Storage, Retrieval, and Analysis of Compacted Shale Data

Dirk J. A. van Zyl, L.E. Wood, and C.W. Lovell, Department of Civil Engineering, Purdue University, West Lafayette, Indiana
W.J. Sisiliano, Indiana State Highway Commission

Large amounts of soil test data are being generated by states testing materials for new highway facilities. Of particular interest to some states is the occurrence of shale as a highway material. This paper describes a storage and retrieval system for compacted Indiana shale data (163 sets). Statistical analyses included frequency analysis, bivariate correlation analysis, and multiple regression analysis. Histograms of the frequency distributions and a summary table of all the different statistical descriptors of the parameters are given. Bivariate correlation analysis (one-to-one correlations) was done for all data lumped and for data grouped by geologic unit. The multiple regressions concentrated on five-parameter models for predicting California bearing ratio values. It was also found by multiple regression that the various values of slake durability indices can be estimated by using second-order equations. The analyses are of potential value to practicing engineers as a first approximation in design and may prove useful to engineers researching compacted shale behavior.

The usefulness of the large amounts of soil test data being generated by testing of materials for new highway facilities was not realized until recently. A number of states are now in the process of introducing storage and retrieval systems (1, 2). The type of system to be used depends on the amount of data available, but any system should be flexible enough to allow changes to be made later with the least amount of effort. The system should therefore be easily administered, economical, dynamic, and adaptable to the changing needs of its users. These qualities can be best incorporated in a computerized system.

The statistical analysis of stored data can lead to a number of very useful results such as proving or disproving what the typical properties and behavior of certain types of material indicated by experience are. Statistical analyses will also indicate the typical range of values to be expected for different parameters of the different members in the system as well as the distribution of these parameters. This can play a very important role in assessing the suitability of a certain sample before elaborate laboratory testing is undertaken. Statistical analyses should be updated regularly as more data become available. Results can be stated with greater confidence as the available data bank grows. An additional benefit is the development of new ideas for more productive research based on empirical relationships and correlation of soil properties (2).

Correlations among different parameters can be developed by using stored data. Regression models can be established to predict the parameters that are difficult to measure by using a set of simple parameters. These models should be reevaluated as new data become available.

This paper describes a simple storage and retrieval system for data available on compacted Indiana shales. Some results of a comprehensive statistical analysis of the data are also presented.

DATA

The data for this study are from testing done by the Indiana State Highway Commission (ISHC) on shales and from three Joint Highway Research Project (JHRP) reports completed at Purdue University (3, 4, 5). A complete summary of all the data available and descriptions of the geology of Indiana shales and the physiography of Indiana are given by van Zyl (6). The table below gives a summary of the data sets according to geological description.

Geological System	Geological Stage	No. of Data Sets
Pennsylvanian	Shelburn formation	3
	Dugger formation	2
	Petersburg formation	2
	Brazil formation	1
	Mansfield formation	23
		31
Mississippian	Core limestone	2
	Palestine sandstone	17
	Watersburg sandstone	5
	Tar Springs formation	15
	Glen Dean limestone	3
	Hardinsburg formation	20
	Haney limestone	3
	Big Clifty formation	3
	Elwren formation	1
	Sample formation	4
	Bethel formation	3
	Borden group	16
	Locust Point formation	4
	New Providence shale	4
	New Albany shale	5
		105
Devonian	Antrim shale	1
	New Albany shale	6
		7
Ordovician	Whitewater formation	2
	Dillsboro formation	6
	Kope formation	12
		20

Each data set consists of a number of parameters obtained during laboratory testing. The parameters and descriptors included in each data set are given in Table 1. Not all parameters are always determined during laboratory testing, which means that not all the data sets are complete and thus that highly reliable conclusions cannot always be drawn. Much of the laboratory testing is done according to AASHTO standard procedures. However, the following parameters are determined by tentative test methods of the ISHC: loss on ignition, slake durability index, slaking index, fissility number, and modified soundness. These test methods are described by van Zyl (6).

STORAGE AND RETRIEVAL OF SHALE DATA

Certain descriptors should be used to identify a shale sample in the storage and retrieval system. The choice is between numerical values and codes or numerical values and words. As the system described here has only 163 data sets, it was decided to use the latter. From Table 1 it is clear that the descriptors used to locate a sample are laboratory number, geographic location (counties), geologic unit (system, series, stage), and physiographic unit (bedrock). The numerical values are the laboratory test results.

The data sets were coded on the four-card series shown in Table 1. The three basic reasons for retrieving data are to get printouts of all the data available, to get printouts based on some descriptor in some order, e.g., geological stage in alphabetical order so that the printed data sets belonging to a certain geological stage can be separated for study, and to use certain sets of data for statistical manipulation.

A program was written for the CDC 6500 system of the Purdue University Computer Center to achieve the first two. The sorting of the data, accomplished by using the sort/merge package on the CDC system, can be done in an ascending or descending order based on any parameter.

Table 1. Coding of four-card series.

Parameter	Symbols (for program and tables and figures)	Format	Columns
Card 1			
Laboratory number	NO	I10	1-10
County	CO	A10	11-20
Geological system	GSY1, GSY2	A10A5	21-35
Geological series	GSE1, GSE2	A10A5	36-50
Geological stage	GST1, GST2	A10A5	51-65
Physiographic unit	PHU1, PHU2	A10A5	66-80
Card 2			
Laboratory number	NO	I10	1-10
Slaking index cycle 1	SDI1	F5.1	11-15
Slaking index cycle 5	SDI2	F5.1	16-20
Slake durability index, 200 revolutions dry	SD2D	F5.1	21-25
Slake durability index, 500 revolutions dry	SD5D	F5.1	26-30
Slake durability index, 200 revolutions soaked	SD2S	F5.1	31-35
Slake durability index, 500 revolutions soaked	SD5S	F5.1	36-40
Fissility number	FN	F5.1	41-45
Modified soundness	MS	F5.1	46-50
Card 3			
Laboratory number	NO	I10	1-10
Textural classification	TEX1, TEX2	A10A5	11-25
AASHTO classification	ASH	A10	26-35
Plastic limit	PL	F5.1	36-40
Liquid limit	QL	F5.1	41-45
Plasticity index	PI	F5.1	46-50
Percent sand	PSA	I5	51-55
Percent silt	PSI	I5	56-60
Percent clay	PCL	I5	61-65
Percent colloids	PCOL	I5	66-70
Card 4			
Laboratory number	NO	I10	1-10
Natural wet density	WD	F5.1	11-15
Natural dry density	DD	F5.1	16-20
Natural moisture content	AMC	F5.1	21-25
Specific gravity	SG	F5.1	26-30
pH	PH	F5.1	31-35
Shrinkage limit	SL	F5.1	36-40
Lineal shrinkage	SH	F5.1	41-45
Loss-on-Ignition	XY	F5.1	46-50
Moisture density[a]	MD	I1	51
Optimum moisture	PM	F5.1	52-55
Maximum wet density	WD1	F5.1	56-60
Maximum dry density	DD1	F5.1	61-65
As Compacted CBR	CBR1	F5.1	66-70
After Soaking CBR	CBR2	F5.1	71-75
Average % Swell	AS	F5.1	76-80

