Regression Analysis of Soil Compaction Parameters Using Support Vector Method

Osman Günaydın[1], Abdurrahman Özbeyaz[2*], Mehmet Söylemez[1]

[1]Faculty of Engineering, Department of Civil Engineering, Adiyaman University, Adiyaman, Turkey
[2]Faculty of Engineering, Department of Electrical Electronics Engineering, Adiyaman University, Adiyaman, Turkey
*aozbeyaz@adiyaman.edu.tr

Abstract
Some challenging studies are experimentally applied for characterizing parameters in Proctor compaction tests. Compression of a fill is mechanically done in Compaction process. Compaction is a physical process which gets the soil into a dense state. Improving the shear strength and decreasing the compressibility and permeability of the soil can be done with this physical process. Support Vector Machine (SVM) is a popular method due to its performance today. This method is commonly employed in the regression analysis as well as being used in the classification process. In this study, SVM was employed to predict of compaction parameters (maximum dry unit weight and optimum moisture content) without making any experiments in a soil laboratory. In the study, more than a hundred compaction data collected from the small dams in central Anatolia region was employed. In the study, R errors are satisfied (0.92 and 0.89) for SVM models. Consequently, the proposed regression analysis with SVM is useful for model design of the projects in where there are limitations as financial and temporal.
Keywords: Compaction, Maximum Dry Weight, Optimum Moisture Content, SVM, Regression.

1. Introduction

Compaction is the process of increasing soil density by applying mechanical energy [1]. Compaction is very important in engineering projects such as road airfields, earth dams, landfill construction, etc. With the help of the compaction process, seats under the fixed and dynamic loads of the floors are reduced, wear ability can be reduced or delayed; liquefaction features can be eliminated, volume changes due to affect such as frost, swelling, and shrinkage can be controlled by permeability is reduced, and a more stable structure can be imparted to the ground.

Two important compaction parameters which are the standard dry weight unit (γk_{max}) and optimum water content (W_{opt}) values can be obtained by standard compaction test in laboratory. In order to estimate compaction parameters, too many different studies have been tried by distinctive researchers. One of the first studies to determine the compaction parameters based on prediction was the correlation study among optimum water content, liquid limit and plasticity index [2]. Jumikis developed different methods to determine the dry unit volume weight (γk_{max}), and optimum water content (W_{opt}) on fine-grained soils [2]. McRae developed compaction classification index to estimate approximate compaction energy with compaction values belonging to different grounds [3]. Joslin developed the compaction curves titled as Ohio compaction curves, and it was stated that calculation of compaction values belonging to any grounds with the help of Ohio compaction curves was possible [4]. Johson and Salberg tried to determine approximate optimum water content for the standard compaction experiment using a chart, including plastic limit and liquid limit values [5]. In the following years different researchers have conducted various researches on Atterberg limit values and the predictability of optimum water content [6–10].

Regression analysis is a method used to measure the relationship between two or more variables. There are two types of regression analysis, one-variable regression and multivariate regression analysis. These analysis types are employed in many applications, especially in the engineering field. For example, in civil engineering, if we want to estimate the maximum dry unit weight or optimum water content from soil index parameters or California bearing ratio from compaction parameters, we can use regression analysis. In such problems, the function's curve/surface between input and output is tried to find. Thanks to this function, we can estimate the unrecorded values like motor powers, bio gas information from recorded input values in an experimental system. Because it is difficult to reach the experimental parameters in soil compaction analysis, outputs can be obtained by regression methods for untreated experiments [10].

Support vector machine (SVM), is a popular method, was developed by Vapnik [11], and it aims to find a special straight line for discriminating different classes. There are many possibilities for drawing a line between two classes. SVM tries to find the farthest line to both classes. Thus, fault tolerance is considerably increased. Boundary line in which SVM is near is tried to be found to distinguish the binary classes from each other. After the boundary, line is found using training data, test data are classified according to which side of the border they are. In SVM, samples which can't be linearly discriminated are moved to a higher-dimensional space. Samples are analyzed in this space. There are lots of methods for doing regression. These methods are single and multiple regression analysis, Artificial Neural Network and SVM. SVM is commonly used in both regression and classification analysis. In this study, we prefer SVM for providing higher estimation performance when compared the others [12, 13].

