important. Such information should prove useful in planning future field studies.

Except for very limited test areas, manual calculation of the usefulness index is impractical. Computer-based techniques are feasible, however. These can compute a single usefulness index for a quadrangle for about $50.00 (including data entry, processing, and display costs, for both salaries and computer time). Most of the data entry and checking can be carried out by technicians. Although, in this study, the computation of the indices was performed by senior staff, this is not necessary because the process can be documented and followed in a routine manner by technicians.

ACKNOWLEDGMENT

The program to establish the mapping standards was undertaken jointly by the Missouri Department of Natural Resources, Rolla, the Resource and Land Information program of the U.S. Department of the Interior, and the Colorado School of Mines, Golden.

REFERENCES

1. D. J. Varnes. The Logic of Geologic Maps, with Reference to Their Interpretation and Use for Engineering Purposes. U.S. Geological Survey, Department of the Interior, Professional Paper 837, 1974, 48 pp.
2. E. E. Lutzen and J. D. Rockaway. Engineering Geology of St. Louis County, Missouri. Missouri Geological Survey, Engineering Geology Series 3, 1971.
3. J. D. Rockaway and E. E. Lutzen. Engineering Geology of the Creve Coeur Quadrangle, St. Louis County, Missouri. Missouri Geological Survey, Engineering Geology Series 2, 1970.
4. J. C. Davis. Statistics and Data Analysis in Geology. Wiley, New York, 1973, 550 pp.
5. J. C. Davis and M. J. McCullagh. Display and Analysis of Spatial Data. Wiley, New York, 1975, 378 pp.
6. A. K. Turner. Computer-Aided Environmental Impact Analysis: Part 1—Procedures. Mineral Industries Bull. (Colorado School of Mines, Golden), March 1976, Vol. 19, No. 2, 23 pp.
7. A. K. Turner. Computer-Aided Environmental Impact Analysis: Part 2—Applications. Minerals Industries Bull. (Colorado School of Mines, Golden), May 1976, Vol. 19, No. 3, 16 pp.

Publication of this paper sponsored by Committee on Engineering Geology.

Computerized Information System for Indiana Soils

G. D. Goldberg, Woodward-Clyde Consultants, Clifton, New Jersey
C. W. Lovell and Robert D. Miles, School of Civil Engineering, Purdue University, West Lafayette, Indiana

A comprehensive information storage system for Indiana soils is being operated on a computer at Purdue University, West Lafayette, Indiana, and at the Division of Materials and Tests of the Indiana State Highway Commission. Information is being collected that includes geotechnical, pedological, and geological data from records of subsurface investigations obtained during the period 1950-1978. Test data from more than 2500 soil samples have been stored and, within the year (1978), it is anticipated that data for an additional 6000 soil samples will be recorded. The data have been evaluated by various statistical methods. Results indicate that the range in values to be expected for a given soil parameter depends on the particular physical property and on the population from which the soil has been sampled. Some soil properties appear to be inherently more variable than others. To illustrate applicability, correlations have been made by using the information relative to physiographic unit and parent material. The grouping of soils by physiographic regions or origin of parent material (or both) suggests that the predictability of some parameters can be improved for certain combinations of parameters and soil groups. Specifically, prediction equations were generated for compression index, compression ratio, and unconfined compressive strength for certain soil populations. It is also possible to predict compaction test values, standard American Association of State Highway and Transportation Officials maximum dry and wet densities, and optimum moisture contents for selected physiographic and parent-material groupings.

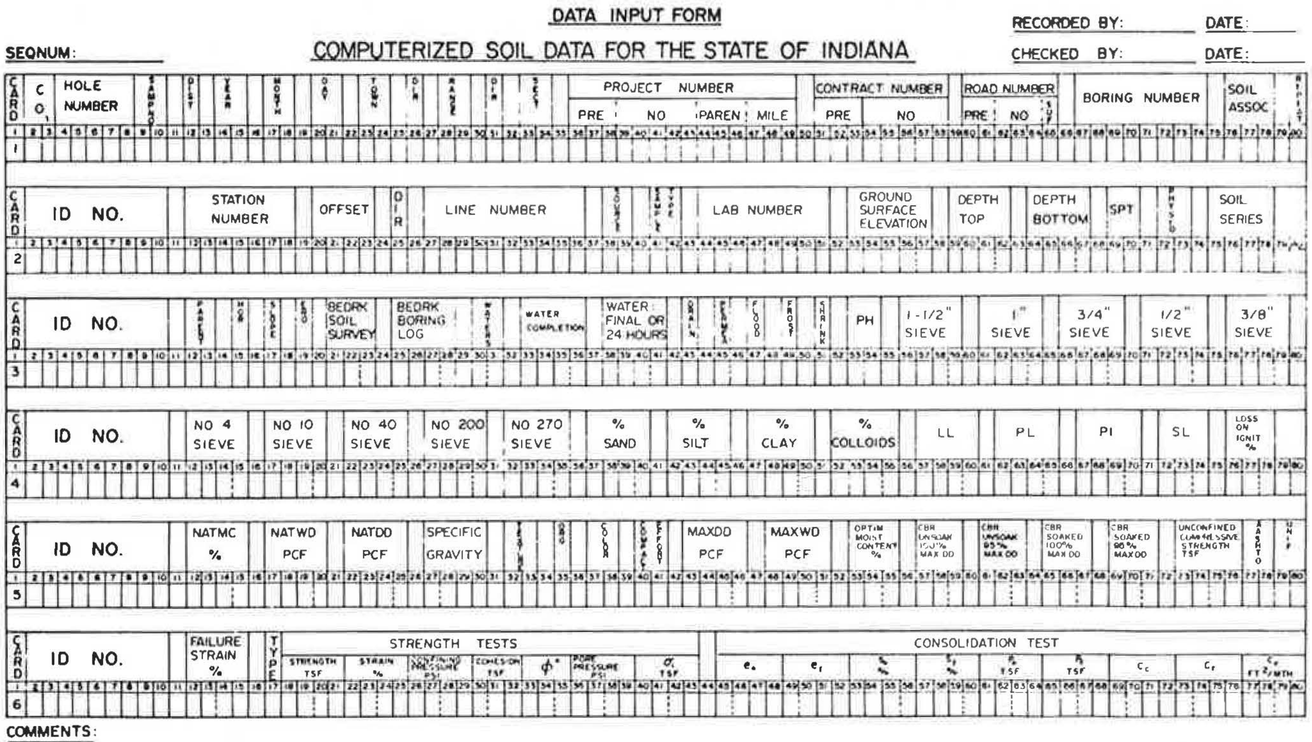
The accumulation of laboratory and field-test data for characterizing the engineering properties of Indiana soils is extensive. An enormous amount of data, collected and stored from highway projects during the period 1950-1978, have been retained in the form of subsurface investigation reports. These reports were prepared by private consulting firms and governmental agencies from routine soil investigations. In their bulky, voluminous form, the majority of these data are not particularly useful for planning and engineering studies.

