AN INVESTIGATION ABOUT PREDICTION OF STABILITY PROPERTY OF BITUMINOUS HOT MIXTURES BY REGRESSION MODELS

Deniz ARSLAN[1]
Gazi University Engineering Faculty, Civil Engineering
Department Maltepe, Ankara
denizarslan@gazi.edu.tr

Hikmet BAYIRTEPE
Gazi University Engineering Faculty, Civil Engineering
Department Maltepe, Ankara
hikmetb@gazi.edu.tr

ABSTRACT

In this study, the stability property which is one of the basic characteristics of the bituminous hot mixtures was aimed to be predicted through the physical properties of the mixtures and the materials used. A data set of 4680 observations relating to the briquette and core samples were collected by this aim. Regression, principal component, cluster and discriminant analyses were performed on the data set. Cluster analysis has shown that the data set can be divided into four main clusters. This outcome was also wholly supported by the discriminant analysis. According to the stepwise regression results, the variance of the stability was determined to be explained up to 73.8 % by the independent variables.

Keywords: *Marshall stability, Regression analysis, Cluster analysis, Principal component analysis, Discriminant analysis*

[1] Corresponding author: Deniz Arslan Tel: +90 312 5823205; Fax: +90 312 2308434

1. INTRODUCTION

It is quite common to use flexible pavements in highway transportation. Flexible pavements can be examined into two groups as low and high standard pavements. Low standard flexible pavements are preferred in roadways with low traffic volume. High standard flexible pavements are composed of wearing, binder, base and subbase layers in general. Wearing and binder layers are constructed by laying and compacting of bituminous hot mixtures which are prepared in asphalt plants. Bituminous hot mixtures are manufactured by mixing of well-graded aggregate combination with bituminous binder (bitumen) at high temperature in an asphalt plant. The performance of the highway pavement in its service life is directly related with the properties of the bituminous hot mixtures one of which is the stability. Stability is the major property of any bituminous hot mixture and can be used as a measure of resistance to permanent deformations [1,2]. In Highway Technical Specifications [3], the lower limit of stability for wearing and binder layers were given as 900 kg and 750 kg, respectively and no upper limit was determined for anyone. Researches for enhancing stability of mixture [4-6] show the importance of that property. Stability of mixture can be defined in laboratory conditions with sensitive and relatively difficult procedures. The properties of the mixture constituents, aggregate and bitumen, is known to be effective on stability. Thus, stability can be determined more easily and effortlessly, if it can be modelled in terms of material and mixture properties. Regression, cluster and discriminant analyses are among the methods used in literature studies about data analyzing and establishing of prediction models [7-9]. Using those methods, unknown variables are tried to be predicted by known parameters. Although successfully results were obtained with small data sets [10,11], obscurity is still available for large data sets. The aim of this study is to predict the stability of bituminous hot mixtures from a large data set by means of regression models based on material and mixture properties. Regression, principal component, cluster and discriminant analyses are performed on the data set for this aim.

2. MATERIAL AND METHOD

2.1. Data Set

In this study, a data set of 4680 observations of bituminous hot mixtures applications for wearing and binder layers of highway pavement within the border of Great Municipality of Ankara in 2001 have been collected. Data set is composed of 2637 samples from wearing layer and 2043 samples from binder layer including the following features: gradation (%), bitumen ratio (%), environmental temperature (°C), stability (kg), compression ratio (%), weights related with specific gravity (g) and specific gravity.

2.2. Method

In this study; regression, principal component, cluster and discriminant analyses were performed on the data set. The methods and their application parameters were explained below;

Regression analysis; improves the prediction equality by using the concept of linearity and relationship among the variables. In regression analysis, variables are either named as dependent or independent. Once the relationship between variables was established, the values of dependent variable can be predicted by the known values of the independent variables. And the dependent variable, can be defined in association with the independent variables. Regression coefficient (R^2), shows the proportion of explanation of dependent variable variance by the independent variables [12]. In this study, the stability argument was designated as the dependent variable and tried to be predicted by using multiple regression analyses. Stepwise regression was performed on the new clusters formed after cluster and discriminant analyses in order to predict stability. Variables are included in the models by the stepwise regression model according to their power of expressing the dependent variable [13]. Variables are entered into and removed from the model by F-to-enter and F-to-remove values, respectively. During standard stepwise regression analyses, F-to-enter and F-to-remove values were taken as 0.15.

