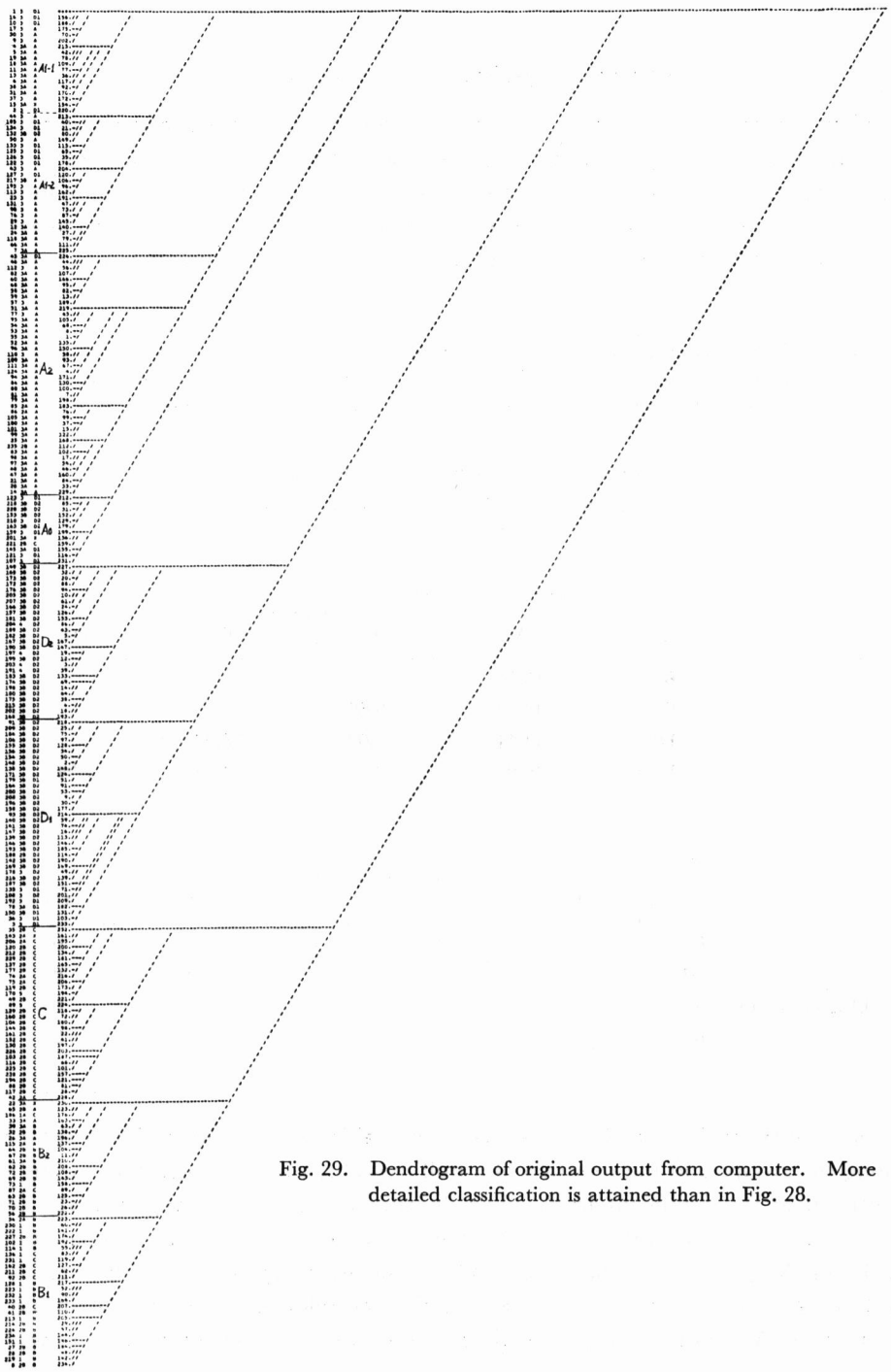


Fig. 29. Dendrogram of original output from computer. More detailed classification is attained than in Fig. 28.

Statistical significance of classification

Classification by the factor analysis: As shown in Tabs. 5a and 5b, the distance between each pair of groups is significant at 1% significant level, and the equality of means among the groups is rejected 1% significant level, so that the multivariate means of the groups are different each other. As shown in Tab. 5c, the discrimination of each sample is excellent, for almost all samples are discriminated into correct group.

Table 5. Discriminant analysis of groups defined by factor analysis with frequency scores and phi values: (a) Test for multivariate equality of means; (b) F-distance matrix (D. F.=14, 172); (c) Re-classification by discriminant analysis.

(a)

U-statistic	0.02927	D. F.=14, 4, 185
Approximate F	17.72478	D. F.=56, 671.22

(b)

		Group			
		A	B	C	D1
Group	B	15.24			
	C	47.29	9.66		
	D1	14.44	30.02	58.37	
	D2	40.05	62.26	104.13	7.94

(c)

		Number of cases classified into group				
		D2	D1	A	B	C
Original group	D2	22	4	0	0	0
	D1	2	21	1	0	0
	A	0	4	67	3	0
	B	0	0	1	28	2
	C	0	1	0	2	32

Classification by the cluster analysis: As shown in Tables. 6a and 6b, both the distances among the groups and the rejection of a null hypothesis on the equalty of means are significant at 1% significant level. The discrimination of each sample is also very good (Table 6c). Several samples of the A_1 group are, however, discriminated into the D_1 or A_2 group. The fact suggests that the A_1 group is situated between the D_1 and A_2 groups in the space, based on the variables used for the classification, and that the A_1 group is a transitional or overlapping of the latter groups in the dis-

tribution pattern of grain-size. Judging from the distance, the groups may be arranged in the order: the D_1, A_1, A_2, B_1, and C group, on the basis of the similarity among them. Examining each neibouring pair of the groups, the distance between the B_2 and B_1 groups is especially small (4.14), in comparison with the others having usually 11.85 to 18.11. The division between the B_2 and B_1 groups is more detailed than those of the other pairs.