The need exists to make this information more accessible to both the engineer interested in detailed information about a site and the engineer interested in general soil characteristics over a large area. A computerized geotechnical data bank was judged to be the most efficient, expedient, and economical way to reduce the accumulated data to a form that could readily be made available to interested individuals.

This paper describes the development of a comprehensive information-storage system for soils data. Geological, pedological, and geotechnical engineering information are being collected and stored in a computerized system. Test data from 2508 soil samples have been stored in conjunction with developing and testing the computer system and, in addition, approximately 5500 other data sets have been stored (for a total of more than 8000 soil test samples).

Various statistical methods have been applied to some of the data. Results indicate that the range in values to be expected for a given soil parameter depends on the particular physical property and on the population from which the soil was sampled. The grouping of

Figure 1. Data input form (DIF).



soils by physiographic regions or the origin of their parent material (or both) suggests that the predictability of certain parameters can be improved for such data populations.

DATA COLLECTION

The information gathered during a roadway soil investigation is generally limited. Usually, only simple laboratory testing for classification purposes is performed on a few soil samples; more-specialized testing is reserved for samples taken from sites where structures are to be constructed or where nontypical or unstable soil conditions are encountered. Information gathered is used in selecting the proper locations for the facility and in making design decisions (1, 2).

The information typically available from roadway soil investigation reports includes

1. Project and sample identification,
2. Sample location,
3. Ground surface elevation,
4. Depth from which the sample has been removed,
5. Depths to groundwater and bedrock,
6. Standard penetration resistance,
7. In situ moisture content,
8. Dry density,
9. Visual textural classification,
10. Gradation characteristics, and
11. Atterberg limits.

In addition, the results from special tests are included for selected soils. These data include the results from

1. Compaction tests,
2. California bearing ratio (CBR) tests,
3. Unconfined compression testing,
4. Triaxial and direct shear tests, and
5. Consolidation tests.

The geotechnical data currently being transferred to the data input form (DIF) are shown in Figure 1 [all figures in this paper are taken from Goldberg (3)]. This form serves as a guideline for card punching and subsequent transferral to magnetic tape for computer storage. A computer program has been developed that uses the information to classify each sample according to the American Association of State Highway and Transportation Officials (AASHTO) and the Unified Soil Classification systems.

The position of each sample hole is located on an agricultural soil-survey map (4). The pedological soil association, the soil series, and the soil horizon represented by each sample are recorded on the DIF. Because such descriptive data (5) cannot be directly recorded, this information has been codified to make the system compatible with computerized storage and retrieval. [Details of the coding are given by Goldberg (3).] Additional pedological information required includes the slope (topographic) class of the soil series, the erosion phase, and the natural soil-drainage class, permeability, flooding potential, frost-heave susceptibility, shrink-swell potential, and pH. These data are determined from published soil-series data sheets and coded onto the DIF. The physiographic unit (see Figure 2) and the parent material (see Figure 3) from which the soil has been derived are also entered on the DIF.

