CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

CLUSTER BASED REGRESSION MODELING FOR PREDICTING CONDITION RATING OF HIGHWAY TUNNELS

Sahar, Hasan[1,2] and Emad, Elwakil[1,3]
[1] School of Construction Management, Purdue University, West Lafayette, IN, 47907, USA.

[2] Construction and Project Management Institute, Housing and Building National Research Center (HBRC), Egypt.

[3] eelwakil@purdue.edu

Abstract: Highway tunnels are long-term projects that committed with higher capital costs and additional levels of maintenance, because of its complex systems. However, fewer deterioration models have been built compared with the other highway components such as bridges. The limited research in this area has raised the need to develop prediction models for the tunnel condition based on tunnel inventory database (NTI); that provides objective decisions for future maintenance plans. This paper has investigated the significant impact of dependent variables on tunnel conditions from three aspects: Geometric, Inspection, Structure Type and Material. The proposed methodology is based on two phase's analysis; two- step cluster analysis and regression modeling. Nine models have been developed with high coefficient of determination (R^2=90.8%), classified in terms of service in tunnel and ground condition. Further analysis using average validity percentage (AVP) method has been used to examine the validity of built models that come out with satisfied results (83%). The developed model will help highway authorities to prioritize the maintenance and make informed investment decisions in an objective manner based on the historical data.

1. INTRODUCTION

Highway Tunnels are critical part to improve the mobility of underground highway. Tunnel structure is a long-term project; with more intensive needs for maintenance and operation than traditional transportation facilities because of its complicated systems (Thompson et al. 2011, Schiessl, Gehlen & Kapteina 2004). Therefore, developing a prediction model for tunnels condition rating to guide operators and owners to prepare the needed improvements, and support the future decisions (Ahluwalia 2008). In previous research work, although, the tunnels are structurally different from the bridges, tunnels have been inspected and incorporated as a branch of bridges highway (ODOT 2016). Therefore at that time, the inspected data were not expressing the tunnel condition, even so these data were scattered and did not contain the related Geometric, Operation, and Condition data of the tunnels' elements.

Recently in 2017, Federal Highway Agency (FHWA) has lunched the 2015 preliminary database for National Tunnels Inventory (NTI), in order to document the tunnel inventory and inspection data of managed tunnels in USA. According to Allen et al. (2015) and Bergeson & Ernst (2015), the Specifications for the National Tunnel Inventory (SNTI) defines the general extent of deficiencies for each of the four condition states: Good, Fair, Poor, and Severe. Where, Quantities are developed for each element and condition states are assigned for the percentage of this quantity in each condition state. However, up to now these inspection data and condition assessment are still not completed in the available preliminary NTI data. Moreover, assessing tunnel conditions under inspected defects (if existing) is considered as corrective not proactive maintenance strategy. Therefore, this paper implemented the clustering analysis as explanatory technique in order to figure out the dummy patterns among the historical data of NTI database for modeling the predicting tunnel condition.

2. BACKGROUND

The Standard for Highway tunnels, Bridges, and Other Limited Access Highways has defined the tunnels as following: “The tunnel is an underground highway that in many cases is equipped with many of interrelated systems and components that need to be properly inspected, maintained, and tested to ensure that the tunnel is performing reliably as designed” (Thompson et al. 2011).

Tunnels deteriorate in various ways due to material used, aging, operation factors, and exposure to harsh environmental conditions .Deterioration models for tunnels have been developed to account their continually or gradually variation over time; by estimating and predicting the performance. Therefore, the successful deterioration modeling should consider the deterioration condition, tunnel structure, used material, and repair history (Wang 2018).

Many researches have been conducted based on the physical mechanism of tunnels deterioration, while neglecting the statistical estimation method. Kobayashi, Kaito & Lethanh (2010) have agreed with this point of view. Where, their study has developed an analytical methodology using the multistage Weibull hazard model for forecasting the deterioration process of infrastructure facilities. The deterioration process has been represented by a transition pattern among multiple condition states, in order to estimate the transition probability of condition states for any arbitrary time intervals. That methodology has based on the observed condition states and characteristic variables. The developed model was validated using database of tunnel lighting reflectors of express highways in Japan. The results have guided tunnel administrators to improve the inspection and maintenance plan.