Table 6. Discriminant analysis of groups defined by cluster analysis with chi-square distance of frequency data: (a) Test of multivariate equality of means; (b) F distance matrix (D. F.=14, 174); (c) Re-classification by discriminant analysis.

(a)

U-statistic	0.02048	D. F.=14, 5, 187
Approximate F	15.02178	D. F.=70, 832.49

(b)

		Group				
		A1	A2	B1	B2	C
Group	A2	11.85				
	B1	30.72	12.68			
	B2	21.39	10.90	4.14		
	C	82.12	54.29	14.41	17.12	
	D1	18.77	34.43	64.34	46.39	127.73

(c)

		Number of cases classified into groups					
		D1	A1	A2	B2	B1	C
Original group	D1	25	1	0	0	0	0
	A1	4	41	8	1	0	0
	A2	0	1	39	2	0	0
	B2	0	0	3	13	2	0
	B1	0	0	0	1	25	0
	C	0	0	0	0	0	27

Classifiication made by Kamada et al. (1973): As shown in Tables. 7a and 7b, a null hypothesis on the multivariate equality of means is rejected at 1% significant level. The distance between each pair of groups is significant at 1% level, except for that between the 2A and 2B groups, which is not significant. Combining the 2A and 2B groups into a group, five groups are tested by discriminant analysis. The result (Table 7c) shows that many of the samples belonging to the 2A and 2B groups are re-classified into the group 1, and that some samples of the group 3 are

re-classified into the group 3A. Therefore, this classification is indistinct especially in the separation between each pair of the groups 1, 2A and 2B.

Table 7. Discriminant analysis of sediment types defined by Kamada et al. (1973): (a) Test of multivariate equality of means; (b) F distance matrix (D. F.=8, 215); (c) Re-classification by discriminant analysis.

(a)

U-statistic	0.03000	D. F.=8, 5, 222
Approximate F	29.03202	D. F.=40, 939.96

(b)

		Group				
		1	2A	2B	3	3A
Group	2A	5.13				
	2B	7.38	2.42			
	3	51.11	27.18	82.21		
	3A	21.14	11.67	35.58	22.88	
	3B	106.51	60.90	216.34	40.25	109.80

(c)

		Number of cases classified into groups				
		1	2A, B	3A	3	3B, 4
Original group	1	14	2	0	0	0
	2A, B	12	40	5	0	1
	3A	0	0	39	1	0
	3	1	0	17	38	2
	3B, 4	0	0	0	1	56

Description of groups

Breakdown: The general statistics of each group, which is classified by KAMADA *et al.* (1973) and the cluster analysis, are shown in Tables 8 and 9. The tables describe the mean and standard deviation of the statistical parameters for each group.

Table 10 illustrates the difference of each parameters among the groups. The median, sand, silt, clay and mud contents are remarkably different among the groups. The difference between two classifications is tested by the F-values of gravel contents (G), sorting coefficient (SO) and water depth (DEP).

Triangle contour diagram: A triangle contour diagram of sand-silt-clay (Fig. 30) is produced for each group, which is defined by the cluster analysis and shown in the original dendrogram (Fig. 29).

All the groups except for the group A_0 show good isolation in the figures. The mean locations are different from each other as shown in Fig. 31.

Table 8. Description of groups defined by cluster analysis, concerning statistical and compositional parameters.