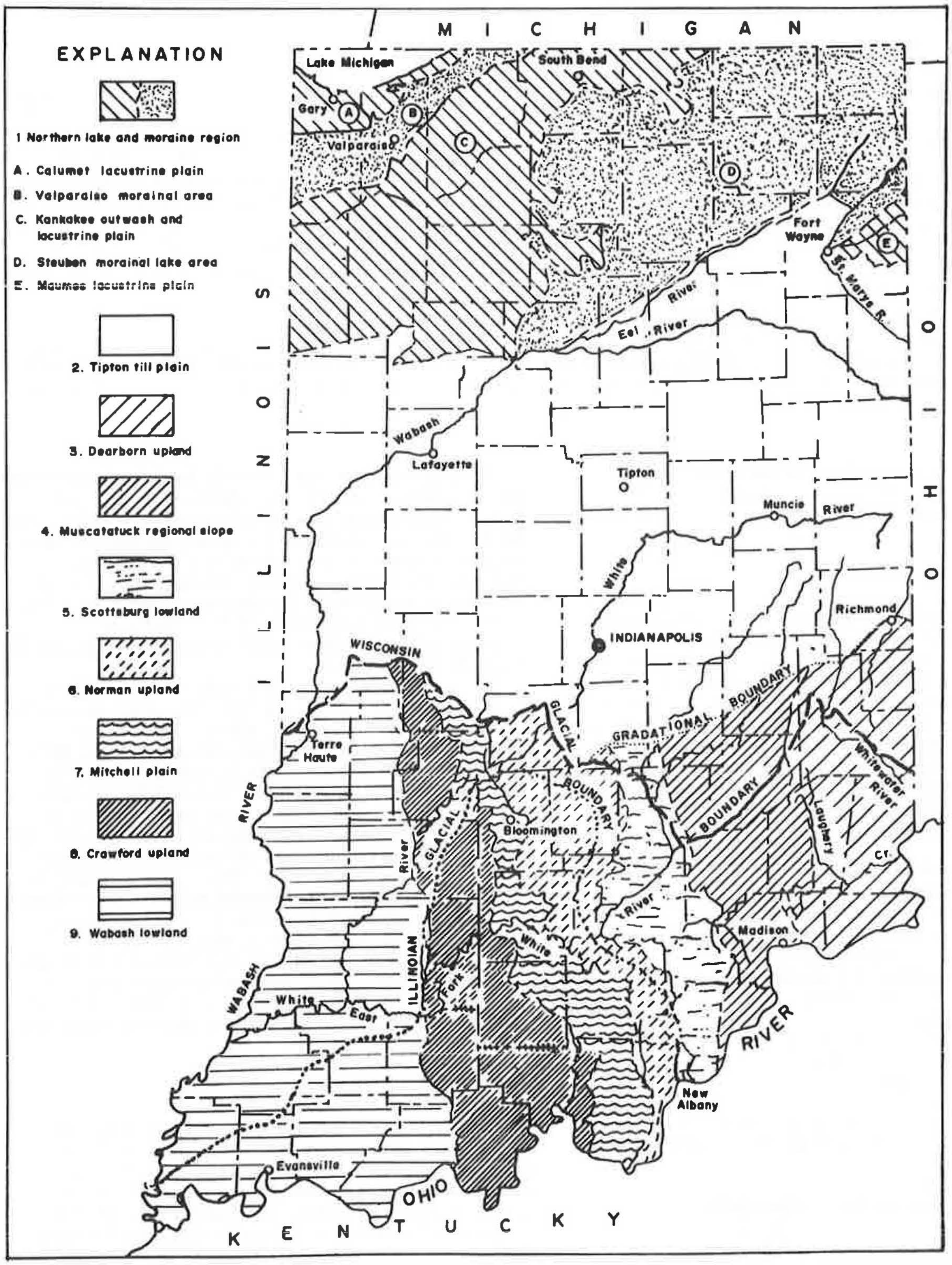
DATA MANAGEMENT

The data are punched onto a series of six cards for each sample. The order of the data cards is fixed, and a number, from one to six, is sequentially assigned to each card for identification. Each sample is also assigned a number, as is each hole. These numbers are assigned sequentially within each county. Each county name is coded. The card number, the county code, the hole number, and the sample number have the purpose of assigning a unique identification number that can be used for internal bookkeeping to each data card.

Because large amounts of data are collected, errors in recording and punching the data are inevitable. Thus, an audit program has been written to identify those errors that can be detected by using the computer and, thereby, allow mistakes to be corrected. For example, if the liquid limit is mistakenly recorded, the liquid limit minus the plastic limit will not equal the plasticity index and the computer will automatically report an error.

A complete instructional User's Manual explaining the operation of the computerized storage and retrieval

Figure 2. Physiographic units based on present topography.



system is described by Goldberg (3). Included in the User's Manual are descriptions of the data items, codification system, formats, card locations, and column locations for each data item, as well as the listing of the programs used to add additional data to the data bank; to check data-input errors, where possible; to use the computer programs for data management and manipulations; and example problems on the use of the data bank.

STATISTICAL ANALYSIS

The collection of large amounts of soil test data and the fact that most natural soil deposits are highly variable in both horizontal and vertical directions require the use of a statistical approach (6-8). By using a computerized geotechnical data bank, an extensive listing of available soil and rock information can be retrieved both quickly and economically. For example, typical ranges of values for different soil parameters in the data system and the distribution of these parameters can be determined as shown in Figures 4-6. Such data are helpful in the selection of suitable sample sites for detailed laboratory testing (9). In addition, the development of correlations among selected soil properties can be helpful to the engineer in reducing the need for extensive laboratory testing (10, 11). This is particularly important to the small engineering unit that needs reliable data but can afford only a small amount of testing.

Regression Analysis

Prediction models usually involve soil parameters that are difficult to determine as dependent variables and more-easily-determined characteristics as independent variables. In the Indiana study (3), the dependent variables of major interest were

1. Compression index (C_c) and compression ratio

Figure 3. Soil regions of Indiana.

Soil Regions, Their Parent Materials and Representative Soil Series

1. Sandy and loamy lacustrine deposits and eolian sand (Maumee, Rensselaer, Plainfield)
2. Silty and clayey lacustrine deposits (McGary, Patton, Hoytville, Dubois)
3. Alluvial and outwash deposits (Fox, Genessee, Warsaw, Wheeling)
4. Eolian sand deposits (Plainfield, Oshtemo, Bloomfield)
5. Thick loess deposits (Alford, Hosmer, Iva)
6. Loamy glacial till (Riddles, Miami, Crosier, Brookston)
7. Clayey glacial till (Blount, Pewamo, Morley)
8. Thin loess over loamy glacial till (Brookston, Crosby, Miami, Parr)
9. Moderately thick loess over loamy glacial till (Fincastle, Russell, Miami, Brookston)
10. Moderately thick loess over weathered loamy glacial till (Cincinnati, Avonburg, Vigo, Ava)
11. Discontinuous loess over weathered sandstone and shale (Zanesville, Berks, Wellston, Muskingrum)
12. Discontinuous loess over weathered limestone (Crider, Frederick, Corydon)
13. Discontinuous loess over weathered limestone and shale (Eden, Switzerland, Pate)

