DEVELOPING STATISTICAL CORRELATIONS OF SOIL PROPERTIES WITH R-VALUE FOR IDAHO PAVEMENT DESIGN

Final Report
N08-11
KLK553
ITD RP185: UI-08-04
September 2009
Prepared by

National Institute for Advanced Transportation Technology

University of Idaho

Primary Author:

Stanley M. Miller, Ph.D., P.E.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Developing Statistical Correlations of Soil Properties with R-Value for Idaho Pavement Design	5. Report Date September 2009 6. Performing Organization Code KLK553
5.Author(s) Stanley M. Miller, Ph.D., P.E.	8. Performing Organization Report No. N08-11
9. Performing Organization Name and Address National Institute for Advanced Transportation Technology University of Idaho PO Box 440901; 115 Engineering Physics Building Moscow, ID 83844-0901	10. Work Unit No. (TRAIS) 11. Contract or Grant No. RP185: UI-08-04
12. Sponsoring Agency Name and Address Idaho Transportation Department PO Box 7129 Boise, ID 83707-7129	13. Type of Report and Period Covered Final Report January 2008 – June 2009 14. Sponsoring Agency Code

Supplementary Notes:

16. Abstract

Historical geotechnical soil testing results have been collected, sorted, and culled from ITD materials reports and soil-profile scrolls to investigate relationships between basic soil properties and R-value. Several thousand data records (dated from 1953 through 2008) representing all 25 classes of soils prescribed by the Unified Soil Classification (USC) system were initially input to Excel® files; after editing, there were 8243 records retained for subsequent statistical analyses. This database is unique in that the exudation pressure used for ITD R-value testing (Idaho T-8) is 200 psi (1,380 kPa), rather than the traditional level of 300 psi (2,070 kPa). Because ITD changed its testing procedure in June of 1971, two-sample statistical tests were used to compare Pre-1971 R-value data with the Post-1971 data to assess whether the earlier results should be adjusted to the Post-1971 levels. The adjustments for the Pre-1971 R-values primarily were needed for the clayey soils. Multiple regression analysis of the culled data indicated that three soil properties had significant influence on R-value: USC code, plasticity index (PI), and percent finer than the No. 200 sieve. Soil resistivity also can be useful in predicting R-value, but primarily for coarse-grained soils. Additional regression analyses were used to develop and recommend predictive models for R-value.

17. Key Words	18. Distribution Statement
R-value, correlation, pavement	Unrestricted; Copies available from http://www.webs1.uidaho.edu/niatt/

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	28	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS				
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa
APPROXIMATE CONVERSIONS FROM SI UNITS				
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

INTRODUCTION

The Materials Sections in all six Districts of the Idaho Transportation Department (ITD) have historical records of numerous R-value tests conducted for the initial strength characterization of subgrade, subbase, and base-course materials used in the design of highway pavements. Much of this historical information is contained in project reports and on soil profile drawings (scrolls) archived in the District offices but has never been analyzed comprehensively to determine if usable statistical correlations can be identified between fundamental soil properties and R-value. Idaho's database is unique in that the exudation pressure used for ITD R-value testing (Idaho T-8) is 200 psi (1380 kPa) rather than the more commonly prescribed level of 300 psi (2070 kPa) per AASHTO T 190 and ASTM D 2844. ITD's testing procedure was modified in June 1971 when the compaction efforts of the kneading compactor were reduced and the exudation pressure was changed from 320 psi to 200 psi.

Thus, to pursue an investigation of potential statistical correlations for R-value using the ITD historical test results, the primary objectives of this study were

1. To review and "mine-out" available historical testing data (R-value and other soil testing results) from each of the six Districts;
2. To organize and sort the data so that statistical comparisons can be made between the pre-1971 and post-1971 R-values to develop adjustment factors to bring the earlier data into "agreement" with the later data; and
3. To investigate and develop usable multiple regression models that can be used to predict R-value, given fundamental soil properties such as soil classification, liquid limit (LL), plasticity index (PI), and percent finer than 0.075 mm (No. 200 sieve).
4. As part of a project work amendment, to add new R-value testing results from 2008 to the original database, and to investigate soil resistivity as another factor to help predict R-value.

DATA COLLECTION

Detailed review of historical soil testing results and the associated data collection at each District office were accomplished during specifically scheduled work periods in 2008. The work was conducted by Jonathan Rush, an undergraduate research assistant in the Department of Civil Engineering at the University of Idaho. He entered the information from ITD project reports and soil profile scrolls directly into Excel files on a laptop computer. Information in Districts 1 and 2 was collected in March 2008; District 3 information was collected in late May 2008; District 4 information was collected in early June, along with initial work at District 5; District 5 was completed in late July; and District 6 information was collected in early August 2008. A summary is presented in Table 1 for the number of R-value testing records obtained from the ITD historical archives deemed useful for subsequent statistical analysis. Not all of the records on file in the District offices contained sufficient soil testing information to be used in subsequent statistical modeling.

Table 1. Summary of Historical R-value Records from ITD (1953-2007)

District	No. of R-value Data (Records)	Pre-1971	Post-1971
1	409	198	211
2	325	99	226
3	2197	1099	1098
4	1115	928	187
5	2408	1483	925
6	1711	988	723
Total	8165	4795	3370

Every reasonable effort was made to obtain as much R-value data as possible from the ITD archives. However, we cannot be sure that all pertinent written documents were obtained from ITD personnel and/or retrieved from storage locations. We do believe that this data set represents a fairly comprehensive and thorough compilation of accessible soil testing and R-value data from ITD historical records available through 2007.

During the initial statistical analysis of this study, and then during multiple regression modeling later in this study, some of the R-value records were culled due to them being duplicate soil samples, whose test results were recorded in new places in the archives for a subsequent highway project. Although such records may have different project names/numbers, the recorded sample numbers (or locations) are identical and the test results are identical. These duplicate records were deleted from the final database used to investigate regression models to predict R-value.

STATISTICAL TESTS FOR ADJUSTING PRE-1971 R-VALUES

The collected soil testing information from each district was sorted into those testing results obtained through June of 1971 (identified as "Pre-1971") and those obtained after June of 1971 (identified as "Post-1971"). As mentioned earlier, ITD changed its R-value testing procedure in June of 1971. Thus, prior to developing any multiple regression models using soil properties to predict R-value, these pre-1971 R-values were adjusted (corrected) as necessary to bring them into the same general agreement with the post-1971 R-values. Such adjustments primarily were required for the clayey soils, whose R-values were affected the most by the change in testing procedure.

Statistical hypothesis testing was used to conduct two-sample tests of means to compare the Pre-1971 R-values with the Post-1971 R-values for each soil type. These two-sample means tests were conducted using the Data Analysis module in Excel®, based a student's "t" statistic to compare each pair of sample means (at a level of significance of $\alpha = 0.05$).

For our initial analysis, we tried to compare means for the soil testing data within each individual District. However, due to a paucity of samples for some soil types and to the high variability observed in R-values for some of these under-sampled soil types in individual Districts, we eventually decided to conduct the means tests on the combined data set for all Districts. After data cleaning and culling, this final combined data set contained 4795 records in the Pre-1971 group and 3370 records in the post-1971 group (total number of clean R-values records was 8165). For a given soil type/class, if the statistical test indicated that the two sample means were significantly different (at a prescribed level of significance of $\alpha = 0.05$), then all the Pre-1971 R-values for that soil type/class would be adjusted by an amount equal to the difference between the two sample means. The only soil class that did not have enough data for the means test was Class 14, SP-SC.

The results of the two-sample means tests for the combined R-value data for all Districts are given in Table 2. As expected, the adjustments for the Pre-1971 R-values primarily applied to the clayey soils. The soil classes or codes identified for Pre-1971 R-value adjustments (i.e., reductions) were Class 3 (CH), Class 5 (CL), Class 6 (CL-ML), and Class 9 (GC). One soil type, Class 13 (GM), showed a statistically significant increase in R-value when the Pre-1971 mean was compared to the Post-1971 mean, so these Pre-1971 R-values were increased (rather than being reduced).

After the appropriate adjustments were made for the identified Pre-1971 R-values to make them consistent with Post-1971 R-values, the data set for each District was visually checked in the Excel files to verify that all duplicate records were culled and to identify misclassifications of soil types when older data records with AASHTO soil classifications had been converted to the Unified Soil Classification (USC) system. For example, in some cases a soil sample classified as AASHTO "A-6" was converted to USC CL instead of SC (due to the AASHTO break between fine and coarse soils being 35 percent finer than the No. 200 sieve, whereas the USC break is at 50 percent finer than the No. 200 sieve). Also, all NP (non-plastic) entries for PI were changed to 0 for subsequent statistical analysis.

ADDITIONAL SOIL TESTING DATA

In early 2009, we obtained recent ITD soil testing results with R-values from laboratory work conducted during 2008. This added 115 new records to the database, including 19 from District 1, 21 from District 2, 36 from District 3, 5 from Districts 4 and 5, and 34 from District 6. These R-value records were checked for consistency and added to the database. During this process, we noted that some of the new testing results included information on soil resistivity reported in ohm-centimeters (Ω-cm) and that the resistivity generally seemed to be less for clayey soils (thus implying that low resistivity may indicate low R-value).