BREAKDOWN OF CHIJIWA MAY 1, 1975 PAGE 6

FILE CHIJIWA (CREATION DATE = MAR.26, 1975) CHIJIWA-BAY SEDIMENT SIZE DATA(WITH TYPE)
SUBFILE ON1000 ON2000

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE MDPH MEDIAN DIAMETER IN PHI SCALE
BROKEN DOWN BY CAG3 GROUP BASED ON FR=CO=CA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			988.9200	4.1231	2.1892	4.7927	(235)
CAG3	A1		245.300	4.543	1.079	1.164	(54)
CAG3	A2		147.100	3.502	0.288	0.083	(42)
CAG3	B1		51.800	1.992	0.515	0.265	(26)
CAG3	B2		54.000	2.700	0.271	0.074	(20)
CAG3	C		30.420	1.014	0.725	0.526	(30)
CAG3	D1		235.100	6.531	0.782	0.611	(36)
CAG3	D2		235.200	7.600	0.309	0.095	(27)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE SO SORTING COEFFICIENT
BROKEN DOWN BY CAG3 GROUP BASED ON FR=CO=CA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			664.3700	2.8272	1.3568	1.8409	(235)
CAG3	A1		195.320	3.617	1.420	2.018	(54)
CAG3	A2		92.310	2.198	0.646	0.417	(42)
CAG3	B1		36.050	1.387	0.138	0.019	(26)
CAG3	B2		37.816	1.890	0.417	0.174	(20)
CAG3	C		53.210	1.774	0.525	0.276	(30)
CAG3	D1		156.090	4.336	1.123	1.262	(36)
CAG3	D2		93.560	3.466	0.324	0.105	(27)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE SK SKEWNESS COEFFICIENT
BROKEN DOWN BY CAG3 GROUP BASED ON FR=CO=CA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			170.6000	0.7260	0.3911	0.1529	(235)
CAG3	A1		34.960	0.647	0.445	0.198	(54)
CAG3	A2		22.430	0.544	0.216	0.047	(42)
CAG3	B1		26.670	1.026	0.113	0.013	(26)
CAG3	B2		17.250	0.862	0.435	0.189	(20)
CAG3	C		30.880	1.029	0.532	0.283	(30)
CAG3	D1		23.350	0.649	0.263	0.069	(36)
CAG3	D2		14.460	0.543	0.137	0.019	(27)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE G GRAVEL RATIO
BROKEN DOWN BY CAG3 GROUP BASED ON FR=CO=CA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			199.1000	0.8472	3.1811	10.1192	(235)
CAG3	A1		0.0	0.0	0.0	0.0	(54)
CAG3	A2		0.0	0.0	0.0	0.0	(42)
CAG3	B1		0.0	0.0	0.0	0.0	(26)
CAG3	B2		26.700	1.335	2.448	6.093	(20)
CAG3	C		146.700	4.867	6.567	42.868	(30)
CAG3	D1		20.100	0.558	3.350	11.222	(36)
CAG3	D2		6.300	0.233	1.000	0.999	(27)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE SA SAND RATIO
BROKEN DOWN BY CAG3 GROUP BASED ON FR=CO=CA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			12593.3999	53.5684	32.3734	1061.0260	(235)
CAG3	A1		2430.400	45.374	13.930	194.041	(54)
CAG3	A2		2723.300	64.952	8.352	69.730	(42)
CAG3	B1		2528.700	97.265	4.933	24.334	(26)
CAG3	B2		1530.200	76.510	7.707	59.399	(20)
CAG3	C		2694.100	89.803	10.860	117.931	(30)
CAG3	D1		532.400	14.789	9.571	91.598	(36)
CAG3	D2		131.700	4.878	1.283	1.647	(27)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE SI SILT RATIO
BROKEN DOWN BY CAG3 GROUP BASED ON FR=CO=CA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			6700.4301	28.5125	20.0982	403.9381	(235)
CAG3	A1		2079.800	38.515	11.144	124.191	(54)
CAG3	A2		947.620	22.562	6.014	36.173	(42)
CAG3	B1		45.700	1.756	3.359	11.280	(26)
CAG3	B2		277.400	13.870	6.902	47.633	(20)
CAG3	C		114.500	3.810	6.357	40.406	(30)
CAG3	D1		1876.500	52.125	8.699	75.678	(36)
CAG3	D2		1359.100	50.337	4.024	16.195	(27)

```
- - - - - - - - - - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - - - - - - - - - -
CRITERION VARIABLE   CL        CLAY RATIO
   BROKEN DOWN BY    CAG3      GROUP BASED ON FR-CU-CA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIABLE                CODE    VALUE LABEL          SUM          MEAN      STD DEV      VARIANCE        N

FOR ENTIRE POPULATION                              4059.3600     17.2569     15.7480     249.1985  (  235)

CAG3                    A1                          869.500     16.102       9.223      85.063   (   54)
CAG3                    A2                          526.700     12.540       3.886      15.101   (   42)
CAG3                    B1                           22.600      0.869       1.728       2.986   (   26)
CAG3                    B2                          151.800      7.590       2.193       4.810   (   20)
CAG3                    C                            54.600      1.820       2.453       8.140   (   30)
CAG3                    D1                         1220.280     33.896      11.798     139.184   (   36)
CAG3                    D2                         1209.900     44.811       4.106      16.859   (   27)

- - - - - - - - - - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - - - - - - - - - -
CRITERION VARIABLE   MOD       MOD RATIO
   BROKEN DOWN BY    CAG3      GROUP BASED ON FR-CU-CA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIABLE                CODE    VALUE LABEL          SUM          MEAN      STD DEV      VARIANCE        N

FOR ENTIRE POPULATION                              10716.0002    45.6000     33.3313    1112.3102  (  235)

CAG3                    A1                         2744.280     50.819      13.958     194.834   (   54)
CAG3                    A2                         1484.200     35.338       8.195      67.154   (   42)
CAG3                    B1                           71.000      2.731       4.929      24.291   (   26)
CAG3                    B2                          429.200     21.460       8.267      68.335   (   20)
CAG3                    C                           148.800      5.027       9.094      82.699   (   30)
CAG3                    D1                         3944.520     84.572      12.768     163.022   (   36)
CAG3                    D2                         2569.300     95.144       1.297       1.681   (   27)

- - - - - - - - - - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - - - - - - - - - -
CRITERION VARIABLE   DEP       DEPTH
   BROKEN DOWN BY    CAG3      GROUP BASED ON FR-CU-CA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIABLE                CODE    VALUE LABEL          SUM          MEAN      STD DEV      VARIANCE        N

FOR ENTIRE POPULATION                              8735.0000     37.1702     16.0196     256.6102  (  235)

CAG3                    A1                         2051.000     37.981      12.319      151.754   (   54)
CAG3                    A2                         1927.000     45.881      13.382      179.083   (   42)
CAG3                    B1                          469.000     18.038      13.595      184.838   (   26)
CAG3                    B2                         1041.000     52.050      22.615      511.418   (   20)
CAG3                    C                           594.000     29.850      20.441      417.834   (   30)
CAG3                    D1                         1311.000     36.417       8.483       71.964   (   36)
CAG3                    D2                         1008.000     37.333       1.922       3.692   (   27)

  TOTAL CASES =     235
```