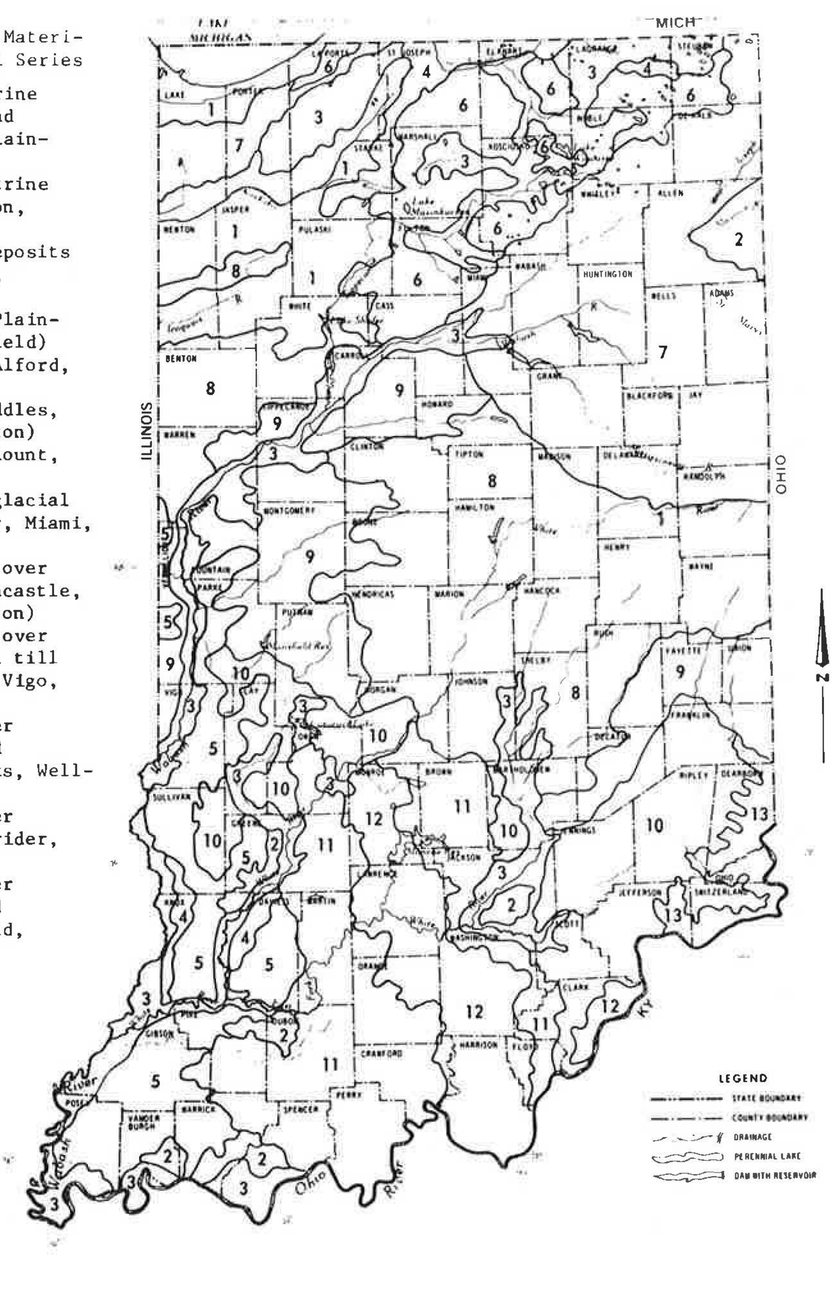


Figure 4. Distributional characterization of percentage passing 2-mm sieve.

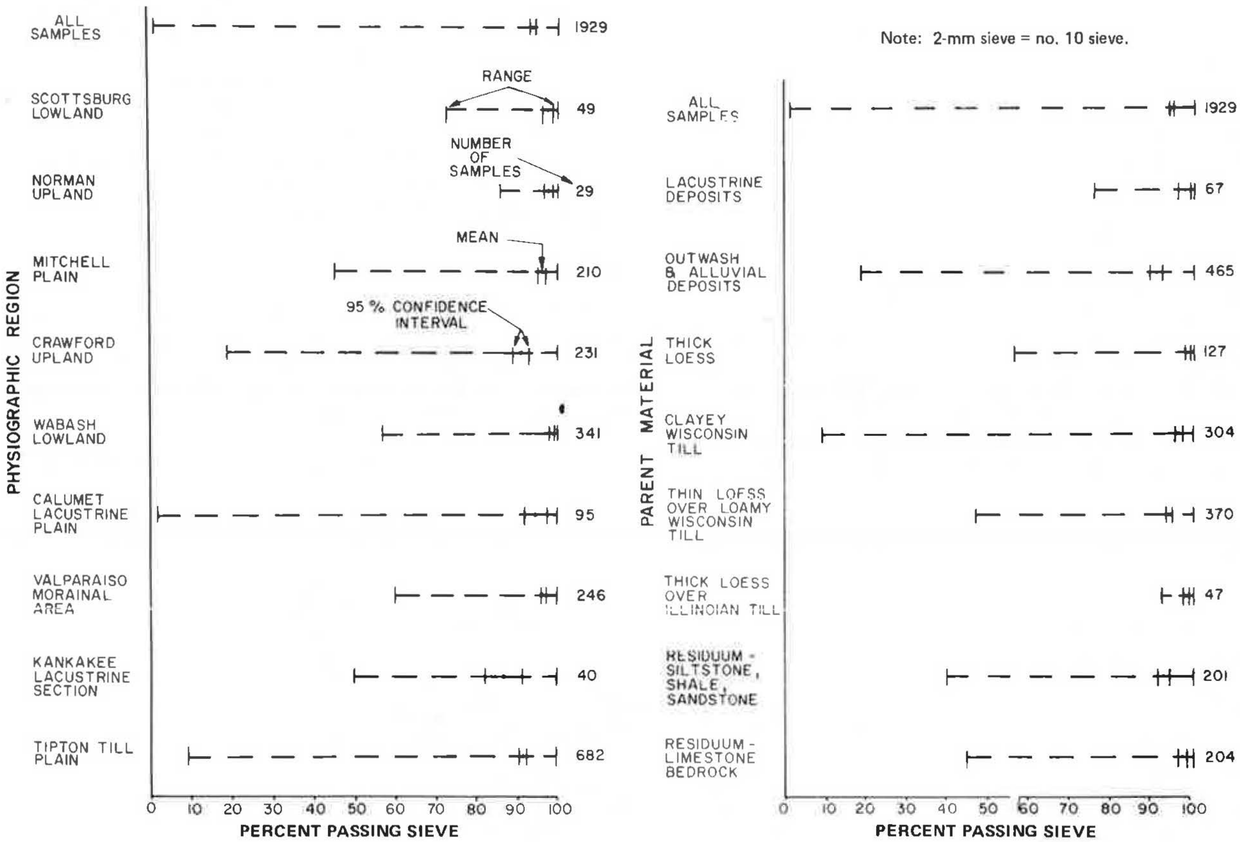


Figure 5. Distributional characterization of liquid limit.

