ACTA TECHNICA CORVINIENSIS – Bulletin of Engineering [e-ISSN: 2067-3809]
TOME XIII [2020] | FASCICULE 2 [April – June]

[1.]Olaolu George FADUGBA

APPLICATION OF MULTIVARIATE DATA ANALYSIS IN THE INTERPRETATION OF GEOTECHNICAL INDEX PARAMETERS – A CASE STUDY OF THE FEDERAL UNIVERSITY OF TECHNOLOGY AKURE, ONDO STATE

[1.]Department of Civil Engineering, The Federal University of Technology, Akure, NIGERIA

Abstract: Geotechnical investigation has evolved from the traditional methods of interpretation to more reliable and improved methods, where their as been an increasing implementation to predictive analysis for the interpretation of geotechnical investigation data and to gain decision making information from these data. The presented work looks into the application of multivariate data analysis for the interpretation of geotechnical index parameters. Uncertainty in geotechnical engineering was discussed, its causes and the use of reliable–based design to achieve a stable design. Data of One hundred and eighty nine soil samples was collected from seven different locations in the Federal University of technology, Akure, and this was what was used for the development of the model. The multivariate data analysis method used for this work was the multiple linear regression method. This method involves a single metric dependent variable that are related to two or more metric independent variables. The objective of multiple linear regression analysis is to predict the changes in the dependent variable in response to changes in the independent variables. This objective is most often achieved through the statistical rule of least squares. Because this research is interested in predicting an amount and size of dependent variable, multiple linear regression method is used. It is a statistical technique that is basically used to analyze the relationship between a single dependent variable and several independent variables.
Keywords: index properties, uncertainty, reliable–based design, multivariate data analysis, multiple linear regressions

INTRODUCTION

Geotechnical engineering over the years has been proven to be very important and critical in construction and civil engineering, because, every structure constructed in one way or the other has a direct or indirect contact with the soil as foundation, and geotechnical engineering deals with the investigation and analysis of the soil. But there has been a need for adequate and reliable geotechnical characterization of the soil because of the much uncertainties involved with geotechnical investigations.

Evaluation of the uncertainties in geotechnical design parameters provides the foundation for reliability–based design (RBD). For example, Baecher and Ladd (1997), Phoon and Kulhawy (1999), Müller (2013) and Mašín (2015) have presented methodologies on how to evaluate uncertainties from a single test procedure. In most cases, the design parameter and associated uncertainty are evaluated using more than one investigation method, which includes both in–situ and laboratory measurements. As such, it is highly beneficial to combine (i.e., cross–validate) data from different types of investigation method to reduce the uncertainty in the design parameter.

Multivariate analysis according to bioimpedance and bioelectricity basics 2015, can be said to be set of operation procedures used for analysis of data sets that consists of more than one variable, and the techniques are especially valuable when working with correlated variables. Olkin et al., 2001 conceptualized multivariate analysis traditionally as the statistical study of experiments in which numerous measurement are made on each experimental unit and for which the relationship among multivariate measurements and their structure are important to the experiment's understanding.

Multivariate analysis is an important tool in data management and it has been widely used in managing and interpreting various data like geochemical characterization of groundwater and soil science. Thus, the use of multivariate methods requires large data sets and therefore its application in geotechnical engineering is not frequently used, since even for large construction projects the number of laboratory analyses usually provides only a relatively limited amount of data in statistical terms. At the same time, it has been known to some extent that data analyses are useful tools in geotechnical engineering. More recent studies have demonstrated the applicability of multivariate data analyses in various fields of engineering geology such as rock engineering soil liquefaction, landslide susceptibility analyses and even in investigating the correlation between clay mineralogy and shear strength of soils. (Kovacs et al., 2015).