Table 2. Two-Sample Means Tests for Pre-1971 and Post-1971 R-Values

(Data from from All Districts for the Period 1953-2007)

Soil Classification	Pre-1971 R-values			Post 1971 R-values			Means Significantly Different? *	Pre-1971 R-Value Recommended Adjustment
	Number of Data	Mean	Standard Deviation	Number of Data	Mean	Standard Deviation		
OH (1)	3	36.3	21.5	2	25.5	6.4	no	0
OL (2)	15	44.3	14.1	17	44.2	13.7	no	0
CH (3)	78	24.3	13.1	63	14.3	9.5	yes	-10.0
MH (4)	5	40.8	23.4	44	26.0	14.9	no	0
CL (5)	1036	33.9	14.4	755	26.6	14.8	yes	-7.3
CL-ML (6)	634	47.3	14.7	407	45.5	13.4	yes	-1.8
ML (7)	905	60.9	13.5	816	59.7	12.9	no	0
SC (8)	170	38.0	17.6	91	36.3	18.5	no	0
GC (9)	184	47.3	17.1	93	37.7	19.0	yes	-9.6
SC-SM (10)	159	52.0	17.3	80	54.6	16.5	no	0
GC-GM (11)	114	58.9	13.7	60	61.8	13.4	no	0
SM (12)	596	67.1	13.2	430	65.5	14.3	no	0
GM (13)	354	69.0	13.2	170	72.0	10.8	yes	+3.0
SP-SC (14)	1	71.0		3	6.3	2.3	**	-30.0
SW-SC (15)	4	70.3	11.3	3	63.7	2.1	no	0
SP-SM (16)	56	74.1	5.7	62	73.4	15.3	no	0
SW-SM (17)	84	77.1	5.5	25	75.0	14.5	no	0
GP-GC (18)	8	63.3	8.3	22	66.0	19.5	no	0
GW-GC (19)	52	66.0	16.9	7	70.4	15.9	no	0
GP-GM (20)	32	77.6	6.3	87	79.3	4.7	no	0
GW-GM (21)	169	79.2	5.6	42	78.0	7.3	no	0
SP (22)	26	68.9	18.5	36	74.9	3.0	no	0
SW (23)	20	74.9	5.6	4	73.0	5.0	no	0
GP (24)	21	75.0	7.0	33	78.2	6.8	no	0
GW (25)	69	78.9	8.6	18	80.7	2.5	no	0

*Based on two-sample t-tests assuming unequal variances (Excel® tool)
**Insufficient Data for two-sample test (assume an adjustment of -30)

Believing that soil resistivity may be useful in predicting R-value, we then reviewed all the original files of the historical ITD soil testing data to search for resistivity values. Those soil testing records that contained a measured resistivity value were separated out and used to generate a new database, which was a subset of the previously established database. As these records were searched and grouped, we noted some additional duplication of a few soil testing records which had to be culled. Thus, the final two R-value databases to be used for multiple regression analysis included one for all soil testing records with basic soil property data (i.e., soil classification, Atterberg limits, percent finer than No. 200 sieve) as summarized in Table 3, and one for the soil testing records that also included measurements of soil resistivity as summarized in Table 4. A follow-up analysis at the end of the project also used a subset of the larger database (from Table 3) that included maximum dry density.

Table 3. Summary of ITD R-Value Database (1953-2008)

District	No. of R-value Records with Basic Soil Properties
1	428
2	346
3	2188
4	1117
5	2409
6	1745
Total	8233 (4808 for 1953 through June 1971) (3425 for July 1971 through 2008)*

*Of these Post-1971 data records, 3167 were obtained prior to 1997, meaning that in the past 12 years (1997-2008) only 258 R-value test results were available.

Table 4. Summary of ITD Data of R-value with Resistivity (1953-2008)

District	No. of R-value Records that Include Resistivity
1	---*
2	20*
3	459
4	170
5	848
6	210
Total	1707

*Resistivity data not recorded during initial review of District archives.

STATISTICAL ASSESSMENTS OF R-VALUE

The larger database of R-values was used to investigate direct relationships between soil properties and R-value. The distribution of R-values for the entire data set is shown by the histogram in Figure 1 and indicates that the largest proportion of data has R-values between 50 and 85. Additional histograms for the common soil types are presented in Figures 2 – 5.

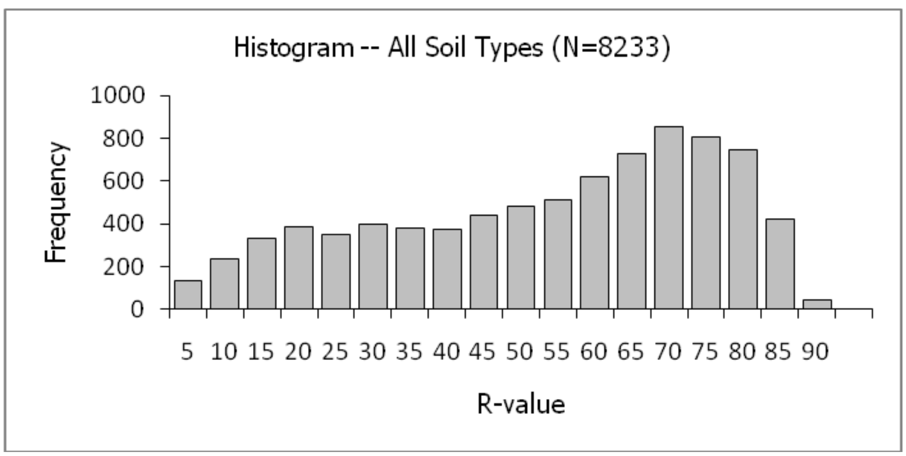


Figure 1. Histogram of R-values for All Soil Types.

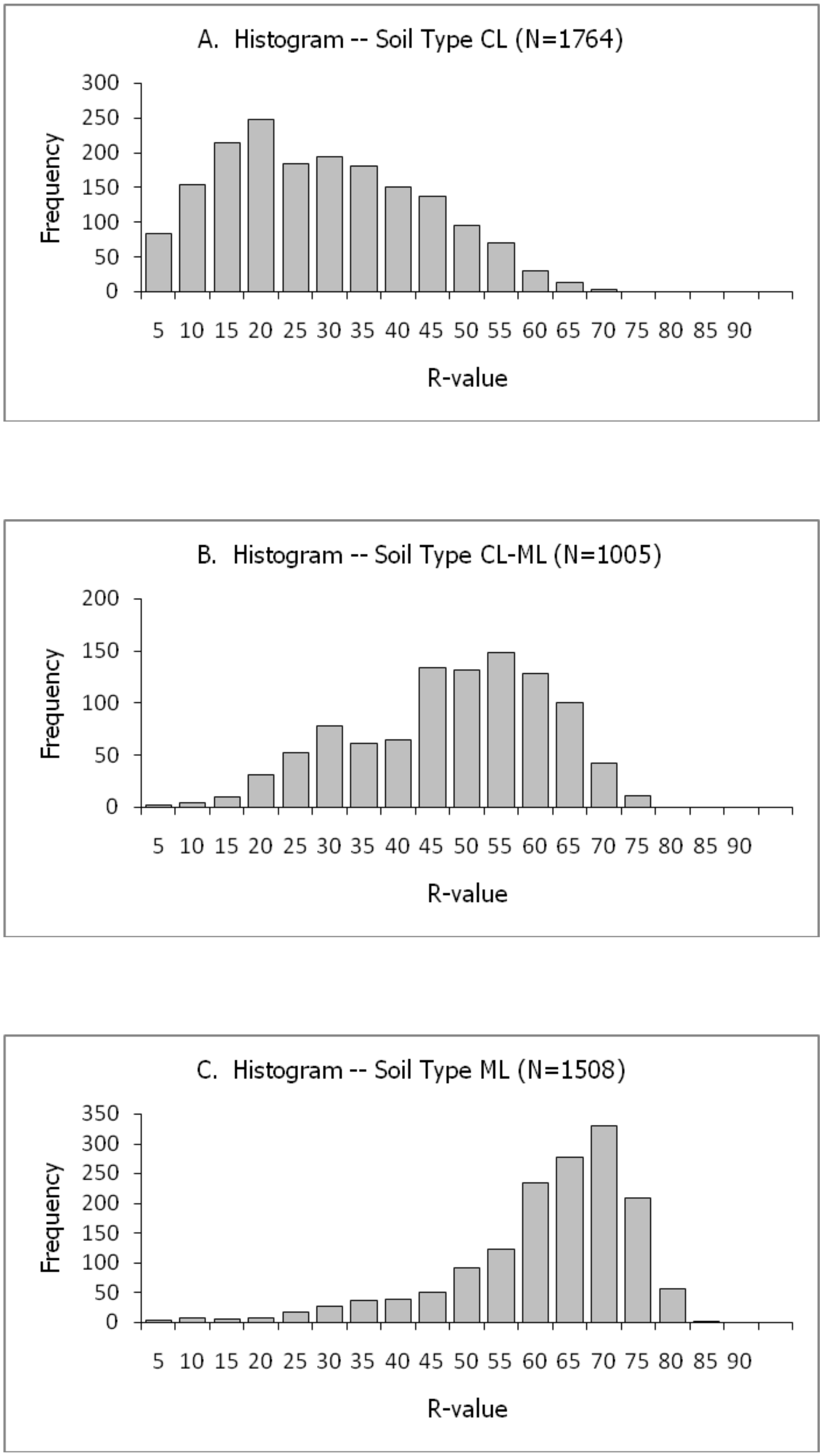


Figure 2. Histograms of R-values for Clay, Silty Clay, and Silt.