Principal component analysis is used to convert the variables having possible correlation to linearly uncorrelated structures called principal components [14]. The number of principal components is less than or equal to the number of original variables. The first principal component accounts for as much of the variability in the data as possible. During analyses, the number of principal component was selected in order to define at least 90% of the general variance.

Cluster analysis separates units and variables into homogenous groups. Group members (observations) are similar in each other whereas differs from other group members. The objective of cluster analysis is to identify groups of observations that are very similar so that the oppurtunity of increasing regression coefficient (reliability factor) occurs with groups (clusters) obtained from the analysis. Nearest neighbor method with 95 % similarity level was used in the analysis.

The functions that provide separation of groups are determined in discriminant analysis. Discriminant function is used to determine which variables discriminate between naturally occurring groups. It attempts to model the difference between the classes of data. A new observation is assigned one of the previously defined groups by means of the calculated functions with a minimum clustering error. Discriminant analyses have been applied in order to control the clustering results.

3. RESULTS AND DISCUSSION

Regression analyses have been applied on the data set of 4680 observations firstly. Tests are realized by taking the stability as the dependent variable while independent variable group includes aggregate gradation properties in 3 different forms as % passing from sieves, % retaining on sieves and aggregate portions of coarse, fine and filler. Independent variable group also includes the measurements related with the specific gravity of briquette and core samples, compression ratio, void ratio, environmental temperature, bitumen percent and surface area properties. The logarithmic and inverse forms of some independent variables can be seen in some applications (Table

1). The regression coefficients obtained from 10 different combinations of the independent variables have been summarized in Table 1.

Table 1. Regression analyses results.

Combination		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
R^2 (%)		37.8	38.4	36.6	38.6	39.5	39.5	38.6	38.6	31.5	31.5
Dependent Variable		St	St	St	St	St	St	St	St	St	St
	3/4"	1	1	1	1	1	1	1			
	1/(3/4")								1	1	1
	1/2"			1	1	1	1	1	1	1	1
	3/8"			1	1	1	1	1	1	1	1
	No 4	1	1	1	1	1	1	1	1	1	1
	No 10	1	1	1	1	1		1	1	1	1
Sieves	No 40	1	1	1	1	1	1	1	1	1	1
	No 80			1	1	1	1	1	1	1	1
	No 200			1		1		1	1	1	1
	Coarse										
	Fine										
	Filler				1	1	1	1			
	1/Filler								1	1	1
% retaining					1	1	1		1	1	1
% passing		1	1	1							
	Height	1	1	1	1	1	1	1	1	1	1
	AD	1	1	1	1	1	1	1	1	1	1
	Weight (in water)	1	1	1	1	1	1	1	1	1	1
Briquette	SSD	1	1	1	1	1	1	1	1	1	1
	Volume	1	1	1	1			1			
	1/Volume					1			1	1	1
	Specific gravity	1	1	1	1	1	1	1	1	1	1
	Height	1	1	1	1	1	1	1	1	1	1
	AD	1	1	1	1	1	1	1	1	1	1
	Weight (in water)	1	1	1	1	1	1	1	1	1	1
	SSD	1	1	1	1	1	1	1	1	1	1
Core	Volume	1	1	1	1			1			
	1/Volume					1	1		1		1
	log(Volume)									1	
	Specific gravity	1	1	1	1	1	1	1	1	1	1
	% compression		1	1	1	1	1	1	1	1	1
	Void		1	1	1	1	1	1	1	1	1
Environmental temperature		1	1	1	1	1	1		1	1	1
Bitumen percent		1	1	1	1	1	1		1	1	1
Surface area					1	1			1	1	1

"1" shows the independent variables used in the analyses while empty cells mean that the variable was not used in the test. "St" expresses stability which was used as dependent variable in the regression analyses. As seen in Table 1, regression coefficients are found to vary between 31.5 % and 39. 5 %. Higher results were obtained in small data sets [10,11]. Regression coefficient is the measure of the explanation of the variance of the dependent variable by the independent variables used in the test. The regression coefficients which are varied between 31.5 % and 39.5 % are found to be insufficient to explain the dependent variable.