[a]Coding for moisture density was 1 for test done on minus No. 4 material, 2 for test done on minus 19-mm (¾-in) material, 3 for when there is a special note on the laboratory report that should be referred to, and 4 for nothing indicated on test report.

STATISTICAL ANALYSIS

The statistical analysis was done by using the routines available in the statistical package for the social sciences (SPSS) (7).

Frequency Analysis

The first step was to do a complete frequency analysis of all the data available. For this the CODEBOOK routine of SPSS was used. The data can be analyzed as single points or as grouped data and were first analyzed as single points in order to obtain minima and maxima of each parameter. The values of each parameter were then grouped into 10 intervals in order to obtain histograms.

The results of the statistics of the different parameters are given in Table 2, where the first line of results for every parameter gives the values obtained by using single points and the second line gives the results for grouped data. The definitions of standard error, kurtosis, skewness, and so on are given in Nie and others (7). The histograms are given in Figure 1. (Note that all density calculations were done in the U.S. customary units of pounds per cubic foot.)

The last two columns of Table 2 give the number of valid values and also the missing values; e.g., for SDI1, 131 data sets had values reported for this parameter but 32 data sets did not. The number of valid values plays a very important role in the regression analysis.

The low values obtained for the skewness for liquid limit (QL) and percent silt (PSI) indicate that the distributions of these two parameters are almost symmetrically bell shaped (see Figure 1). Low values of coefficient of variance were also obtained for these two variables; this confirms the small relative spread of values. Low values of the coefficient of variation are to be expected for specific gravity (SG) and pH, while the low values for the coefficient of variation obtained for the different densities reflect the high average values and small spread of these parameters.

It is interesting to note the relatively small values of coefficient of variation for the Atterberg limits, which can therefore be estimated with reasonable confidence if a few test results are available. Assuming that the plastic limit (PL) and QL are normally distributed (which is reasonable, as shown in Figure 1), it can be said that 68 percent of all values will be within one standard deviation of the mean, i.e., the ranges of values are 19.9-24.1 for the PL and 25.4-41.4 for the QL.

It is clear from Figure 1 that some distinction based on the histograms can be made between low- and high-durability shales. The dispersion of the data however is significant. It is possible that further analysis based on separate geologic groupings will distinguish between strong and weak or durable and nondurable shales.

Bivariate Correlation Analysis

In order to determine whether any simple correlations can be established between the different parameters, bivariate correlation analyses were performed using the SCATTERGRAM routine of the SPSS package.

The above analysis was performed for each combination of variables, and plots were obtained. It was decided to include complete data on all the analyses that gave a good correlation, or an R value ≥ 0.7. These results are given in Table 3. Three potential uses of the plots are apparent.

Table 2. Statistics obtained by CODEBOOK.

