International Journal of GEOMATE, Aug., 2018 Vol.15, Issue 48, pp.86-90
Geotec., Const. Mat. & Env., DOI: https://doi.org/10.21660/2018.48.7257
ISSN: 2186-2982 (Print), 2186-2990 (Online), Japan

CORRELATION AMONG THE SOIL PARAMETERS OF THE KARNAPHULI RIVER TUNNEL PROJECT

*Khondoker Istiak Ahmad, Masnun Abrar, Sultan Al Shafian and Hossain Md. Shahin,

Islamic University of Technology, Bangladesh

*Corresponding Author, Received: 1 Dec. 2017, Revised: 13 Feb. 2018, Accepted: 13 April 2018

ABSTRACT: Defining relationships among the soil parameters and strength properties have become a popular research topic among the soil mechanics researchers in recent years. As conducting laboratory tests are quite tiresome and time-consuming, investigating interconnection and correlation among the soil parameters using collected data will help to understand and predict the characteristics of soil effortlessly. This research aims at stating regression and product moment correlation analysis between the strength parameters and index properties of the soils of Karnaphuli river tunnel project site. Around 120 borings were done in that site to determine both the index and strength properties of the soil in the laboratories. In this paper, multiple linear regression analysis was developed for the cohesion and liquid limit along with the collected explanatory variables and correlation analysis was done to quantify the direction and strength of the linear association**.** This equation supports to adopt different characteristics of the soils of Karnaphuli River and additionally a statistical significance of correlation among different soil parameters have also been unfolded. Finally, the equation is compared with some of commonly used classical soil mechanics equation to determine the acceptability of this correlation.

Keywords: Index Properties, Strength Parameters, Regression Analysis, Correlation.

1. INTRODUCTION

Determining strength parameters and index properties of a soil sample is a time-consuming process. By knowing the strength parameters (e.g. shear strength) will help to determine soil characteristics and to be useful for designing future underground structures. Researchers have tried long for stating correlation between strength parameters and index properties of soil. Smith et al. [1] established a correlation among the index properties and other parameters of soil in Israel. They established relationships among shrinkage, Atterberg limits, and physical and chemical properties of soil. Correlation analysis for rock and other marine soils have been established many time by the researchers. But in the developing countries like Bangladesh where soils are characterized as soft soil in a bulk, there are little attempts have been made on correlation analysis. As far as Authors' knowledge none of the studies have conducted correlation analysis among different parameters of soil in context of Bangladesh.

This study aims at establishing a correlation between the strength parameters and other index properties of the soft soil of Karnaphuli River tunnel project. Around one hundred and twenty (120) bore hole samples have been collected for testing and determining the various strength and index properties of the soil. In this research work, a mathematical model has been adopted to establish the correlation among different soil parameters. The aim of this model establishment is to reduce the number of tests in the laboratory and save time. However, it is almost difficult to illustrate the variability of soil parameters, the statistical mathematical model has been adopted by many researchers to establish the correlation of soil samples. Ching et al. [7] [8] have used multivariate probability distribution model to establish correlation among various clay parameters. These researches made a contribution in technical terms of statistics when it is used in terms of geotechnical perspective. Ching [8] constructed multivariate probability distribution that would be used as a prior distribution to make the joint distribution of design parameters based on limited but site-specific field data. Ching [7] worked for the transformation of the calibrated model developed from soil parameters earlier. They recalibrated the biases and transformation uncertainties of those transformation models with respect to the global database. It is found that more recent transformation models are less biased and that the transformation uncertainties are typically not so large.

There are some other studies performed in probabilistic reliability analysis of soil parameters. Some literature are: Christian and Urzua [12] evaluated the earthquake-induced slope stability failure probabilistically, Dasaka [13] performed reliability analysis using probability and statistical techniques on slope stability and site characterization, Lacasse and Nadim [14] analyzed the uncertainties in characterizing soil properties. Some other studies performed geotechnical variability and geotechnical property variability using statistical technique [15, 16]. Earlier attempted was also made by Vanmarcke [17] to model soil profiles using probability theory.