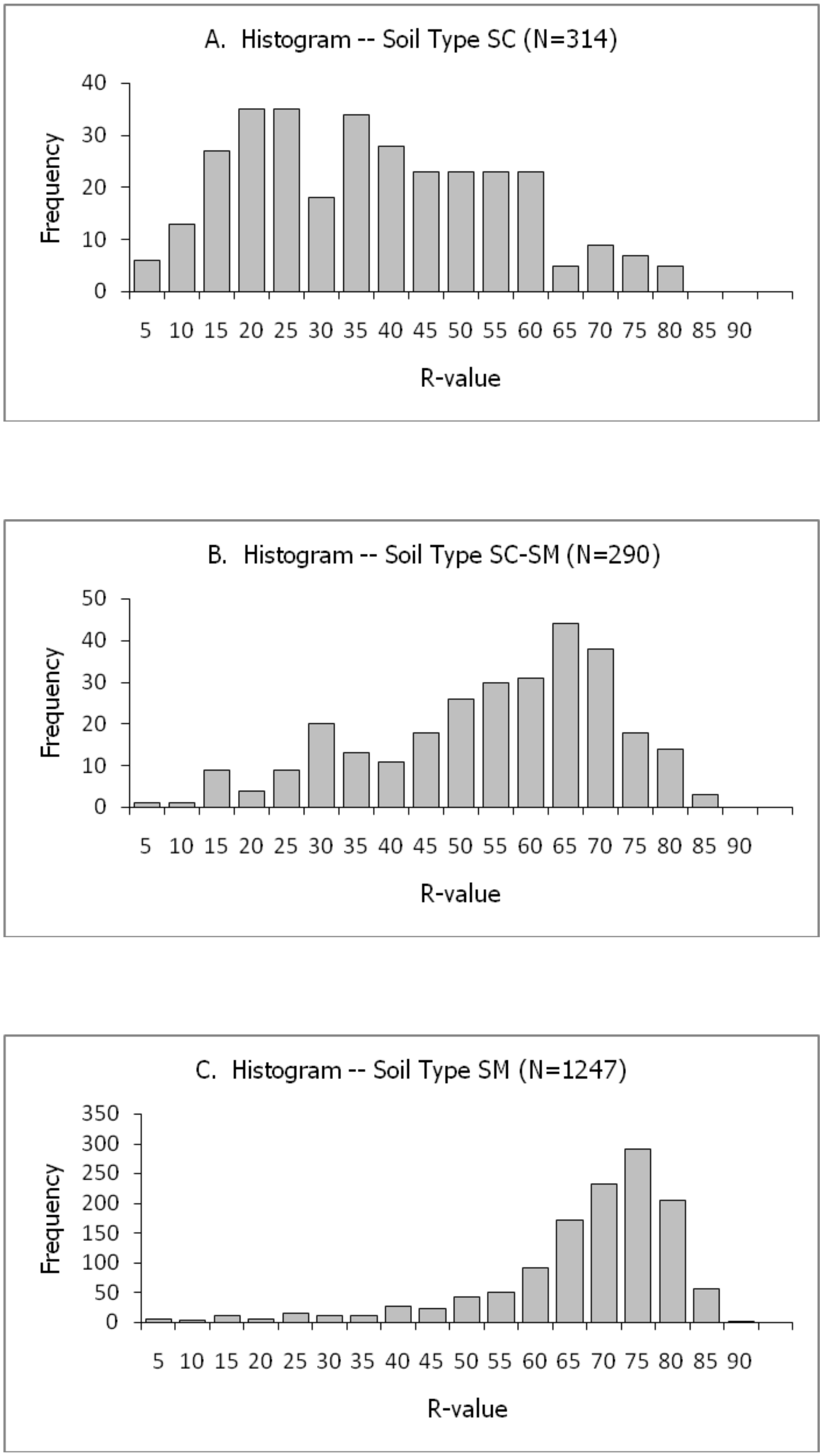


Figure 3. Histograms of R-values for Common Sandy Soils.

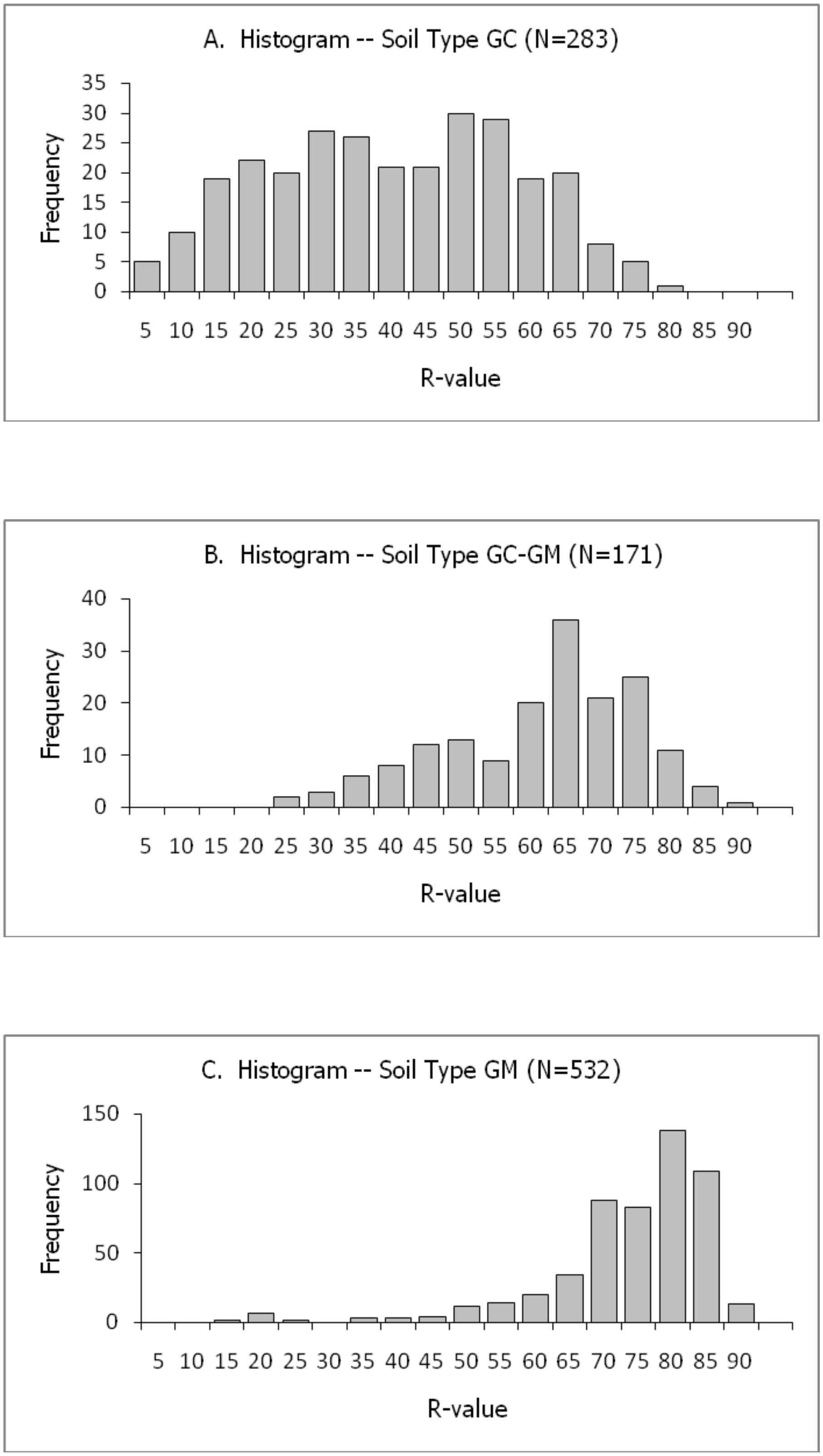


Figure 4. Histograms of R-values for Common Gravelly Soils.