Symbol	Mean	Median	Mode	Variance	SD	SE	Coefficient of Variation	Kurtosis	Skewness	Min	Max	Range	Valid V.	Missing V.
SDI1	30.537	19.775	0.800	967.33	31.102	2.717	101.85	-1.117	0.586	0	99.0	99.0	131	32
	31.565	20.5	5.0	877.73	29.626	2.588	93.86	-1.102	0.613	0	95.0	95.0	131	32
SDI2	57.247	71.575	0.500	1 419.14	37.671	3.291	65.8	-1.526	-0.338	0	100.0	100.0	131	32
	57.405	71.154	95.0	1 285.9	35.86	3.133	62.47	-1.532	-0.345	0	95.0	95.0	131	32
SD2D	70.482	80.638	99.5	809.26	28.45	3.33	40.36	-0.326	-0.919	0.2	99.7	99.5	73	90
	69.93	81.786	95.0	800.34	28.29	3.311	40.45	-0.47	-0.925	5.0	95.0	90.0	73	90
SD5D	58.625	71.275	93.8	1 116.48	33.41	3.013	56.99	-1.363	-0.399	0	99.5	99.5	123	40
	58.94	72.06	95.0	1 086.17	32.96	2.972	55.92	-1.377	-0.413	0	95.0	95.0	123	40
SD2S	52.96	57.35	99.4	1 091.5	33.04	3.867	62.39	-1.353	-0.107	0	99.5	99.5	73	90
	52.33	55.0	95.0	1 051.45	32.43	3.795	61.97	-1.376	-0.122	0	95.0	95.0	73	90
SD5S	44.23	40.65	99.0	1 105.38	33.25	2.81	75.18	-1.376	0.242	0	99.1	99.1	140	23
	44.25	41.67	5.0	1 075.87	32.8	2.772	74.12	-1.395	0.252	0	95.0	95.0	140	23
FN	42.45	34.5	100.0	676.6	26.01	2.872	61.27	-0.482	0.654	0	100.0	100.0	82	81
	41.65	35.0	25.0	656.21	25.617	2.829	61.51	-0.67	0.558	0	95.0	95.0	82	81
MS	41.78	18.0	100.0	1 605.81	40.07	7.859	95.91	-1.515	0.479	0.5	100.0	99.5	26	137
	40.77	18.33	5.0	1 473.39	38.385	7.528	94.15	-1.549	0.486	5.0	95.0	90.0	26	137
PL	20.54	20.85	18.5	12.827	3.582	0.345	17.7	8.604	-1.608	0	28.40	28.40	108	55
	20.486	20.55	19.5	14.01	3.743	0.360	18.27	7.060	-1.386	0	28.50	28.50	108	55
QL	33.536	32.72	23.0	62.01	7.875	0.758	23.48	2.632	-0.053	0	55.30	55.30	108	55
	33.538	33.136	33.0	67.70	8.228	0.792	24.5	2.075	-0.106	0	57.00	57.00	108	55
PI	12.995	12.55	10.00	36.07	6.006	0.578	46.22	0.462	0.697	0	31.50	31.50	108	55
	12.792	12.25	10.50	36.44	6.037	0.581	47.19	0.097	0.565	0	28.50	28.50	108	55
PSA	14.361	9.167	2.00	189.43	13.763	1.324	95.84	1.207	1.362	1.00	60.00	59.00	108	55
	13.889	9.20	2.50	167.212	12.931	1.244	93.10	0.428	1.212	2.50	47.50	45.0	108	55
PSI	48.972	48.9	49.0	144.31	12.013	1.156	24.53	-0.362	-0.130	16.00	75.00	59.00	108	55
	48.611	48.71	55.0	152.259	12.339	1.187	25.38	-0.302	-0.053	15.0	75.0	60.0	108	55
PCL	21.657	19.357	17.00	100.994	10.05	0.967	46.41	1.874	1.167	4.00	61.00	57.00	108	55
	21.389	19.464	17.50	96.417	9.819	0.945	45.91	0.578	0.863	2.5	47.5	45.0	108	55
PCOL	17.269	16.40	17.00	58.0	7.62	0.79	44.13	0.007	0.533	1.00	38.00	37.00	93	70
	16.64	15.938	17.5	60.122	7.754	0.804	46.6	0.113	0.570	2.5	37.5	35.0	93	70
WD[a]	149.16	150.9	151.1	61.734	7.857	0.756	5.27	0.851	-0.783	122.3	164.9	42.6	108	55
	149.306	150.875	152.5	64.233	8.015	0.771	5.37	0.584	-0.738	122.5	162.5	40.0	108	55
DD[a]	138.192	141.93	145.0	146.28	12.095	1.091	8.71	0.249	-0.806	101.5	159.7	58.2	123	40
	139.49	141.5	147.5	110.14	10.495	0.906	7.52	-0.902	-0.343	122.5	157.5	35.0	123	40
AMC	6.132	4.575	3.4	24.332	4.933	0.418	80.45	2.469	1.464	0.4	27.2	26.8	139	24
	6.183	5.092	1.5	24.54	4.953	0.42	80.11	2.827	1.486	1.5	28.5	27.0	139	24
SG	2.752	2.756	2.77	0.001	0.034	0.004	1.24	-0.465	-0.078	2.68	2.84	0.16	93	70
	2.745	2.747	2.75	0.002	0.040	0.004	1.46	3.078	-0.445	2.65	2.85	0.20	93	70
PH	7.141	7.328	7.30	0.933	0.966	0.101	13.53	4.197	-1.80	3.4	8.6	5.2	92	71
	7.065	7.255	7.5	0.974	0.987	0.103	13.97	3.917	-1.666	3.5	8.5	5.0	92	71
SL	16.784	16.075	16.50	21.645	4.652	0.499	27.72	2.226	1.386	10.6	35.3	24.7	87	76
	16.845	15.983	16.50	19.973	4.469	0.479	26.53	0.360	0.970	10.5	28.5	18.0	87	76
SH	5.586	5.175	4.100	6.932	2.633	0.282	47.14	-0.290	0.400	0	12.70	12.70	87	76
	5.506	5.056	4.50	6.515	2.552	0.274	46.35	-0.679	0.235	0	10.50	10.50	87	76
XY	7.222	6.067	4.80	19.65	4.433	0.455	61.38	5.627	2.262	1.20	27.10	25.90	95	68
	7.184	6.044	4.50	20.77	4.56	0.468	63.47	5.938	2.265	1.50	28.50	27.0	95	68
PM	12.365	12.325	11.30	8.304	2.882	0.363	23.31	-0.348	0.045	6.90	19.20	12.30	63	100
	12.63	12.29	12.5	6.05	2.46	0.31	19.48	-0.463	0.607	9.50	18.50	9.0	63	100
WD1[a]	132.68	134.12	134.2	55.59	7.46	1.076	5.62	9.899	-2.811	97.5	141.4	43.9	48	115
	132.71	133.89	132.5	57.40	7.58	1.094	5.71	8.647	-2.481	97.5	142.5	45.0	48	115
DD1[a]	116.82	117.9	117.9	43.27	6.58	0.829	5.63	3.356	-1.43	90.1	126.7	36.6	63	100
	116.79	117.5	117.5	46.66	6.83	0.861	5.85	2.013	-1.041	92.5	127.5	35.0	63	100
CBR1	11.77	9.0	9.6	61.66	7.85	1.079	66.69	0.761	1.44	2.10	31.80	29.70	53	110
	11.632	8.929	7.5	59.271	7.699	1.058	66.19	0.549	1.346	1.50	28.50	27.0	53	110
CBR2	6.02	3.625	2.1	33.22	5.76	0.792	95.68	0.164	1.108	0	21.80	21.80	53	110
	6.057	3.938	1.5	33.401	5.779	0.794	95.41	0.441	1.189	0	22.50	22.50	53	110
AS	1.442	0.863	0	2.788	1.67	0.23	115.81	2.857	1.658	0	7.80	7.80	53	110
	1.325	0.750	0.25	2.074	1.440	0.198	108.66	0.090	1.115	0	4.75	4.75	53	110

Note: The first line for every parameter gives the values obtained by using single points, while the second line gives the results for grouped data (10 equal groups in range of values).
[a]Density values are in pounds per cubic foot; 1 kg/m^3 = 0.062 lb/ft^3.

1. Figure 2 illustrates a good straight-line correlation between parameters, well enough defined over a wide enough range so that results of the one can be used to estimate values for the other.