From the above discussion, it is clear that compared to other research on geotechnical engineering, in the field of statistics the number of research papers is less. So, there is room to continue multidimensional research in this field because with the advancement of technology more data are generating. So, statistical relationships with certain tools can be beneficial to unveil interesting relationships among soil variables. Also, from the literature it is evident that most of the studies explored the correlation among variables but gave less attention on generating equations with a certain level of statistical significance.

In this study, multivariate statistical analysis has been used. Considering the depth and other geotechnical properties of soil, first a multivariate statistical analysis i.e. Multiple Linear Regression is established which comprises the statistically significant variables and then a correlation analysis is established among the soil parameters with proper statistical techniques. Multiple Linear Regression analysis is mainly used to recognize the consequence that the independent variables have on a dependent variable. In this study, this regression is mainly used to identify the effect of cohesion and liquid limit over other common soil parameters.

2. STUDY AREA

The soil samples studied for this project have been collected from the *Karnaphuli River tunnel Project Site of Chittagong*, the first ever underwater tunnel project going to be constructed in Bangladesh. This Multi-lane Road Tunnel project site is located at the sea entrance of River Karnaphuli of Chittagong suburb, and its west coast starting point is connected with Coastal Road. Geologically the area falls within the Bengal Foredeep of Bengal Basin where Neogene sediments with alternation of shale and sandstone are well developed. The general elevation of the investigated area ranges from -8.8 to 4.395 m.

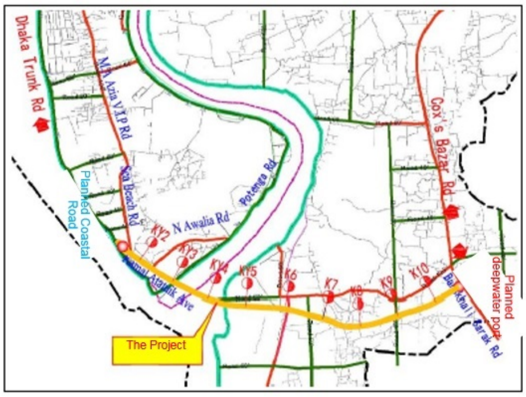

Fig. 1 Location of the Karnaphuli River Tunnel Project Site [2]

The geological characteristics of the surface area of Chittagong mainly showed two basic distinct patterns. The drainage pattern, their distribution, landform features and their position in tidal flat areas make the area as a complex geomorphic region [4]. Soil samples have been collected from various parts of this project site. The depth of common boreholes is 20 m under the tunnel design bottom surface, while for controlling boreholes, they should be 40 m under the design bottom surface or 5 m into the complete bedrock at least. The number of controlling boreholes accounts for 2/5 of the total number of boreholes [2].

3. METHODOLOGY

Multiple Linear Regression (MLR) has been used to establish a multivariate equation with statistical significant explanatory variables. MLR has been preferred by many other researchers for establishing suitable relationships among various parameters of both rocks and soils. [3] [7]

3.1 Multiple Linear Regression

The most widely used regression model to analyze a dependent variable on the basis of change in more than one independent variables is the Multiple Linear Regression (MLR) model which is given by

$$y = x\beta + \varepsilon \quad (1)$$

in Eq. (1), where 'y' is the dependent or response variable, 'X' is the vector of explanatory variables; '$\boldsymbol{\beta}$' is a vector of parameters to be estimated and 'ε' is an error term. In this research, an effort has been made to create a relationship among different geotechnical parameters with cohesion and liquid limit. Around 15 variables are selected to formulate this model. The variables that we used in this research are – void ratio, dry density, liquid limit, natural mass density, water content, plastic limit, plasticity index, stress, strain, soil unit weight, specific gravity, dthe egree of saturation, depth etc.

In statistics, it is very common to conduct a hypothesis test to find the statistically significant variables and put it into the model. For this reason, t-test was applied with 5% significance level which means, if the p-value of any variable is less than 0.05, than the null hypothesis will be rejected, which means that variable will not be inserted into the statistical model. The model equation will consist of only statistically significant variables [5].

Table 1 represents the summary results of the statistical model for two dependent variables (Liquid limit and Cohesion). The model is developed in 'R' - open source statistical software.

