# USE THE GEOTECHNICAL DATA BANK!

Gary D. Goldberg, Woodward-Clyde Consultants
C. W. Lovell and R. D. Miles, Purdue University

## Abstract

Computerized systems of data storage, retrieval and analysis for information about the soils and rocks within a state are being used relatively frequently. These systems have particular and special potential uses in the design of low-volume roads, where the funds to generate original geotechnical data are very limited. This paper briefly describes the data likely to exist in such a system, how to access them, and what kinds of predictions can likely be made from them. Of special interest are: (1) frequency distribution analyses of particular soil characteristics to determine typical magnitudes and variabilities within a given area, and (2) correlation of simple and easy-to-measure soil characteristics with parameters that require complex and costly tests. Either approach may supply appropriate presumptive values for the structural roadway design, which can be verified by original subsurface investigation if the budget permits. Specific examples are drawn from the Indiana bank of geotechnical data.

The need for geological, pedological and geotechnical engineering information for use in site selection, planning, design, construction, and maintenance of transportation facilities and of most engineering structures is widely realized. Much of the information initially required by the engineer is used in preliminary construction planning, site selection and for guidance in further soil investigations. Unfortunately, most of these data are necessarily limited in quantity due to economic and time constraints.

The engineer is therefore faced with the problem of determining the location, sequence, thickness, and areal extent of each soil stratum, including a description and classification of the soils and their structure, by extrapolating the data from a few selected sites to an area many times greater than that which has been sampled. Even though large amounts of detailed soils data are available from previous work performed during planning and construction of nearby projects, these data are usually not readily accessible for use, or their existence is unknown.

The need therefore exists to make this information more accessible both for the engineer interested in detailed information of a site and the engineer interested in general soil characteristics over a large area. A computerized geotechnical data bank is the most efficient, expedient, and economical way to reduce the accumulated data to a form which can readily be made available to interested individuals, such as highway engineers, geotechnical engineers, contractors, and land use planners.

This paper briefly describes the development of a computerized geotechnical data bank, including the data likely to exist in such a system, and the kinds of predictions that can likely be made from them. Of special interest are: (1) frequency distribution analyses of particular soil characteristics to determine typical magnitudes and variabilities within a given area, and (2) correlation of simple and easy-to-measure soil characteristics with parameters that require complex and costly tests. Both uses of the data are particularly appropriate to low volume roads, where the geotechnical data generation for a particular job is quite limited.

## Data Bank

Large amounts of geotechnical information for transportation projects are accumulated each year by highway departments throughout the United States and abroad. Geotechnical investigations are conducted to provide surface and subsurface information relative to soil, rock and water. This information is used in selecting the proper locations for the project and in making design decisions (3, 11). Subsequent use of this information after the design and construction of a project from which soil samples are taken and geotechnical data generated has been limited (4, 10).

Recognizing this, geotechnical data banks have been developed in a number of geographic locations, e.g., South Dakota (1), Kentucky (9), Indiana (2) and Sweden (5). The authors' experience is with the Indiana data bank, where data have been collected from private consulting firms, private soil testing firms, and from tests conducted by the Indiana State Highway Commission (ISHC).

What data are likely to be available from such sources? The minimum information stored for a particular boring is:

1. Location,
2. Gradational characteristics based on standard sieve sizes and hydrometer analysis,
3. Atterberg limits,
4. Visual textural classifications,
5. Color based on moist condition.

It is simple to write a computer program utilizing the above information to classify the samples by the American Association of State Highway and Transportation Officials (AASHTO) and Unified Soil Classification (USC) systems (2).

Other information to be stored, if available, includes:

1. Organic content (loss on ignition),
2. In situ moisture content,
3. In situ dry and wet densities,
4. Specific gravity,
5. Compaction test results,
6. California bearing ratio (CBR)
7. Unconfined compressive strength and failure strain,
8. Strength data from triaxial and direct shear tests,
9. Consolidation test results.