2. Figure 3 shows a definite trend between two parameters, although the correlation is low. This type of correlation is useful in determining qualitative relationships that may assist in future research.

3. After plotting the data, certain boundaries may be established within which all the data presently available fall. This type of diagram may be used in the future to determine the reliability of laboratory testing. The boundaries may change when more data become available and new groupings may be identified. An example of this type of plot is given in Figure 4.

Very few good correlations were obtained on a one-to-one basis for all the shale data. It should be remembered, however, that the Indiana shales cover a wide range of characteristics. With this in mind it was decided to analyze some of the data after dividing the data sets according to the geology as follows:

1. Ordovician, all formations;
2. Devonian, all formations plus New Albany of Mississippian;
3. Mississippian, Borden group, Locust Point, New Providence;
4. Mississippian, rest of formations; and
5. Pennsylvanian, all formations.

Only those parameters containing some notion of durability and strength were analyzed. The best correlations were observed between the different slake durability

Figure 1. Histograms for grouped data.

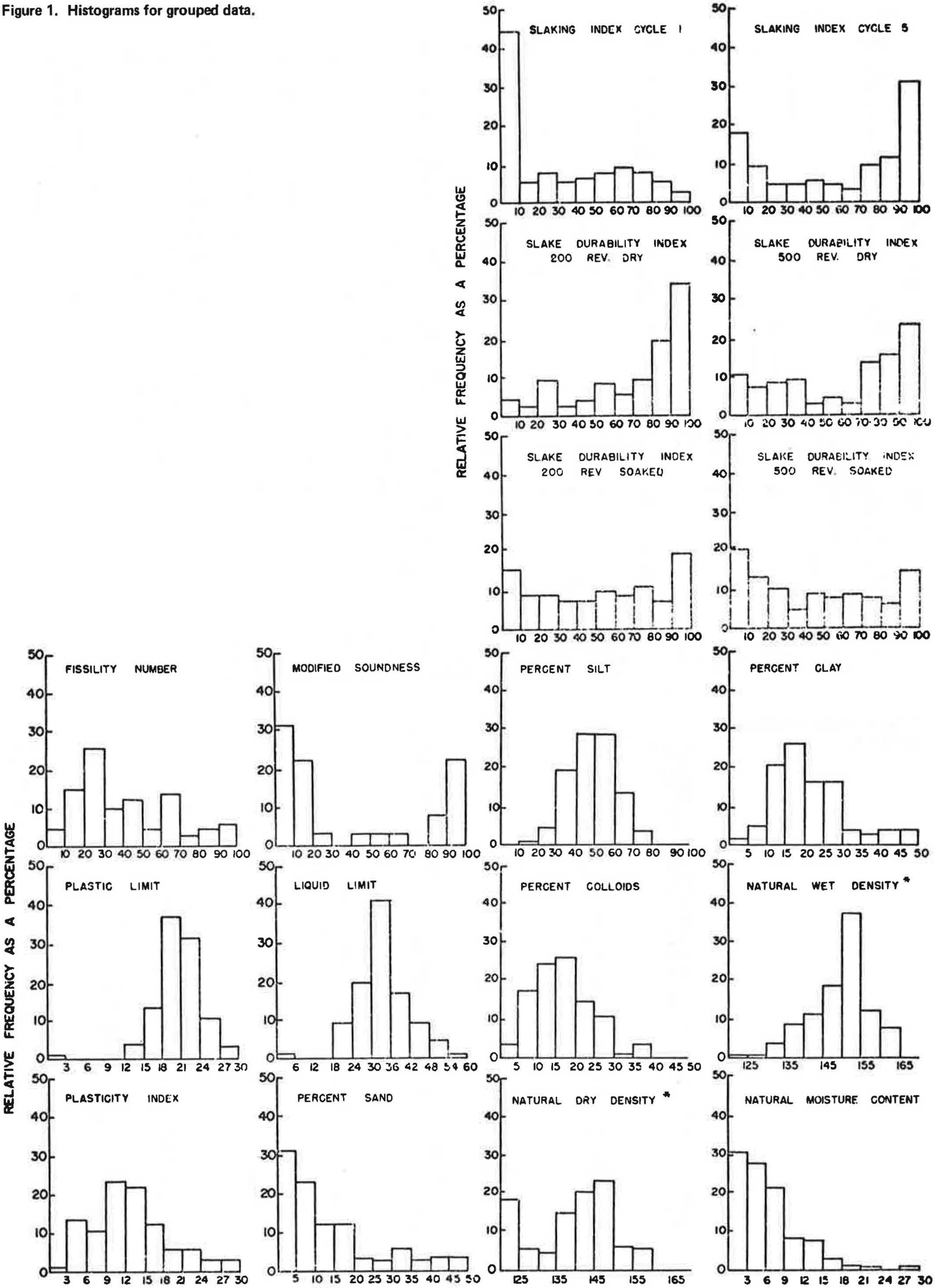

Figure 1. Continued.