LITERATURE REVIEW

— Reliability–based design in geotechnical uncertainties

Reliability–based designs can be simply and best explained as statistical approaches to provide stable and reliable design. Considering the large number of uncertainty involved and associated with geotechnical investigation, reliability–based designs are needed for cost effective and reliable site investigation (Anders et al., 2017). Reliability designs aims at assessing to what extent uniformity, consistency and stability can be achieved in design. Failure probability and reliability index are used to quantify the reliability of a design. The lower the failure probability the higher the reliability index, the more reliable a design is. To fully adhere the scope of geotechnical site–investigation with the reliability of a construction, the uncertainty related to inherent variability of soil (aleatory uncertainty) must be accompanied by the epistemic uncertainties related to the method of soil characterization, which is typically divided into measurement error, statistical uncertainty and transformation uncertainty (Prästings et al. 2016; Ching et al. 2014; Ching et al. 2016). A major approach to a reliable based design is the multivariate approach, which was considered in this research.

— Multivariate data analysis

Multivariate data analysis (MVDA) describes the practice of using mathematical and statistical tools to extract information from data tables where each observation contains a large number of variables. In such cases, the desired information lies in the correlation structure between variables, this often leads to erroneous results when tested independently. Multivariate data analysis by means of projection methods is able to analyze data where challenges such as multidimensionality of the data set, multicollinearity, missing data and variation introduced by deviating factors such as experimental error.(Darren et al, 2014)

The increasing use of multivariate data analysis (MVDA) in both basic research and applied scientific fields has enabled the diagnostic evaluation of parameter interactions that were previously undefined. Therefore univariate or bivariate analysis is often inefficient resulting in misleading conclusions (Kourti, 2004). Key information in such cases lies in the correlation structure between variables and can lead to spurious results when tested independently.

MATERIALS AND METHOD

The type of data used were secondary data from previous subsoil investigation conducted in FUTA, the data comprised of various index parameters extracted from the subsoil investigation of these selected places: New 1000 capacity multipurpose hall, the new postgraduate school, the new staff, block of classroom A and B, and the indoor sports hall. Only a few selected index parameters where used, and they are: the depth, the natural moisture content (NMC), specific gravity (Gs), average pocket penetrometer value (kPa), cohesion c, and Atterberg limits LL – Liquid Limit, PL – Plastic Limit, PI – Plasticity Index, LS – Linear Shrinkage.

The area used for the research is The Federal University of Technology, Akure. FUTA, located in Akure, the capital city of Ondo state. About 250m north of highway 364 (Ilesa–Owo Expressway), southwestern Nigeria. The campus falls within latitudes 7°17'03"N – 7°19'06"N and longitudes 5°07'02"E – 5°09'05"E

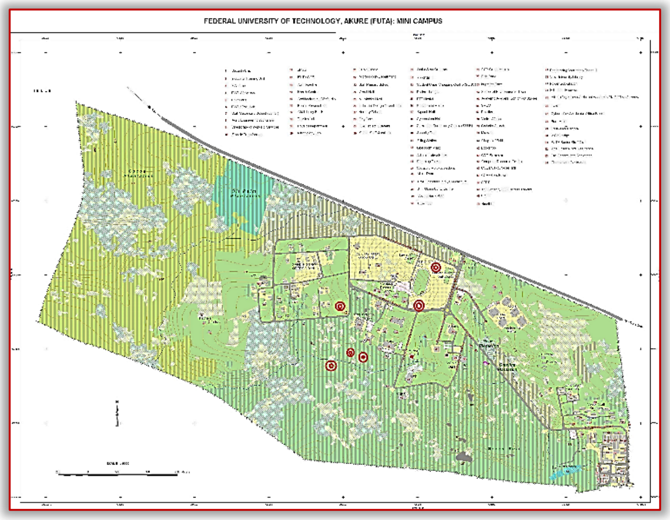

Figure 1. Map of Study Area

RESULT AND DISCUSSION

To evaluate the geotechnical parameters and to generate the regression model by identifying the combinations of the inputs ad their linearization, the MATLAB software was used. To analyze the random relationship that exists between the variables, a matrix correlation was used. The correlation coefficient (R) and its square, the coefficient of determination (R^2) describe the linear connection. The correlation is strong, when $|R| \geq 0.7$ and weak, when $|R| \leq 0.5$.

The models were developed taking Average Pocket Penetrometer value as the dependent variable while other parameters were considered as independent variables. The following abbreviations were used in the tables.

Table 1. Summary of Correlation Adequacy Check.