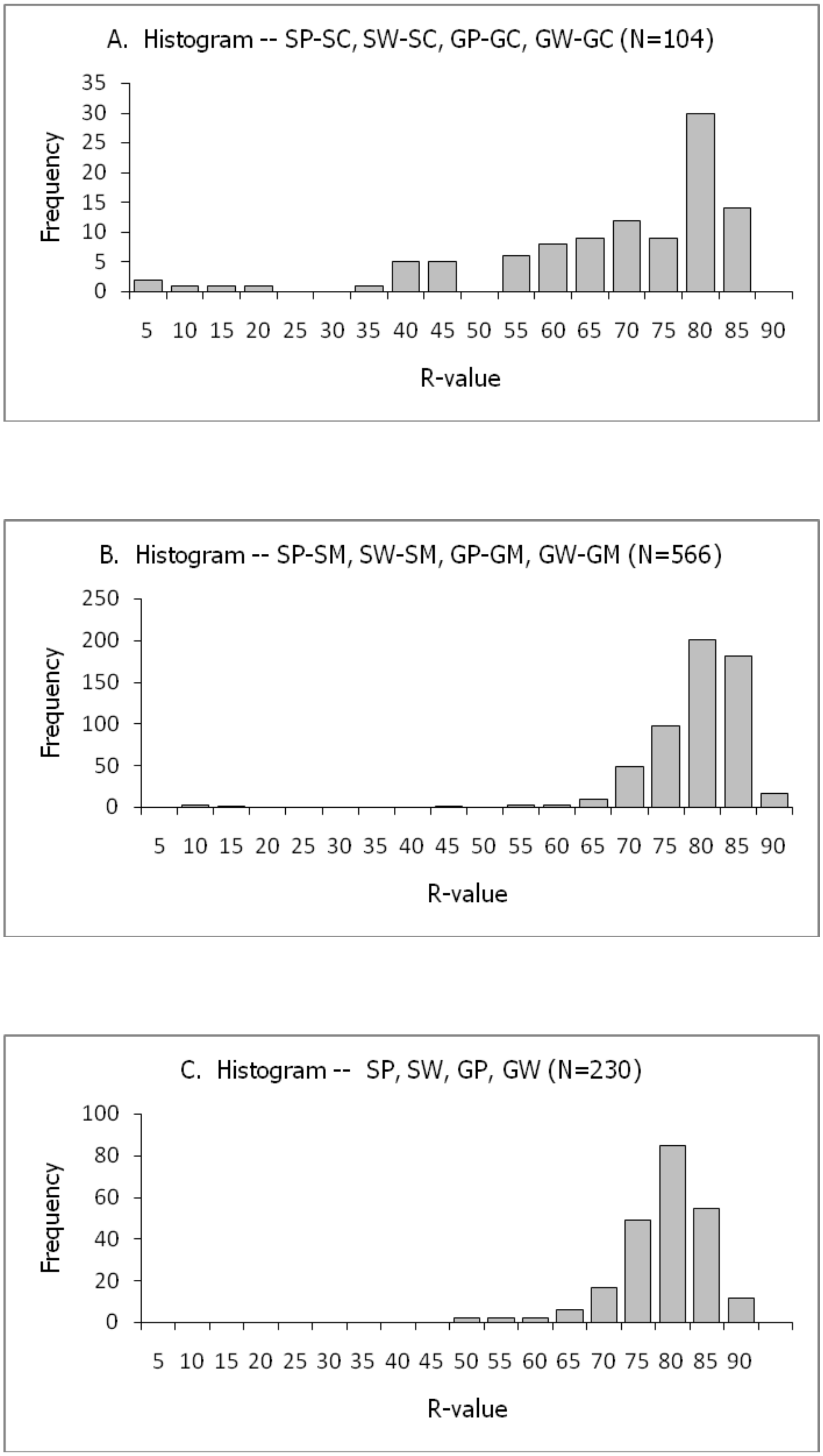


Figure 5. Histograms of R-values for Coarse-Grained Soils with ≤ 12 Percent Fines.

The distribution of R-values clearly shows some relation to the PF (percent fines) and the PI (plasticity index) of the soils, because they are used in the USC System to differentiate the soil classes. Figures 2, 3, and 4 show that the more plastic soils (i.e., the clayey soils) tend to have widespread R-values, but they also show higher R-values for coarser soils (i.e., less percent fines). Figure 5 shows that the coarsest soils generally have high R-values, but they can have low R-values if they contain small amounts of plastic (clayey) fines.

The relationships between R-value and other soil properties are illustrated using scatterplots in Figures 6 – 10. Although R-value generally is proportional to USC code, resistivity, and maximum dry density (units of pcf), and it is inversely proportional to percent fines (PF) and to plasticity index (PI), all the plots show considerable scatter especially in regard to the fine-grained soils that have high percent fines. General observations include the following: 1) coarse-grained soils with 12 percent fines or less (USC codes 14 – 25) typically have R-values greater than 40; 2) soils with PI's greater than 50 generally have R-values less than 20; 3) nonplastic and low-plasticity soils have R-values spread across a wide range; 4) soils with resistivity exceeding 8,000 ohm-cm almost always have R-values greater than 60.

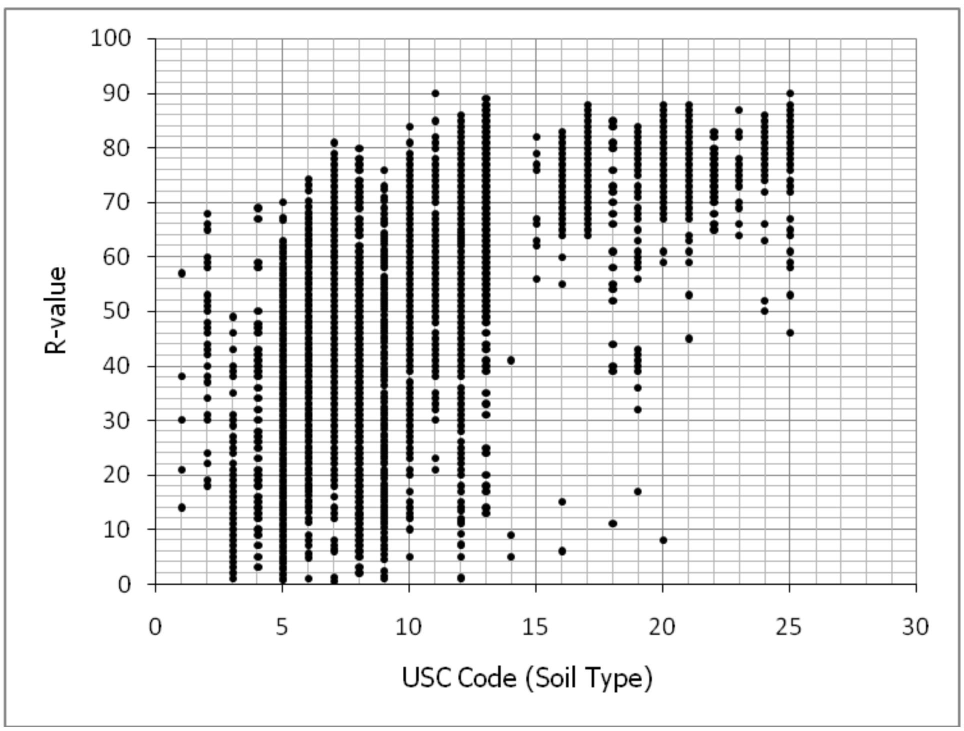


Figure 6. Scatterplot of R-value as a Function of USC Code (N=8233).

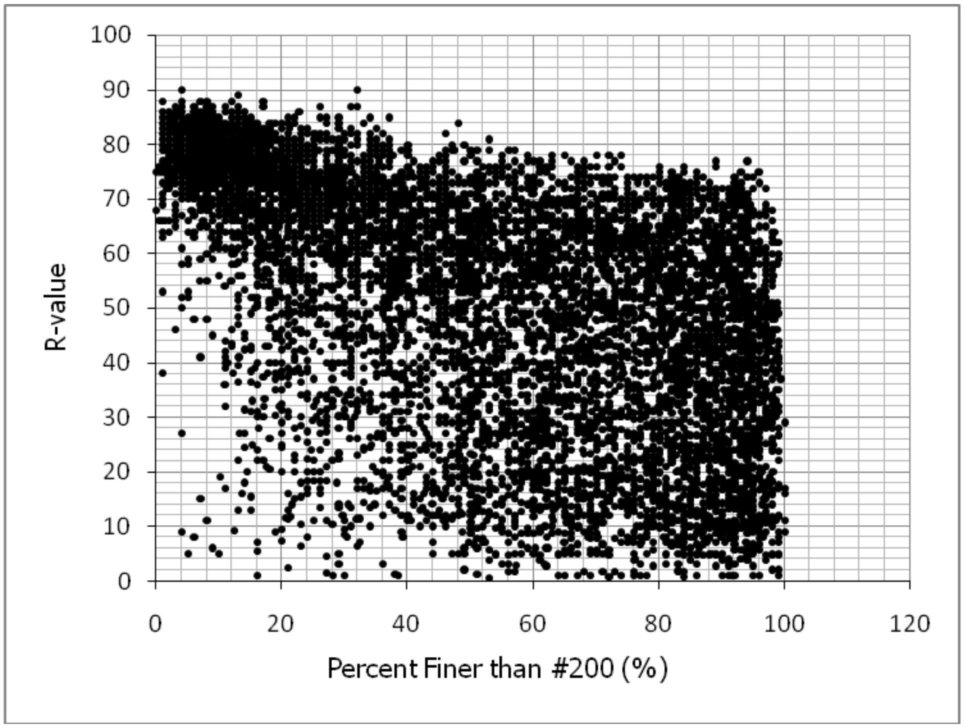


Figure 7. Scatterplot of R-value as a Function of Percent Fines (N=8233).