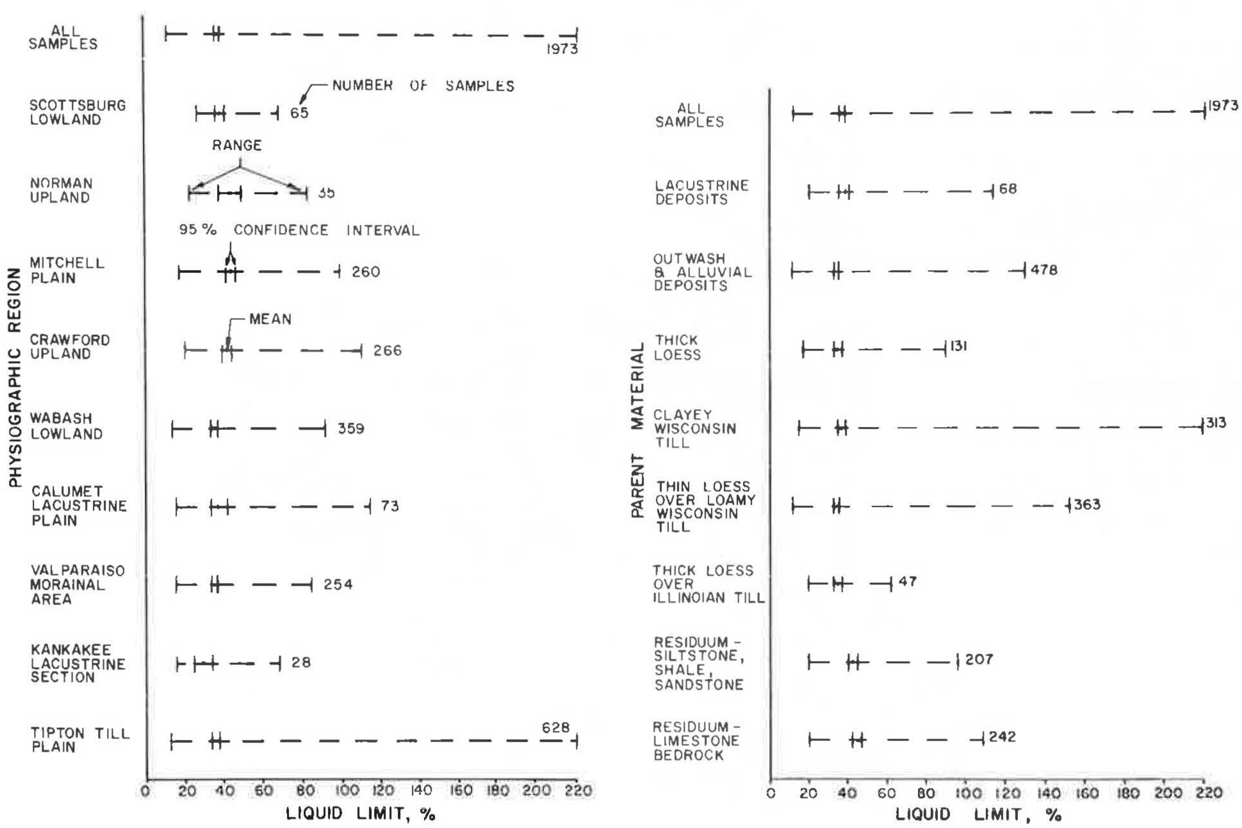
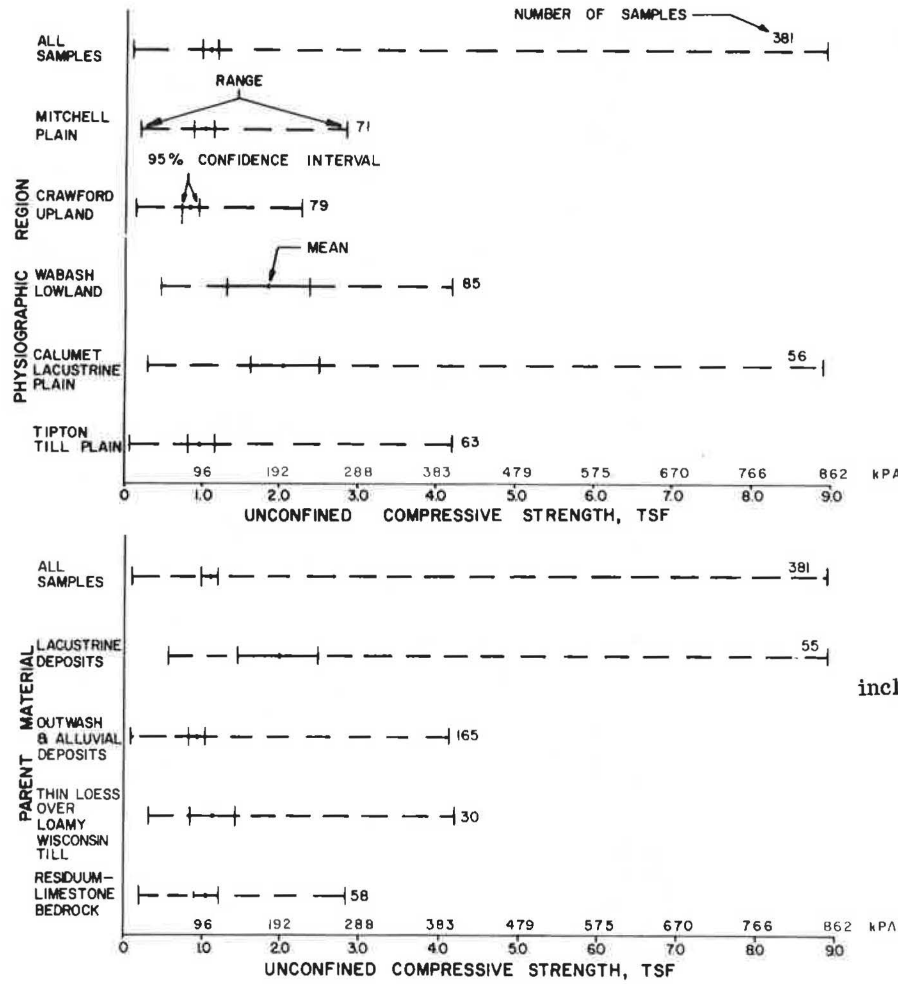


Figure 6. Distributional characterization of unconfined compressive strength.



include

(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 AASHTO maximum dry ($\gamma_{d_{max}}$) and wet ($\gamma_{m_{max}}$) densities and optimum moisture content (w_{opt}); and
4. Soaked CBRs at 100 and 95 percent of $\gamma_{d_{max}}$.