Table 1 Parameter Estimation from Multiple linear regression (Only statistical significant variables are included)

Variables	Liquid Limit	Cohesion
Stress	0.0091 (<0.001)	0.61 (<0.001)
Void ratio	120.55 (<0.001)	-
*N	102.16 (0.03)	97.37 (0.02)
Dry density	102.16 (0.03)	-
Liquid limit	-	-16.19 (0.017)
Sample	75	75
R^2	66%	65%
Adjusted R^2	64%	63%
F-statistics	33.33	41.54

*N = Natural mass density
(All the parameters are of 95% of the confidence interval, in bracket p-values are given)

To measure the suitability of fit of the model, squared R is taken to show how close the data are to the fitted regression line. Squared R assumes that every predictor in the model explains the variation in the response variable. It gives the percentage of explained variation as if all predictors in the model affect the response variable, whereas the adjusted R-squared gives the percentage of variation explained by only those independent variables or predictor that in reality affect the response variables.

To measure the model overall significance F-statistics value was tested and evaluated. It can assess multiple coefficients simultaneously.

3.2 Correlation

Correlation describes the degree of relationship between two variables. Product moment correlation efficient (r) can be measured to identify the correlation among different important geotechnical parameters. The equation is [5]:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (2)$$

where, $S_{xy} = \left(\sum xy\right) - nx_{avg}y_{avg}$

$$S_{xx} = \left(\sum x^2\right) - nx_{avg}^2$$

$$S_{yy} = \left(\sum y^2\right) - ny_{avg}^2$$

It is also tested whether the correlation between two variables is statistically significant or not. So, a hypothesis test is introduced with 5% significance level following two-tailed test. Table 3.2 shows the correlation matrix among different independent variables [6]

Table 2 represents the correlation values of the sample data that has been used for establishing the correlation among the soil parameter

Table 2 Correlation matrix of independent variables

	σ	W	N	e	LL	PL	PI
σ	1						
W	-0.54*	1					
N	0.62*	-0.64*	1				
e	-0.61*	0.93*	-0.86*	1			
LL	0.13	0.56*	-0.21	.5	1		
PL	0.14	0.54*	-0.20	0.49	0.99	1	
PI	0.11	0.57*	-0.21	0.50	0.99	0.98	1

(* represents statistical significant with 5% significance level)

In the following Table 2, the abbreviations of the short terms are: Ϭ = Stress, W= Water Content, N= Natural mass density, e= Void ratio, LL = Liquid limit, PL= Plastic limit, PI =Plasticity index

4. RESULTS AND DISCUSSION

Using proper regression analysis, following two equations has been derived.

Cohesion=-158.01-16.19(LL)+ 0.61(Ϭ)+ 97.37(N) (3)

LL= -296.82 + 0.0091(Ϭ) + 120.55 (e) + 38.33 (N) + 102.16 (Dry density) (4)

In Eq. 3 for one unit increase in LL, the cohesion will decrease by 16.19 unit. For one unit increase of stress, the cohesion will increase by 0.61 unit.For one unit increase in natural mass density, cohesion will increase by 97.37 unit. In Eq. 4, for one unit increase in stress liquid limit, will increase by 0.0091 unit. For one unit increase in void ratio, the liquid limit will increase by 120.55 unit. For one unit increase in natural mass density, the liquid limit will increase by 38.33 unit. And lastly, for one unit increase in dry density, the liquid limit will increase by 102.16 unit. From this equation, it is observed that void ratio has great influence on the liquid limit. It is natural that if void ratio increases, the liquid limit will also increase.

From table 2, we can observe that the best positive relationships exist between liquid limit – plastic limit, liquid limit-plasticity index, plastic limit-plasticity index and void ratio-water content for obvious reasons as there are theoretical relationships among these parameters. The marginal positive degree of relationship exists between natural mass density-stress, liquid limit-water content, plastic limit-water content, plastic limit-water content, the void ratio with plastic limit and liquid limit. There are some negative correlation among variables, which means if one parameter increases in a specific unit others will decrease and vice-versa. There are some statistical insignificant variables i.e. they do belong to the null hypothesis, cannot be rejected.

5. MODEL ADEQUACY

To check how adequate and effective the model is, residual plots have been introduced. It is important to investigate how well the multiple linear regression models fit the data. This job is done using squared (R) and adjusted squared (R). But, the residual plot is an efficient tool for visualization of the fitted data,

Residuals,

$$e_i = y_i - \overline{y_i} \quad (5)$$

where " y_i " is number of observation (here, Cohesion and Liquid Limit) and " $\overline{y_i}$ " is corresponding fitted value coming from the model with statistical significant variables.