In this study, it is aimed to estimate the compaction parameters without performing laboratory experiments using SVM. For this purpose, 126 compactions and ground classification experiments were employed from different grounds (CH, GC, GM, MH, MI, ML SC) throughout Turkey. In the study, we used Matlab programming environment while SVM models were generated. We employed 126 rows data and 37 of 126 were chosen as test and the rest of data was used as train data. In result of the study, we measured correlation coefficients (R) as success in the regression process with the help of SVM models.

2. Materials and Methods

2.1 Characteristics of the studied soils

The concepts of grading, consistency limits (liquid limit and plastic limit), density and compaction parameters (γk_{max}, W_{opt}) have been the oldest and most fundamental in soil mechanics. These characteristics are commonly used to identify, classify and assess the soils properties. Grading can be readily appreciated by even the most untrained eye that gravel is a somewhat different material from sand. Likewise, silt and clay are also different too. Perhaps not quite so obvious is that it is not just the particle size is important but the distribution of sizes that make up a particular soil. Thus, the grading of a soil determines many of its characteristics. Specific density is usually of primary consideration where density values are used directly; to calculate earth pressures behind retaining walls or basements, for example, since it is the combined mass of soil and water that determines the pressure. The notion of soil consistency limits steams from the concept that soil can exist in any of four states, depending on its moisture content. However, the study has been investigated for only two states (the liquid and plastic limits). The liquid and plastic limits represent the moisture contents at the borderline between plastic and liquid phases and semisolid and solid phases (Figure 1).
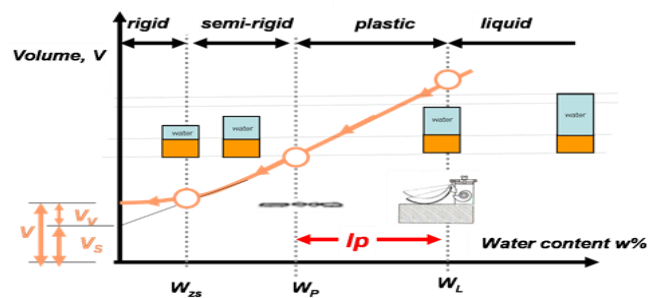

Figure 1. Atterberg (Consistency) limits.

The compaction parameters which are the maximum dry unit volume and the optimum water content are determined by the proctor test under the laboratory conditions. The optimum water content providing to obtain the maximum dry unit weight is called as the water content (Figure 2). Assessing optimum water content using maximum dry unit weight is important in the engineering projects and the required engineering parameters are determined for compressing the soils using these data.
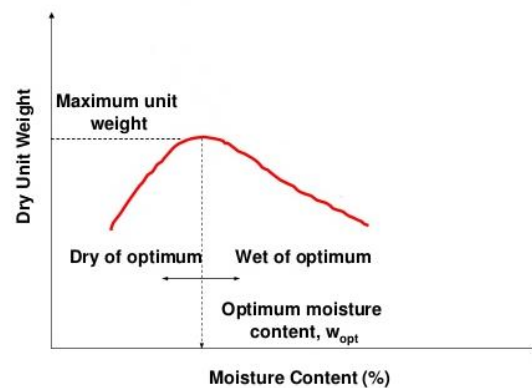

Figure 2. Compaction curve.

2.2 Data

Data obtained from ground tests made in different regions of Turkey were used in Support Vector Machine (SVM) experiments. The statistical values of the data used in analyzing section are shown in Table 1. According to the unified soil classification system (USCS), grounds are segmented as CH, CI, CL, GC, GM, MH, MI and ML. 126 data set are developed in regression models using SVM.

Table 1. Statistical values of the data used in the study.

