Canadian Journal of Civil Engineering

Operational Based Stochastic Cluster Regression-Based Modeling for Predicting Condition Rating of Highway Tunnels

Journal:	<i>Canadian Journal of Civil Engineering</i>
Manuscript ID	cjce-2020-0081.R3
Manuscript Type:	Article
Date Submitted by the Author:	11-Dec-2020
Complete List of Authors:	Hassan, Sahar; Purdue University, School of Construction Management Elwakil, Emad; Purdue University, School of Construction Management
Keyword:	Highway/Roadway Tunnels, Tunnel Deterioration Models, Infrastructure Assessment
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Operational Based Stochastic Cluster Regression-Based Modeling for Predicting Condition Rating of Highway Tunnels

Authors:

Sahar Hasan, Lecturer, Construction and Project Management Institute, Housing and Building National Research Center (HBRC).

Emad Elwakil, Associate Professor, School of Construction Management, Purdue University, West Lafayette, IN, 47907, USA.

Draft

Corresponding Author:

Emad Elwakil Ph.D., P.E., C.C.P., P.M.P.

Purdue University

School of Construction Management

401 N. Grant St.

West Lafayette, IN, 47907

Phone: 1(765)494-6452

Email: eelwakil@purdue.edu

Operational Based Stochastic Cluster Regression-Based Modeling for Predicting Condition Rating of Highway Tunnels

Abstract

Despite the higher capital costs of tunnels and supplementary cost of maintenance, fewer deterioration models have been built compared with other highway components. Most of these models were limited to structural defects, not operational conditions. Moreover, there are inherent subjectivity and inaccuracy of the developed models, which may affect the maintenance process. The aim of this research and proposed contribution are investigating and modeling the impact of explanatory variables and non-periodic maintenance effect on highway tunnels condition. The stochastic regression analysis has been conducted to come out with a realistic tunnels condition through the Monte Carlo simulation methodology. The research methodology consists of three phases: cluster analysis, regression modeling, and stochastic analysis. A data set of 473 highway tunnels along 41 American states from the National Tunnels Inventory (NTI) has been used. Nine models have been developed with a high coefficient of determination ($R^2=90.8\%$). The obtained results and the models could push the wheel towards developing tunnel deterioration models from a management perspective, not only from the structural view. The developed models help highway authorities to prioritize the maintenance and objectively make informed investment decisions based on the historical data. This research has come as a response to the significant problems facing the highway authorities regarding tunnels and asset management.

Keywords: Highway/Roadway Tunnels; Tunnel Deterioration Models; Maintenance of Highway Tunnels; Infrastructure Assessment; Clustering Analysis; Regression Analysis; Monte Carlo Simulation; Sensitivity Analysis.

1. Introduction

Highway tunnels are the main element of the infrastructure system and a critical component to improving the highway's mobility. Tunnel structure is a long-term project; its service life varies from 50 to 125 years and requires more intensive maintenance and operation than traditional transportation facilities (Thompson et al. 2011; Schiessl et al. 2004). Inadequately maintained tunnels usually need more costly and extensive repairs (English 2016; Richards 1998). Tunnels are a very complex system with structural, civil, mechanical, and electrical systems.

Therefore, establishing a condition and developing a predicting model for condition rating for tunnels will guide operators and owners to make rehabilitation decisions. Then, develop a proactive financial plan for maintenance and rehabilitation. Ahluwalia (2008) stated that predicting the condition of infrastructure is a crucial aspect in any asset management tool since it helps infrastructure ranking for enhancing the maintenance, repair, and rehabilitation planning. In 2016, the earlier inventory data for tunnels had been recorded. However, it was assigned and incorporated as a part of the bridge's highway (ODOT 2016), according to the bridge's inventory data classification. However, the tunnels are structurally different from the bridges. Therefore, the investigated data did not express the tunnel condition; even so, these data were scattered and did not contain the tunnel elements' related geometric, operation, and condition data. However, later, in 2017, the Federal Highway Agency (FHWA) had launched the 2015 preliminary separate database for National Tunnels Inventory (NTI) to document the initial tunnel inventory and inspection data for 473 inspected; tunnels. The tunnel condition is assessed by maintenance staff, engineers, consultants using the guidelines and specifications for the National Tunnel Inspection Standards (NTIS). The NTIS

defines the general extent of deficiencies for each of the four condition states: good, fair, poor, and severe. (Allen et al. 2015; Bergeson & Ernst 2015).