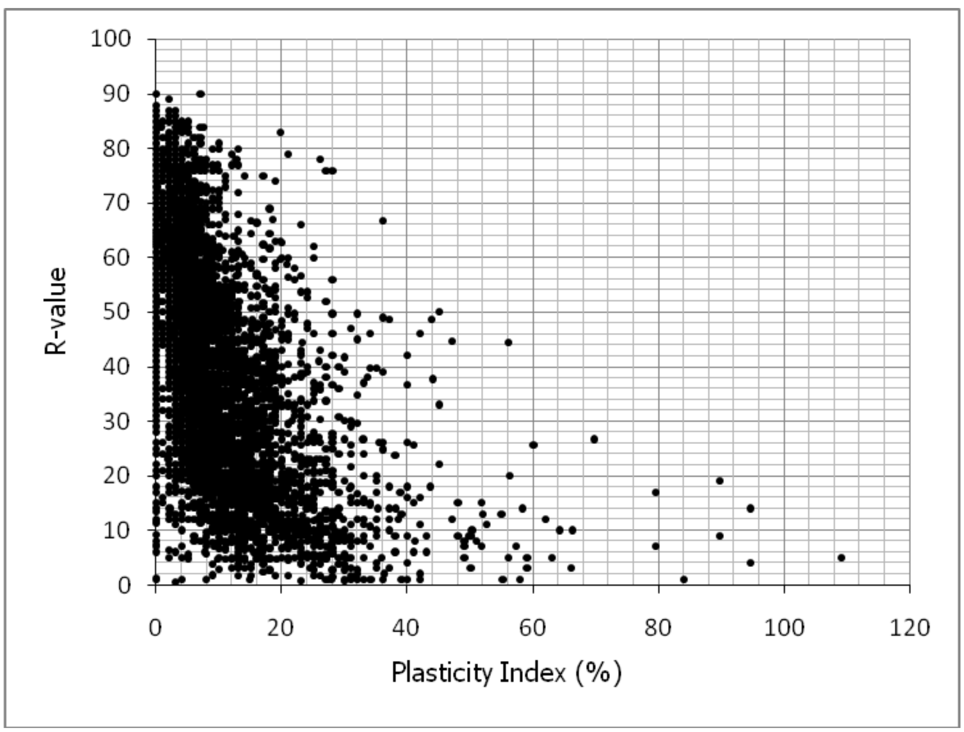
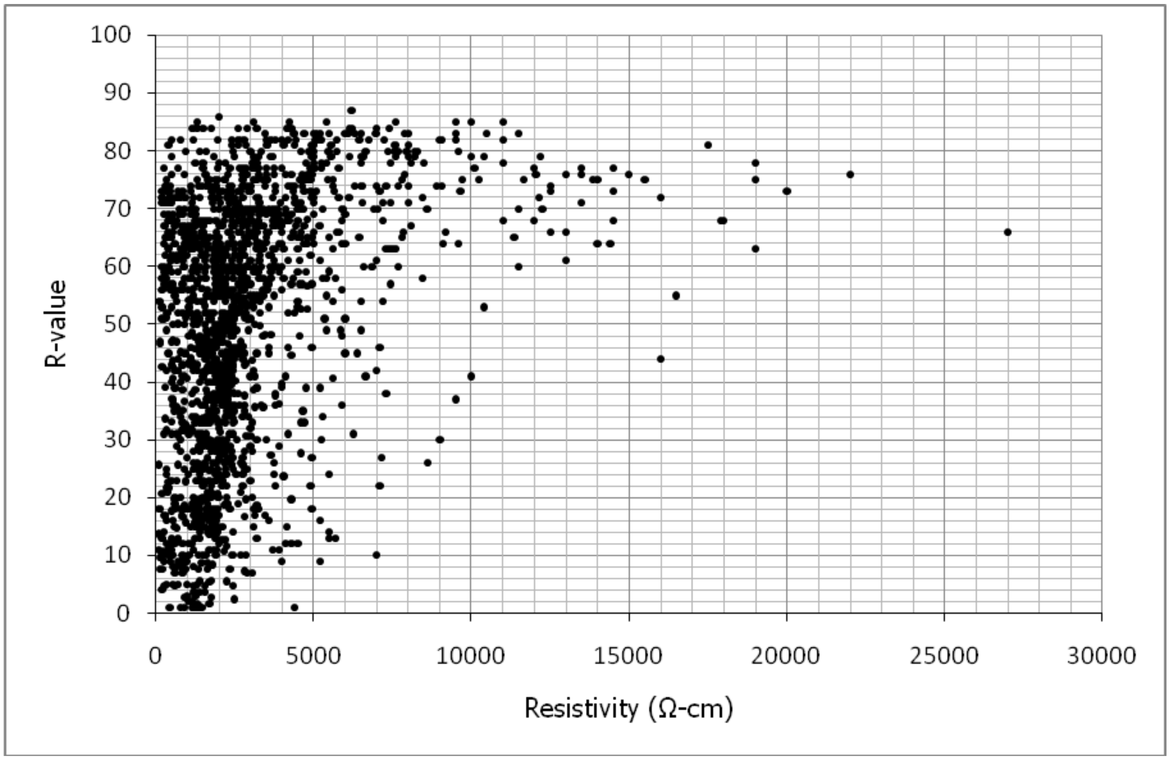
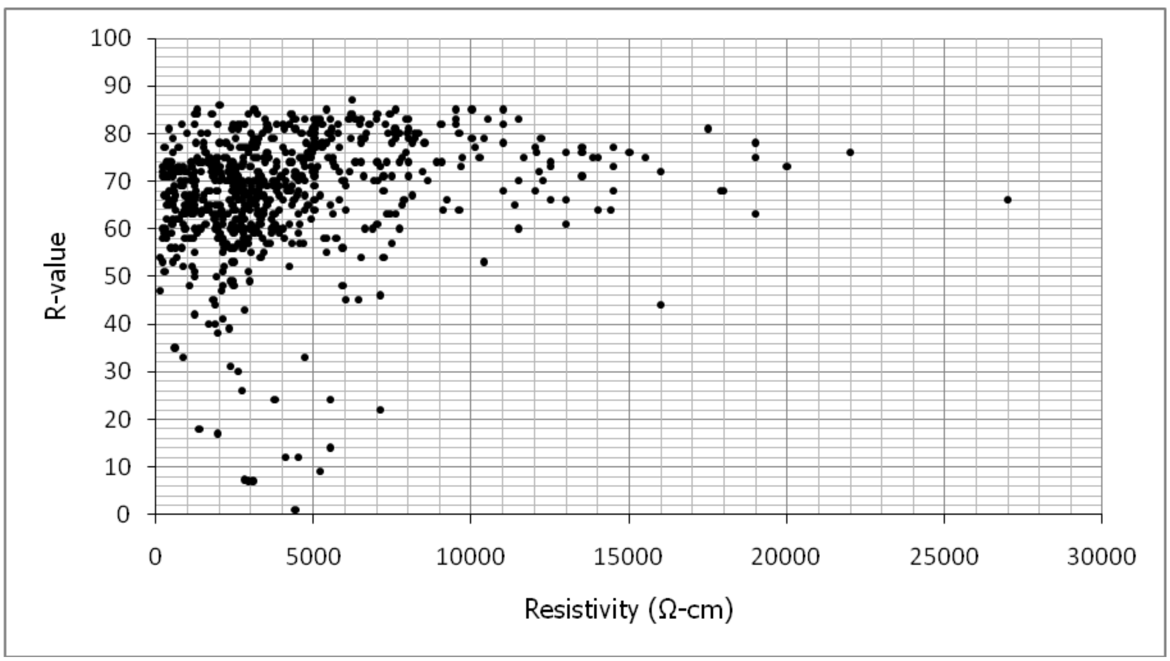


Figure 8. Scatterplot of R-value as a Function of Plasticity Index (N=8233).



A. Data for All Soils (N=1707).



B. Data for Nonplastic Soils (N=685).

Figure 9. Scatterplots of R-value as a Function of Soil Resistivity.