The set of easier-to-measure independent variables includes

1. e_o, natural moisture content (w_n), natural dry density (γ_d), liquid limit (w_L), plastic limit (w_p), plasticity index (I_p), and percent clay for the consolidation test data;
2. w_L, w_p, w_n, γ_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 statewide regression modeling, or if the data were available in amounts large enough to justify modeling on smaller units (that is, physiographic regions, parent-material areas and, in some cases, soil types), the data were grouped accordingly to determine whether the prediction models could be significantly improved.

The soil test data have been collected throughout the entire state as shown in Figure 7, but no attempt was made to collect equal numbers of samples from all parts of the state. Therefore, the correlation and prediction equations presented are applicable for only those areas from which data have been collected.

Each dependent variable was first plotted against each independent variable to investigate the nature of the dependence. These plots indicate whether linear terms, quadratic terms, or transformations of the variables are appropriate. In addition, the coefficient of determination (r^2) was examined to determine the degree of relationship between the dependent and independent variables in each case. Those sets of independent and dependent variables that appeared to exhibit some dependence were then examined by multiple regression analysis methods.

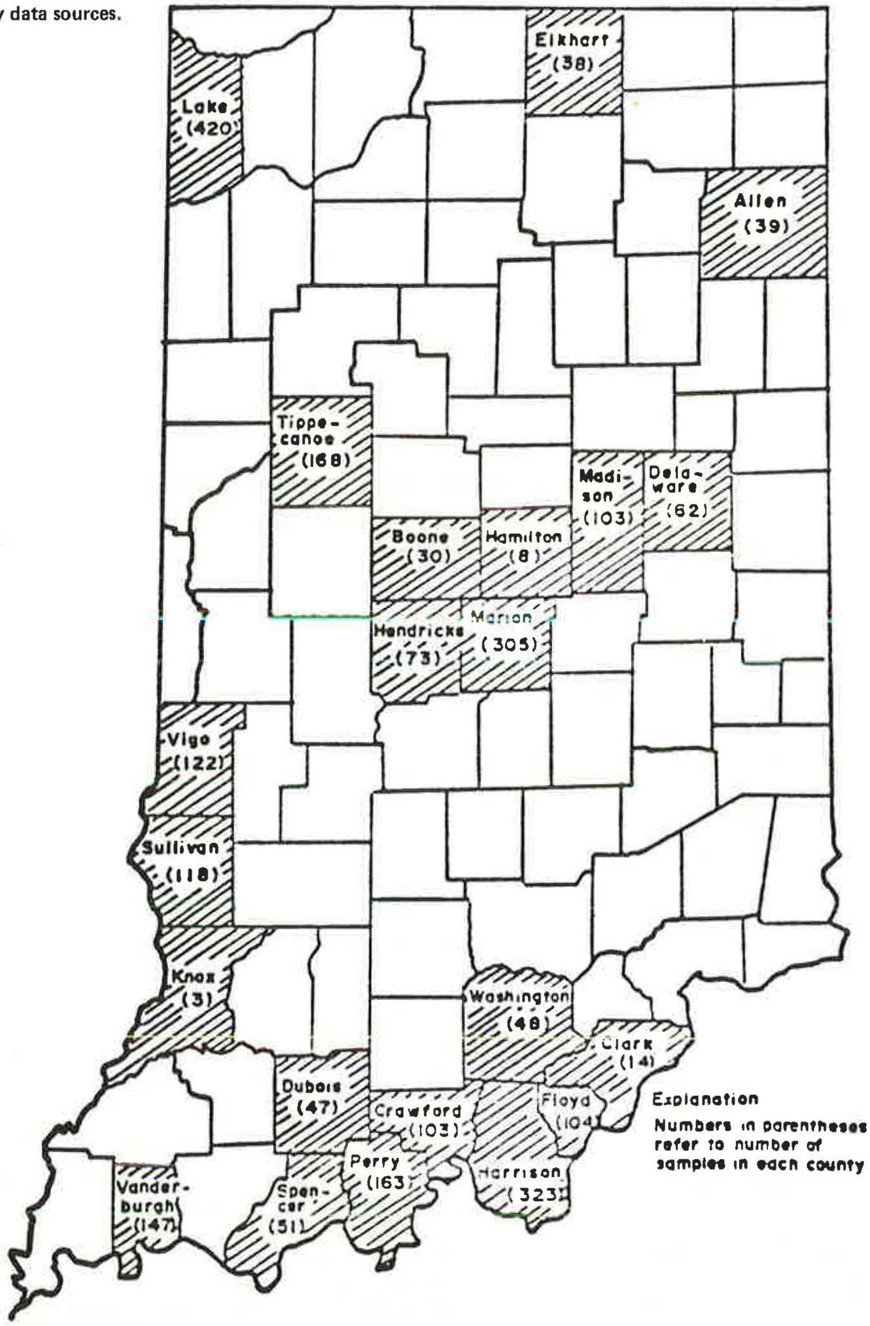
The potential terms in each model were linear independent variables, squared independent variables, reciprocal transformations of independent variables, and linear interactions of independent variables. Logarithmic transformations of dependent and independent variables were attempted when it appeared that the model was intrinsically linear by suitable transformation.