The National Tunnels Inventory (NTI) established a preliminary database, but it still does not contain the related data of assessed or inspected tunnel conditions. Therefore, this paper implemented the clustering analysis as an analytical technique to figure out the historical data patterns. There is a lack of research and concern for practitioners and inspectors to have more realistic highway authorities models. Moreover, there is a need for a model to help highway authorities to prioritize the maintenance and objectively make informed investment decisions based on the historical data.

This research has come as a response to the significant problems facing the highway authorities regarding tunnels and asset management that are summarized as follows:

- i. Required an initial investigation about the abilities of the recently issued National Tunnel Inventory database NTI;
- ii. Lack of research of management perspective that study factors affecting the tunnel condition based on historical data; and
- iii. Lack of developed models for condition rating prediction of highway tunnels.

2. Research Significance and Objectives

According to the stated problem of highway tunnels' asset management, this study aims to contribute to the asset management of highway tunnels in four ways. First, the research has focused on exploring the factors affecting the highway tunnels' condition. Secondly, using the National Tunnels Inventory (NTI) preliminary database and exploring its current abilities. The third contribution presents an integrated methodology composed of two-step cluster analysis, Regression modeling, and Monte Carlo simulation as a solution for uncompleted data. As a final output, the paper

presented a validated model for predicting the highway tunnel condition based on historical data, with a new corresponding condition rating for highway tunnels based on the clustered groups.

3. Background

3.1. Defects (Deterioration features) of highway Tunnels

The simplest definition for the highway tunnel is defined by Caltrans Structure Maintenance & Investigations (SM&I) as "*a structure mined or bored through undisturbed material*" to distinguish between tunnels and bridges. While the standard for road tunnels, bridges, and other limited-access highways has defined the tunnels as "*an enclosed road for motor vehicle traffic with vehicle access that is limited to portals*." (Thompson et al. 2011). The material used, aging, operation factors, and exposure to harsh environmental conditions are the main reasons for the deterioration (Wang 2018). The National Tunnel Inspection Standards (NTIS) has identified the primary defects that affect the common tunnel structure materials (steel, concrete, and timber). These defects included corrosion and cracks in the steel structure. While concrete defects included scaling, cracking, delamination, spalling, water leakage, and steel or reinforcement corrosion. On the other hand, timber structures could be deteriorated by decay, insects, splits, fire damage, hollow, and leakage (Bergeson & Ernst 2015).

3.2. Deterioration models and Factors Contribute to Tunnel Deterioration

Deterioration models for tunnels have been developed to consider the continual or gradual variation over time through estimating and predicting the performance. Therefore, the successful deterioration models should consider the deterioration condition, tunnel structure, used material, and repair history (Wang 2018).

Extensive research related to tunnel design, construction, rehabilitation, and inspection has been conducted from the physical mechanism of the deterioration of structures; conversely, neglecting the statistical estimation method because of a historical tunnel database's unavailability. Kobayashi et al. (2010) developed an analytical methodology using the multistage Weibull hazard model for forecasting the deterioration of infrastructure facilities. The developed methodology was validated using a database of tunnel lighting reflectors. The results have guided tunnel administrators to improve the inspection and maintenance plan.

Because of scarce literature about the deterioration of tunnels using inspection data, subway structures were reviewed in this study as the closest infrastructure modeling to highway tunnels. Yamany and Elwakil (2020) developed a model to estimate and predict cast-in-place concrete tunnel liners' overall structural condition across the United States based on tunnel characteristics.

According to Meguid and Dang (2009), tunnel structural degradation is affected by ground movement and sudden leakage. These impacts were varied according to the surrounding ground condition of the tunnel. In this context, Andersson & Söderhäll (2001) have studied tunnel maintenance from a different perspective to address the impact of geology and geometry (tunnel shape) on tunnel maintenance.

Schiessl et al. (2004) have considered the inspection data of measured carbonation depth as one of the primarily collected data to update existing tunnels' service life. The applied method was the full probabilistic to assess the reliability of the existing tunnel lining against carbonation-induced corrosion. They concluded that the structure's actual behavior within the environment could be considered directly by integrating inspection results into the calculation procedure and, consequently, improving the quality of predicting service life.

