International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 17251-17260

http://www.ripublication.com

Geotechnical Parameters Prediction Models For Infrastructure Developments

Arulkumaran. S[1]
[1] *Corresponding author & Research Scholar, Public Works Department, Government of Tamil Nadu, Email: geo_arul@yahoo.com*
Sukumar. S[2]
[2] *Professor and Head, Department of Civil Engineering, Government College of Engineering, Salem, TamilNadu. Email: Email:sukumar_237@yahoo.co.in*

Abstract

In Civil Engineering projects, study of available subsoil data and estimation of parameters at desired location helps the planners in preliminary stage. Now a day, Mathematical models are useful to generate the data by various interpolation techniques from the measured properties. Regression is one of the fundamental models in statistics used to determine the relationship between dependent and independent variables. MATLAB software has rich toolboxes that cover almost every aspect of mathematical and statistical engineering applications. The objective of this study is to introduce regression analysis by MS Excel and MATLAB programming to estimate the required subsoil properties of unvisited locations. Salem block of Tamilnadu has been taken as the study area. The validity of model is checked with other area data sets for its adequacy and found that the resulted Chi Square goodness of fit value is less than the tabulated value (140.2 for more than 100 degree of freedom). The results imply that the developed modelsare useful one in preliminary stage of a project planning.

Key Words: Site investigation, Geotechnical Data, Prediction of soil parameters,MS Excel, MATLAB Regression model.

Introduction

Reliable estimation models to predict subsoil properties will be very valuable in the preliminary stage of project planning. Geotechnical professionals are always looking out for a tool, which can improve and help better to handle the large uncertainties and variations inherent in soil properties.

Site investigation seems to be the area for introducing statistics to geotechnical engineers [1]. Computing techniques such as artificial neural networks, fuzzy

inference systems, evolutionary computation, etc. and their hybrids have been successfully employed for developing predictive models to estimate the parameters [2]-[6]. Though these techniques predict the parameters within acceptable accuracy, simple regression models can estimate the parameters with equal accuracy [7],[8].Multiple regressions are used to represent the relationship between a dependent variable and several independent variables. Spreadsheet software is one of the powerful and practical tools for developing model since it has a good user interface and optimization capability[9]. Methods of using Microsoft Excel to fit non-linear functions has the advantage that Excel is probably included in the computer package as part of Microsoft Office, and thus no additional expense is required.Previous studies based on the linear regression analysis using MS Excel,predicted the unknown parameters for Coimbatore and Chennai cities[10],[11].

Matrix Laboratory (MATLAB) is numerical computing, user friendly software which is developed by Math work Inc. Authors focuses on expressing the multiple linear regression models and analyse using a script approach with MATLAB[12] ,and uses MATLAB to solve geotechnical problems involved in underground mine and open pit design[13]. MATLAB also used for various analyses and visualise the results easily [14], [15].Though numbers of techniques are used in prediction models, studies shows that, there is not one single prediction method that can produce chief results for the generation of continuous soil property maps all of the time for the studied data set[16],[17].

This article focuses on expressing the multiple linear regression models using MS Excel and MATLAB for predicting values from non-sampled locations. The developed models are also validated with study area data and other area data.

Site Condition In Salem Block

In this study, data from 104 soil investigation bore wells has been collected from various public and private organisations including Public Works Department, Highways Department, and Technical institutions. Soils were tested for determination of grain size distribution, Atterberg limits, Differential Free Swell, Standard Penetration Test (SPT) Nvalues and other parameters according to the procedure suggested by Indian standards. Based on the properties and testing methods, the collected soil parameters are grouped as shown in Table1.

Table 1: Details of collected subsoil Properties

I. Field and Index Properties		X7	% Gravel, %G
X1	Latitude	X8	% Fines (Silt and Clay), %F
X2	Longitude	**II Properties from Lab and Field Tests**	
X3	Elevation, MSL in 'm'		
X4	Depth of Exploration, D_{SF} in 'm'	X9	Liquid limit, LL in %
X5	Thickness of poor Overburden , 'T' in 'm'	X10	Plastic limit, PL in %
		X11	Differential Free Swell, DFS in %
X6	% Sand, %S	X12	SPT-N

Multiplelinear Regression Model Using Ms Excel

The collected parameters as listed in table 1 is correlated with each other's to derive a relationship among the parameters. In multiple regression analysis, the model is of the type,

$$Y = b_1 x_1 + b_2 x_2 + b_3 x_3 + .. + b_n x_{n+} a \quad (1)$$

Where Y is the dependent variable, x_1, x_2 ... x_n are the independent variables, b_1, b_2... b_n are the coefficients of the respective independent variableswhich will be determined from the input data and "a" is a constant, where the regression line intercepts the y axis. Method of least square technique is used to develop this model.