The regression equations that had adjusted coefficients of multiple determination (R_a^2) (12) greater than 0.65 were examined to determine whether the relationships were statistically useful. The criteria included (a) small confidence intervals (at the 95 percent confidence level) and (b) confidence intervals that did not cross zero.

Evaluation of Models

After obtaining good prediction models, it is important to evaluate them to determine whether they are ap-

Figure 7. County data sources.



propriate for the particular data being examined. Certain assumptions are inherent in the formulation of regression models; an examination of the residuals (differences between the observed and corresponding predicted values) will suggest whether any of the usual assumptions are invalid. The usual assumptions are that the errors are independent, have zero mean, a constant variance, and follow a normal distribution.

The residuals of each model were plotted against each independent variable in it, in addition to the dependent variable and its predicted value. The residuals were tested for normality by the Purdue University computer program called NORP. The models that had residual plots that did not display systematic tendencies to be positive or negative, but tended to fall within horizontal bands centered around 0, and also satisfied the normality criterion at the 90 percent confidence level were selected as the final models (see Tables 1-3) (3).

SUMMARY

The computerized data bank described in this paper should facilitate efficient and economical handling of geotechnical information in Indiana. Soils information that was essentially lost after a project was completed can now be used for future highway projects and improvements. The data bank will be maintained by the Indiana State Highway Commission for all potential users.

The application of statistical methods to the geotechnical data stored to January 1978 is promising. However, no soil group studied can be said to produce better correlation equations than any other, overall. The grouping of soils by physiographic regions and parent-material areas certainly appears to be justified for

Table 1. Regression equations for prediction of compression index and compression ratio.

Unit	Dependent Variable	R^2_a	Regression Equation	N
All	C_c	0.800	$C_c = 0.5363(e_o - 0.4110)$	96
samples		0.792	$C_c = 0.0002(w_n^2 - 106.2727)$	
		0.783	$C_c = 0.0129(w_n + 0.1015w_L - 16.1875)$	
	C_r	0.691	$C_r = 0.2037(e_o - 0.2465)$	
Wabash	C_c	0.838	$C_c = 0.5673(e_o - 0.4422)$	29
lowland	$\log C_c$	0.831	$\log C_c = 2.7904(e_o - 0.3346e_o^2 - 0.8449)$	
	C_r	0.750	$C_r = 0.221(e_o - 0.3074)$	
		0.748	$C_r = 0.0065(w_n - 11.6361)$	
		0.735	$C_r = 0.0034[(e_o \times w_n) + 8.3647]$	
Crawford	C_c	0.859	$C_c = 0.0101[(e_o \times w_L) - 0.5765w_L + 12.665]$	28
upland		0.833	$C_c = 0.0114(w_n + 0.2491w_L - 18.8134)$	
		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	$C_r = 0.0033[(e_o \times w_n) + 12.5168]$	
		0.721	$C_r = 0.1164(e_o^2 + 0.3594)$	
Outwash	C_c	0.842	$C_c = 0.5621(e_o - 0.4215)$	63
and alluvial		0.822	$C_c = 0.0153(w_n + 0.1022w_L - 0.3104w_p - 11.6123)$	
deposits	$\log C_c$	0.772	$\log C_c = 2.1389(e_o - 0.2967e_o^2 - 0.9374)$	

Table 2. Regression equations for prediction of unconfined compressive strength.

Unit	Dependent Variable	R^2_a	Regression Equation	N
Calumet	q_u	0.756	$q_u = 0.0003644(\gamma_d^2 - 2518883.9)$	40
lacustrine plain	$\log q_u$	0.750	$\log q_u = 0.3804 \times 10^{-6}(\gamma_d^2 + 2.401 \times 10^6)$	
Lacustrine deposits	$\log q_u$	0.699	$\log q_u = 0.3804 \times 10^{-6}(\gamma_d^2 + 2.570 \times 10^6)$	48

Note: These coefficients were derived for q_u expressed in kilonewtons per square meter and γ_d expressed in kilograms per cubic meter.

Table 3. Regression equations for prediction of standard Proctor maximum dry and wet densities and optimum moisture content.

Unit	Dependent Variable	R^2_a	Regression Equation	N
All samples	w_{opt}	0.894	$w_{opt} = -0.03062(\gamma_{d_{max}} -2340.3644)$	138
Valparaiso	$\log \gamma_{d_{max}}$	0.816	$\log \gamma_{d_{max}} = -3.683[(1/w_L) + 0.127\log w_L - 1.109]$	
morainal		0.785	$\log \gamma_{d_{max}} = 0.2239(\log w_L - 16.097)$	
area	$\gamma_{m_{max}}$	0.790	$\gamma_{m_{max}} = -1849.7498[\log w_L + 9.9623(1/w_L) - 2.9758]$	26
	$\log \gamma_{m_{max}}$	0.694	$\log \gamma_{m_{max}} = -0.1348(\log w_L - 26.2080)$	
	w_{opt}	0.972	$w_{opt} = 0.04482(\gamma_{m_{max}} -1.2985\gamma_{d_{max}} + 604.899)$	
		0.870	$w_{opt} = -0.0260(\gamma_{d_{max}} -2432.7188)$	
		0.810	$w_{opt} = 23.0357 + 0.002(w_L \times w_p) - 285.9386(1/w_L)$	
Residuum	$\gamma_{d_{max}}$	0.772	$\gamma_{d_{max}} = -1841.0591[\log w_L + 14.0953(1/w_L) - 2.9063]$	22
of limestone bedrock	$\log w_{opt}$	0.781	$\log w_{opt} = 0.0042(w_L + 259.0381)$	

Note: These coefficients were derived for γ expressed in kilograms per cubic meter.

some dependent variables and for certain groups of soils.

From the statistical analysis to date, the following preliminary conclusions are drawn.