Data set includes observations from different layers of the pavement. It is required to cluster observations with common features after having insufficient regression results performed on the whole data set. Thus, it will be possible to obtain higher regression coefficients in clusters containing homogeneous observations. Cluster analysis was applied to the data set followed by the principal component analysis. Such an approach can be preferred in data analysing studies [7]. Principal component analysis was performed due to the possible correlation between independent variables which breaks up full independence concept of the variables. The sieves have been divided into two parts as coarse-fine and % retaining on each sieve have been used in determination of principal components for gradation. The variables used for coarse aggregate, fine aggregate, briquette and core samples together with principal component coefficients (PCC) have been shown in Table 2. AD and SSD represent air-dry and saturated surface dry weigths, respectively.

The standardized variables that have been used in cluster analysis are the principal components accounting for at least 90 % of the variability in the data set for the coarse aggregate, fine aggregate, briquette and core samples. Filler ratio, environmental temperature, bitumen percent, compressing and void ratios of core samples were also used. As seen in Table 2, it was sufficient to use 2 PCC for coarse aggregate, 3 PCC for fine aggregate, 2 PPC for briquette and 1 PCC for core samples in cluster analysis.

Table 2. Variables used for coarse-fine aggregate, briquette and core samples with their principal component coefficients.

Variables	Principal Component Coefficients		
Coarse aggregate	*PCC1(%87,6)*	*PCC2(%7,5)*	*PCC3*
	1"= 0	0	
	3/4"= -0,28	-0,146	
	1/2"= -0,567	0,588	
	3/8"= -0,032	-0,715	
	No4= 0,774	0,348	
Fine aggregate	*PCC1(%58,5)*	*PCC2(%25,1)*	*PCC3(%10,8)*
	No10= -0,883	-0,457	0,036
	No40= 0,465	-0,881	0,002
	No80= 0,005	0,061	-0,793
	No200= 0,057	0,11	0,608
Briquette	*PCC1(%90,1)*	*PCC2(%8,8)*	*PCC3*
	Height= -0,027	0,022	
	AD= -0,598	0,55	
	Weight (in water) = -0,375	0,361	
	SSD= -0,678	-0,734	
	Volume= -0,202	0,165	
	Specific gravity= 0	0	
Core	*PCC1(%99,4)*	*PCC2*	*PCC3*
	Height= -0,024		
	AD= -0,63		
	Weight (in water) = -0,37		
	SSD= -0,63		
	Volume= -0,261		
	Specific gravity= 0		

146 clusters were obtained from the cluster analysis. 4 of those clusters vary from others in terms of the number of observations they contain. They contain 798, 1104, 279 and 1218 observations respectively while the number of observation of the other 142 clusters vary between 3 and 44.

Discriminant analyses have been performed on dominant 4 clusters. The principal components of coarse aggregate, fine aggregate, briquette and core samples together with filler ratio, environmental temperature, bitumen percent, compression ratio, void ratio were used as variables in the analyses. The success rate was found to be 100 % in the test. In other words, observations of the clusters stay unchanged thus, cluster analyses results were verified completely. The distances between clusters have been given in Table 3.

Table 3. Distance between clusters.

Cluster	1	2	3	4
1	0	625.11	98.209	818.09
2	625.11	0	668.5	119.57
3	98.209	668.5	0	658.72
4	818.09	119.57	658.72	0

After examining observations of dominant clusters, it has been understood that the observations in Clusters 1+2 and in Clusters 3+4 conform wearing and binder gradings, respectively. Additional assessment related to the dominant clusters was carried out for the stability values which varies between 875 kg and 2009 kg. Histogram of the stability values were presented in Fig 1.