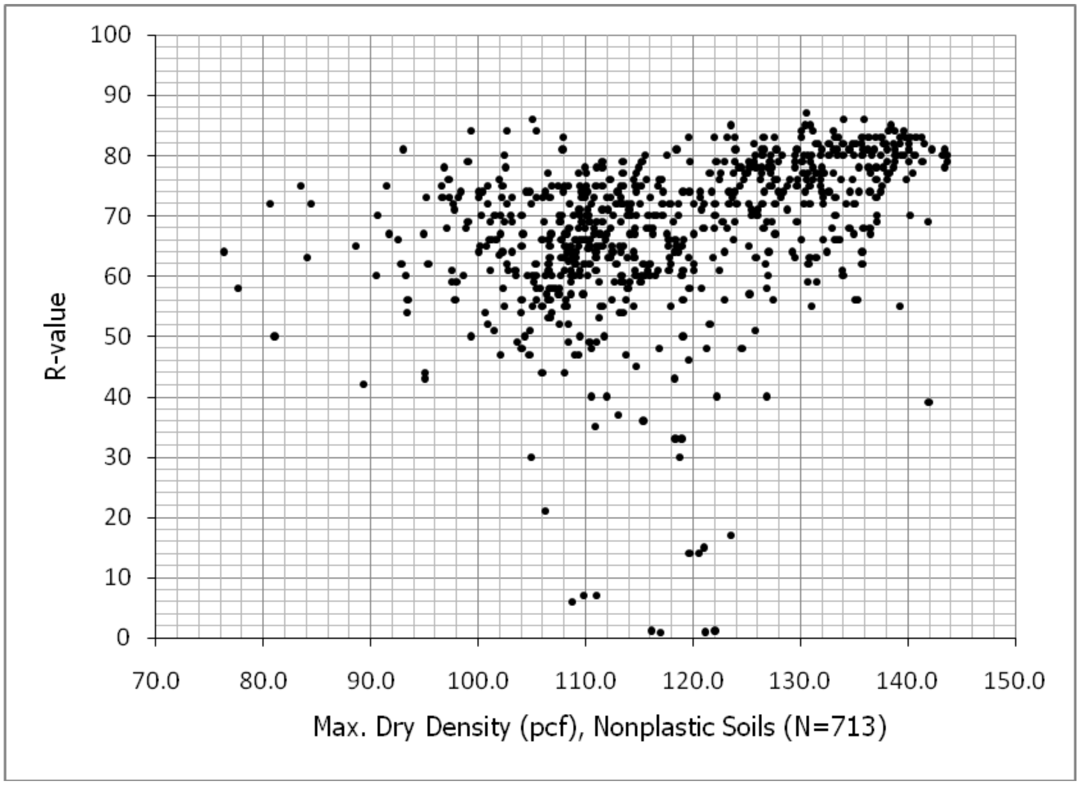


Figure 10. Scatterplot of R-value as a Function of Maximum Dry Density for Nonplastic Soils (Database 1978-2008).

The distribution of soil types in the R-value database is summarized by District in Table 5. This information indicates that the largest portion (nearly 70 percent) of the testing was conducted on fine-grained soils and silty sand soils (SM). Based on the R-value histograms (Figures 1 – 5), it appears that the higher percentage of R-values in the 50 to 85 range is due to the large numbers of low-plasticity, fine-grained soil samples (CL-ML, ML) along with the silty sands (SM) and other coarse-grained soils, all of which tend to have large proportions of high R-values. To be more specific, the percentage of samples with R-values of 50 or greater is 45 percent for CL-ML soils, 83 percent for ML soils, and 89 percent for SM soils.

Table 5. Distribution of Soil Types in the Database by District (Values are Approximate Percentages of the Database Totals)

District	CL	ML	CL-ML	Other Fine Soils	SC	SM	SC-SM	GC	GM	GC-GM	Other Coarse Soils
1	18	21	7	2	2	17	3	3	9	< 1	18
2	32	8	6	18	8	17	1	3	3	1	3
3	20	15	9	4	6	23	5	2	7	1	8
4	16	35	17	< 1	3	13	2	1	6	< 1	5
5	27	18	14	2	2	8	2	6	6	3	11
6	17	14	12	< 1	4	16	5	4	7	3	18
All	21	18	12	3	4	15	4	4	6	2	11

MULTIPLE REGRESSION MODELING

Initial multiple regression modeling of the entire R-value database with Excel® included four soil attributes as the independent variables used to predict the dependent variable, R-value. These four attributes were USC classification code (i.e., assigned a numerical code from 1 through 25; see Table 2), liquid limit (LL), plasticity index (PI), and percent fines (PF) which is the percent by weight passing the No. 200 sieve. After applying several different multiple regression models, it became clear that liquid limit was not adding any significant information to the regression models, so it was dropped from further consideration. The remaining three soil attributes all appeared to be adding significance, with PI and USC code having a greater influence than PF.

Because the Excel Data Analysis module with the multiple regression option requires that all data fields have numerical entries (i.e., no blank spaces), only those historical R-value samples with accompanying test results for USC code, PI, and PF were used in the multiple regression model building process. This resulted in fewer R-values (as compared to the original data sets) being used in the regression models; the total number of records in the complete database for regression modeling was 8,233 (see Table 3.)

A separate database was generated using only those soil testing records that contained measured values of soil resistivity. We did not include any records with reported resistivity values of less than 100 Ω-cm or greater than 27,000 Ω-cm.

The following two regression models were used in the initial trials for the R-value data sets. The first was a simple linear combination of USC code, PI, and PF200. The second was based on a $\log_{10}$ model proposed by the Arizona DOT in a Materials Design Manual from January 1985 (current availability unknown). Note that the b_i terms are the regression coefficients.

$$Rval = b_0 + b_1(USC) + b_2(PI) + b_3(PF)$$

Equation 1. Initial Multiple Regression Model, Type 1.

$$\log_{10}(Rval) = b_0 + b_1(PI) + b_2(PF)$$

$$\text{or} \quad Rval = 10^{[b0 + b1(PI) + b2(PF)]}$$

Equation 2. Initial Multiple Regression Model, Type 2.

For initial trials conducted on the data sets from Districts 1 through 3, the first model clearly out-performed the second one. Therefore, various new combinations of the three soil terms (e.g., products and quotients of any two terms, square roots and cube roots of any of the terms) were investigated. After many trials and evaluations of different combinations of soil terms, the multiple regression model shown below as Equation 3 provided the greatest value of R^2 (multiple coefficient of determination), and it consistently out-performed other regression models with their various combinations and cross terms. The closer R^2 is to 1.0, the better the regression model.

$$Rval = b_0 + b_1\,(USC) + b_2\,(PI) + b_3\,(PI \times PF)^{0.333}$$

$$\text{or} \quad Rval = b_0 + b_1\,(USC) + b_2\,(PI) + b_3 \cdot \sqrt[3]{(PI \times PF)}$$

Equation 3. Preferred Multiple Regression Model for All Soils Combined.

Typical results for trial multiple regression models are summarized in Table A1 in the Appendix. The cube-root term clearly improved the regressions results, though we have no mathematical explanation why it should be better than a square-root or logarithmic term, both of which also were tried in the regression modeling. Overall, even the best of the regression models did not perform exceedingly well, with R^2 values near 0.70. The regression models for District 4 were especially poor with R^2 values less than 0.50, which perhaps can be explained by the unusually high proportion of ML soils tested in that District (see Table 5), causing more statistical variability in the R-values.

Because the better-quality regression models seemed to depend strongly on the PI, the effectiveness of applying such models to nonplastic soils may by questionable. Thus, we turned to the database with soil resistivity values to investigate other regression models perhaps more suited to nonplastic and low-plasticity soils. Unfortunately, even the results from this modeling did not provide R^2 values greater than 0.23 (Table A4 in the Appendix). When the nonplastic soils were analyzed separately from the larger database (a new subsample of 685 records), results of the regression modeling were poor, with R^2 values near 0.20 (Table A4).

Assuming the statistical "noise" in the multiple regression analysis may be due to the large proportion of fine-grained soils and silty sands, we also investigated a reduced database that contained subsamples from these soil types to bring their proportions more in line with other soil types. The highest R^2 obtained from several different models of this subsample was approximately 0.62, which was little or no improvement over previous models.

Thus, using the three variables, USC, PI, and PF, the recommended regression models for predicting R-value are given below. Equation 4 is for all the statewide data combined. Equation 5 is based on the best performing model obtained by combining R-value data from Districts 1 and 5.

The best performing regression model using the resistivity database (exclusive of the PI information) was based on three variables: USC, PF, and Resistivity (Res). It is given by Equation 6. If PI is included as a variable, the regression model is improved significantly (Equation 7). However, if only data from the nonplastic soils in this database are used in a regression model, the R^2 value is quite low (Equation 8). If maximum dry density (per AASHTO T 99) is used as an independent variable instead of resistivity, the regression results are slightly improved for nonplastic soils (Equation 9 and Table A5).

$$Rval = 55.91 + 1.10(USC) - 0.41(PI) - 2.49[\sqrt[3]{PI \times PF}] \qquad (R^2 = 0.6353)$$

Equation 4. Three-Parameter R-value Regression Model from Statewide Data.

$$Rval = 57.35 + 1.11(USC) - 0.86(PI) - 1.98[\sqrt[3]{PI \times PF}] \qquad (R^2 = 0.6968)$$

Equation 5. Three-Parameter R-value Regression Model from Districts 1 and 5.

$$Rval = 20.15 + 2.27(USC) + 0.51(PF) - 2.68(PF/USC) + 0.48[\sqrt[3]{Res}] \qquad (R^2 = 0.4965)$$

Equation 6. Three-Parameter R-value Regression Model Using Resistivity Database.

$$Rval = 51.38 + 1.53(USC) - 0.05(PF) - 0.21[\sqrt[3]{Res}] - 1.32(PI) \qquad (R^2 = 0.6279)$$

Equation 7. Four-Parameter R-value Regression Model Using Resistivity Database.

$$Rval = 64.60 + 0.78(USC) - 0.15(PF) + 0.51(PF/USC) - 0.18[\sqrt[3]{Res}] \qquad (R^2 = 0.2064)$$

Equation 8. Three-Parameter R-value Regression Model for Nonplastic Soils.

$$Rval = 63.95 + 0.54(USC) - 0.31(PF) + 1.00(PF/USC) + 0.03\gamma_{dmax} \qquad (R^2 = 0.3160)$$

Equation 9. Three-Parameter R-value Regression Model for Nonplastic Soils Based on Maximum Dry Density Database.