In addition to laboratory test data, field information to be stored should include:

1. Project identification,
    a. project number
    b. contract number
    c. road number
    d. data collection agency
2. Sample location,
    a. county
    b. district
    c. township
    d. range
    e. section
    f. line number
    g. station number
    h. offset and the left or right direction from the centerline
3. Sample identification,
    a. boring number
    b. laboratory number
    c. sampling procedure
4. Date the sample was taken from the hole,
5. Physiographic region,
6. Parent material from which the soil has been derived,
7. Ground surface elevation,
8. Depth from which the sample has been removed,
9. Depth to bedrock,
10. Depth to groundwater,
11. Standard penetration resistance (SPT),
12. Pedological soils information,
    a. soil association name
    b. soil series name
    c. horizon
    d. slope (topographic) class
    e. erosion class
    f. natural soil drainage class
    g. permeability
    h. flooding potential
    i. frost heave susceptibility
    j. shrink-swell potential
    k. pH

All of this information is transferred to a Data Input Form (DIF) such as is shown in Figure 1.

A User's Manual, explaining in detail the operation of the computerized data system, must be prepared (2). Included in the User's Manual are descriptions of the data items, the codification scheme to make the system compatible with computerized storage and retrieval, card formats, and card and column locations for each data item. Also needed is a listing of the programs used to add additional data to the data bank, to check data input errors, to use the computer programs for data management and manipulations, as well as example problems on the use of the data bank.

## Benefits of the Data Bank

The benefits which can be obtained from the development of a computerized geotechnical data storage and retrieval system can be divided into two major categories: (a) the direct use of raw data; and (b) the use of statistical methods to reduce the data to a usable form via distribution characterizations, correlations, and predictions. Either approach may supply appropriate presumptive values for the structural design of low volume roadways, which can be verified by original subsurface investigation if the budget permits.

### Direct Use of Raw Data

A computerized geotechnical data bank provides the capability of retrieving an extensive listing of available soil and rock information both quickly and economically. For example, the location of possible sources of granular and select borrow materials could be facilitated, along with route selection studies and right-of-way appraisals. Problem soil areas may be identified. In addition, the compilation of large scale engineering soil maps and profiles based on engineering characteristics is possible (1). This information would be of particular value in locating low volume roads.

### Statistical Methods of Data Reduction

Statistical methods are used to study the variability of soil properties, to compare one soil type to another, and to group soil types with similar soil characteristics. Various correlations among selected soil properties can also be useful to the engineer when extensive laboratory testing is not possible (8, 12).

The first step in assessing the variability and typical magnitudes of selected soil characteristics is to develop frequency distribution characterizations of selected soil properties. In an attempt to explain the variation in the data, the soil samples may be grouped according to physiographic regions and parent material area. Figures 2 through 4 graphically illustrate the range, 95% confidence interval, and the mean of selected soil parameters, based upon such groupings for Indiana. These values will help the engineer to obtain an idea of the expected values of the soil parameters.

Prediction models of parameters that are difficult to measure and therefore require complex and costly tests are potentially of value if correlations can be made with simple and easy-to-measure soil characteristics. This was attempted for the Indiana data with dependent variables of:

(1) coefficient of consolidation ($C_c$) and compression ratio ($C_r$, which equals $C_c/1+e_o$, where $e_o$ is the initial void ratio), (2) unconfined compressive strength ($q_u$), (3) standard Proctor maximum dry ($\gamma_{d_{max}}$) and wet ($\gamma_{m_{max}}$) densities and optimum moisture content ($w_{opt}$), and (4) soaked California bearing ratios (CBR) at 100 (CBRS01) and 95 (CBRS02) percent of standard Proctor maximum dry densities.

The independent variables included: (1) initial void ratio ($e_o$), natural moisture content ($w_n$), natural dry density ($\gamma_d$), liquid limit ($w_L$), plastic limit ($w_p$), plasticity index ($I_p$), percent clay, overburden pressure ($p_o$), and preconsolidation pressure ($p_c$) for the consolidation test data; (2) $w_L$, $w_p$, $w_n$, $\gamma_d$, and liquidity index ($L_I$) for the unconfined compressive strength data; and (3) $w_L$, $w_p$, $I_p$, and shrinkage limit ($w_s$) for the compaction and CBR test data.

If a particular dependent variable resisted state-wide modelling, or if data were contained in significant quantities to justify modelling on smaller units, that is, physiographic regions, parent material areas, and in some cases on soil types, the data were grouped accordingly to determine if the prediction models could be significantly improved.