Coefficient of determination (R^2)represents the fraction of the overall variance of the 'dependent' variable that is explained by the 'independent' variable. It is a measurement of how well the multiple regression line fits the data. It is calculated from the sum of the squares of the residuals and the sum of the squares of regression. The sum of the squares of the residuals captures the error between the estimate and the actual. Data from 104 borehole locations are analyzed to determine LL,PL,DFS and SPT N values from the other measured parameters.The results of regression analysis are summarized in Table 2.

Table 2: Value of Regression coefficients

X	LL	PL	DFS	SPT N
X1	-81.42	17.37	116.85	-52.56
X2	-28.23	20.07	31.81	-24.34
X3	0.16	-0.05	-0.20	0.07
X4	0.35	0.38	1.28	1.67
X5	-0.74	0.43	-1.35	-1.45
X6	0.53	-0.46	-0.71	-0.46
X7	0.44	-0.39	-0.44	-0.14
X8	0.52	-0.35	-0.29	-0.47
X9	-	0.51	0.67	0.13
X10	0.81	-	0.07	-0.01
X11	0.25	0.01	-	-0.22
X12	0.04	0.0	-0.19	-
a	3066	-1718.8	-3731	2570
R^2	0.83	0.80	0.77	0.50

where X are parameters and 'a' is Constant .

Using the relationship given in Table 2, one can predict the required parameters by measuring the other parameters value. This will considerably save the resources (Man,Material and Cost) at the preliminary stage of a project. Most of the geotechnical design parameters are correlated with the Standard Penetration Test (SPT) N value. The following relationship between the SPT N values and the remaining properties is derived as shown in equation (2). This will save the resources in preliminary stage of a project.

N-Value= - 52.56 * X1 –24.34 * X2 + 0.073*X3 + 1.67 * X4 - 1.45 * X5 -0.46 * X6 -0.14 * X7 -0.47 * X8 + 0.13 * X9 - 0.01 * X10 - 0.22 * X11 + 2570.7 (2)

A very powerful test for testing the significance of discrepancy between theory and experiment is known as Chi-square test of goodness of fit. The value of the test-statistic is

$$\chi^2 = \sum \frac{(O_i - P_i)^2}{P_i} \quad (3)$$

Where, χ^2 = the test statistic that asymptotically approaches a $\chi 2$ distributionO_i= an observed frequency and P_i= an expected frequency, asserted by the null

hypothesis.Table 3, shows the Chi Square Goodness of fit value of the predicted parameters. The calculated values for laboratory and field test parameters are less than the tabulated value (140.2 for more than 100 degree of freedom). It shows that the obtained relationships were found to be statistically significantaccording to the Chi square test. Hence, the model developed is an adequate one.

Table 3: Test Value of Chi-Square Goodness of Fit

Description	**LL**	**PL**	**DFS**	**N**
Chi-square goodness of fit	41	41	118	136

Quadratic Regressionanalysis

In this study, quadratic regression analysis is tried to correlate the variables. In Quadratic regression analysis, the model is of the type, $Y= b_1 x_1 +b_1x_1^2+ b_2 x_2 + b_2x_2^2$

$..+ b_n x_n^2 + a.$ (4)

An example of correlating the SPT 'N' value to the dependent parameters is briefed here. Since the maximum number of independent variables is limited to 8, different trialshas been tried by considering various parameters as listed in Table 4 to find out the set of parameters which give maximum R^2 value. From the trails, the independent parameter listed in trial T4, gives higher R^2 may be used to derive the quadratic regression equation as that of linear regression equations.

Comparison With Other Studies

The model developed to predict SPT'N' value is compared with the previous studies presented by Gandhi mathi et al [18] for Coimbatore area and Sakundaladevi et al [11] for Chennai area as shown in Table 5.

Table 4: Impact of Independent Variables on Quadratic Regression Analysis

Trials	Parameters considered	R^2
T1	X1,X2,X3,X4, X5,X6,X7,X8	0.505
T2	X1,X2,X4,X5, X6,X7,X8,X11	0.543
T3	X1,X2,X3,X5, X7,X9,X10,X11	0.539
T4	X1,X2,X4,X5, X6,X8,X10,X11	0.57
T5	X1,X2,X3,X5, X7,X8,X10,X11	0.54
T6	X1,X2,X4,X5, X6,X8,X9,X10	0.52

It is cleared that, in all studies, thetested quadratic regression model produce high R^2 values. Similarly more number of observations and more number of independent

parameters predicts the dependable parameters accurately. Hence the developed model with increasing future data in the study area of Salem block will serve as a better tool to predict the parameters in unvisited locations.