According to Meguid and Dang (2009), the tunnel structural degradation is affected by ground movement; appears as transverse or longitudinal cracks and crushing at the inside surface of a tunnel, as well as sudden leakage. These impacts were variated according to the surrounding ground condition of the tunnel. In this context, Andersson & Söderhäll (2001) have studied the tunnel maintenance from different perspective. Where it addressed the impact of geology and geometry (tunnel shape) on tunnel maintenance. The study concluded that there is an indication that the tunnel alignment to the stress field has an impact on the need for maintenance. Where, the maintenance work in the horseshoe shaped tunnel TASF has been compared to the maintenance work in the TBM tunnel. It was noticed that no maintenance has been performed in the TBM tunnel but TASF have been scaled and bolted several time and was finally shotcreted.

Because of scarce of literature addressing tunnel’s deterioration using inspection data, subway structures were reviewed in this study as the closest infrastructure models to highway tunnels. Gkountis (2014) has developed an assessment model for subway network performance, which included tunnel as one of its main components. The followed methodology has based on identified actual defects to develop a condition assessment model as first step. It utilized The Analytic Hierarchy and Network Processes (ANP) for weights and the Technique of Order Preference by Similarity to Ideal Solution (TOPSIS) for the aggregation. Secondly, a condition prediction model was developed using Weibull theory, which has built deterioration curves for all the network levels. However, his study has focused on electrical and mechanical defects for tunnel as part of subway network while neglected the structural system. Earlier, Semaan (2011) has developed a model assessing the condition of structural components based on defects of stations, tunnels and auxiliary structures. The different defects were weighted with the AHP. The distribution of defects of tunnels was related to the main tunnel elements; dome, sidewalls, and bottom slab. That system has depended on the structural behavior of the tunnel components themselves.

It can be revealed that prediction models for highway tunnel condition is uncovered area. Where most of studies have concentrated on studying the tunnel performance based on defects from the structural behavior. Therefore, there is a crucial need to develop deterioration models based on inventory database, where they reflect a real picture of tunnel performance affected by the actual operational and environmental items not only the designed ones. Thus, developing the prediction models would help highway authorities to improve design and proactive maintenance plans.

This paper has worked on NTI database, where it proposed to study the influence of the related variables of Service, Geometric Data, and Structure Type and Material items; as they express the operational and environmental parameters. These variables composed of Age, Annual Average Daily Traffic, Annual Average Daily Truck Traffic, Tunnel Length, Road Width, and Service in Tunnel, Tunnel Shape, Number of Bores, and Ground Conditions.

3. METHODOLOGY:

Based on the literature review and data in hand, this study has proposed a methodology composed of two phases; the clustering analysis as the first statistical analysis to figure out the hidden patterns among the collected tunnel data. Then followed by the regression modelling based on the satisfied clustered groups to examine the relevance between the significant independent variables and tunnel condition. This methodology has been illustrated graphically in Figure 1.