1. The prediction of C_c values by using simpler soil measures is reasonable on a statewide basis. Soils investigated in the Wabash lowland and the Crawford upland physiographic regions also produce regression equations for C_c that have relatively high correlation coefficients. Furthermore, equations generated for soils derived from outwash and alluvial deposits also are statistically significant for the prediction of C_c.
2. The prediction of q_u by the method of regression analysis is possible for soils found in the Calumet lacustrine plain and soils derived from lacustrine deposits.
3. The prediction of $\gamma_{d_{max}}$, $\gamma_{m_{max}}$, and w_{opt} is possible from simpler-to-determine independent variables for the soils from the Valparaiso morainal area. Soils derived from residuum of limestone bedrock also produced satisfactory regression equations.

As the size of the data base is increased, more complete analyses will be undertaken with the expectation that more meaningful and valid correlations can be achieved. Other data groupings, including the pedologic hierarchy, will also be examined.

ACKNOWLEDGMENT

We are grateful to the Joint Highway Research Project of Purdue University, the Indiana State Highway Commission, and the Federal Highway Administration for their financial support of the research reported herein. Special thanks are due Robert Ringholz for his assistance in the collection of the project data.

The contents of this paper reflect our views; we are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official veiws or policies of the Federal Highway Administration or the Indiana State Highway Commission. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. D. J. Hagerty, N. G. Schmitt, and W. J. Pfalzer. Retrieval and Use of Geotechnical Information. In Geotechnics in Transportation Engineering, Ohio River Valley Soil Seminar, Lexington, KY, Oct. 1973 (ASCE, Kentucky section), pp. 1-9.

2. Acquisition and Use of Geotechnical Information. NCHRP, Synthesis of Highway Practice 33, 1976, 40 pp.
3. G. D. Goldberg. Development of the Computerized Geotechnical Data Bank for the State of Indiana. Purdue Univ., West Lafayette, IN, M.S.C.E. thesis; Joint Highway Research Project, Rept. 78-6, May 1978, 163 pp.
4. Soil Survey Manuals. Soil Conservation Service, U.S. Department of Agriculture, 1958-present.
5. D. Spradling. A Soils Data System for Kentucky. Bureau of Highways, Kentucky Department of Transportation, Frankfort, Res. Rept. 441, 1976, 58 pp.
6. T. K. Liu and T. H. Thornburn. Engineering Index Properties of Some Surficial Soils in Illinois. College of Engineering, Univ. of Illinois at Urbana, Bull. 477, 1965.
7. P. Lumb. The Variability of Natural Soils. Canadian Geotechnical Journal, Vol. 3, No. 2, May 1966, pp. 74-97.
8. P. Lumb. Application of Statistics in Soil Mechanics. In Soil Mechanics—New Horizons (I. K. Lee, ed.), Elsevier, New York, 1974, pp. 44-111.
9. D. J. A. van Zyl. Storage, Retrieval, and Statistical Analysis of Indiana Shale Data. Joint Highway Research Project, Purdue Univ., West Lafayette, IN, 140 pp.
10. V. Rizkallah and E. El Nimr. Applicability of Regression Analysis in Soil Mechanics with the Help of Data Banks. Proc., 2nd International Conference on the Application of Statistics and Probability in Soil and Structural Engineering, Aachen, Federal Republic of Germany, Vol. 2, 1975, pp. 423-438.
11. M. Van Rooyen. Intercorrelation of Engineering Properties of Soil. Proc., 4th Regional Conference for Africa on Soil Mechanics and Foundation Engineering, Cape Town, South Africa, 1967, pp. 203-208.
12. J. Neter and W. Wasserman. Applied Linear Statistical Models. Irwin, Homewood, IL, 1974, 842 pp.

Publication of this paper sponsored by Committee on Engineering Geology.

Design of Subsurface Drainage Systems for Control of Groundwater

Lyle K. Moulton, Department of Civil Engineering, West Virginia University, Morgantown

In recent years, awareness has grown of the need for subsurface drainage systems that can drain water from the pavement structural system. Much of the emphasis associated with studies of this subject has been on the removal of the moisture that infiltrates through the surface of the pavement, but it has also been recognized that the control of groundwater is an essential part of any effective highway subsurface drainage system. In this paper, rational analytical methods for the design of subsurface drainage systems for the control of groundwater are developed and presented. Although these methods are, in general, approximate in nature, they are soundly based on fundamental seepage theory. The resulting solutions have been used to prepare graphical design aids that can be readily applied by the highway designer. The use of these design aids is illustrated by a series of examples, and the results are compared with more-exact flow-net solutions obtained by the use of electric analogs. On the basis of this comparison, it concluded that the proposed design procedures, although approximate, do permit the development of good practical designs for subsurface drainage systems for the removal and/or control of groundwater in highway applications.

In recent years, there has been a growing awareness of the need for subsurface drainage systems that can drain water from a pavement structural system and thus minimize detrimental effects. Workshops dealing with water in pavements (1) have been conducted, and guidelines for the design of subsurface drainage systems for pavement structural sections have been published (2, 3). Although much of the emphasis of these activities has been on the removal of the moisture that infiltrates through the surface of the pavement, it has also been recognized (3) that the control of groundwater is an essential part of any effective highway subsurface drainage system.

Commonly, the design of groundwater drainage systems is based on empirical rules of thumb that have been developed by trial and error over a period of years or on rather tedious graphical techniques involving the use of flow nets (4). The purpose of this paper is to present some rational, approximate analytical methods for the design of groundwater control systems such as the interceptor drains shown in Figures 1 and 2 and the symmetrical drawdown drains shown in Figure 3. Although, at present, it is not possible to eliminate all elements of empiricism, the methods presented are based on fundamental seepage theory.

LONGITUDINAL INTERCEPTOR DRAINS

Calculation Method

Let us consider the case of the unconfined flow of groundwater over a sloping impervious boundary toward a single interceptor drain, as illustrated in Figure 4. A solution for the shape of the drawdown curve for this situation, which was developed by R. E. Glover of the U.S. Bureau of Reclamation, is given by Donnan (5). This solution, which is based on an adaptation (6) of Dupuit theory, has the form

$$x = \{H \ell n[(H - H_0)/(H - y)] - (y - H_0)\}/S \qquad (1)$$

where

x and y = coordinates of a point on the drawdown