Table 9. Description of sediment types defined by Kamada et al. (1973), concerning statistical and compositional parameters.

```
BREAKDOWN OF CHIJIWA                                                    MAY  1, 1975        PAGE    4

FILE   CHIJIWA  (CREATION DATE = MAR.26, 1975)   CHIJIWA-BAY SEDIMENT SIZE DATA(WITH TYPE)
SUBFILE   ON1000     ON2000

- - - - - - - - - - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - - - - - - - - - -
CRITERION VARIABLE   MDPH      MEDIAN DIAMETER IN PHI SCALE
   BROKEN DOWN BY    TYPE      SEDIMENT TYPE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIABLE                CODE    VALUE LABEL          SUM          MEAN      STD DEV      VARIANCE        N

FOR ENTIRE POPULATION                               963.9203     4.1233      2.1892      4.7927  (  233)

TYPE                    1                            33.900      2.200       0.323       0.104   (   15)
TYPE                    1A                            2.700      2.700       0.0         0.0     (    1)
TYPE                    2A                           21.900      2.190       0.725       0.519   (   10)
TYPE                    2B                           81.420      1.696       1.193       1.424   (   48)
TYPE                    3                           191.900      4.795       0.660       0.436   (   40)
TYPE                    3A                          208.200      3.521       0.334       0.112   (   59)
TYPE                    3B                          402.600      7.063       0.550       0.302   (   57)
TYPE                    4                            32.400      8.100       0.115       0.013   (    4)
TYPE                    5                            -0.300     -0.150       0.071       0.005   (    2)

- - - - - - - - - - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - - - - - - - - - -
CRITERION VARIABLE   SO        SORTING COEFFICIENT
   BROKEN DOWN BY    TYPE      SEDIMENT TYPE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIABLE                CODE    VALUE LABEL          SUM          MEAN      STD DEV      VARIANCE        N

FOR ENTIRE POPULATION                               664.3700     2.8271      1.3568      1.8409  (  235)

TYPE                    1                            20.310      1.341       0.110       0.012   (   15)
TYPE                    1A                            1.830      1.830       0.0         0.0     (    1)
TYPE                    2A                           18.950      1.895       0.559       0.312   (   10)
TYPE                    2B                           81.880      1.706       0.546       0.296   (   48)
TYPE                    3                           130.120      3.253       1.352       1.827   (   40)
TYPE                    3A                          138.450      2.387       0.844       0.712   (   58)
TYPE                    3B                          236.340      4.130       0.969       0.901   (   57)
TYPE                    4                            13.020      3.255       0.094       0.009   (    4)
TYPE                    5                             3.470      1.735       0.021       0.000   (    2)

- - - - - - - - - - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - - - - - - - - - -
CRITERION VARIABLE   SK        SKEWNESS COEFFICIENT
   BROKEN DOWN BY    TYPE      SEDIMENT TYPE
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

VARIABLE                CODE    VALUE LABEL          SUM          MEAN      STD DEV      VARIANCE        N

FOR ENTIRE POPULATION                               170.6000     0.7260      0.3911      0.1529  (  235)

TYPE                    1                            10.490      1.663       0.110       0.012   (   15)
TYPE                    1A                            0.830      0.830       0.0         0.0     (    1)
TYPE                    2A                           11.690      1.169       0.851       0.725   (   10)
TYPE                    2B                           46.500      0.969       0.310       0.096   (   48)
TYPE                    3                            27.030      0.676       0.511       0.261   (   40)
TYPE                    3A                           30.220      0.521       0.193       0.037   (   58)
TYPE                    3B                           33.660      0.591       0.195       0.038   (   57)
TYPE                    4                             2.970      0.743       0.034       0.001   (    4)
TYPE                    5                             1.740      0.870       0.184       0.034   (    2)
```

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE G GRAVEL RATIO
BROKEN DOWN BY TYPE SEDIMENT TYPE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			199.1000	.84472	3.1811	10.1192	(235)
TYPE	1		0.900	0.060	0.232	0.054	(15)
TYPE	1A		0.0	0.0	0.0	0.0	(13)
TYPE	2A		37.700	3.770	6.578	43.269	(10)
TYPE	2B		86.400	1.800	3.129	9.791	(48)
TYPE	3		0.0	0.0	0.0	0.0	(40)
TYPE	3A		4.200	0.072	0.408	0.167	(58)
TYPE	3B		20.400	0.463	2.736	7.486	(57)
TYPE	4		0.0	0.0	0.0	0.0	(4)
TYPE	5		43.500	21.750	1.061	1.125	(2)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE SA SAND RATIO
BROKEN DOWN BY TYPE SEDIMENT TYPE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			12593.4004	53.3889	32.5734	1061.0256	(235)
TYPE	1		1478.900	98.593	3.197	10.218	(15)
TYPE	1A		77.400	77.400	0.0	0.0	(1)
TYPE	2A		824.500	82.450	10.775	116.096	(10)
TYPE	2B		4223.100	87.981	15.412	237.534	(48)
TYPE	3		2514.700	37.968	10.814	116.987	(40)
TYPE	3A		3637.400	62.714	8.374	70.127	(58)
TYPE	3B		670.300	11.760	11.000	121.008	(57)
TYPE	4		19.600	4.900	1.577	2.487	(4)
TYPE	5		143.500	71.750	8.132	66.125	(2)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE SI SILT RATIO
BROKEN DOWN BY TYPE SEDIMENT TYPE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			4700.4001	20.0023	20.0932	403.9381	(235)
TYPE	1		10.700	0.713	1.494	3.567	(15)
TYPE	1A		15.300	15.300	0.0	0.0	(1)
TYPE	2A		82.700	8.270	6.424	41.271	(10)
TYPE	2B		307.700	6.410	9.522	90.675	(48)
TYPE	3		1749.100	43.728	12.055	145.319	(40)
TYPE	3A		1498.100	25.829	7.723	59.646	(58)
TYPE	3B		2470.800	50.365	7.208	51.958	(57)
TYPE	4		177.900	44.475	0.477	0.769	(4)
TYPE	5		8.100	4.050	5.728	32.805	(2)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE CL CLAY RATIO
BROKEN DOWN BY TYPE SEDIMENT TYPE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			4055.3500	17.2569	15.7860	249.1984	(235)
TYPE	1		6.700	0.447	1.232	1.518	(15)
TYPE	1A		7.300	7.300	0.0	0.0	(1)
TYPE	2A		51.400	5.140	4.006	16.045	(10)
TYPE	2B		181.800	3.788	6.384	40.744	(48)
TYPE	3		753.600	18.923	9.097	82.758	(40)
TYPE	3A		658.400	11.359	3.926	15.414	(58)
TYPE	3B		2144.960	38.403	9.552	91.250	(57)
TYPE	4		202.500	50.625	1.565	2.449	(4)
TYPE	5		9.900	4.950	3.465	12.005	(2)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE MUD MUD RATIO
BROKEN DOWN BY TYPE SEDIMENT TYPE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			10740.0000	45.9000	33.3313	1112.3103	(235)
TYPE	1		20.100	1.340	3.011	9.067	(15)
TYPE	1A		22.600	22.600	0.0	0.0	(1)
TYPE	2A		134.100	13.410	10.194	103.915	(10)
TYPE	2B		489.600	10.196	15.609	243.703	(48)
TYPE	3		2482.200	62.055	10.797	116.578	(40)
TYPE	3A		2166.700	37.357	8.249	68.046	(58)
TYPE	3B		5007.500	87.851	12.974	168.426	(57)
TYPE	4		380.400	95.100	1.657	2.747	(4)
TYPE	5		13.500	6.500	9.192	84.500	(2)