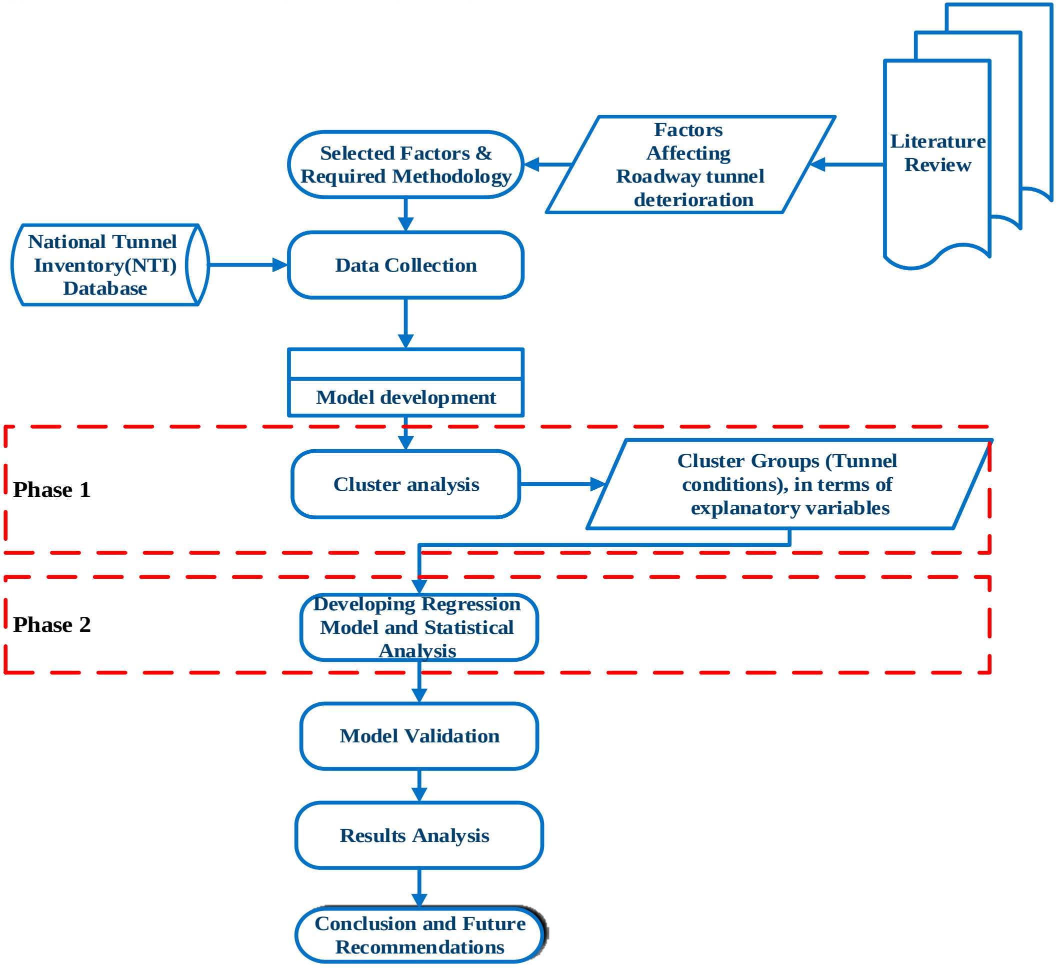


Figure 1. Research Methodology

4. DATA COLLECTION:

NTI database contains data for 473 tunnels along 41 American states (Bergeson and Ernst 2015). The inspected inventory items included**:** Identification items, Age and Service, Classification, Geometric Data, Inspection items, Load Rating and Posting, Navigation, Structure Type and Material. However, non-recording of the inspected tunnels condition has resulted in hindering the development of deterioration or prediction models for highway tunnels. This study has based on the NTI database (FHWA 2015) to be analyzed for deterioration modeling. Nine independent variables were selected from three aspects: Geometric, Inspection, and Structure Type & Material. The related variables were summarized in Table 1 according to National Tunnel Inspection Standard (NTIS) (FHWA 2015). An earlier screening has been conducted to remove both of the uncompleted records and recently reconstructed tunnels in 2015. Subsequently, data for 464 tunnels have been transferred to cluster analysis.

Table 1. Variables description

Items	Suggested Variables	Code	Description
Service	Age	AGE	Years since initial tunnel construction up to 2015.
	Annual Average Daily Traffic	AADT	Annual Average Daily Traffic (Vehicle/day).
	Annual Average Daily Truck Traffic	AADTT	Annual Average Daily Truck Traffic (Truck/day).
	Service in Tunnel	ST	1)Highway , 2)Highway and Railroad , 3)Highway & Pedestrian, 4)Highway, Railroad & Pedestrian , and 5)Other .
Geometric	Tunnel Length (Feet)	TL	Length of the tunnel to the nearest foot. The length shall be measured along the centerline of the road.
	Road Width(Feet)	RW	The most restrictive minimum distance between curbs or rails on the mainline tunnel road.
Structure type and material	Tunnel Shape	TS	1) Oval, 2) Horseshoe, 3) Rectangular, 4) Circular.
	Number of Bores	NB	One digit number is defining the number of bores in a tunnel.
	Ground Conditions	GC	1) *Soil*: consisting primarily of clay, silt, sand, gravel or a mixture, 2) *Rock*, 3) *Mixed Face* usually refers to a situation where the soil conditions vary along the length and/or height of the tunnel.