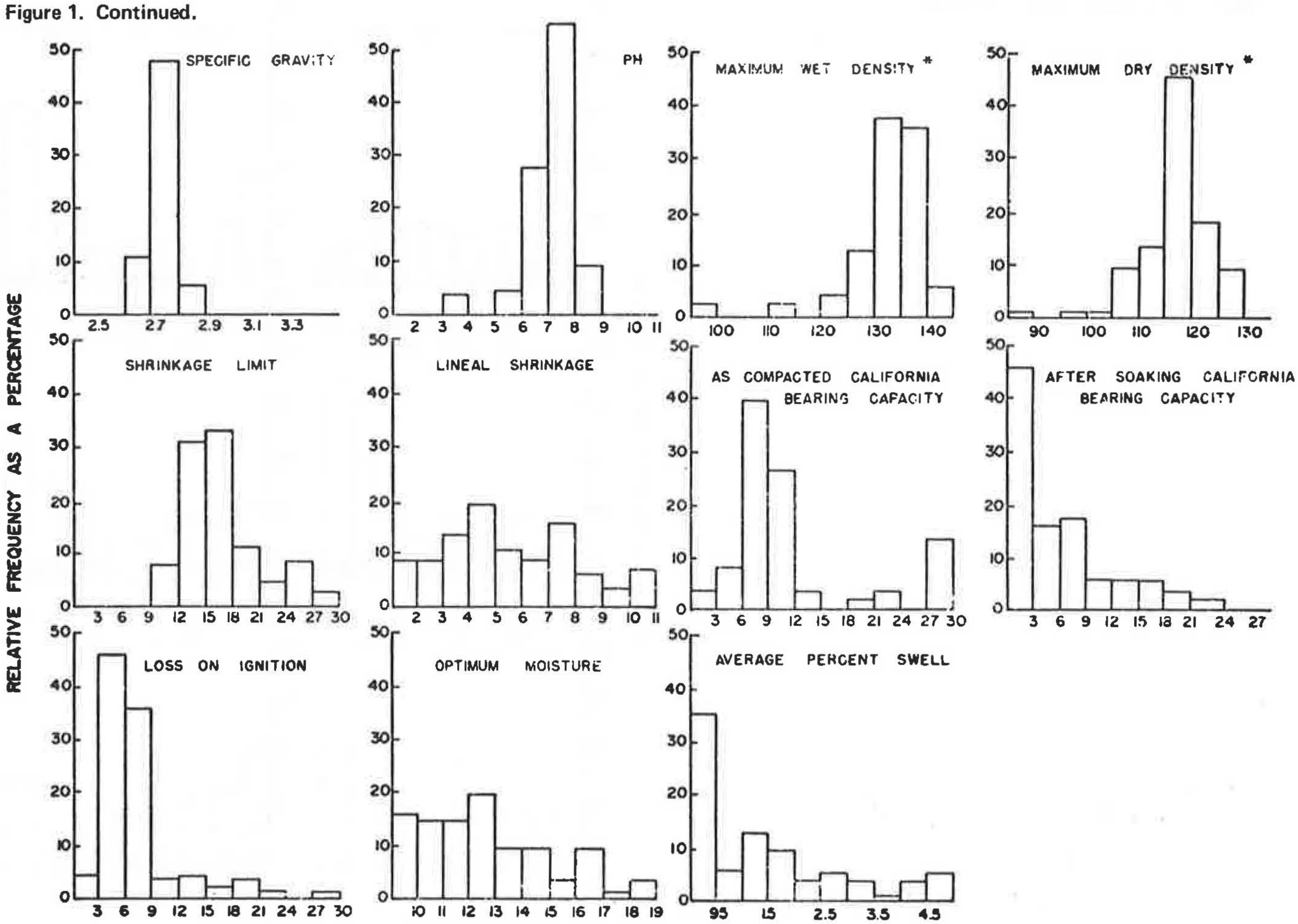

Table 3. Bivariate correlations for R > 0.7.

Independent Variable (down)	Dependent Variable (across)	R	Standard Error of Estimate	Intercept A	Standard Error of A	Slope B	Standard Error of B	Plotted Values
SDI1	SDI2	0.827 3	17.539	-8.565	2.795	0.683	0.0408	131
SDI1	SD2D	-0.700 6	19.83	71.436	6.373	-0.685	0.0834	72
SDI1	SD5D	-0.729 1	21.07	62.773	3.959	-0.673	0.062	107
SDI1	SD2S	-0.710 6	19.554	54.449	4.405	-0.594	0.0703	72
SDI2	SD2D	-0.73	24.609	115.143	7.909	-0.9253	0.1035	72
SDI2	SD5D	-0.839 9	20.638	103.835	3.877	-0.9578	0.0604	107
SDI2	SD2S	-0.86	18.375	99.129	4.139	-0.9311	0.066	72
SDI2	SD5S	-0.8	22.595	92.151	3.45	-0.9445	0.066	117
SD2D	SD5D	0.789	17.593	32.273	4.086	0.685	0.0632	73
SD2D	SD2S	0.719 7	19.89	37.667	4.42	0.6197	0.071	73
SD5D	SD2S	0.775 2	20.867	15.069	4.637	0.7695	0.074	73
SD5D	SD5S	0.823 8	19.018	20.885	2.918	0.8189	0.0512	123
SD5D	PM	0.699 8	22.015	156.324	12.608	-7.5146	1.0072	60
SD2S	SD5S	0.810 1	19.5067	18.579	3.733	0.7983	0.0686	73
SD5S	MS	-0.727 8	17.369	95.44	4.972	-0.4506	0.0867	26
SD5S	PI	-0.685 2	24.319	89.66	5.749	-3.832	0.4095	101
SD5S	PM	-0.678	24.874	148.354	14.246	-7.994	1.133	60
MS	AMC	0.919 5	18.381	-26.439	12.679	17.999	2.565	11
MS	SL	-0.723 5	32.278	201.748	48.125	-7.342	2.335	11
MS	XY	-0.758 3	30.482	115.705	20.04	-7.209	2.066	11
MS	PM	0.769 4	26.803	-68.162	20.07	10.635	1.883	24
MS	WD1	0.814 1	28.437	-247.38	83.466	2.403	0.648	9
MS	CBR1	-0.678	29.579	82.51	12.415	-2.629	0.637	22
MS	CBR2	-0.720 3	27.916	84.674	11.61	-4.484	0.966	22
PL	QL	0.687 3	2.614	10.058	1.105	0.3126	0.0321	108
QL	PI	0.901 3	3.427	18.179	0.789	1.1818	0.0552	108
QL	PCOL	0.686 8	5.612	22.513	1.449	0.6925	0.0768	93
QL	SH	0.711 55	5.473	22.835	1.383	2.093	0.224	87
PI	PCOL	0.697 8	4.256	4.419	1.099	0.5414	0.0583	93
PI	SH	0.721 4	4.0936	4.914	1.034	1.6104	0.1677	87
WD	DD	0.876 8	3.796	42.528	5.692	0.7513	0.04	108
PM	CBR2	-0.769 5	1.875	14.467	0.3743	-0.3883	0.045	53
WD1	DD1	0.938 5	2.6027	19.283	0.163	0.9697	0.0526	48
CBR1	CBR2	0.802 5	4.731	5.1903	0.9442	1.0935	0.1138	53