	Silt (%)	Sand (%)	Gravel (%)	Gs (%)	LL (%)	PL (%)	W_opt (%)	Max DUW
Max	83,30	71,26	67,10	2,85	56,75	29,84	26,00	2,09
Min	13,00	15,49	0,05	2,58	23,10	13,68	7,60	1,43
Aver.	50,43	37,10	12,47	2,72	39,44	21,35	16,19	1,79
Hydr.	51,00	35,75	9,32	2,72	37,20	21,50	16,05	1,79
Skewn.	-0,16	0,54	1,44	-0,06	0,42	0,06	0,38	-0,06
Kurtosis	-0,57	-0,11	2,35	-0,33	-0,54	-1,12	0,75	0,22
Varyans	224,91	135,46	153,24	0,00	53,10	16,92	10,76	0,01
STD	14,97	11,98	12,37	0,06	7,20	4,06	3,27	0,12

2.3 Support Vector Machine Regression (SVM-R)

Support vector machine is a kernel based method and it is perfectly applies for regression and classification problems [14, 15]. This method was developed in 1998 by Vapnik firstly. SVM is robust, powerful and based on principles when compare the other machine learning tools. Linear SVM tries to find a regression function stated as $f(x) = w^T\varphi(x) + b$ in hyper plane. This function is obtained using "ε *insensitive*" loss function. Nonlinear SVM is obtained by solving the Squared Programming Problem (SPP) given in the following.

$$\min_{w,b,\varepsilon,\varepsilon*} \frac{1}{2}\|w\|^2 + C(e^T\varepsilon + e^T\varepsilon^*)$$
$$\text{s.t. } (\varphi(A)w + eb) - Y \le e\epsilon + \varepsilon, \varepsilon \ge 0e \quad (1)$$
$$Y - (\varphi(A)w + eb) \le e\epsilon + \varepsilon^*, \varepsilon^* \ge 0e$$

In this equation, C is both a predetermined parameter and an edit parameter maintaining balance between adaptation of mistakes and evenness of regression function. ε and ε^* are factitious variables showing whether samples enter or not to ϵ- tube. Dual of SPP are calculated as in the following equations using Lagrange multipliers (a and a^*).

$$\max_{a,a*} -\frac{1}{2}(a - a^*)^T K(A, A^T)(a^* - a) + Y^T(a^* - a)$$
$$+ \epsilon e^T(a^* + a) \quad (2)$$
$$\text{s.t. } (\varphi(A)w + eb) - Y \le e\epsilon + \varepsilon, \varepsilon \ge 0e$$
$$Y - (\varphi(A)w + eb) \le e\epsilon + \varepsilon^*, \varepsilon^* \ge 0e$$

We can find $a^{(*)} = (a_1, a_1^*, a_2, a_2^*, \dots, a_n, a_n^*)$ and b threshold values after SPP (2) is solved and then we obtain regression function.

$$f(x) = \sum_{i=1}^{n}(a^* - a) K(x_i, x) + b \quad (3)$$

In this statement, $K(x_i, x) = (\varphi(x_i) . \varphi(x))$ represents the kernel function and gives the dot product in hyper plane. a and a^* are Lagrange multipliers and provide $a_i a_i^* = 0, i = 1, 2, \dots, n$. Function $f(x)$ is stated by $a_i \ne 0$ and $a_i^* \ne 0$ samples (Support Vectors) which are Lagrange multipliers. Moreover, $A = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$ are input and output of training data respectively [15].

3. Results

In this study, the compaction parameters (γk_{max}, W_{opt}) were determined by SVM Regression method using experimental data collected from different regions of Turkey vicinity of Niğde. We employed cross-validates a support vector machine (SVM) regression model. SVM regression model was based on non-linear system so the model used radial basis kernel function (RBF). Thanks to the RBF function we carried system to multi-dimensional space with SVM regression, and we used six inputs and two outputs in the model. And we had two prediction models and finally we had two outputs for both γk_{max} and W_{opt} outputs. Moreover, two different models in which the number of input variables weren't changed, and inputs were preferred as six in the analyses. The six input variables used for Model-I and Model-II. The variables were fine-grained (FG, %), sand (S, %), gravel (G, %), specific density (G_s), liquid limit (W_L, %) and plastic limit (W_P, %). Hence, regression model was obtained according to the two outputs. We employed Matlab platform for regression analysis. Our SVM regression model was obtained and given for each model in Table 2.