5. MODEL DEVELOPMENT

5.1. Clustering Data Analysis- Two step clustering

Two-Step Cluster analysis has been used in order to investigate the tunnel conditions as dependent variable. This approach has been selected for early analysis because of its ability to manipulate categorical and continuous variables simultaneously with auto selected correct number of clusters; on contrary of Hierarchical, and K-means cluster (Sarstedt & Mooi 2014, Shih, Jheng & Lai 2010). Two-Step Cluster analysis has been conducted using Statistical Package for Social Sciences (SPSS) Software. The followed analysis processes have been illustrated in Figure 2. As a result of many iterations to obtain the highest Silhouette coefficient (S), cluster analysis has reduced the significant variables to five variables; 1) service in tunnel, 2) ground condition, 3)annual average daily traffic, 4)annual average daily truck traffic, and 5) road width. These variables have been clustered in 6 groups with high Silhouette coefficient (S= 0.644) as shown in Figure 3 and 4. These resulted cluster groups showed that the most important variables in clustering was service in tunnel and ground condition, and then followed by the Annual average daily traffic and Annual average daily truck traffic, While Road width has a little importance for shaping the groups. Technically, ground condition is responsible for selecting the method of construction, the shape of a tunnel, and consequently it affects the tunnel deterioration (FHWA 2015).

The characteristics of each group have been represented in Figure 5. It can be noticed that the highway as service in tunnel has represented the majority of clusters. 100% of the soil as ground condition and highway as service in tunnel formed Cluster 6 that represented 26% of groups, with relatively low AADT, AADTT and little high road width. Cluster 5 represented the highest percentages of groups 32%. It contained 100%

of rock, and highway with the smallest values of AADT, AADTT and RW. While, cluster 1 and 3 have been formed with 100% of mixed soil (which is the weakest soil), but different service in tunnel. Noticeably group 1 contain high values of AADT, AADTT and RW.

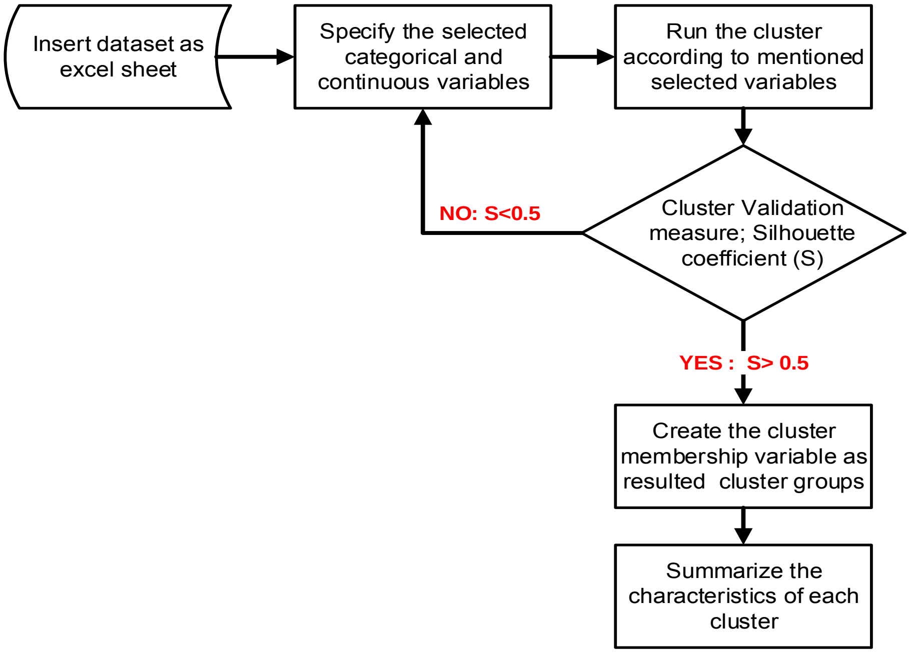
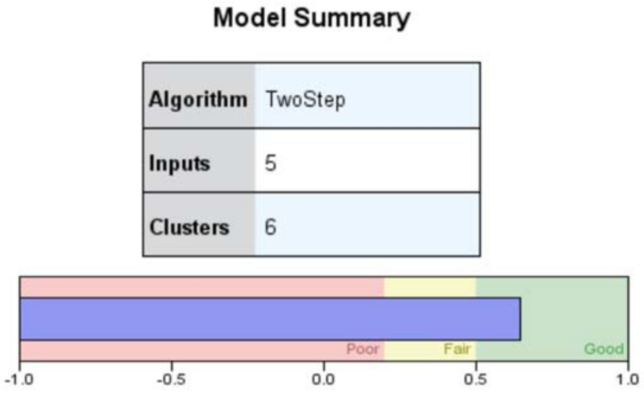