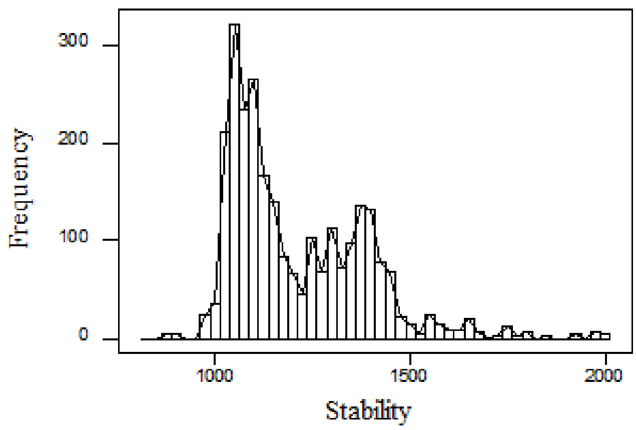

Fig 1. Histogram of the stability values.

The histogram of the stability values was seem to be formed by the combination of two normal distributions. Considering totally 3399 observations from 4 dominant clusters which belong to the two different layers of pavement as wearing and binder, such a histogram has already been expected. But; when the other variables had been incorporated to the analysis, 4 clusters were obtained which means that observations related to wearing and binder layers have been divided into 2 clusters in each other. In order to examine this situation, normal probability plot of the stability values has been generated and shown in Fig 2.

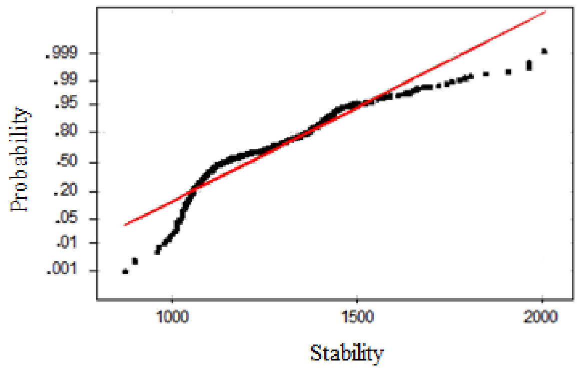

Fig 2. Normal probability plot of stability values.

Each linear part of the normal probability plot shows a normal distribution. Thus, stability values have been estimated to have 4 normal distribution due to the 4 linear parts of the curve in Fig 2. The findings of cluster analysis of 4 dominant cluster have been found to be supported by normal distribution approach by this way. In the following steps, 4 dominant clusters and their combinations shown on Table 4 have been used to predict the stability parameter by the established regression models.

Table 4. Groups subjected to analysing and the clusters they occur.

Group	Cluster(s) in Group
All groups	Cl 1+Cl 2+Cl 3+Cl 4
All wearing groups	Cl 1+Cl 2
All binder groups	Cl 3+Cl 4
1. Wearing group	Cl 1
2. Wearing group	Cl 2
1. Binder group	Cl 3
2. Binder group	Cl 4

Regression analyses have been performed on the clusters or combinations of clusters using the independent variables of filler ratio, environmental temperature, bitumen percent, compression ratio and void ratio together with the principal components created for each cluster combination which are subjected to the regression analyses. The insufficient regression coefficients obtained in the previous analyses indicate that there are non-linear relationships between the dependent and independent variables. This necessity can be provided by linearizing the relationship through different mathematical forms of independent variables. (1/x), x^2, x^3, e^x forms of the independent variables have been calculated and added to the analyses for this aim. But, regression models have been considered to become complicated by such a large number of independent variables in this case. Stepwise regression analyses have been performed on the clusters using independent variables and their different mathematical forms so that stability have been aimed to modeling with the optimal independent variables and with their optimal forms. Totally 60 independent variables have been used in the stepwise regression analyses with F-to-enter and F-to-remove statistics of 0.15. The results of the analyses have been summarized in Table 5.

Table 5. Standard stepwise regression analyses results for different cluster combinations.

Cluster combinations	Regression coefficient, R^2 (%)	# of independent variable in regression equation
Cl 1+Cl 2+Cl 3+Cl 4	60.9	35
Cl 1+Cl 2	61.3	20
Cl 3+Cl 4	41.5	33
Cl 1	53.1	33
Cl 2	33.8	27
Cl 3	73.8	17
Cl 4	37.2	40

The incorporation of independent variables to the regression equation from the data pool continues as long as F-to-enter and F-to-remove conditions are provided. When F-to-enter and F-to-remove conditions are not provided, stepwise regression analysis is terminated for the cluster it is applied. The number of the variables incorporated to the regression equation can change according to F values used. As seen in Table 5, number of independent variables in the regression equations is different for different combinations of clusters. In other words, it has not been able to develop a common regression model for each cluster. Besides, regression coefficient was found to be improved by the further studies up to 73.8 %. However, the aim of describing the variance of the stability with common and minimum number of independent variable has not been reached at a sufficient level.