DESCRIPTION OF SUBPOPULATIONS

CRITERION VARIABLE DEP DEPTH
BROKEN DOWN BY TYPE SEDIMENT TYPE

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			8739.0000	37.1702	16.6196	276.2102	(235)
TYPE	1		265.000	17.667	20.811	433.810	(15)
TYPE	1A		33.000	33.000	0.0	0.0	(1)
TYPE	2A		395.000	39.500	30.725	944.056	(10)
TYPE	2B		1636.000	34.083	21.776	474.208	(48)
TYPE	3		1604.000	40.100	11.422	130.451	(40)
TYPE	3A		2554.000	44.121	14.015	196.424	(58)
TYPE	3B		2027.000	35.596	5.979	35.751	(57)
TYPE	4		156.000	38.500	1.291	1.667	(4)
TYPE	5		62.000	31.000	0.0	0.0	(2)

TOTAL CASES = 235

Table 10. Univariate variance analysis of statistical and compositional parameters.

Procedure	TYPE	CAG3
Degrees of Freedom	8 226	6 228
MDPH	251.1	341.8
SO	39.6	51.6
SK	11.1	11.5
G	24.0	12.6
SA	224.4	399.8
SI	131.7	228.2
CL	117.5	166.4
MUD	213.2	389.6
DEP	4.9	17.2

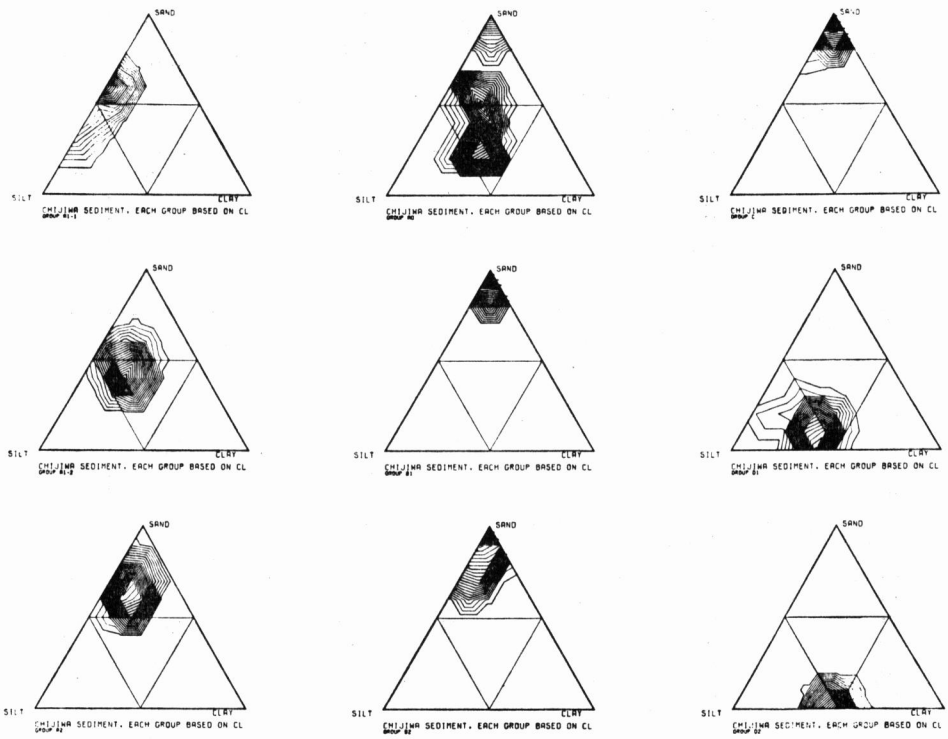


Fig. 30. Triangle contour diagrams, showing concentration pattern of samples on a triangle diagram of sand-silt-clay. Produced for each group defined by cluster analysis: A1–1; A1–2; A2; A0; B1; B2; C; D1; D2 group.

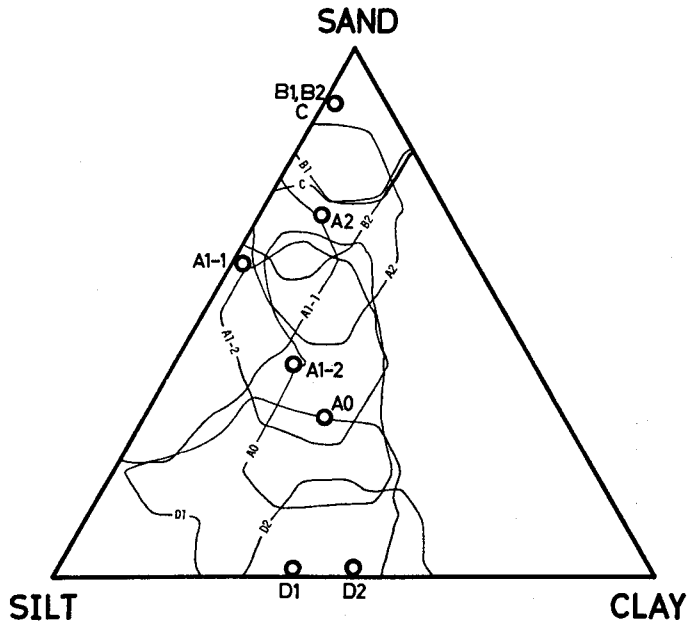


Fig. 31. Configuration of concentration centers of groups defined by cluster analysis, see also Fig. 30.