Figure 2. Two – step Cluster analysis processes



Model Summary

Algorithm	TwoStep
Inputs	5
Clusters	6

Figure 3. Cluster quality

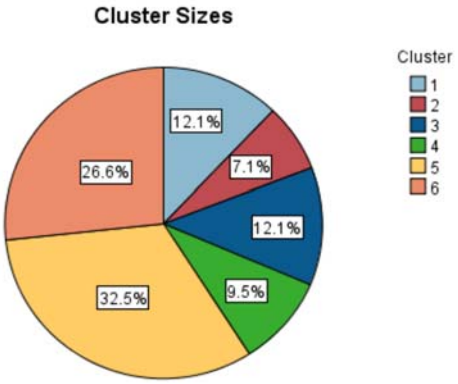


Figure 4. Distribution of cases in clusters

On the other hand, group 4 contained the highest traffic service for AADT, AADTT and RW, but it contained good ground condition “soil”, and “highway service in tunnel”, this formation has enabled this group to survive to be in good conditions. Based on the resulted features in each cluster, this paper has assumed that clustering groups were representing the tunnel condition, ranged from 1 to 6, where 1 is the worst condition and 6 is the best condition.

5.2. Regression modeling

Regression modeling is widely applied method for the deterministic approach for predictive analysis. It aims to investigate the relation between the forecasted condition and significant variables (Inkoom & Sobanjo, 2018; Bektas et al., 2012). Based on the satisfied resulted of cluster analysis, further analysis has been conducted to examine the relevance between the significant independent variables and the dependent variable by regression analysis. Herewith, the regression analysis was committed with both of the resulted significant variables and cluster groups as tunnel conditions.

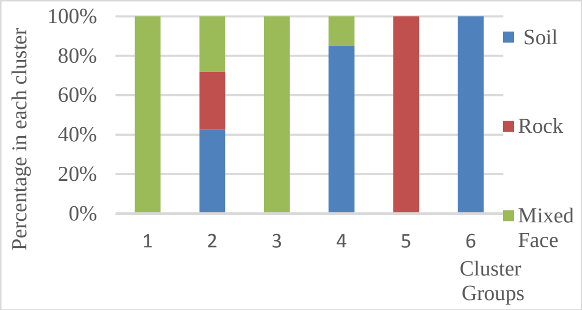


Figure 5-A. Groups' characteristics in terms of Ground conditions

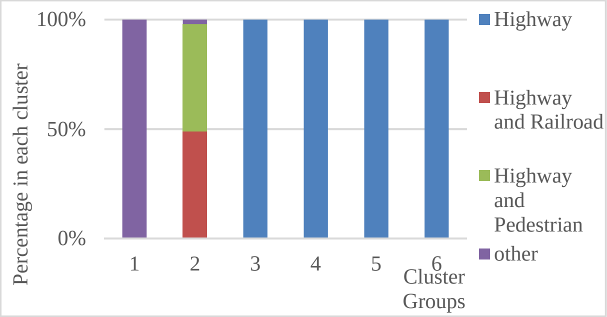


Figure 5-B. Groups' characteristics in terms of Service in the Tunnel

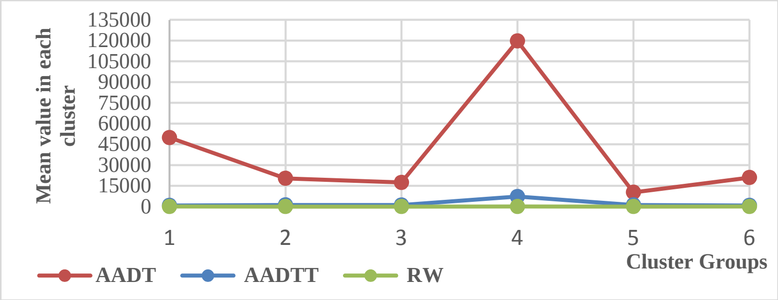


Figure 5-C. Groups' characteristics in terms of continuous variables

Figure 5. Characteristics of cluster groups in terms of significant variables

Because of clustering, about 85 tunnels were classified as non-clustered tunnels, so they have been eliminated, and then the rest of data were exported to Minitab 18 as statistical software for regression analysis. Where, 80% of 379 tunnels aged between 1 to 132 years have proceeded as modeling data, while the rest have been excluded earlier as data validation. The regression model has examined five independent variables against clustered tunnel condition. They composed of two categorical variables; Service in tunnel and Ground condition, and three continuous variables; Annual average daily traffic, annual average daily truck traffic, and road width.