Regression analysis was used to establish the prediction models for each dependent variable. The method of least squares was used to find "good" estimates of the regression parameters and to isolate the effects of the independent variables on the chosen dependent variables. The REGRESSION routine of the Statistical Package for the Social Sciences (SPSS) developed by Nie et al (7) was used. The regression results give adjusted coefficients of multiple determination ($R^2_a$). Usually, only the coefficient of multiple determination, denoted by $R^2$, is used as a measure of the proportionate reduction of the total variation in the dependent variable associated with the use of a set of independent variables. Since $R^2$ can be made large by increasing the independent variables in the model, $R^2_a$ was used as a criterion for selecting a good model. $R^2_a$ recognizes the number of independent variables in the model and may actually become smaller when another independent variable is introduced into the model (6). $R^2_a$ takes on values between 0 and 1. The larger $R^2_a$, the better the fitted equation explains the variation in the data.

The regression results summarized in Tables 1 through 3 are examples of analyses of stored data, with $R^2_a$ values greater than 0.65.

## Summary

The geotechnical data bank has substantial potential for aiding in the design and construction of low volume roads. Frequency distribution analysis permits prediction of the range of values which may be expected for a given soil parameter, although these vary with the particular physical property and the population from which the soil has been sampled. The grouping of soils by physiographic regions and the origin of their parent materials shows that the predicability of some soil properties can be thereby improved.

Where budgets strictly limit the amount of geotechnical sampling and testing, prediction equations can be generated. We have shown how this was done for compressibility, strength, and compaction parameters. In some cases valid regression equations could be produced for an entire state, but generally the relations were stronger when the population of samples was from a smaller geologic or pedologic unit.

As new data become available for incorporation into the data bank, they should be used for validating the existing prediction models. The reliability of the equations can subsequently be improved. Soils information which was essentially "lost" after a project was completed, can now be utilized for future highway projects and improvements. The data bank should be maintained for all potential users, particularly the designers of low volume roads. Initial access to the bank for the latter group would undoubtedly be on a manual basis, with possible computerization later.

## Acknowledgments

The authors are grateful to the Joint Highway Research Project of Purdue University, the Indiana State Highway Commission, and the U. S. Federal Highway Administration for their financial support of the research reported herein.

## References

1. Crawford, R. A., Thomas, J. B. and Stout, M. Jr. Computerized Soil Test Data for Highway Design. Physical Research Section, South Dakota Department of Highways, Pierre, South Dakota, 1972, 24 pp.
2. Goldberg, G. D. Development of the Computerized Geotechnical Data Bank for the State of Indiana. MSCE Thesis, Purdue University, May 1978, 163 pp. See also Joint Highway Research Project Report No. 78-6.
3. Hagerty, D. J., Schmitt, N. G. and Phalzer, W. J. Retrieval and Use of Geotechnical Information. Ohio River Valley Soil Seminar, Geotechnics in Transportation Engineering, October 1973, pp. 1-9.
4. Humphreys, A. P. Data Presentation, Preservation and Recovery. Ground Engineering, Vol. 9, No. 8, November 1976, pp. 33-38.
5. Lundin, S. E., Stephansson, O. and Zetterlund, P. Geoteknisk Databank. Rapport R70, National Swedish Building Research Institute, 1973, 158 pp.
6. Neter, J. and Wasserman, W. Applied Linear Statistical Models. Richard D. Irwin, Inc., 1974, 842 pp.
7. Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. and Bent, D. H. Statistical Package for the Social Sciences. 2nd Edition, McGraw-Hill Book Co., 1975, 675 pp.
8. Rizkallah, V. and El Nimr, E. Applicability of Regression Analysis in Soil Mechanics with the Help of Data Banks. Application of Statistics and Probability in Soil and Structural Engineering, 2nd International Conference, Aachen, F. R. Germany, Vol. 2, 1975, pp. 423-438.
9. Spradling, D. A Soils Data System for Kentucky. Research Report 441, Division of Research, Bureau of Highways, Kentucky Department of Transportation, 1976, 58 pp.
10. Tomlinson, M. J. and Francis, H. W. A. Site Investigation for Motorways. The Highway Engineer, Vol. XX, No. 6, June 1973, pp. 23-33.

11. Transportation Research Board. Acquisition and Use of Geotechnical Information. Synthesis of Highway Practice 33, National Cooperative Highway Research Program, National Research Council, Washington, D.C., 1976, 40 pp.
12. Van Rooyen, M. Intercorrelation of Engineering Properties of Soil. Proceedings of the 4th Regional Conference for Africa on Soil Mechanics and Foundation Engineering, Cape Town, 1967, pp. 203-208.