Aoki et al. (2007) have forecasted the expected life length of highway facilities, equipment, and the inspection/replacement processes of tunnel lighting systems using the Weibull hazard model. Gkountis (2014) has developed an assessment model for subway network performance, which included a tunnel as one of its main components based on identified actual defects using Analytic Network Processes (ANP) for weights and the technique for order preference for the aggregation. However, this study neglected the structural system. Semaan (2011) has developed a model assessing structural components' condition based on defects of stations, tunnels, and supporting structures. The different defects were weighted using the AHP. The distribution of defects of tunnels was related to the main tunnel structural elements; dome, sidewalls, and bottom slab.

The bridge studies have been considered different aspects rather than condition rating and assessment. Antoniou et al. (2018), Markiz and Jrade (2019), and Safi et al. (2013) deal with the cost estimate, life cycle costing, scheduling for construction, and budget allocation. It can be revealed that most of the prediction models for highway tunnel condition studies have concentrated on studying the tunnel performance based on defects from the structural perspective. Therefore, there is a crucial need to develop deterioration models based on the inventory database. They reflect a real picture of tunnel performance affected by the operational and environmental items, not only the designed ones. Thus, developing the prediction models is crucial to improve highway authorities' design and proactive maintenance plans.

According to this literature, the following limitations have been revealed and summarized as follows:

- Most of the models have been based on the measured defect-based approach, which has inherent subjectivity and uncertainty and it does not consider the operational and environmental conditions;
- Most of the developed models are subway tunnels, which have a constant daily traffic loading rate with fewer environmental impacts. Besides, they are working on a network system connected by stations, so their models could be insufficient to be applied on highway tunnels;
- There is a lack of developed models focusing on condition prediction of highway tunnels; and
- The developed models have not included relevant factors that affect the tunnel condition in actual operation and environment status.

Considering the above literature and revealed limitations, this paper has suggested some variables to model the relation between the inspected variables and predicted tunnel conditions. These variables include the service type, geometric data, structure type, and material type.

3.3. Cluster Analysis

Clustering is considered one of the most essential applied techniques in data analysis, organizing a dataset into different groups based on similar patterns (Beringer & Hüllermeier 2006). Clustering is defined as categorizing similar data into a group but different from the other groups, where the resulted groups have similar patterns in terms of selected parameters. The applications of clustering usually deal with large datasets having many attributes. Cluster methods are utilized for model development in order to recognize existing hidden patterns among the historical data or as a preprocessing step with the clustering results used further in other data mining tasks, such as classification, prediction, and correlation analysis (Salem & Moselhi 2020; Salem 2018; Elwakil &

Zayed 2018; Sarstedt & Mooi 2014; Mirahadi et al. 2013; Berkhin 2006; Kogan et al. 2006). Cluster analysis is an iterative discovery process, which is necessary to modify data preprocessing and model parameters until the result achieves the desired properties.

Three different methods could be used to cluster data; hierarchical, k-means, and two-step cluster analysis. Hierarchical cluster analysis is one of the most straightforward techniques. It requires a distance or similarity matrix between all pairs of cases, so it is more suitable for a small dataset. On the contrary, the K-means cluster does not require computation of all possible distances and can be applied for a moderately dataset sized. Nevertheless, it requires shuffling cases in and out of groups, and it is necessary to know in advance the number of groups.

On the other hand, Chiu et al. (2001) have developed the two-step cluster analysis to handle previous types' problems. Two-step clustering has been recommended in clustering large data sets with a mixture of continuous and categorical variables. It can also solve with the auto-selected correct number of clusters (Sarstedt & Mooi 2014; Norusis 2003; Chiu et al. 2001).