5.3. Model building and statistical analysis:

The regression equations for the developed models were listed in Table 2, where the model is classified in terms of ground condition and service type in the tunnel.

Table 2. Developed Regression model for Predicting Highway tunnel condition

Ground Condition	Service in Tunnel	Regression Equation for predicting Tunnel Condition
Soil	Highway	5.7142 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Soil	Highway & Pedestrian	3.216 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Soil	Other	3.571 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Rock	Highway	5.0617 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Rock	Highway & Pedestrian	2.563 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Rock	Other	2.918 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Mixed face	Highway	3.3338 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW
Mixed face	Highway & Pedestrian	0.835 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW

Mixed face	Other	1.190 - 0.000004 AADT- 0.000066 AADTT + 0.00258 RW

According to the results of regression analysis in terms of assigned variables, R^2 is 90.84%; indicating that the predictors explain 90.84% of the variance in “response variable- Highway tunnel condition rating”. The P-value has been used for testing the significance of the developed models. *P-value* (statistical significance) for the developed model is 0.000 (P-value<0.001), meaning that that the estimated model is sound and significant at α-level of 0.001 and 0.01, and the null hypothesis is rejected. This means that none of the regression coefficients is zero.

5.4. Residual analysis:

In order to verify the linear Regression assumptions, the residuals and their patterns were analyzed in the following description: Figure 6 shows the normal probability plot for the residuals of the developed models. The normal probability plot shows that error terms are nearly normal. The results could be satisfied and interpreted by the possibility of outliers. Figure 7 shows the residuals vs. the order of data plot for the model under consideration, where points around the regression line should be independent for each value of predictors. The results show the residuals were located at inner bands of X values, except for some points of outliers; that lies on the outer negative and positive bands.

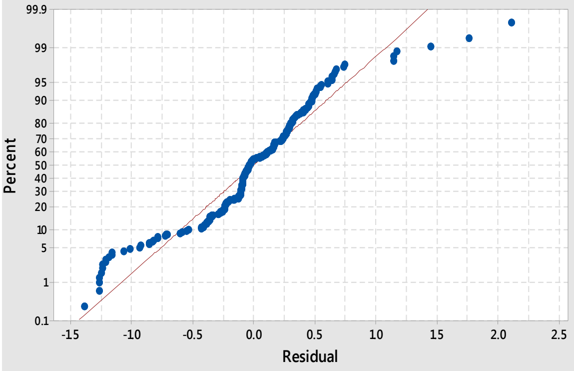


Figure 6. Normal probability of residual plots for super structure condition model

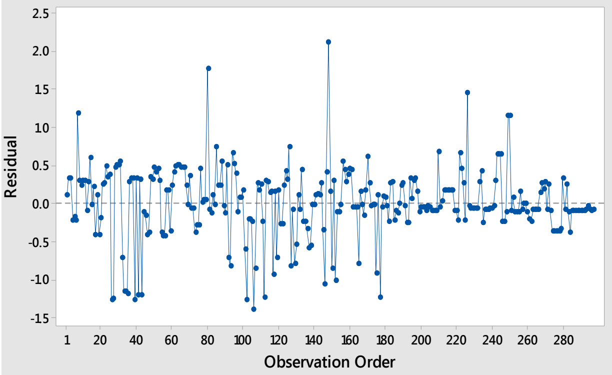


Figure 7. Residual vs. order of data plot for superstructure condition model

6. MODEL VALIDATION

To demonstrate the applicability and validity of the developed models, the validation process has been implemented using approximately 20% of collected data that are picked randomly and excluded earlier from model building. The predicted results of the models with actual data (observed conditions in NBI) have been compared for checking the model capability for predicting the condition rating. Fares et al.(2012), and Zayed& Halpin (2005) have used two terms for model validation : 1)The Average Invalidity and Validity Percent for the prediction error (AIP , AVP%) in Equations (1&2) which shows the validation as a percentage ; and 2) the Root Mean Square Error (RMSE) in Equation (3). If the value is closer to zero, the model is sound; and a value closer to 1 shows that the model is not appropriate.