Table 5: Comparison With Previous Studies

Description	**Studies by**		
	1	**2**	**3**
Regression statistics for Linear relationship			
No of Parameters	10	6	11
Multiple R	-	0.627	0.713
R^2	0.680	0.394	0.508
Adjusted R Square	0.660	0.361	0.443
Standard Error	5.257	28.523	7.236
No of Observations	138	111	104
Regression statistics for quadratic relationship			
No of Parameters	-	6	8
Multiple R	-	0.707	0.755
R Square	-	0.500	0.570
Adjusted R Square	-	0.427	0.491
Standard Error	-	27.220	6.555
No of Observations	-	111	104

Note: 1 Gandhimathi et al ,2 Sakundaladevi et aland 3 Arulkumaran et al

Regression Analysis Using Matlab

Matrix Laboratory (MATLAB) is numerical computing, user friendly software which is developed by Math work. This is a language of technical computation tool useful for matrix form of analysis. The collected data form a matrix of 12 parameter from 104 locations (104 x 12). MATLAB function ‘regress' addresses the multiple linear regression problems based on the least squares approach, estimates the model parameters and performing the regression model automatically. The calculation of R^2 using MATLAB Regression Model and MS Excel models are compared as shown in Fig 1.Though the R^2 value for both models are in equal range, the predicted value for each site is different ,thus gives the range of parameters to the user.

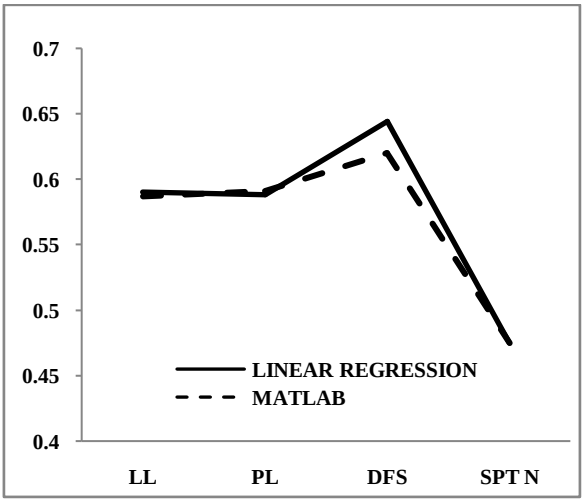

Figure 1: Comparison of R^2values of Developed Models

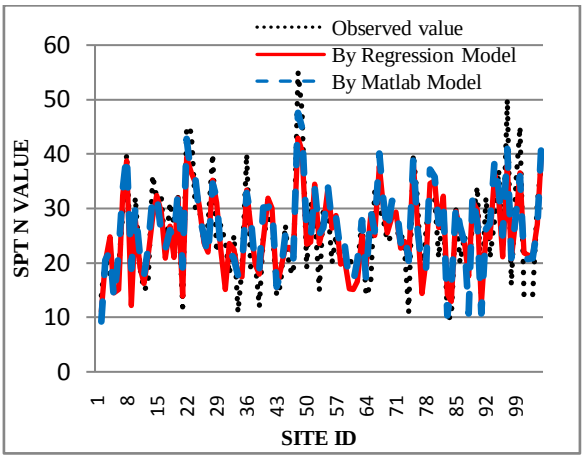

Figure 2: Prediction of SPT N Value For Coimbatore Data

Prediction of SPT N value for the Coimbatore data using old regression model developed by Arumairaj[10] and new MATLAB model is compared as shown in Fig 2. Table 6 and Table 7compares the performance of developed models for Coimbatore and Salem Data.

Table 6: Performance of MATLAB Model

Description	Chi Square Value
In prediction of SPT N Value for Coimbatore data	112.90
In prediction of SPT N Value for Salem data	123.66

Table 7: Comparison of Regression models with observed values in Salem block

Selected Properties	Observed Values	Predicted values bybegression Analysis	
		Excel	MATLAB
Site1- Latitude-11.6825, Longitude-78.1392			
LL (%)	21	-2.02	20.26
PL (%)	10	0.74	9.4
DFS (%)	16	15.62	19.03
SPT-N	40	39.63	38.15
Site 2- Latitude-11.6611, Longitude-78.1967			
LL (%)	15	6.28	21.30
PL (%)	0	-0.57	10.86
DFS (%)	0	18.48	19.37
SPT-N	41	38.02	38.02
Site 3- Latitude-11.6570, Longitude-78.1834			
LL (%)	23	5.90	20.6
PL (%)	12	-0.79	10.28
DFS (%)	15	16.52	18.39
SPT-N	42	40.09	38.75