Evaluation of analytical results

The crosstabulation table between the two classification by KAMADA *et al.* (1973) and by the multivariate techniques demonstrated in this study are given in Table 11. According to this table, the sum of the groups D_1 and D_2 in this study corresponds to the group 3B by KAMADA *et al.*, the sum of the groups A_1 and A_2 to that of the groups 3 and 3A. But the groups containing coarser grains have no clear correspondence between the two classifications.

This may influence the interprepation of the sedimentary environments of the Chijiwa Bay.

As shown in this case study, the cluster analysis and factor analysis can classify the samples easily, definitely and in more detail, and the multivariate disciminant analysis can reveal the characterisrics of classified groups and the relation among them.

Table 11. Crosstabulation table between two classification by Kamada et al. (1973) (across side) and by cluster analysis (down side). Numbers in each cell indicate count, row percent, column percent and total percent in descending order.

CHIJIWA GROUP BASED ON CA, VAR ADDITION MAR.26, 1975 PAGE 57

FILE CHIJIWA (CREATION DATE = FEB. 1, 1975) CHIJIWA-BAY SEDIMENT SIZE DATA(WITH TYPE)
SUBFILE GN1000 GN2000

* * * * * * * * * * * * * * * * * * * CROSSTABULATION OF * * * * * * * * * * * * * * * * * * *
CAG3 GROUP BASED ON FR-CD-CA BY TYPE SEDIMENT TYPE
* PAGE 1 OF 1

CAG3	TYPE / COUNT / ROW PCT / COL PCT / TOT PCT	1	1A	2A	2B	3	3A	3B	4	5	ROW TOTAL
A1		0	0	0	1	30	17	6	0	0	54
		0.0	0.0	0.0	1.9	55.6	31.5	11.1	0.0	0.0	23.0
		0.0	0.0	0.0	2.1	75.0	29.3	10.5	0.0	0.0	
		0.0	0.0	0.0	0.4	12.8	7.2	2.6	0.0	0.0	
A2		0	0	4	1	4	35	0	0	0	42
		0.0	0.0	9.5	2.4	9.5	83.3	0.0	0.0	0.0	17.9
		0.0	0.0	20.0	2.1	10.0	60.3	0.0	0.0	0.0	
		0.0	0.0	0.9	0.4	1.7	14.9	0.0	0.0	0.0	
B1		14	0	1	11	0	0	0	0	0	26
		53.8	0.0	3.8	42.3	0.0	0.0	0.0	0.0	0.0	11.1
		93.3	0.0	10.0	22.9	0.0	0.0	0.0	0.0	0.0	
		6.0	0.0	0.4	4.7	0.0	0.0	0.0	0.0	0.0	
B2		1	1	2	11	0	5	0	0	0	20
		5.0	5.0	10.0	55.0	0.0	25.0	0.0	0.0	0.0	8.5
		6.7	100.0	20.0	22.9	0.0	8.6	0.0	0.0	0.0	
		0.4	0.4	0.9	4.7	0.0	2.1	0.0	0.0	0.0	
C		0	0	5	23	0	0	0	0	2	30
		0.0	0.0	16.7	76.7	0.0	0.0	0.0	0.0	6.7	12.8
		0.0	0.0	50.0	47.9	0.0	0.0	0.0	0.0	100.0	
		0.0	0.0	2.1	9.8	0.0	0.0	0.0	0.0	0.9	
D1		0	0	0	1	6	1	28	0	0	36
		0.0	0.0	0.0	2.8	16.7	2.8	77.8	0.0	0.0	15.3
		0.0	0.0	0.0	2.1	15.0	1.7	49.1	0.0	0.0	
		0.0	0.0	0.0	0.4	2.6	0.4	11.9	0.0	0.0	
D2		0	0	0	0	0	0	23	4	0	27
		0.0	0.0	0.0	0.0	0.0	0.0	85.2	14.8	0.0	11.5
		0.0	0.0	0.0	0.0	0.0	0.0	40.4	100.0	0.0	
		0.0	0.0	0.0	0.0	0.0	0.0	9.8	1.7	0.0	
	COLUMN TOTAL	15	1	10	48	40	58	57	4	2	235
		6.4	0.4	4.3	20.4	17.0	24.7	24.3	1.7	0.9	100.0

CHI SQUARE = 536.99706 WITH 48 DEGREES OF FREEDOM SIGNIFICANCE = 0.0
CRAMER'S V = 0.61713
CONTINGENCY COEFFICIENT = 0.83402
KENDALL'S TAU B = 0.20361 SIGNIFICANCE = 0.0000
KENDALL'S TAU C = 0.19545 SIGNIFICANCE = 0.0000
GAMMA = 0.22817
SOMER'S D = 0.20856

Conclusion

In the case study mentioned above, the grain-size distribution data are analyzed using the computer supported system. Namely, factors are extracted and their meanings are examined by R-mode factor analysis; samples are classified into several groups by the Q-mode factor analysis and cluster analysis; of their significance, the given classifications are tested by the discriminant analysis; and the results are also shown in several types of diagrams using the display programs. In the course of these studies the data file system is used for the storage, retrieval, selection, transformation, and display of data.

The results are well concordant with those which are concluded from the ex-

cellent studies of ordinal geologic techniques. Therefore, the validity of the data analysis in geology is confirmed. Simultaneously the usefulness of the computer supported system presented here is shown in the smooth processing of thses studies.

The size of geologic data is generally larger than those used in this study, and the data analysis for such data is hardly performed without the help of computer. The computer supported system presented here can treat such a large amount of data, and is available for any geologists with a little knowledge on the computer.