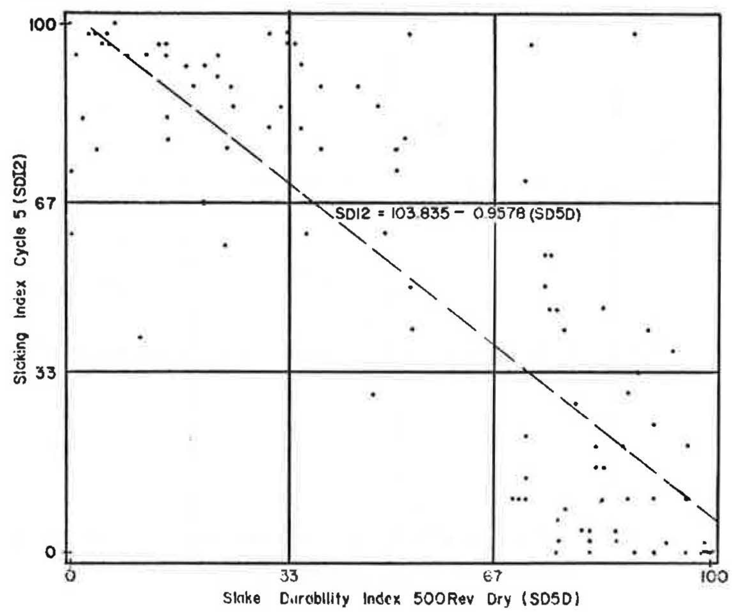

Figure 2. Slake durability index 500 revolutions dry versus slaking index cycle 5 (R = 0.84).

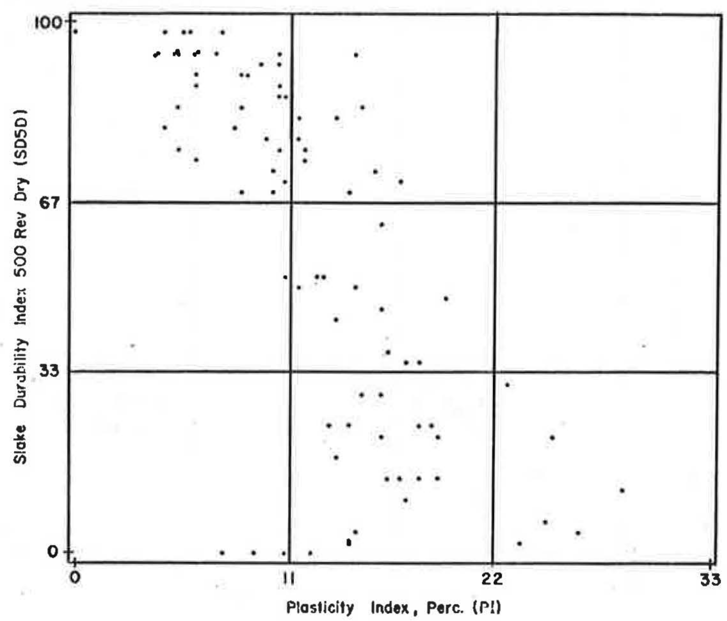

Figure 3. Plasticity index versus slake durability index 500 revolutions dry (R = 0.68).

measures. The basic trends were the same, but some differences do exist between the different geological formations. Generally the slake durability measures for point 2 show better one-to-one correlations than those obtained for the other geological stages. However, the number of sample points was smaller (2, Table 14). It is considered that the division according to geological stages will prove to be a helpful tool in correlating different parameters when more complete data sets are available.

Multiple Regression Analysis

The REGRESSION routine of the SPSS package was used to carry out the multiple regression analysis. For this study, attempts were made to predict California bearing ratio (CBR) from some of the other parameters because CBR is considered to be the most time-consuming test. Although some other correlations are available in the literature to determine CBR, this study concentrated on the parameters available from the laboratory reports, those given in Table 1.

First it was necessary to pick out the 24 data sets with points for all the different parameters.

Three different regression models were obtained to calculate as-compacted CBRs and after-soaking CBRs. Stepwise regression was performed, and the first regression model was obtained by choosing the five parameters with the highest correlation on a one-to-one basis with as-compacted CBR and after-soaking CBR. However, determining the values of the five parameters would likely take as long as the testing, which means that the model has no practical value.

The second regression model was set up with parameters reported in a larger number of data sets. The model was set up with the same 24 sets as before, while it was tested against the data of Deo (3) (i.e., 15 different

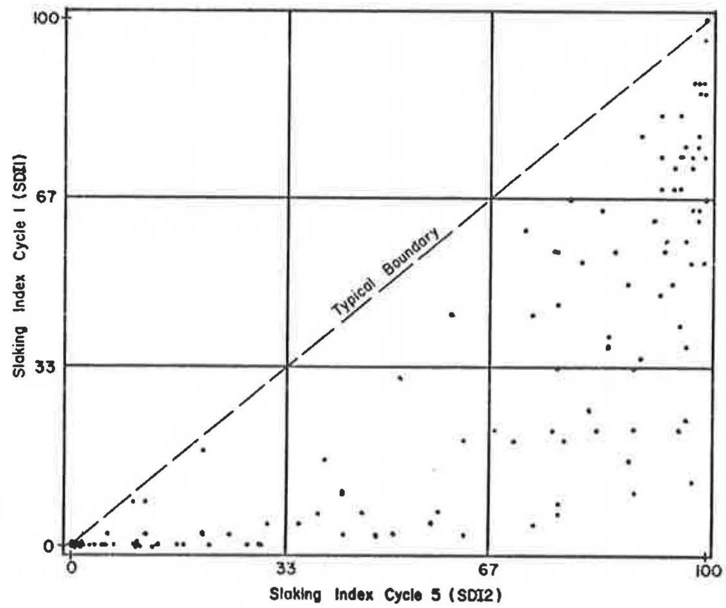

Figure 4. Slaking index, cycle 5 versus cycle 1 (R = 0.83).