4. Research Methodology

An extensive literature review has been performed to identify the factors that affect highway tunnel deterioration and review the developed models for assessing and predicting tunnel conditions. The revealed limitations led this study to apply a methodology composed of three-phased analysis processes. The clustering analysis is the first statistical analysis to determine the hidden patterns among the collected tunnel data. Then followed by regression modeling and stochastic analysis. This methodology comprises the below steps, as shown in Figure 1:

- Literature review to survey the factors affecting tunnel deterioration and the previously developed models;
- Data collection;
- Model development:
 - Investigate the tunnel condition using Cluster Analysis; where the output is unknown, and creating groups with similar characteristic is required;
 - Build a deterministic model to predict the highway tunnel condition; and
 - Build a stochastic model to develop a more realistic tunnel condition.
- Model validation;
- Sensitivity analysis to analyze the effect of each factor on tunnel deterioration; and
- Results Analysis.

5. Data Collection

In 2015, the Federal Highway Administration (FHWA) had established a preliminary national tunnel inventory (NTI) database for tunnel projects in the USA. It contains data for 473 tunnels along 41 American states, which planned to be inspected annually or biennially (Bergeson and Ernst 2015). The inspected inventory items include age and service, classification, geometric data, inspection items, load rating and posting, navigation, structure type, and material. However, non-recording of the inspected condition and current defects for the inspected highway tunnels resulted in retarding the development of deterioration or prediction models for highway tunnels. This study is based on the NTI database (FHWA 2015). Earlier screening has been conducted to remove both of the uncompleted records and recently reconstructed tunnels in 2015. Subsequently, data of 464 tunnels were categorized using cluster analysis. The selected

variables for analysis are summarized in Table 1, according to the National Tunnel Inspection Standard (NTIS) (FHWA 2015).

6. Model Development

The applied methodology in this research has presented three-phased analysis processes; cluster analysis, regression modeling, and stochastic analysis to develop a model for predicting tunnel conditions. This paper has selected nine independent variables from four aspects: Geometric, Inspection, Structure Type, and Material to investigate the tunnel conditions as a dependent variable using Two-Step Cluster analysis. The results of cluster analysis have reduced the significant variables to five variables. A further study was conducted based on the satisfying results in clustered groups to examine the relevance between the significant independent variables and the dependent variable by regression and stochastic analysis to come out with a probabilistic prediction of tunnel condition. These processes were graphically represented in Figure 2.

6.1. Clustering Data Analysis- Two-Step Clustering

Using the NTI database, the assessed variable for condition tunnels was absent; thus, the clustering analysis is required to group the tunnels in clusters with similar behavior or trend. Moreover, the previous literature did not provide a guide for prior knowledge regarding selecting the number of clusters "tunnel condition rating." Consequently, the two-step cluster has been selected for the initial analysis. One of the advantages of the two-step cluster method is deploying categorical and continuous variables simultaneously and identifying the correct number of clusters (Sarstedt & Mooi 2014; Shih et al. 2010). Two-Step Cluster analysis has been conducted using Statistical Package for Social Sciences (SPSS) Software © in the following steps:

Step (1): Insert dataset as an excel sheet;

Step (2): Specify the selected categorical and continuous variables as listed in Table 1;

Step (3): Number of clusters: use automated cluster selection to select the optimal number of clusters using the Schwarz Bayesian criterion or the Akaike information criterion (Sarstedt & Mooi 2014);

Step (4): Run the cluster according to the mentioned selected options;

Step (5): Cluster Validation measure: To verify the "goodness" of a cluster solution, extract the model summary and cluster quality. It displays the number of output cluster groups and the Silhouette coefficient. It is a measure of both cohesion and separation (Silhouette coefficient (overall average silhouette)). The silhouette coefficient is calculated according to equation 1 (Liu et al. 2010; Saitta et al. 2007).

$$\text{Silhouette Coefficient} = S = \frac{1}{NC} \sum_i \left\{ \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x) - a(x)}{\max[b(x), a(x)]} \right\} \quad (1)$$

The highest S is taken to be optimal. Where:

$$a(x) = \frac{1}{n_i - 1} \sum_{y \in C_i, y \neq x} d(x, y) \quad , \quad b(x) = \min_{j, j \neq i} \left[\frac{1}{n_j} \sum_{y \in C_j} d(x, y) \right]$$

n: Number of objects in data set; NC: number of clusters;

C_i : i th cluster; n_i : number of objects in C_i ; $d(x, y)$: the distance between x and y.

Step (6): For the right cluster solution, the within-cluster distances are small, and between clusters, distances are large (Sarstedt & Mooi 2014; Liu et al. 2010).