Table 1 Summary of Regression Equations for Prediction of Compression Index ($C_c$) and Compression Ratio ($C_r$).

| Unit | Dependent Variable | $R_a^2$ | Regression Equation | Number of Samples, N |
|---|---|---|---|---|
| All Samples | $C_c$ | 0.856 | $C_c = 0.5684\ (e_o + 0.0033\ w_L - 0.0082\ w_p + 0.0329\ p_c - 0.4322)$ | 96 |
| | | 0.800 | $C_c = 0.5363\ (e_o - 0.4110)$ | |
| | | 0.792 | $C_c = 0.0002\ (w_n^2 - 106.2727)$ | |
| | | 0.783 | $C_c = 0.0129\ (w_n + 0.1015\ w_L - 16.1875)$ | |
| | $C_r$ | 0.691 | $C_r = 0.2037\ (e_o - 0.2465)$ | |
| Wabash Lowland | $C_c$ | 0.838 | $C_c = 0.5673\ (e_o - 0.4422)$ | 29 |
| | log $C_c$ | 0.831 | $\log C_c = 2.7904\ (e_o - 0.3346\ e_o^2 - 0.8449)$ | |
| | $C_r$ | 0.750 | $C_r = 0.221\ (e_o - 0.3074)$ | |
| | | 0.748 | $C_r = 0.0065\ (w_n - 11.6361)$ | |
| | | 0.735 | $C_r = 0.0034((e_o)\ (w_n) + 8.3647)$ | |
| Crawford Upland | $C_c$ | 0.859 | $C_c = 0.0101\ ((e_o)\ (w_L) - 0.5765\ w_L + 12.665)$ | 28 |
| | | 0.833 | $C_c = 0.0114\ (w_n + 0.2491\ w_L - 18.8134)$ | |
| | | 0.788 | $C_c = 0.4941\ (e_o - 0.3507)$ | |
| | | 0.777 | $C_c = 0.0133\ (w_n - 12.1886)$ | |
| | $C_r$ | 0.740 | $C_r = 0.0001\ (w_n^2 + 455.8889)$ | |
| | | 0.736 | $C_r = 0.0033\ ((e_o)\ (w_n) + 12.5168)$ | |
| | | 0.721 | $C_r = 0.1164\ (e_o^2 + 0.3594)$ | |
| Outwash and Alluvial Deposits | $C_c$ | 0.894 | $C_c = 0.6076\ (e_o + 0.003\ w_L - 0.0095\ w_p + 0.0430\ p_c - 0.4186)$ | 63 |
| | | 0.842 | $C_c = 0.5621\ (e_o - 0.4215)$ | |
| | | 0.822 | $C_c = 0.0153\ (w_n + 0.1022\ w_L - 0.3104\ w_p - 11.6123)$ | |
| | log $C_c$ | 0.772 | $\log C_c = 2.1389\ (e_o - 0.2967\ e_o^2 - 0.9374)$ | |

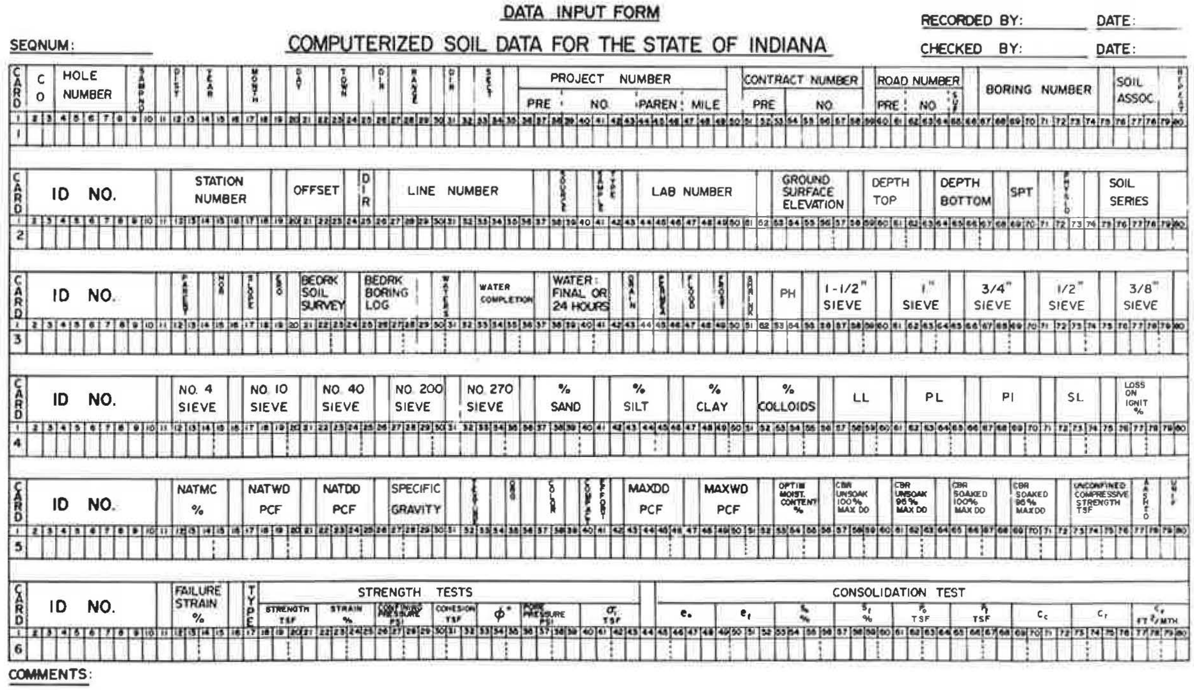
Table 2 Summary of Regression Equations for Prediction of Unconfined Compressive Strength ($q_u$).
($q_u$ in kPa; $\gamma_d$ in kg/m$^3$)