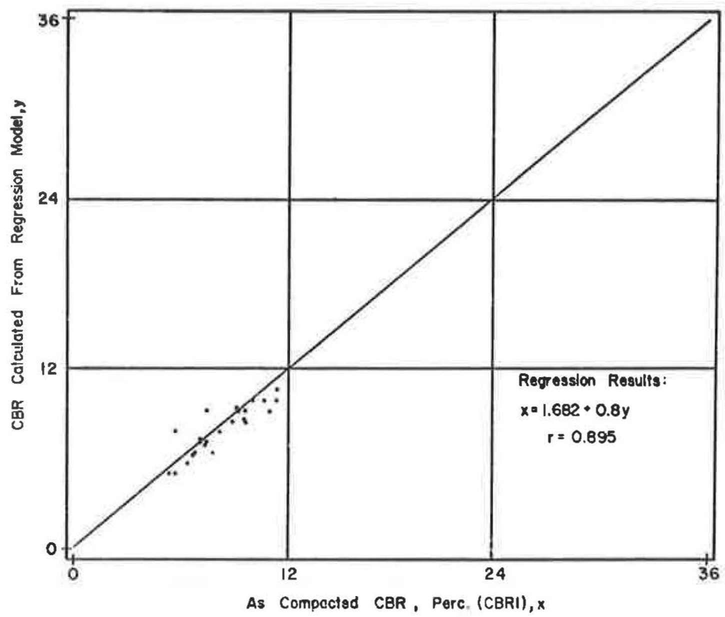

Figure 5. Goodness of fit of regression model for data used to set up the model, 24 data sets.

sets). Figures 5 and 6 indicate how well the model for as-compacted CBRs fits the data of the 24 data sets (used to set up the model) and also the data of Deo. The fit of the latter plot is not as good as the first, although it has a wider range of values and may be considered as a reasonable model for planning purposes. This model still includes parameters that are difficult to determine and might therefore be impractical.

A final model based on simple tests was tried. The results are given in Table 4. It can be seen that this model is very good for the 24 sets of data, but it must still be validated.

With the good correlations obtained between the different slake durability parameters on a one-to-one basis, it was decided to analyze these correlations, again using second-order equations and the regression program. For example, the relationship between SDI1 and SDI2 would be determined by the equation

$$SDI2 = C_1(SDI1)^2 + C_2(SDI2)(SDI1) + C_3(SDI1) + C_4 \quad (1)$$

The results of the two different analyses for five parameters are given below.

The simple-to-perform tests yield the following multiple regressions (a) for as-compacted CBR (R = 0.8657)

$$\begin{aligned}\text{As-completed CBR} = {} & 0.058\,55(SDI2)(PH) + 0.017\,099(SDI2)(XY) \\ & + 0.002\,860\,4(DD)^2 - 0.026\,762(DD)(AMC) \\ & - 14.8706(PH) - 0.003\,068\,7(SDI2)(DD) \\ & + 0.842\,56(PH)^2 + 10.9822(AMC) \\ & - 0.263\,75(XY)(AMC) + 0.064\,04(XY)^2 \\ & - 0.023\,976(DD)(XY) - 0.259\,51(PH)(AMC) \\ & - 0.015\,17(SDI2)(AMC) + 0.467\,54(PH)(XY) \\ & - 0.142\,685(AMC)^2 + 0.000\,486\,4(SDI2)^2 \\ & - 6.406\,005\,2 \quad (2)\end{aligned}$$

Figure 6. Goodness of fit of regression model for data used to verify the model, 15 data sets.

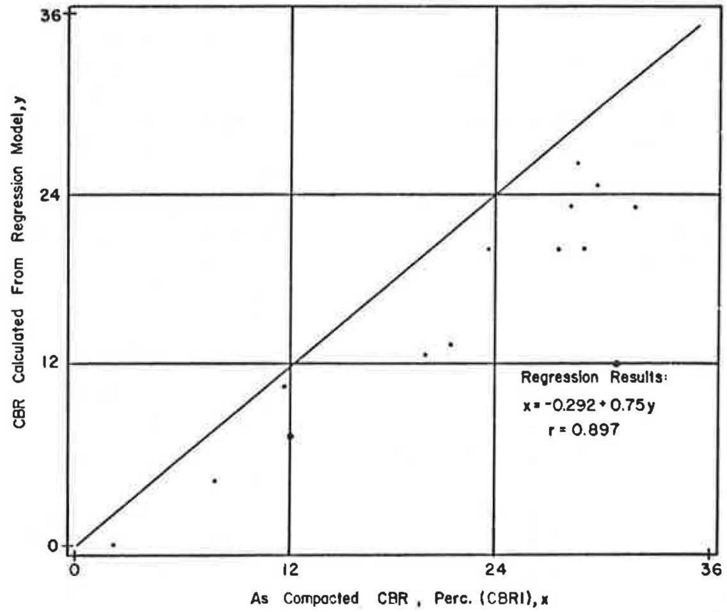

Table 4. Results of regression on second-order equations for durability parameters.

Dependent Variable (y)	Independent Variable (x)	C_1	C_2	C_3	C_4	Multiple R
SDI2[a]	SDI1	-0.031 01	0.032 47	0.367 89	19.1696	0.9018
SD2D	SDI1	-0.017 63	0.017 84	-2.687 4	90.5737	0.8984
SD5D	SDI1	0.020 93	0.017 59	-2.842 7	83.325	0.8998
SD2S	SDI1	0.024 14	0.021 77	-3.009 6	78.699	0.8981
SD5S	SDI1	0.025 07	0.028 2	-2.993	69.224	0.8399
SD2D	SDI2	0.006 14	0.012 05	-1.667 2	97.144	0.9798
SD5D	SDI2	0.009 56	0.014 2	-1.970 98	92.5352	0.9759
SD2S	SDI2	0.012 03	0.013 74	-2.166	93.406	0.9808
SD5S	SDI2	0.014 78	0.016 38	-2.341 3	85.7386	0.9495
SD5D	SD2D	-0.008 88	0.012 23	0.663 3	3.2595	0.9899
SD2S	SD2D	-0.009 78	0.012 66	0.731 9	2.063	0.9808
SD5S	SD2D	-0.005 98	0.011 81	0.433 4	3.901	0.9732
SD2S	SD5D	-0.006 96	0.014 4	0.118 4	20.87	0.9045
SD5S	SD5D	-0.007 53	0.013 09	0.431 2	7.8161	0.9626
SD5S	SD2S	-0.009 14	0.012 36	0.692 8	2.7344	0.9795