In this study mainly numeric data are used, though large amounts of geologic data are nonmetric, nominal, and nonquantitative. The computer supported system can treat only a part of such data. By the improvement of the system, however, nonquantitative geologic data will be treated in near future.

I have studied not only the computer supported system, but also the mathematical and statistical techniques to treat geologic data. Some of the contributions on these aspects will be reported in future.

References

ANDERBERG, M. R. (1973): *Cluster Analysis for Applications,* Academic Press, New York and London, 359 p.

BLACKITH, R. E., and REYMENT, R. A. (1971): *Multivariate Morphometrics,* Academic Press, New York and London, 412 p.

BONHAM-CARTER, G. F. (1967): FORTRAN IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers. *Computer Contr.* 17, Kansas Geol. Surv., 28 p.

CHEETHAM, A. H., and HAZEL, J. E. (1969): Binary (presence-absence) similarity coefficients. *Jour. Palaeont.,* v. 43, n. 5, pp. 1130–1136.

COLLYER, P. L., and MERRIAM, D. F. (1973): An application of cluster anaysis in mineral exploration. *Jour. Intern. Assoc. Mathematical Geol.,* v. 5, n. 3, pp. 213–223.

DAVIS, J. C. (1970): Information contained in sediment-size analysis. *Jour. Intern. Assoc. Mathematical Geol.,* v. 2, n. 2, pp. 105–112.

——— (1973): *Statistics and Data Analysis in Geology,* John Wiely and Sons, Inc., New York, 550 p.

DIXON, W. J. (ed.) (1973): *BMD Biomedical Computer Programs,* Univ. California Press, Berkley, 773 p.

——— (ed.) (1975): *BMDP Manual,* Univ. California Press, Berkley, 785 pp.

DRAPEAU, G. (1973): Factor analysis: How it copes with complex geological problems. *Jour. Intern. Assoc. Mathematical Geol.,* v. 5, n. 4, pp. 351–363.

FOLK, R. L., and WARD, W. C. (1957): Brozos River bar, a study in the significance of grain size parameter. *Jour. Sed. Petrol.,* v. 27, n. 1, pp. 3–26.

HARBAUGH, J. W., and BONHAM-CARTER, G. F.(1970): *Computer Simulation in Geology,* Wiley-Interscience, New York, 575 p.

———, and MERRIAM, D. F. (1968): *Computer applications in Stratigraphic analysis,* John Wiley and Sons Inc., New York, 282 p.

HATTORI, I, and MIZUTANI, S. (1971): Computer simulation of fracturing of layered rock. *Engineer. Geol.,* v. 5, n. 5, pp. 253–269.

HAYAMI, I. and NAKANO, M. (1968): A consideration on the numerical taxonomy in palaeontology, taking the Trigoniidae as example. *Sci. Rep. Fac. Sci. Kyushu Univ. (Geol.),* v. 8, n. 4, pp. 191–236 (in Japanese).

HAZEL, J. E. (1970): Binary coefficients and clustering in biostratigraphy. *Geol. Soc. Amer. Bull.,*

v. 81, pp. 3237–3252.

HOWARTH, R. J. (1973): Preliminary assessment of a non-linear mapping algorithm in a geological context. *Jour. Intern. Assoc. Mathematical Geol.*, v. 5, n. 1, pp. 39–57.

IMBRIE, J. and PURDY, E. G. (1962): Classification of modern Bahamian carbonate sediments. *Amer. Assoc. Petrol. Geol. Mem.* 1, pp. 253–272.

INMAN, D. L., and CHAMBERLAIN, T. K. (1955): Particle-size distribution in nearshore sediments. *Soc. Eco. Paleont. Mineral. Spec. Pub.* n. 3, pp. 106–129.

INOUE, N. (1970): Submarine topography and grain-size distributions of sediments in Tachibana Bay. In Report of investigations on the hydrographic conditions and fish shoal distributions in Tachibana Bay. (*Contr. Seikai Region Fish. Res. Lab.*, n. 256) pp. 23–32 (in Japanese).

KAMADA, Y. (1966): Coarse fraction studies of marine sediments along the Mogi Coast in Chijiwa Bay, Nagasaki Prefecture. *Sci. Bull. Fac. Lib. Arts Educ. Nagasaki Univ.*, n. 17, pp. 45–54 (in Japanese).

————, and HORIGUCHI, Y., INOUE, M., and WATANABE, H. (1973): Bottom sediments of Chijiwa Bay, Nagasaki Prefecture, with special reference to the distribution of muddy sediments. *Bull. Fac. Lib. Arts Nagasaki Univ. Natural Sci.*, n. 24, pp. 61–79 (in Japanese).

KLOVAN, J. E. (1966): The use of factor analysis in determining depositional environments from grain-size distributions. *Jour. Sed. Petrol.*, v. 36, n. 1, pp. 115–125.

KLOVAN, J. E., and IMBRIE, J. (1971): An algorithm and FORTRAN-IV program for large-scale Q-mode factor analysis and calculation factor scores. *Jour. Intern. Assoc. Mathematical Geol.*, vol. 3, n. 1, pp. 61–77.

KOMAR, P. D. (1973): Computer models of delta growth due to sediment input from rivers and longshore transport. *Geol. Soc. Amer. Bull.*, v. 84, n. 7, pp. 2217–2226.

KRUSKAL, J. B. (1964a): Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrica*, v. 29, n. 1, pp. 1–26.

———— (1964b): Nonmetric multidimensional scaling: A numeric method. *Psychometrica*, v. 29, n. 2, pp. 115–129.

LUMSDEN, D. N. (1973): TRI: A FORTRAN subroutine to plot points on a triangular diagram. *Geol. Soc. Amer. Bull.*, v. 84, n. 5, pp. 1765–1768.

MCCAMMON, R. B. (1966): Principal component analysis and its application in large-scale correlation studies. *Jour. Geol.*, v. 74, n. 5, pt. 2, pp. 721–733.