| Unit | Dependent Variable | $R_a^2$ | Regression Equation | Number of Samples, N |
|---|---|---|---|---|
| Calumet Lacustrine Plain | $q_u$ | 0.756 | $q_u = 0.00268\ (\gamma_d^2 - 37.333)$ | 40 |
| | $\log q_u$ | 0.750 | $\log q_u = 0.0257\ (\gamma_d^2 - 116.265)$ | |
| Lacustrine Deposits | $\log q_u$ | 0.699 | $\log q_u = -\ 0.0257\ (\gamma_d^2 + 116.150)$ | 48 |

Table 3 Summary of Regression Equations for Prediction of Standard Proctor Maximum Dry ($\gamma_{d_{max}}$) and Wet ($\gamma_{m_{max}}$) Densities and Optimum Moisture Content ($w_{opt}$).
($\gamma$'s in kg/m$^3$)

| Unit | Dependent Variable | $R_a^2$ | Regression Equation | Number of Samples, N |
|---|---|---|---|---|
| All Samples | $w_{opt}$ | 0.894 | $w_{opt} = -\ 7.958\ (\gamma_{d_{max}} - 9.005)$ | 138 |
| Valparaiso Morainal Area | $\log \gamma_{d_{max}}$ | 0.816 | $\log \gamma_{d_{max}} = -\ 3.683\ (1/w_L + 0.127 \log w_L - 0.454)$ | 26 |
| | | 0.785 | $\log \gamma_{d_{max}} = -\ 0.224\ (\log w_L - 5.269)$ | |
| | $\gamma_{m_{max}}$ | 0.790 | $\gamma_{m_{max}} = -7.118\ (\log w_L + 9.962\ (1/w_L) - 2.976)$ | |
| | $\log \gamma_{m_{max}}$ | 0.694 | $\log \gamma_{m_{max}} = -\ 0.135\ (\log w_L - 8.294)$ | |
| | $w_{opt}$ | 0.972 | $w_{opt} = 11.649\ (\gamma_{m_{max}} - 1.298\ \gamma_{d_{max}} + 3.203)$ | |
| | | 0.870 | $w_{opt} = 6.769\ (\gamma_{d_{max}} - 9.355)$ | |
| | | 0.810 | $w_{opt} = 23.0357 + 0.002\ (w_L)\ (w_p) - 285.939\ (1/w_L)$ | |
| Residuum of Limestone Bedrock | $\gamma_{d_{max}}$ | 0.772 | $\gamma_{d_{max}} = -\ 7.0843\ (\log w_L + 14.095\ (1/w_L) - 2.906)$ | 22 |
| | $\log w_{opt}$ | 0.781 | $\log w_{opt} = 0.0042\ (w_L + 259.0381)$ | |

Figure 1. Data input form (DIF).



DATA INPUT FORM

SEQNUM:

COMPUTERIZED SOIL DATA FOR THE STATE OF INDIANA

RECORDED BY: DATE:

CHECKED BY: DATE:

CARD 1: C O · HOLE NUMBER · PROJECT NUMBER (PRE · NO · PAREN · MILE) · CONTRACT NUMBER (PRE · NO) · ROAD NUMBER (PRE · NO) · BORING NUMBER · SOIL ASSOC.