We tested the convergence of the model and we had 1 and this value 1 indicated that the two models did converge. Iteration number was 126 when we set up the model. We used k fold cross validation method for modeling. In this study, k was generated as 30% of the data. We constructed a partitioned regression model, and examined the cross-validation losses for the folds and loss costs were obtained as 3,37 for Model-I and 0,45 for Model-II. At the end of the analyses, we obtained two predicted outputs. We compared the outputs with the true outputs for each model. And result values are given in Table 3 and Table 4 for each model respectively. Errors were calculated, and these values are given in third column in the tables. Mean Square Errors (MSE) were calculated as 0.96 for Model-I and 0.44 for Model-II respectively.

Table 2. The Model of Regression SVM.

Alpha:	[99x1] for Model 1 [112x1] for Model 2
Bias:	15.53 for Model 1 17.96 for Model 2
KernelParameters:	Function:'gaussian' Scale: 2.52 *for each model*
Mu:	[50,72 36,95 12,33 2,72 39,70 21,39]
Sigma:	[15.23 11.62 12.21 0.05 7.77 4.09]
NumObservations:	126 for each model
ConvergenceInfo:	1 for each model
Solver:	'SMO'for each model

Table 3. True labels and Predicted Labels and differences for Model I.

Iter.	True Labels (Outputs)	Predicted Labels (Fit Outputs)	Erros (True-Predicts)
1	13,80	14,31	0,51
2	14,60	14,27	0,33
…	…	…	…
36	20,40	18,75	1,65
37	17,85	17,63	0,22

Table 4. True labels and Predicted Labels and differences for Model II.

Iter.	True Labels (Outputs)	Predicted Labels (Fit Outputs)	Erros (True-Predicts)
1	18,60	18,63	0,03
2	18,37	18,85	0,48
…	…	…	…
36	16,89	17,22	0,33
37	17,02	16,86	0,16

In the study, SVM regression technic was employed to predict outputs across to six inputs. Two different models are investigated in the study. Optimum water content and the maximum dry unit weights were used as outputs in the first and second model respectively. R errors are satisfied (0.92 and 0.89) for two models in the study.

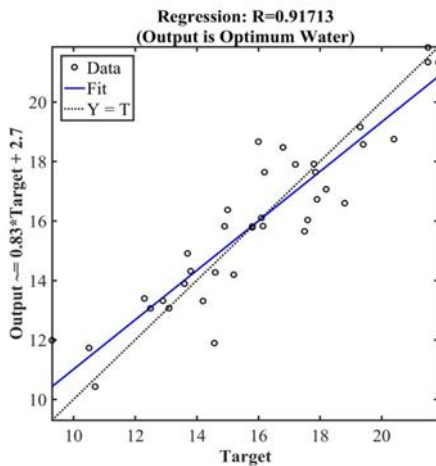In Figure 3 and Figure 4, R error graphics are given respectively.

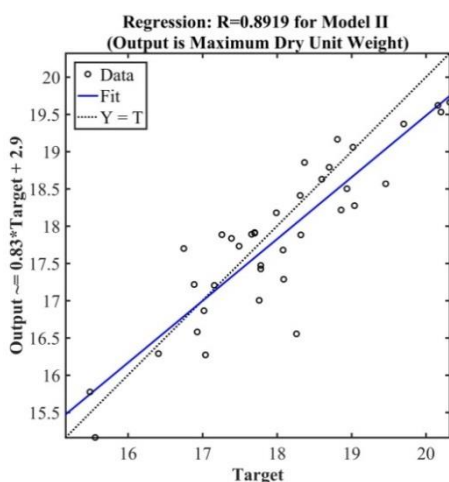**Figure 3.** Measured value of compaction parameters for Model-I.

Figure 4. Measured value of compaction parameters for Model-II.

Mathematical models of the data were obtained for two outputs, and the expressions were generated with some coefficients. Coefficients of mathematical models are given in the Table 5 for each model. Regression models for these mathematical models are given in the following figures for optimum water and maximum dry unit weight in Figure 5 and Figure 6 respectively.

Table 5. Coefficients belonging to mathematical equations for each model in Regression analysis.