4. CONCLUSION

A data set of 4680 observations of bituminous hot mixtures applications for wearing and binder layers of the highway pavement within the border of Great Municipality of Ankara in 2001 were collected in this study. Stability has been chosen as the dependent variable and tried to be predicted by using regression analyses on the basis of independent variables derived from material and mixture properties. According to the findings, regression models can’t able to reach sufficient level of describing the dependent variable as the maximum regression coefficient obtained in the models have been determined to be 73.8 % for the sub-cluster. It has been observed that the independent variables are not significantly effective on the production results meaning that material and mixture properties are not sufficient enough to define the variance of the stability.

ABBREVIATIONS AND SYMBOLS

AD	: Air-dry weight
Cl 1	: Cluster 1
Cl 2	: Cluster 2
Cl 3	: Cluster 3
Cl 4	: Cluster 4
PCC	: Principal component coefficient
SSD	: Surface-saturated weight
St	: Stability

REFERENCES

[1] Tunç, A., "Yol Malzemeleri ve Uygulamaları", ***Atlas Yayıncılık***, İstanbul, (2001).

[2] Hınıslıoğlu, S., Ağar, E., "Use of waste high density polyethylene as bitumen modifier in asphalt concrete mix" ***Materials Letters*** 58: 267-271, (2004).

[3] KGM, "Karayolları Teknik Şartnamesi", Karayolları Genel Müdürlüğü, Ankara, (2013).

[4] Sangita, Khan, T.A., Sabina, Sharma, D.K., "Effect of waste polymer modifier on the properties of bituminous concrete mixes" ***Constr Build Mater*** 25: 3841-3848, (2011).

[5] Arslan, D., Gürü, M., Çubuk, M.K., "Performance assessment of organic-based synthetic calcium and boric acid modified bitumens" ***Fuel*** 102: 766-772, (2012).

[6] Arslan, D., Gürü, M., Çubuk, M.K., "Improvement of hot mix asphalt performance in cold regions by organic-based synthetic compounds" ***Cold Regions Science and Technology*** 85: 250-255, (2013).

[7] Ghosh, S., Chatterjee, R., Shanker, P., "Estimation of ash, moisture content and detection of coal lithofacies from well logs using regression and artificial neural network modelling" ***Fuel*** 177: 279-287, (2016).

[8] Mirzahosseini, M.R., Aghaeifar, A., Alavi, Gandomi, A.H.,Seyednour, R., "Permanent deformation analysis of asphalt mixtures using soft computing techniques" ***Expert Systems with Applications*** 38: 6081-6100, (2011).

[9] Busch, J.M., Fortney, W.G., Berry, L.N., "Determination of lithology from well logs by statistical analysis" ***Formation Evaluation*** 2: 412-418, (1987).

[10] Aksoy, A., Iskender, E., Kahraman, H.T., "Application of the intuitive k-NN Estimator for prediction of the Marshall Test (ASTM D1559) results for asphalt mixtures" ***Constr Build Mater*** 34: 561-569, (2012).

[11] Özgan, E., Serin, S., Kap, T., "Multi variety investigation into the effects of hot-mix asphalt parameters on the marshall stability" ***6th International Advanced Technologies Symposium***, Elazığ-Turkey, 101-108, (2011).

[12] Ünver, Ö., Gamgam, H., "Uygulamalı İstatistik Yöntemler", ***Siyasal Kitabevi***, Ankara, (1996).

[13] Özdamar, K., "Paket Programlar ile İstatistiksel Veri Analizi", 1, ***Kaan Kitabevi***, Eskişehir, (1999).

[14] Özdamar, K., "Paket Programlar ile İstatistiksel Veri Analizi", 2, ***Kaan Kitabevi***, Eskişehir, (2002).