CARD 2: ID NO. · STATION NUMBER · OFFSET · DIR · LINE NUMBER · LAB NUMBER · GROUND SURFACE ELEVATION · DEPTH TOP · DEPTH BOTTOM · SPT · SOIL SERIES

CARD 3: ID NO. · BEDRK SOIL SURVEY · BEDRK BORING LOG · WATER COMPLETION · WATER: FINAL OR 24 HOURS · PH · 1-1/2" SIEVE · 1" SIEVE · 3/4" SIEVE · 1/2" SIEVE · 3/8" SIEVE

CARD 4: ID NO. · NO. 4 SIEVE · NO. 10 SIEVE · NO. 40 SIEVE · NO. 200 SIEVE · NO. 270 SIEVE · % SAND · % SILT · % CLAY · % COLLOIDS · LL · PL · PI · SL · LOSS ON IGNIT %

CARD 5: ID NO. · NATMC % · NATWD PCF · NATDD PCF · SPECIFIC GRAVITY · MAXDD PCF · MAXWD PCF · OPTIM MOIST. CONTENT % · CBR UNSOAK 100% MAX DD · CBR UNSOAK 95% MAX DD · CBR SOAKED 100% MAX DD · CBR SOAKED 95% MAX DD · UNCONFINED COMPRESSIVE STRENGTH TSF

CARD 6: ID NO. · FAILURE STRAIN % · TYPE · STRENGTH TESTS (STRENGTH TSF · STRAIN % · CONFINING PRESSURE PSI · COHESION TSF · $\phi°$ · PORE PRESSURE PSI · $\sigma_1$ TSF) · CONSOLIDATION TEST ($e_o$ · $e_f$ · $C_c$ · $C_r$ · $C_v$ FT²/MTH)

COMMENTS:

Figure 2. Distributional characterization of natural moisture content.