To see the fitted model and to check whether the model is influenced by any of the outliers, scaled residuals are proved to be effective. In this paper, standardized residuals will be used as it follows an important property of MLR assumption that it has mean zero and unit variance. The standardized residuals are calculated by following:

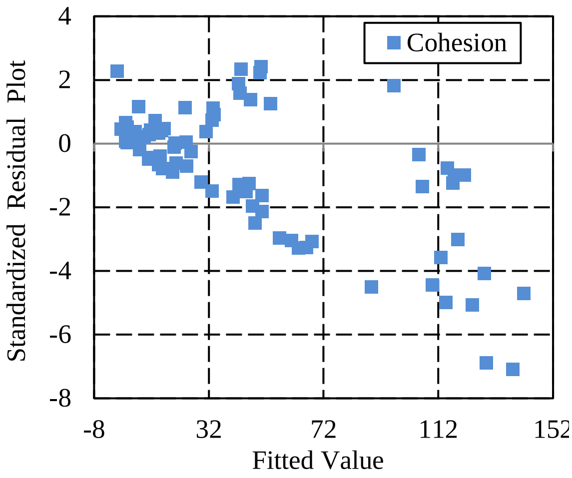

Fig 2: Residuals (e_i) versus fitted value ($\overline{y_i}$) of the cohesion model

$$d_i = \frac{e_i}{\sqrt{MS_{res}}} \quad (6)$$

where e_i is residuals, MS_{res} is residual mean square, and d_i is standardized residuals

Figure 2 represents the relationship between standardized residuals of the model and the fitted data availed from the model. In Figure 2, all of the plots show a non-constant variance of the errors. The figure shows dispersed data points after selecting statistical significant variables for the model. But, the majority of the plot in Figure 2 shows nonlinearity. The patterns indicate that the model needs more predictors as a variable to come up with the better predictable equation of cohesion model.

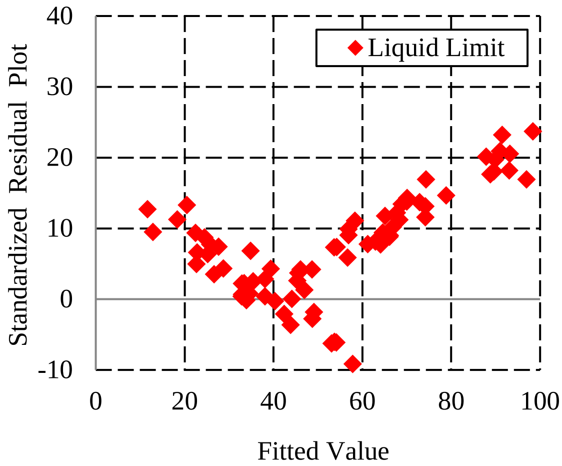

Fig 3: Residuals (e_i) versus fitted value ($\overline{y_i}$) of the Liquid Limit model

In Figure 3, the liquid limit model shows a semi-double bow pattern. This model suggests that the equation needs a transformational model to come up with the better fit model. But, it is worth to notice that the plots in Figure 3 suggest a close satisfactory model. From Figure 2 and 3 we can observe some data points that are observed as outliers but it is not possible to take data points as outlier before applying scaled residuals. Here this research each and every data points are treated and concluded as that those data points as outliers if standardized residuals are greater than 2. From Figure 2 and 3 we can conclude that we did not find any of these data points which show its standardized residual greater than the value 2. So, some points that have been observed far from the cluster of other data points are not outliers.