[a] $y = C_1x^2 + C_2xy + C_3x + C_4$; for example $SDI2 = C_1(SDI1)^2 + C_2(SDI2)(SDI1) + C_3(SDI1) + C_4$.

and (b) for after-soaking CBR (R = 0.9231)

$$
\begin{aligned}
\text{After-soaking CBR} = {} & 0.041\,95(SDI2)(PH) + 0.005\,662(SDI2)^2 \\
& + 8.120\,46(PH) - 0.125\,84(XY)^2 \\
& + 0.021\,707(SDI2)(XY) + 0.021\,051(SDI2) \\
& \times (AMC) + 12.2994(XY) - 0.694\,905(SDI2) \\
& - 0.577\,57(PH)(AMC) - 0.448\,96(PH)(XY) \\
& + 0.086\,209(AMC)^2 - 0.264\,604(PH)^2 \\
& - 0.244\,53(XY)(AMC) - 0.045\,907(DD)(XY) \\
& + 0.018\,094(DD)(AMC) + 0.130\,03(DD) \\
& - 49.9787 \qquad (3)
\end{aligned}
$$

where

SDI2 = slaking index cycle 5,
PH = pH,
XY = loss on ignition,
DD = natural dry density, and
AMC = natural moisture content.

The correlations are much higher than those obtained during the bivariate analyses. It is therefore recommended that these second-order equations be used for predicting the different durability parameters when one is known. It should be remembered that these models are based on all 163 sets of data and were not validated. New data that become available should be used for validating them.

CONCLUSIONS

The storage and retrieval system described in this report is only a first attempt and should be improved in the future by the inclusion of additional data as they become available. The shales of Indiana cover a wide range of characteristics and it is necessary to have complete data sets on as many different shales as possible in order to obtain good statistical correlations.

The values of all the slake durability indices can be estimated if one is known, by using the second-order equations above (Equation 1). Various relations between the data on a one-to-one basis give reasonable correlations; details of the most significant relations are given in Table 3.

Various regression models can be obtained to predict some parameters by using others. Good models for estimating CBR were obtained by using five parameters. It should be remembered that the data used for the models come from all the different geological series. Better models may result in the future when more complete data sets are available for separate geological formations, but results obtained should not be extrapolated beyond the range of test conditions of this study.

This partial analysis indicates that, when the data are divided into groups based on geological origin, the one-to-one correlation can be increased over that of all the data lumped together.

ACKNOWLEDGMENTS

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 here. We also appreciate the advice of our colleague Aubrey Abeyesekera. The contents of this paper reflect our views, and we are responsible for the facts and the accuracy of the data presented here. The contents do not necessarily reflect the official views or policies of the research sponsors, who were the Federal Highway Administration and the Indiana State Highway Commission. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. R.A. Crawford, J.B. Thomas, and M. Stout, Jr. Computerized Soil Test Data for Highway Design. Physical Research Section, South Dakota Department of Highways, 1972.
2. D. Spradling. A Soils Data System for Kentucky. Division of Research, Bureau of Highways, Kentucky Department of Transportation, Research Rept. 441, 1976.
3. P. Deo. Shale as Embankment Materials. Joint Highway Research Project, Purdue University, West Lafayette, IN, Rept. No. 14, 1972.
4. D.R. Chapman. Shale Classification Tests and Systems: A Comparative Study. Joint Highway Research Project, Purdue University, West Lafayette, IN, Rept. No. JHRP-75-11, 1975.
5. M.J. Bailey. Shale Degradation and Other Parameters Related to the Construction of Compacted Embankments. Joint Highway Research Project, Purdue University, West Lafayette, IN, Rept. No. JHRP-76-23, 1976.
6. D.J.A. van Zyl. Storage, Retrieval and Statistical Analysis of Indiana Shale Data. Joint Highway Research Project, Purdue University, West Lafayette, IN, Rept. No. JHRP-77-11, 1977.
7. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent. Statistical Package for the Social Sciences. McGraw-Hill, New York, 2nd Ed., 1975.

Publication of this paper sponsored by Committee on Soil and Rock Properties.

Using Indicative Properties to Predict the Density-Moisture Relationship of Soils

Moshe Livneh and Ilan Ishai, Transportation Research Institute, Technion—Israel Institute of Technology, Haifa

The need for predictions of density-moisture relationships by means of the index properties of a soil is obvious because an engineer first becomes acquainted with a soil by determining these properties. Preliminary engineering reports normally use information from soil classification surveys to obtain findings relevant to earthwork, slope design, and structural pavement design. It is, therefore, natural to analyze the accumulated density test results taken from various sites in order to obtain correlations between the index properties and the engineering properties of the soil. This paper presents a method for predicting the optimum line—the curve that connects the peaks of the density-moisture curves obtained at different levels of compaction effort. The first stage in the prediction process is to make a qualitative and quantitative acquaintance with the soil compaction mechanism governing the characteristics of the typical density-moisture curve. The next step is to predict maximum densities and optimum moisture contents for given levels of compaction effort (standard and modified AASHTO). Graphs and regression equations based on the plastic and liquid limits of the soil and using the suction criterion are presented. The predicted maximum density and optimum moisture content are also related to the critical voids percentage. The method does not substitute the execution of the tests themselves but enables one to obtain reliable preliminary information.

The density-moisture relationship of soils is a well-known criterion for the compaction design of subgrades and embankments of flexible pavements. Such compaction design can be expressed in terms of (a) compaction moisture content, (b) recommended compaction degree, and (c) type of compaction effort.

The need for predictions of density-moisture relationships by means of the index properties is obvious, because one's first acquaintance with soils is made by determining these properties. Stated differently, one of the objectives of soil classification surveys is to produce general information about the expected engineering properties. Preliminary engineering reports usually use this information to determine earthwork, slope design, and structural pavement design.

The usefulness of predicting density-moisture relationships is expressed mainly in the preliminary design phase and feasibility studies of highways and airports when major information is needed for the evaluation of the earthwork and design parameters in terms of degree of natural density, maximum density, required molding water content, etc. It is also very important in cases where soil types are variable along the alignment. In this case