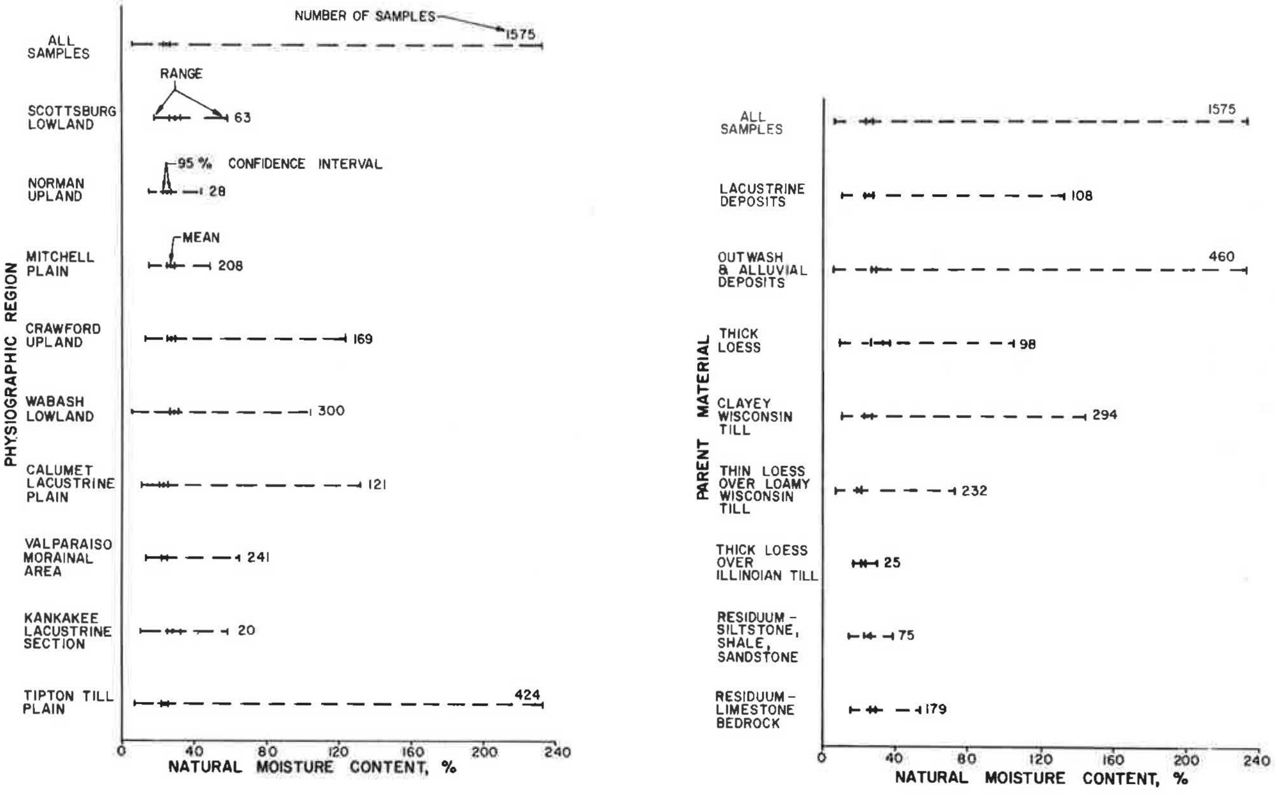


Figure 3. Distributional characterization of plasticity index.

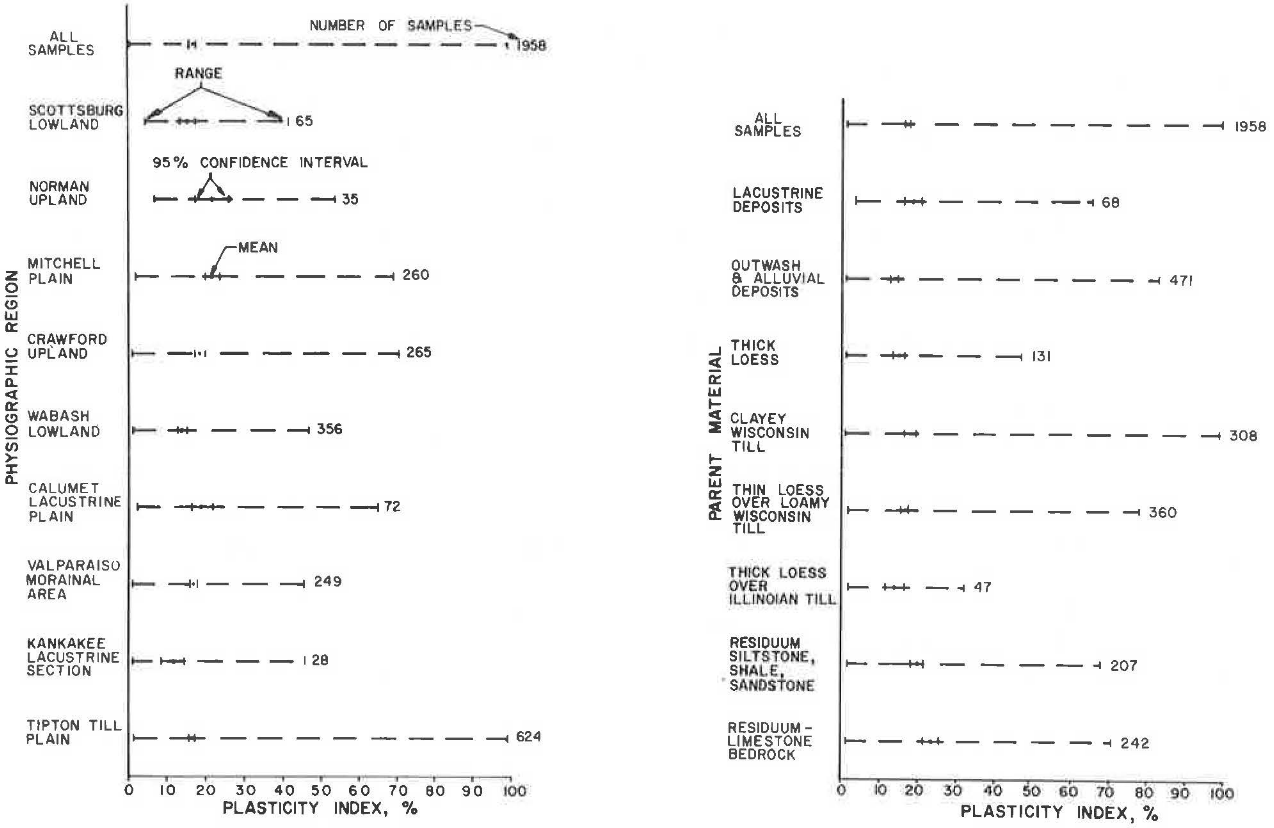


Figure 4. Distributional characterization of percent clay.