$$[1] \quad \text{AIP} = \frac{\sum_{i=1}^{n}|1-(E_i/C_i|}{n}$$

$$[2] \quad \text{AVP}= 1 - AIP$$

$$[3] \quad \text{RMSE}= \sqrt{\sum_{i=1}^{n}(\mathrm{C}_i - \mathrm{E}_i)^2\,/n}$$

Where, E_i: Predicted value for case i; C_i: Actual value; and N: Number of bridges for validation

The estimated values for these terms for model validation based on the randomly selected data for validation (80 Tunnel =20% of collected data). Where it showed that, the AVP=83%, and RMSE=0.591<1;

as satisfied values to prove that the proposed models were able to forecast the highway tunnel condition. Figure 8 represented the actual and predicted tunnel condition in a scatter diagram where most of actual and predicted points were close, and have the same trend.

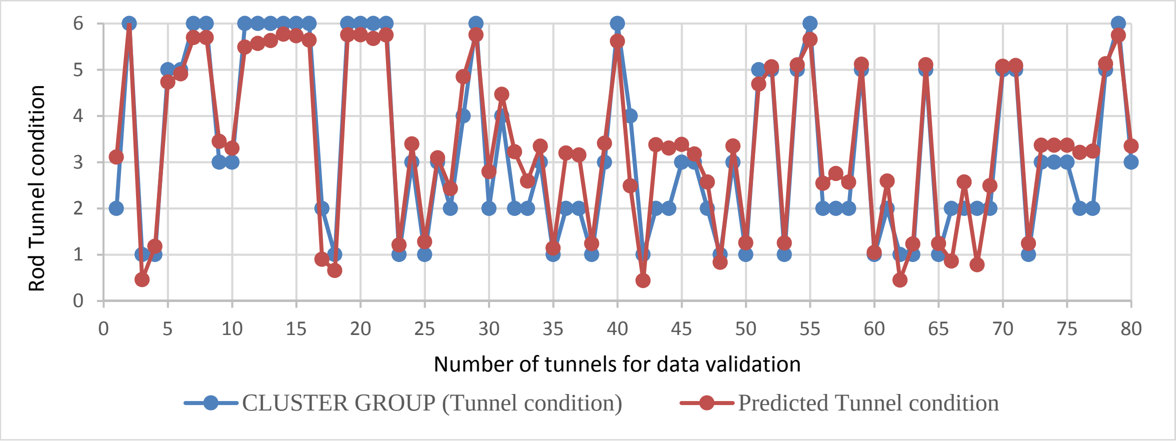


Figure 8. Actual vs predicted tunnel condition for developed models

7. RESULTS

This paper has used cluster-based regression for predicting the tunnel condition based on NTI database. The analysis processes has revealed the following results:

I. The proposed methodology allowed covering the unavailability of inspected tunnel conditions through exploring the dummy patterns among the historical data of NTI database for modeling the predicting tunnel condition;
II. The analysis processes found that Ground condition, Service in tunnel, Annual Average Daily Traffic, Annual Average Daily Truck Traffic, and Road Width are the most significant factors related to the clustered tunnel condition;
III. Two-step cluster analysis has shown ability to cluster NTI database in six groups. Where group 5 and 6 filled 32.5%, 26.6% respectively; where contained the majority of sample;
IV. The assumed conditions of clustered groups has depended on the characteristics of each of them; and
V. The regression analysis has showed significant relation between the selected independent variables and tunnel condition.

8. LIMITATIONS

This study has developed nine models for predicting tunnel condition. The results could push the wheel towards developing tunnel deterioration models from management perspective; not only from the structural view. However, two limitations should be considered to provide a reliable objective discussion: 1) using preliminary tunnel inventory data, which did not include the inspected tunnel condition and not always have completed data, and 2) The developed models have depended on the generated groups of tunnel conditions using cluster analysis, which could be vary from the actual observations of inspected tunnels, so using different approach such as Artificial Neural Network (ANN) might be useful to generalize these results.