Conclusions

Multiple regression (MR) is a powerful technique for standardizing data analysis.The results of analysis shows that MR models using MS Excel and MATLAB predict the geotechnical parameters within the acceptable accuracy. The Chi Square Goodness value for the parameters predictions are less than the tabulated value (140.2 for more than 100 degree of freedom). Accuracy of Quadratic regression is greater than linear regression. When the order of the relationship is increased from first order (linear) to second order (quadratic), thecoefficient of determination $R^2$for the SPT N value increases from 50.8 % to 57.0 %.Linear regression analysis using MATLAB model have an advantage of calculating and plotting the analysis result in a single programme. The models can be used to analysis the data from any study area. The authors' emphasis that the procedure used here is to estimate the probable value for the preliminary planning purpose of a project,and not for omission of detailed site investigation at any circumstances. The developed models are useful in the field for practicing engineers in predicting the required geotechnical properties across any study area data. The prediction value from different models gives the range of parameters.

Referances:

[1] Hammah, R. Eand Curran, J.H. ," Geostatisticsin Geotechnical Engineering: Fad or Empowering?", Geo Congress Conference, Atlanta, USA, pp1-5, 2006.

[2] Mohamed A. Shahin, Mark B. Jaksa and Holger R. Maier,"Recent Advances and Future Challenges for Artificial Neural Systems in Geotechnical Engineering Applications", Advances in Artificial Neural Systems, Article ID 308239, 2009.

[3] Nihat Sinan Isik ," Estimation of swell index of fine grained soils using regression equations and artificial neural networks", Scientific Research and Essay, Vol.4 (10), pp. 1047-1056,2009.

[4] Isık Yilmaz and Oguz Kaynar," Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils", Expert Systems with Applications,Vol 38,pp5958-5966,2011.

[5] Pijush Samui, Thallak G. Sitharam ,"Applicationof Geo statistical models for estimating spatial variability of rock depth', Journal of Engineering, Vol 3,886-894,2011

[6] Gokcen Yonter and Huriye Uysal, "The determination of the relationships between physical and chemical properties of soils to water erosion and crust strengths in Menemen Plain Soils, Turkey", African Journal of Agricultural Research ,Vol. 7(2), pp 183-193,2012.

[7] Miroslov Miroslav Radovanovic and Milos Madic," Modelling of surface roughness using MRA and ANN method', ANNALS of the Ordea University.- Fascicle of Management and Technological Engineering, Volume X (XX),PP 281-286,2010

[8] ZohrehIzadifar," Modeling and Analysis of Actual Evapotranspiration using Data Driven and Wavelet Techniques", - M.S Thesis, Department of Civil and Geological Engineering University of Saskatchewan, Canada,2010.

[9] Angus M. Brown ," Step-by-step guide to non-linear regression analysis of experimental data using Microsoft Excel spreadsheet", Journal of ComputerMethods and Programs in Biomedicine,65,pp191–200,2001.

[10] Arumairaj P. D , "GIS based spatial analysis and Modeling of soil in Coimbatore for Geotechnical engineering purposes", Ph.D Thesis, Bharathiyar University, Coimbatore,2006.

[11] Sakunthala Devi. S., and Stalin V,K,," Development of soil suitability map for Geotechnical applications using GIS approach", Indian Geotechnical Conference,pp 797-800,2011.

[12] Scott H. Brown," Multiple Linear Regression Analysis: A Matrix Approach with MATLAB",Alabama Journal of Mathematics,2009.

[13] Esteban Hormazabal Zuniga, Pamela Ortubia Fernandez and Francisco Rovira Frez," Some Applications to mine Geotechnical Design using MATLAB", Geotechnia, 2007.

[14] Hassan A.N. Hejase, Ali H. Assi,"MATLAB-assisted regression Modeling of Mean Daily Global Solar Radiation in Al-Ain, UAE." In: Assi AH (Ed) Engineering Education and Research using MATLAB, ISBN 978-953-307-656 Chapter 8, pp. 195-218, http://cdn.intechopen.com/pdfs-wm/21381.pdf,2011

[15] Margaret Segou," Ground Motion Estimation During Strong Seismic Events Using MATLAB" , Scientific and Engineering Applications using MATLAB,2011.

[16] T.P. Robinson and G. Metternicht," Testing the performance of spatial interpolation techniques for mapping soil properties", Journal of computers and electronics in agriculture ,50,pp97-108,2006.

[17] Teruhisa Masada," Shear Strength of Clay and Silt Embankments ", Report No. FHWA/OH-2009/7, The Ohio Department of Transportation (ODOT) and the U.S. Department of Transportation, Federal Highway Administration, 2009.

[18] Gandhimathi A, Arumairaj P.D, Lakshmipriya L. and Meenambal.T ," Spatial Analysis of Soil in Coimbatore for Geotechnical Engineering Purposes", International Journal of Engineering Science and Technology, Vol. 2(7), pp 2982-2996,2010.