We tried to improve the performance of Equations 4 and 5 by removing the testing records for nonplastic soils and working only with the data for plastic soils (PI > 0). The resulting R^2 values were 0.4773 and 0.5390, respectively (Table A2), indicating the nonplastic soils actually do contribute significantly to the overall regression models for combined soil types.

For the nonplastic soils, we compared the prediction quality of Equations 8 and 9 by computing R-value regression errors based on the 685 data records in the resistivity data-base and the 713 records in the maximum dry density database, respectively. Histograms of these errors are shown in Figure 11, indicating that both of these regression models produce

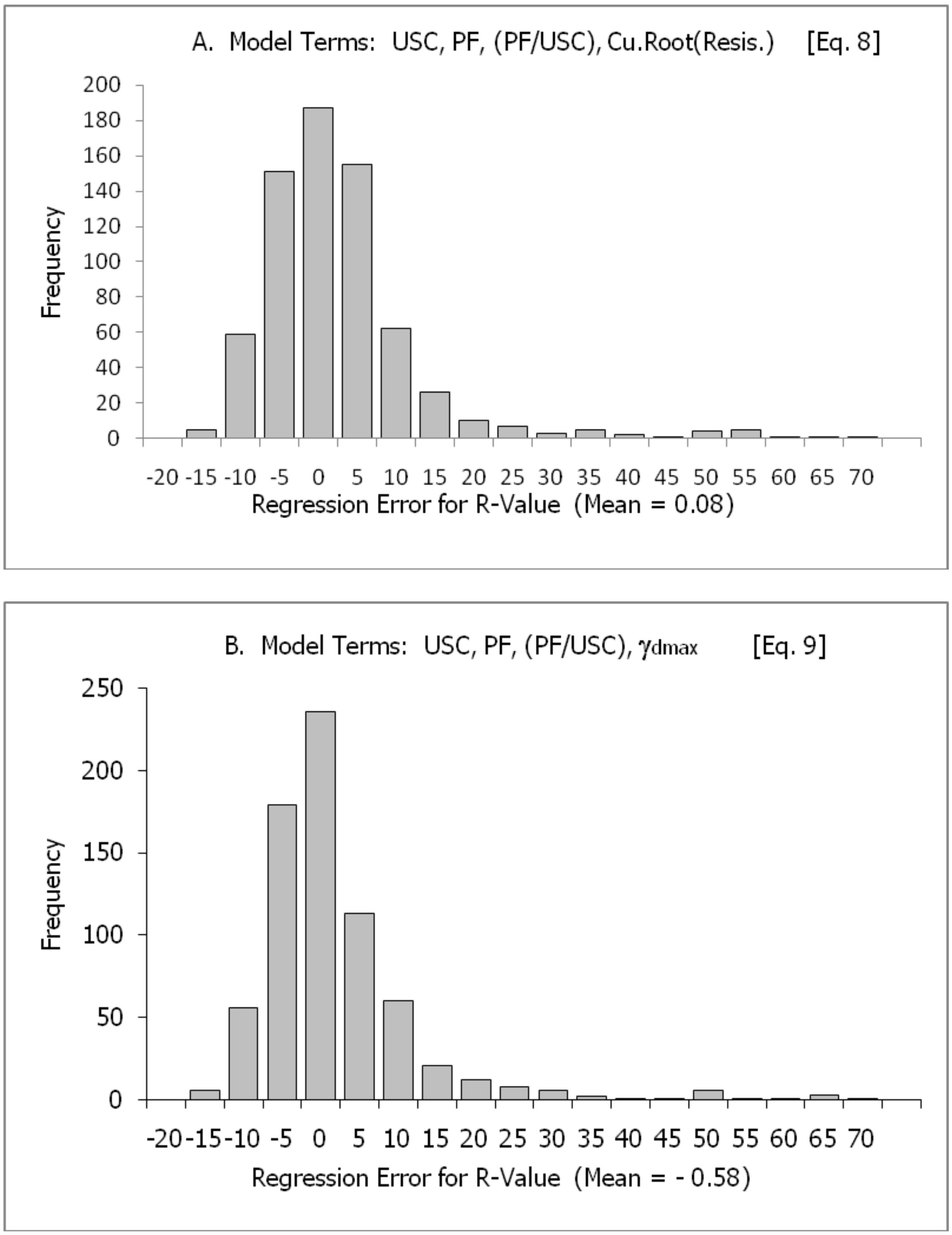


Figure 11. Histograms of Errors for R-value Regression Models for Nonplastic Soils.

similar spreads in the errors, but the regression model based on resistivity (Equation 8) tends to overestimate R-values while the regression model based on maximum dry density (Equation 9) tends to underestimate R-values. Another possible model for nonplastic soils is Equation 5 (based on all soil types), but regression errors based on this model have a greater spread than those for the other two previous models, and the mean error is 3.07 (significant overestimation). Another option for only nonplastic soils is Equation 6, but regression errors based on this model have an even greater spread than the previous models, and the mean error is -5.77.

DISTRIBUTION QUANTILES FOR ESTIMATING MINIMUM R-VALUES

The R-value database can provide other statistical information to help estimate minimum R-values for conservative initial design of pavement subgrades. R-value distributions by soil type already have been presented in histograms, and these can provide useful information on the most likely R-values to be expected from given soil types.

In addition, quantile values (also known as percentile values) of R-value data distributions can provide a rational basis for selecting reasonable, conservative R-values for initial pavement design. For example, a conservative pavement design might involve the use of a relatively low quantile value from the R-value distribution for the specified soil type. For example, the 0.05 quantile value is that R-value at which 5 percent of the available R-value data are less than or equal to it. Selected quantile values (abbreviated by Q) are reported in Table 6, according to the most common soil types.

The quantile values for a given soil type can provide guidance in predicting initial R-values for use in the preliminary design of pavement subgrade or subbase. For example, if an engineer is dealing with a CL-ML soil, a very conservative estimate of the R-value would be 16 (i.e., the 0.02Q value), a conservative estimate would be 20 (the 0.05Q value), and a slightly conservative estimate would be 25 (the 0.10Q value).

At this stage, until practitioners gain experience in applying the R-value regression models and statistical information to the preliminary design of pavement sections, it is suggested that one option be to use the 0.05Q value for silty soils and the 0.10Q value for other soils.

Table 6. R-Value Quantiles for Common Soil Types in the Statewide Database.

Soil Type	0.02Q	0.05Q	0.10Q	0.25Q
CH	1	2	3	7
MH	5	8	10	14
CL	4	6	9	15
CL-ML	16	20	25	36
ML	24	33	43	55
SC	6	10	13	21
SC-SM	13	21	28	43
SM	22	39	49	62
SP-SC, SW-SC	5	6	8	43
SP-SM	56	65	69	72
SW-SM	66	68	71	74
SP	66	69	70	73
SW	65	67	69	73
GC	6	10	14	25
GC-GM	31	34	40	51
GM	31	50	58	68
GP-GC, GW-GC	31	39	42	59
GP-GM	64	69	71	77
GW-GM	63	68	71	77
GP	69	73	75	77
GW	57	63	72	79

CONCLUSIONS

Historical soil testing information has been collected from archived documents in all six ITD Districts. The data records were digitally recorded and stored using Excel® files. Then, these digital records were carefully inspected, cleaned, and culled to provide complete data sets with R-value, USC soil class codes, plasticity index (PI), and percent fines (PF) for the fraction passing the No. 200 sieve. Those original records that included soil resistivity results also were identified, so a subsample of the large database could be generated that would include records with all the typical soil properties plus the resistivity data. Some of the R-value test results dating to Pre-1971 (primarily for the clayey soil types) were adjusted to Post-1971 conditions in order to account for changes in the ITD R-value testing procedures that were implemented in June of 1971.

R-value histograms and scatterplots by soil type were generated to investigate the distributions of R-value and its relationship to other soil properties. Quantile values of the R-value for the common soil types also were computed to provide guidance in estimating conservative initial R-values, given a specific soil type.

Multiple regression models then were developed to predict R-value, given three input variables: USC, PI, and PF. Even with considerable scatter in the data, the regression models for all of the Districts except for District 4, showed reasonable significance (merit) with R^2 values on the order of 0.61 to 0.70 (coefficient of multiple determination). Regression results for the nonplastic soils using typical soil properties and resistivity were poor, with the best R^2 values being less than 0.25. Slightly better results were obtained using maximum dry density instead of resistivity ($R^2 = 0.3160$).

Though some practitioners may be tempted to use the R-value summary information in the histograms or scatterplots to forego any soil testing at all for preliminary pavement design, due to the scatter in much of this historical data (i.e., the high standard deviations observed in R-value for most of the soil types), we recommend that site-specific basic soil testing be conducted to provide Atterberg limits (LL and PI) and particle-size distribution curves to accurately classify each soil using ASTM D 2487. Then, the appropriate multiple regression model can be applied or a conservative quantile value used to generate an estimate for the R-value to be considered for the preliminary design of pavement sections. Based on known dispersions of R-values for different soil types, such estimates for fine-grained soils and clayey coarse-grained soils will have less reliability than those estimates for other coarse-grained soils. Roughly speaking, R-value estimates in the 5-40 range would likely have errors of 15-20 percent, while estimates in the 60-90 range would have errors of 5-10 percent; mid-range R-values (40 to 60) would have errors of approximately 10-15 percent.