6. CONCLUSION

From Table 1, it can be observed that the both of the results for liquid limit and cohesion exert a marginal relationship with the significant independent variables. Liquid limit and cohesion show 66% and 64% of R2 value respectively. As R2 value represent the total goodness of fit of the model, so it can be stated that this two models can predict liquid limit and cohesion with the stated variables with 66% and 64% accuracy. In statistical analysis, it is common to get marginal fit. Because there involves probability of nature in the selection of variables while constructing a model. The relation that has been established among the soil parameters is completely based on a statistical approach which will be very much helpful for determining the proper characteristics of the soils of Karnaphuli River Tunnel project in future. This research work is completely statistical approach considering the laboratory tests of soils. There are some rooms to improve this model as a further study. The mathematical model can be further improvised by considering more variables and sample number. Multiple linear regression relies on some assumptions. It assumes that the equation will show a linear relationship between variables. But, a model can often show the better goodness of fit when the nonlinear equation is introduced. So, this assumption is not checked in this model. As the samples size was limited and small, so model validation is not checked. So, transferability of this model by introducing different data of the same variables from a different site is not checked. So, transferability of a statistical geotechnical model would be beneficial to come up with the more robust equation. Apart from statistical technique, data mining techniques, artificial intelligence (e.g., Artificial Neural Network) can be incorporated using this data that may uncover some rational facts in the tunnel study.

7. REFERENCES

[1] Smith C.W., Hadas A., Dan J. and Koyumdjisky H., 1985, "Shrinkage and Atterberg limits in relation to other properties of principal soil types in Israel'' ,Geoderma, 35:47-65.

[2] China Communications Second Highway Survey, Design and Research Institute Co. Ltd., ''Detailed Investigation Report of Engineering Geology-Multi lane Road Tunnel Under River Karnaphuli Chittagong Bangladesh'', Volume I and II , October 2016

[3] Mishra D.A., Basu, A., "Estimation of the uniaxial compressive strength of rock materials by index tests using regression analysis and fuzzy inference system", Engineering Geology, Vol. 160, 2013, pp. 54-68

[4] Feasibility Study for Multi-LaneRoad Tunnel under the River Karnaphuli, Chittagong, Bangladesh, Final Report, Issue 1 , April 2013, Bangladesh Bridge Authority.

[5] Montgomery D.C., Peck E.A., Vining, G.G. (2012). Introduction to Linear Regression Analysis (5th Ed.), John Wiley and Sons Ltd

[6] Kottegoda, N.T., Rosso, R. (2008). Applied Statistics for Civil and Environmental Engineers (Second Edition), Blackwell Publishing, John Wiley and Sons Ltd

[7] Ulusay R. , Tureli K., Ider M.H. , Prediction of engineering properties of a selected litharenite sandstone from its petrographic characteristics using correlation and multivariate statistical techniques, Engineering Geology, Volume 38, Issues 1–2, December 1994, Pages 135-157

[8] Ching Jianye and Phoon Kok-Kwang, Transformations and correlations among some clay parameters the global database, Canadian Geotechnical Journal 51 (2014), no. 6, 663–685.

[9] Ching Jianye and Phoon Kok-Kwang, Correlations among some clay parameters- the multivariate distribution, Canadian Geotechnical Journal 51 (2014), no. 6, Pages 686–704.

[10] Michael Carter & Bentley, Stephen P (1991). Correlations of soil properties. Page 77-96,chapter 7

[11] Cao Z, Wang Y. Bayesian model comparison and selection of spatial correlation functions for soil parameters. Struct Saf(2014)

[12] Baecher GB, Christian JT. Reliability and statistics in geotechnical engineering. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2003.

[13] Christian, J.T. and Urzua, A. (1998). Probabilistic evaluation of earthquake-induced slope failure, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, Vol. 124, No.11, pp.1140-1143.

[13] Dasaka S N Murthy (2006) Probabilistic site characterization and reliability analysis of shallow foundations and slopes (Doctoral thesis, unpublished) Indian Institute of Science, Bangalore.

[14] Lacasse, S. and Nadim, F. (1996). Uncertainties in characterizing soil properties, Plenary paper for ASCE Conference on Uncertainties, Madison, Wisconsin, USA.

[15] Phoon, K.K. and Kulhawy, F.H. (1999a). Characterization of geotechnical variability, Canadian Geotechnical Journal, Vol. 36, pp. 612-624.

[16] Phoon, K.K. and Kulhawy, F.H. (1999b). Evaluation of geotechnical property variability, Canadian Geotechnical Journal, Vol. 36, pp. 625-639.

[17] Vanmarcke, E.H. (1977). Probabilistic modeling of soil profiles, Journal of the Geotechnical Engineering Division, ASCE, Vol.103, No.GT11, pp.1227-1246.