Location	Degree of freedom	Coefficient of determination (R^2)	F–Statistic	P–Value
Block of Classrooms Block A	15	0.597	2.221	0.136
Block of Classrooms Block B	19	0.164	0.335	0.922
Indoor Sports Hall	23	0.314	1.046	0.439
New 1000 Capacity Multipurpose Hall	23	0.429	2.131	0.103
New Staff Offices	17	0.266	0.663	0.682
New Postgraduate School	18	0.255	0.741	0.626

Natural Moisture Content (NMC), Specific gravity (SG), Cohesion (C), Average Pocket Penetrometer Value (APPV), Liquid Limit (LL), Plastic Limit (PL), Plasticity Index (PI) and Liquid shrinkage (LS). The models try to predict the desired Average Pocket Penetrometer for a given set of data.

For data evaluation, it is important that no missing data occur in the matrix. The samples with missing data points were not used in the analyses.

The data provided required computerized statistical analysis to determine any meaningful correlations between the variables and to establish relevant predictive models. In this study, the statistical analyses performed include:

≡ descriptive statistics (mean, standard deviation, etc. of each variable),

≡ linear and nonlinear regression analyses between all variables, and

≡ multivariant analyses between selected variables.

The statistical runs included F–tests and the use of correlation coefficients before drawing inferences. Six functions were fitted using the data at various data collection points. The main objective is to extract, in a condensed form, the largest possible information contained in the data, whether related to links between variables or between individuals (tests).

In order to determine the overall significance of the regression tests, the F–test was also performed. The coefficient of multiple determination, R^2, provides an overall measure of the adequacy of the equation to predict. If the line is a good estimator of the data, this coefficient will be near unity. Based on the correlation model generated only the data obtained from the block of classrooms showed a strong correlation of the model with R^2 value of 0.597, however, it doesn't show a measure of statistical significance because probability of F–values is greater than 0.05 significance level and the regression can therefore be considered as insignificant. This could possibly be attributed to the lack of sufficient data collected or lack of cohesion in the test data provided.

The Pearson Correlation table obtained from the data analysis was also studied with an aim to understand several correlation that may exist between the independent variables: The Pearson Correlation table shows a weak negative correlation between specific gravity and Atterberg limit values. Also a relatively strong positive correlation between cohesion and APPV. Further as observed from the analysis, the Pearson Table suggests a moderately strong positive correlation between the natural moisture content and the Atterberg tests and rightfully so since Atterberg tests are measures of the critical water contents of a fine–grained soil

The model generated from the multivariate analysis indicates a significant negative linear relationship between cohesion and Average pocket penetrometer value. Also it shows a negative correlation for natural moisture content

A plot of the parameters against each other allowed the following observations to be made: the natural moisture content has a positive correlation with the depth as a 60% deduction in soil depth yielded a 27% decrease in Natural Moisture Content, but the cohesion has a weak positive non–significant correlation with the depth, that is as a 50% depth decrease results in an average 6% increase in cohesion values. This, thereby confirms the findings of Netra R et al., 2015 that cohesion strongly depends on the bonding of fine–grained particles and some physical properties of the soil such as moisture content, that that an increasing soil moisture content reduces the cohesion and thus shear strength.

Table 2. Summary of Models

LOCATION	MODEL EQUATIONS
Block of Classrooms Block A	APPV = –210.271 – 1.000NMC + 40.184SG + 0.162C + 2.408PL – 0.342PI + 3.170LS
Block of Classrooms Block B	APPV = –1647.996 – 3.774NMC – 480.083SG – 0.106C + 4.247LL – 1.014PL + 0.285PI – 0.223LS
Indoor Sports Hall	APPV = –47.766 – 2.939NMC + 110.409SG + 0.020C – 93.514LL + 95.018PL + 97.911PI + 2.752LS
New 1000 Capacity Multipurpose Hall	APPV = 56.037 – 1.951NMC + 144.498SG – 0.027C + 2.796PL + 2.100PI + 1.754LS
New Staff Offices	APPV = –963.077 – 2.521NMC – 248.606SG – 0.387C + 9.745PL + 5.387PI – 14.728LS
New Postgraduate School	APPV = –777.925 – 2.291NMC – 189.916SG – 0.191C + 9.363PL + 5.330PI – 15.133LS

CONCLUSIONS

From the study, the following conclusions were drawn:

— There is a strong correlation between the index properties and the APPV of the soil.