If several samples of the same soil type have been tested, then a range of predicted R-values can be obtained readily using the selected regression model. The use of such regression models, which are based on over 50 years of ITD testing data, certainly provides the basis for a rational, objective, and defensible way for geotechnical and materials engineers to develop and apply R-value estimates for pavement design.

Predictive regression models for R-value and/or estimated values using soil-type-specific quantile values for R-value can be quite useful in the preliminary (or ITD "Phase I Tentative") ballast section design of highway pavements. Such models also could be used for final design of pavements in low-volume traffic zones, provided that appropriate soil classification testing is performed on sufficient numbers of subgrade or subbase samples.

One of the greatest outcomes provided by this project is the database itself, which will serve as a valuable resource for ITD personnel and geotechnical consultants working with ITD. Besides having Excel® files with basic soil property data and R-values, the original raw files also contain some information on compaction testing (maximum dry density and optimum moisture content) and linear shrinkage.

ACKNOWLEDGEMENTS

This project could not have been completed without the valuable assistance of ITD personnel in each of the six Districts who provided access and guidance to the historical soil testing data in their office archives; we very much appreciated their patience and help. In addition, we are grateful for the technical review comments received from Stanley Crawforth, P.E.

We also extend our thanks to Mr. Jonathan Rush, for his time-intensive and painstaking work to collect all these data at the District offices, and to Ms. Rimi Kim, for her careful and efficient data processing work on the Excel® files.

APPENDIX A

Guide for R-value Multiple Regression Modeling Using the Excel® Data Analysis Option

Excel file formatting

The cleaned and culled R-value data files used for the multiple regression analysis are listed below. A brief description of the color coding for cells in the spreadsheets also is given.

Dist1RvalMay09.xlsx
Dist2RvalMay09.xlsx
Dist3RvalMay09.xlsx
Dist4RvalMay09.xlsx
Dist5RvalMay09.xlsx
Dist6RvalMay09.xlsx
DistAllRvalMay09Total.xlsx (all Districts combined)
DistAllRvalMay09Total_histog.xlsx (histograms for all Districts combined)
DistAllRvalResistivTotal.xlsx (all District soil records that have resistivity data)
DistAllRvalNPdrydens.xlsx (all District nonplastic soil records with γ_{dmax}; 1978-2008)
DistALLRval_t-tests.xlsx (all Districts combined; results of two-sample t-tests for each of the 25 soil types is reported)

Color codes

Purple: Individual purple cells indicate a value was edited during data processing.
Red: Column heading H: Red indicates the final adjusted R-values.
Yellow: Column headings I-M: Yellow indicates data used for scatter plots or reference information.
Green:Column headings N-Q: Green indicates data used for regression analysis.
Blue: Column headings R-T: Blue indicates preservation of original data and the R-value adjustment based on t-test results.
Orange: Column heading U: Orange indicates District Number (this was used specifically when all Districts were combined into one file).

Regression Analysis

Prior to conducting the regression analysis in Excel®, the user must first be sure that the Data Analysis “ToolPak” (module) has been installed within Excel®. Directions are as follows:

The Analysis ToolPak is a Microsoft Office Excel add-in program that is available when you install Microsoft Office or Excel. To use it in Excel, you need to load it first.

Click the Microsoft Office Button, and then click Excel Options.
Click Add-Ins, and then in the Manage box, select Excel Add-ins.
Click Go.

In the Add-Ins available box, select the Analysis ToolPak check box, and then click OK. *Tip* If Analysis ToolPak is not listed in the Add-Ins available box, click Browse to locate it. If you get prompted that the Analysis ToolPak is not currently installed on your computer, click Yes to install it.After you load the Analysis ToolPak, the Data Analysis command is available in the Analysis group on the Data tab.

When you have opened the Excel file of choice, select the Data tab, then click on the Data Analysis option (usually located at the right-hand side of the listed options). Then, select the Regression option from the Analysis Tools list, and click the OK button. This will display a new dialog box where you will specify the data you want to analyze.

Input Y Range – Here is where you enter the data for your dependent variable (R-value). You can click on the little spreadsheet icon, then select the desired column within your spreadsheet (Col. H for the Adjusted R-values), or you can type in the window the following: $H2:$H410

Example (for District 1, which has 409 records in the spreadsheet):

Input X Range – Here is where you enter the data for your independent variable(s). If you only want one variable (say, PI), then repeat the process used for selecting the Y range, except select the proper column ($O2:$O410) for the PI data.

Important: When you want two or more X variables, you have to be sure they are stacked in adjacent columns and have the same number of rows as the Y variable. For the three variables with green column headings, type in the window the following: $N2:$P410. (This will include the data from all three columns, N, O, and P).

If you want to modify the X variables and use other combinations, be sure to stack the new variables in adjacent columns prior to using this regression analysis.

After specifying the Y and X ranges, leave all other options at their default values, and then click the OK button. The regression results/statistics will be displayed in a new tab at the bottom of the spreadsheet.

Note: If the following message is displayed, a blank cell occurred in your data range. You will need to locate it and either delete that record from the database or enter a value in that cell, such as zero, if appropriate:

LINEST() function error. Please check input ranges again.

APPENDIX B

Reference Tables for Multiple Regression Results

Table A1. Coefficients of Multiple Regression Models for Estimating R-value Using All Relevant Soil Testing Records.

District	N	b_o	b_i(USC)	b_i(PF)	b_i(PI)	$b_i\sqrt[3]{(PI \bullet PF)}$	R^2
1	428	74.33	0.17	-0.27	-1.98	~~	0.6763
	428	**57.62**	**0.92**	~~	**-0.51**	**-2.99**	**0.6923**
2	346	60.96	0.52	-0.24	-1.02	~~	0.5886
	346	**57.99**	**0.43**	~~	**-0.18**	**-2.96**	**0.6246**
3	2188	45.52	1.83	-0.12	-0.87	~~	0.5559
	2188	**52.09**	**1.32**	~~	**-0.11**	**-2.78**	**0.6118**
4	1117	52.84	1.31	-0.04	-1.45	~~	0.4057
	1117	**59.03**	**0.85**	~~	**-0.34**	**-2.36**	**0.4636**
5	2409	59.05	1.10	-0.10	-1.69	~~	0.6839
	2409	**57.32**	**1.61**	~~	**-0.90**	**-1.89**	**0.7039**
6	1745	53.60	1.21	-0.09	-1.72	~~	0.6491
	1745	**54.66**	**1.12**	~~	**-0.83**	**-2.10**	**0.6719**
All	8233	51.12	1.42	-0.09	-1.29	~~	0.5879
	8233	**55.91**	**1.10**	~~	**-0.41**	**-2.49**	**0.6353**
1&5	**2837**	**57.35**	**1.11**	~~	**-0.86**	**-1.98**	**0.6968**

Table A2. Coefficients of Multiple Regression Models for Estimating R-value Using Only the Testing Records of Plastic Soils (PI>0).

District	N	b_o	b_i(USC)	b_i(PF)	b_i(PI)	$b_i\sqrt[3]{(PI \bullet PF)}$	R^2
All	4952	67.78	1.05	0.21	0.45	-6.64	0.4773
1&5	1810	66.00	1.11	0.05	-0.43	-3.93	0.5390

Table A3. Coefficients of Multiple Regression Models for Estimating R-value Using Soil Testing Records that Include Resistivity.

N	b_o	b_i(USC)	b_i(PF)	b_i (PF/USC)	$b_i \sqrt[3]{(\text{Resis.})}$	b_i(PI)	R^2
1707	21.26	3.05	0.04	~~	~~	~~	0.3835
1707	14.12	2.88	0.05	~~	0.59	~~	0.3900
1707	25.96	2.40	0.51	-2.70	~~	~~	0.4923
1707	20.15	2.27	0.51	-2.68	0.48	~~	0.4965
1707	51.38	1.53	-0.05	~~	-0.21	-1.32	0.6279

Table A4. Coefficients of Multiple Regression Models for Estimating R-value for Nonplastic and Low-Plasticity Soils (0<PI<4) for Records that Include Resistivity.

Soils	N	b_o	b_i(USC)	b_i(PF)	b_i (PF/USC)	$b_i \sqrt[3]{(\text{Resis.})}$	R^2
NonPlas.	685	61.34	0.74	-0.06	~~	~~	0.2036
NonPlas.	685	61.96	0.74	-0.14	0.48	~~	0.2048
NonPlas.	685	64.60	0.78	-0.15	0.51	-0.18	0.2059
NonPlas.+ Low Plas.	800	61.45	0.74	-0.08	~~	~~	0.2228
NonPlas.+ Low Plas.	800	62.25	0.74	-0.18	0.60	~~	0.2251
NonPlas.+ Low Plas.	800	64.83	0.78	-0.19	0.64	-0.18	0.2266

Table A5. Coefficients of Multiple Regression Models for Estimating R-value for Nonplastic Soils using Records that Include Maximum Dry Density (pcf).

N	b_o	b_i(USC)	b_i(PF)	b_i (PF/USC)	$b_i (\gamma_{dmax})$	R^2
713	62.13	0.54	-0.15	~~	0.04	0.3086
713	63.95	0.54	-0.31	1.00	0.03	0.3160