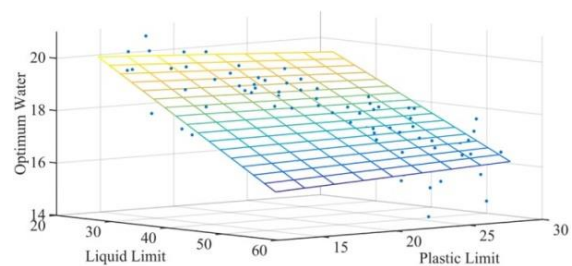| Outputs	b(0)	b(1)	b(2)	b(3)
W_{opt}	23,72	-0,17	-0,12	0,005
γk_{max}	24,25	-0,16	-0,08	0,002

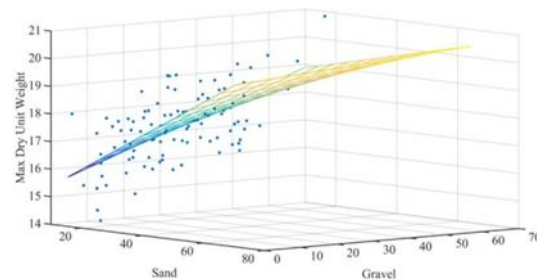**Figure 5.** Regression model is generated when optimum water is used as output.

Figure 6. Regression model is generated when maximum dry unit weight is used as output.

4. Conclusion

Analyzing of compaction parameters (γk_{max} and W_{opt}) from soil characteristics in the regression analysis using Support Vector Machine (SVM) algorithm was evaluated in the study. SVM method was tried in the models, and regression analysis estimation indicated strong correlation as 0.92 and 0.89 for two different models respectively. Consequently, the proposed regression analysis with SVM should be useful for a preliminary design of the project where there are financial and time limitations.

Acknowledgments

References
1. Holtz R.D, Kovacs W.D, Compaction, An Introduction to Geotechnical Engineering, New Jersey, USA: Prentice Hall, 1981, pp 109–161.

2. Jumikis A.R, Geology of Soils of the Newark (N.J.) Metropolitan Area, *Journal of the Soil Mechanics and Foundations Division*, 1958, 84(2), 1–41.

3. McRae J.L, Index of Compaction Characteristics, Symposium on Application of Soil Testing in Highway Design and Construction, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959: ASTM International, 1959, pp 119-127.

4. Joslin J, Ohio's Typical Moisture-Density Curves, Symposium on Application of Soil Testing in Highway Design and Construction, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959: ASTM International, 1958, 111-118.

5. Johnson A and Sallberg J, Factors influencing compaction results, Washington, DC: Highway research board bulletin no. 319, 1962; pp 171.

6. Al-Khafaji A.N, Estimation of soil compaction parameters by means of Atterberg limits, *Quarterly Journal of Engineering Geology and Hydrogeology*, 1987, 26(4), 359–368.

7. Sridharan A, Nagaraj H.B, Plastic limit and compaction characteristics of finegrained soils, *Proceedings of the Institution of Civil Engineers - Ground Improvement*, 2005, 9(1), 17–22.

8. Sridharan A, Gurtug Y, Compressibility characteristics of soils, *Geotechnical & Geological Engineering*, 2005, 23(5), 615–634.

9. Kayadelen C, Estimation of effective stress parameter of unsaturated soils by using artificial neural networks, *International Journal for Numerical and Analytical Methods in Geomechanics*, 2008, 32(9), 1087–1106.

10. Günaydın O, Estimation of soil compaction parameters by using statistical analyses and artificial neural networks, *Environmental Geology*, 2009, 57(1), 203–215.

11. Vapnik V, An Overview of Statistical Learning Theory, *IEEE Transactions on Neural Networks*, 1999, 10(5), 988–999.

12. Amasyalı M.F, Makine Öğrenmesine Giriş, https://slideplayer.biz.tr/slide/2285047/, 2006 (accessed July 15, 2018).

13. Smola A.J, Schölkopf B, A tutorial on support vector regression, *Statistics and Computing*, 2004, 14(3), 199–222.

14. Burges C.J.C, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, 1998, 2(2), 121–167.

15. İnce H, İmamoglu S.Z, Destek Vektör Regresyon ve İkiz Destek Vektör Regresyon Yöntemi ile Tedarikçi Seçimi, *Doğuş Üniversitesi Dergisi*, 2016, 17(2), 241–253.