— The cohesion value is an essential soil property for foundation design. However, it can't be measured as accurately from the linear regression models provided as it shows a negative correlation with the soils Atterberg limit.

— The Average Pocket Penetrometer Value and the Cohesion value are closely correlated, both being important in foundation design problems and as such, appropriate means should be devised to measure or estimate their values.

— The application of multivariate analysis in geotechnical engineering designs offers a reliable guide and solution to the interpretation of geotechnical index parameters, problems and geotechnical uncertainty

References

[1] Prästings A, (2016). Aspects on probabilistic approach to design: from uncertainties in pre–investigation to final design Licentiate Thesis.
[2] Baecher, G., Ladd, C., (1997). Formal observational approach to staged loading. Transportation Research Record: Journal of the Transportation Research Board, 1582, pp.49–52.
[3] Baecher, G.B. Christian, T. (2003). Reliability and statistics in geotechnical engineering. Wiley
[4] Bradák B., Kovács J. (2014). Quaternary surface processes indicated by the magnetic fabric of undisturbed, reworked and fine–layered loess in Hungary. Quatern. Int., 319, 76–87.
[5] Christian, J.T., Ladd, C.C., Baecher, G.B. (1994). Reliability applied to slope stability analysis. Journal of Geotechnical Engineering, ASCE: Vol. 120, No. 12, pp. 2180–2207.
[6] Cameron, G. (2002). Uncertainty, reliability and risk in geotechnical engineering. Proc. of the British Geotechnical Association: 7th Young Geotechnical Engineers Symposium – Diversity in Geotechnics, Dundee, pp. 19–20.
[7] Contreras, L.F, Ruest, M. (2016), unconventional methods to treat geotechnical uncertainty in slope design, Australian center for Geomechanics.
[8] Dalu T., Richoux N.B., Froneman P.W., (2014). Using multivariate analysis and stable isotopes to assess the effects of substrate type on phytobenthos communities. Journal of the International Society of Limnology, Inland Waters, 4, 397–412.
[9] Elkateb, T., Chalaturnyk, R., Robertson, P.K. (2003a). An Overview of Soil Heterogeneity: Quantification and Implications on Geotechnical Field Problems, Canadian Geo
[10] Griffiths, D.V., Fenton, G. A., Manoharan, N. (2002). Bearing capacity of rough rigid strip footing on cohesive soil: Probabilistic study. Journal of Geotechnical and Geoenvironmental Engineering, ASCE: Vol. 128, No. 9, pp. 743–755.
[11] Harr, M. E. (1977). Mechanics of particulate media – a probabilistic approach. McGrawHill, New York.
[12] Kovács, J, Bodnár, N., Török, Á. (2016). The application of multivariate data analysis in the interpretation of engineering geological parameters: Journal of Open Geosci.8, 52–61.
[13] Lacasse, S., Nadim, F. (1996). Uncertainties in Characterizing Soil Properties. In Uncertainty in the Geological Environment: From Theory to Practice, Geotechnical Special Publication No. 58, C.D. Shackleford, P.P. Nelson, and M.J.S. Roth (eds.), pp 49–75.
[14] Lumb, P. (1966). The Variability of Natural Soils, Canadian Geotechnical Journal, Vol 3, No. 2, pp. 74–79.
[15] Muller, R., (2013). Probabilistic stability analysis of embankments founded on clay. KTH, Royal Institute of Technology.
[16] Phoon, K.–K., Kulhawy, F.H., (1999). Characterization of geotechnical variability. Canadian Geotechnical Journal, 36, pp.612–624
[17] Phoon, K. K. (1995). Reliability–based design of foundations for transmission line structures. A PhD dissertation submitted to Cornell University.

ACTA TECHNICA CORVINIENSIS – Bulletin of Engineering
ISSN: 2067-3809