9. CONCLUSIONS AND RECOMMENDATIONS

This study has investigated the impact of Geometric, Inspection, Structure Type and Material items on tunnel condition, and concluded that Ground condition, service in tunnel, Annual average daily traffic (AADT), Annual average daily truck traffic (AADTT), and Road width (RW) are the most significant variables. It concluded with six condition ratings; affected by main five variables through nine regression models; based on data in hand. The statistical analysis has shown satisfied results for model significance and validation. This study has paved the road for academic researchers to model the tunnel deterioration based on statistical method. As well as, it benefits the highway authorities to prepare future maintenance plans based on objective decisions. For future studies, it is recommended to examine the impact of each variable on the tunnel condition in order to maximize the trends that improve the tunnel condition with minimum maintenance cost. In addition, it is recommended to examine the impact of weather conditions, and maintenance history on tunnel condition, which have not been available for current data.

10. REFERENCES

Ahluwalia, S. S. 2008. A Framework for Efficient Condition Assessment of the Building Infrastructure. PhD Thesis, University of Waterloo, Waterloo.

Allen, C., Averso, M., Hargraves, C., & McNeil, S. 2015. Guide for the Preservation of Highway Tunnel Systems. Transportation Research Board.

Andersson, C., & Söderhäll, J. 2001. Rock mechanical conditions at the Äspö HRL. A study of the correlation between geology, tunnel maintenance and tunnel shape (No. SKB-R--01-53). Swedish Nuclear Fuel and Waste Management Co.

Bektas, B. A., Carriquiry, A., & Smadi, O. 2012. Using classification trees for predicting national bridge inventory condition ratings. Journal of Infrastructure Systems, 19(4), 425–433. doi:10.1061/(asce)is.1943-555x.0000143.

Bergeson, W., & Ernst, S. 2015. Tunnel Operations, Maintenance, Inspection, and Evaluation (TOMIE) Manual (No. FHWA-HIF-15-005). Office of Bridges and Structures, FHWA, 2015.

Fares, H., Shahata, K., Elwakil, E., Eweda, A., Zayed, T., Abdelrahman, M., & Basha, I. 2012. Modelling the performance of pavement marking in cold weather conditions. Structure and Infrastructure Engineering, 8(11), 1067-1079.

Federal Highway Administration (FHWA). 2015. Specifications for the National Tunnel Inventory. July 2015 Publication No. FHWA-HIF-15-006 . Available at : https://www.fhwa.dot.gov/bridge/inspection/tunnel/inventory.cfm

Gkountis, I. 2014. Infrastructure performance assessment of subway networks (Doctoral dissertation, Concordia University).

Inkoom, S., & Sobanjo, J. 2018. Availability function as bridge element's importance weight in computing overall bridge health index. Structure and Infrastructure Engineering, 14(12), 1598-1610.

Kobayashi, K., Kaito, K., & Lethanh, N. 2010. Deterioration forecasting model with multistage weibull hazard functions. Journal of infrastructure systems, 16(4), 282-291.

Meguid, M. A., & Dang, H. K. 2009. The effect of erosion voids on existing tunnel linings. Tunnelling and Underground Space Technology, 24(3), 278-286.

Oregon Department of Transportation (ODOT). 2016. Bridge Condition Report & Tunnel Data. Available at : file:///U:/Personal/Downloads/Final_2016_bridge_tunnel_report_091316.pdf

Sarstedt, M., & Mooi, E. 2014. Cluster analysis. In A concise guide to market research (pp. 273-324). Springer, Berlin, Heidelberg.

Schiessl, P., Gehlen, C., & Kapteina, G. 2004. Assessment and service life updating of existing tunnels. In First International Symposium:" Safe & Reliable Tunnels (pp. 189-198).

Semaan, N. (2011). Structural performance model for subway networks (Doctoral dissertation, Concordia University).

Shih, M. Y., Jheng, J. W., & Lai, L. F. 2010. A two-step method for clustering mixed categorical and numeric data. Tamkang Journal of Science and Engineering, 13,11-19.

Thompson, K., Rohena, J., Bardow, A. K., Brecto, B. B., Khaleghi, B., Ruzzi, L., ... & Ralls, M. L. 2011. *Best Practices For Roadway Tunnel Design, Construction, Maintenance, Inspection, And Operations* (No. NCHRP Project 20-68A, Scan 09-05).

Wang, J. H. 2018. Lifecycle cost and performance analysis for repair of concrete tunnels. In Eco-Efficient Repair and Rehabilitation of Concrete Infrastructures (pp. 637-672). Woodhead Publishing Series in Civil and Structural Engineering.

Zayed, T. M., & Halpin, D. W. 2005. Pile construction productivity assessment. Journal of Construction Engineering and Management, 131(6), 705-714.