MERRIAM, D. F. (ed.) (1969): Symposium on computer applications in petroleum exploration. *Computer Contr.* 40, Kansas Geol. Surv., 41 p.

MILLER, R. L., and KAHN, J. S. (1962): *Statistical analysis in the geological sciences*, John Wiley and Sons, Inc., New York, 483 p.

MIZUTANI, S. (1974): Mathematical geology. *Earth Science* ("*Chikyu Kagaku*"), v. 28, n. 4, pp. 128–149 (in Japanese).

NIE, N. H., BENT, D. H., and HULL, C. H. (1975): *Statistical Package for the Social Sciences*. McGraw-Hill Book Company, New York, 675 p.

NISHIWAKI, N., and YAMAMOTO, K. (1975): Computer processing of geologic data specially on compositional data of sandstone. *Jour. Geography* v. 84, n. 6, pp. 317–335.

NOBLE, D. C., and EBERLY, S. W. (1964): Discussion; A digital computer procedure for preparing beta diagrams. Amer. Jour. Sci., v. 262, n. 9, pp. 1124–1129.

PARKS, J. M. (1966): Cluster analysis applied to multivariate geologic problems. *Jour. Geol.*, v. 74, n. 5, pt. 2, pp. 703–715.

———— (1969): Multivariate facies maps. *Computer Contr.* 40, Kansas Geol. Surv., pp. 6–12.

———— (1970): FORTRAN IV program for Q-mode cluster analysis on distance function with printed dendrogram. *Computer Contr.* 46, Kansas Geol. Surv., 32 p.

ROBINSON, P. (1963): Preparation of beta diagrams in structural geology by a digital computer. *Amer. Jour. Sci.*, v. 261, n. 10, pp. 913–928.

SAMMON, J. W. Jr. (1969): A nonlinear mapping for data structure analysis. *IEEE Trans. Com-*

puters C-18, pp. 401–409.

SHEPARD, F. P. (1954): Nomenclature based on sand-silt-clay ratios. *Jour. Sed. Petrol.*, v. 24, n. 3, pp. 151–158.

SOKAL, R. R., and SNEATH, P. H. A. (1963): *Principles of Numerical Taxonomy*, Freeman and Co., San Francisco, 359 p.

SPENCER, D. W. (1963): The interpretation of grain size distribution curves of clastic sediments. *Jour. Sed. Petrol.*, v. 33, n. 1, pp. 180–190.

TRASK, P. D. (1932): *Origin and environment of source sediments of petroleum.* Gulf Pub. Co., Huston, Texas.

UJIIÉ, H. (1973): Sedimentation of planktonic foraminiferal shells in the Tsushima and Korea Straits between Japan and Korea. *Micropaleont.*, v. 19, n. 4, pp. 444–460.

VISTELIUS, A. B. (1967): *Studies in Mathematical Geology*, Consultants Bureau, New York, 294 p.

WARNER, J. (1969): FORTRAN IV programs for construction of Pi diagrams with the Univac 1108 computer. *Computer Contr.* 33, Kansas Geol. Surv., 38 p.

WEBB, W. M., and BRIGGS, L. I. (1966): The use of principal component analysis to screen mineralogical data. *Jour. Geol.*, v. 74, n. 5, pt. 2, pp. 716–720.

YAMAMOTO, K. (1973a): The trend surface analysis by computer and its application. *Jour. Geol. Soc. Japan*, v. 79, n. 5, pp. 349–362 (in Japanese).

——— (1973b): Computer program for displaying samples in a three-dimensional factor space by using an X–Y plotter. *Kyoto Univ. Data Proc. Center Bull.*, v. 6, n. 8, pp. 300–302 (in Japanese).

YAMAMOTO, K. (1974a): Computer simulation for a sedimentary process. *Proc. Kansai Branch, Geol. Soc. Japan*, no. 76, pp. 6–7 (Abstract).

——— (1974b): An application of factor analysis to sedimentary environments. *Jour. Geol. Soc. Japan*, v. 80, n. 4, pp. 165–177 (in Japanese).

——— (1974c): Explanation of BMDP programs converted for FACOM 230-60/75 (1). *Kyoto Univ. Data Proc. Center Bull.*, v. 7, n. 4, pp. 186–193 (in Japanese).

——— (1975a): Explanation of BMDP programs converted for FACOM 230-60/75 (2). *Kyoto Univ. Data Proc. Center Bull.*, v. 8, n. 1. (in Japanese).

——— (1976): Preparing a contour map with a line-printer from irregularly spaced data. *Jour. Inform. Proc. Soc. Japan*, v. 7, n. 5, pp. 448–450. (in Japanese).

———, and NAKAGAWA, Y. (1974): Time trand analysis of the Plio-Pleistocene sequence in the central part of the Kinki District, Japan. *Mem. Fac. Sci., Kyoto Univ., Ser. Geol. Mineral.*, v. 40, n. 2, pp. 45–65.

———, and NISHIWAKI, N. (1975a): FORTRAN program of preparing contour maps for geologic use. *Mem. Fac. Sci., Kyoto Univ., Ser. Geol. Mineral.*, v. 41, n. 1, pp. 1–34.

———, and ——— (1975b): Computer processing of oceanographic data, by using SPSS "Statistical Package for the Social Sciences". *Marine Sci. Month.*, v. 7, n. 6, pp. 53–59. (in Japanese).

YAMAMOTO, K., and NISHIWAKI, N. (1975c): Computer programs of trend surface analysis by polynomial approximation and moving average, and of contouring with an X–Y plotter. *Kyoto Univ. Data Proc. Center Bull.*, v. 8, n. 3, pp. 121–134 (in Japanese).

———, and ——— (1975d): Automatic analysis of the geologic structure from dip-strike data. *Computers and Geosciences*, v. 1, n. 4, pp. 309–323.