A larger Silhouette value, close to the maximum value of 1, indicates a better quality of a clustering result (Chen et al. 2002). Figure S1 and Figure S2 displayed the results of cluster validation;

Step (7): Repeat step 2 with another variable If the Silhouette coefficient is lower than 0.5. The silhouette coefficient has based on the average distances between the objects and can vary between -1 and +1;

Step (8): Extract the Range of solutions: When getting a satisfied Silhouette index, create the cluster membership variable as evaluated field of resulted clustering;

Step (9): View the Importance of variables as shown in Figure S3; and

Step (10): Summarize the characteristics of each cluster as represented in Figure S4.

The characteristics of cluster groups in terms of each independent variable were illustrated in figures (S4-a), (S4-b), and (S4-c).

Two-step cluster analysis was performed on the NTI database according to the mentioned steps and analyzed as follows:

Many iterations have been conducted with a different mixture of categorical and continuous variables. As a result, five inputs (factors) have achieved a high Silhouette coefficient; service in the tunnel, ground condition, annual average daily traffic, annual average daily truck traffic, and road width.

Two-step cluster analysis resulted in 6 groups, with a reasonable degree of Silhouette coefficient ($0.644 > 0.5$, as explained in step 7), based on equation 1. It indicated that the cluster groups are distinct, as shown in Figure S1 and Figure S2.

Figure S3 ranked the variables from most to least importance. The most crucial variable is service in tunnel and ground condition, which represented 1 and 0.99 as the importance weight. Then followed by equal importance for both variables of Annual average daily traffic and Annual average daily truck traffic with importance "0.35". While Road width has almost no effect on cluster formation "0.064". Technically, the ground condition is responsible for selecting the construction method, the shape of a tunnel, and, consequently, affecting tunnel deterioration (FHWA 2015).

Figure S4 summarizes each cluster's characteristics, where it is divided into four figures from Figure S4-a to Figure S4-d. Each Figure of them represented how each variable has shaped the nature of each cluster.

It can be noticed that the highway as service in the tunnel represents the majority in clusters. Moreover, the 100% of soil as the ground condition and highway as service in a tunnel formed Cluster 6 representing 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 the rock and highway with the smallest values of AADT, AADTT, and RW. Cluster 1 and 3 have been formed with 100% of mixed soil (which is the weakest soil) but different service in the tunnel. Noticeably group 1 contains high values of AADT, AADTT, and RW.

On the other hand, group 4 has proved the paper assumption identifying the clustered groups; where it contained the highest traffic service for AADT, AADTT, and RW, but it contained good ground condition "soil" and "highway service in the tunnel," that enabled this group to survive to be in good conditions.

Based on each cluster's features, it is supposed that clustering groups were representing the tunnel condition, ranged from 1 to 6, as depicted in Figure S5, where 1 is the worst condition, and 6 is the best condition. For further analysis, the resulted cluster groups have been exported to Minitab software to examine the relation and correlation between the produced tunnel condition and studied predictors.

6.2. Regression Modeling

Regression models are one of several methodologies that have been used to predict future deterioration, such as curve-fitting models, Markovian models, and reliability-based models (Elhakeem 2005). The regression technique investigates the relation

between the forecasted condition and significant variables (Inkoom & Sobanjo 2018; Bektas et al. 2012).

Based on the completed clustering analysis results, this study has come out with the assumption that the cluster groups were representing the tunnel conditions. Further investigation is required to identify the relation between the related significant variables and the predicted variable using the regression method. Hence, the regression analysis was based on the cluster results and committed with the resulted significant variables. As a result of the cluster, about 85 tunnels were classified as uncluttered tunnels, so they were removed. The rest of the 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 authors used (80%) out of 379 tunnels data points to build regression models, and the other 20% (76 points) have been used for validation. The regression model has categorized the models into nine models.

The regression model has examined five independent variables against clustered tunnel conditions. They are composed of two categorical variables; Service in tunnel and Ground condition, three continuous variables; Annual average daily traffic, annual average daily truck traffic, and road width. The descriptive analysis for these variables was summarized in Table 2.

6.2.1. Model building and statistical analysis

The regression models are listed in Table 3 based on ground condition and service type in the tunnel. The R^2 is **90.84 %**, and R^2_{adjusted} is **90.62%**, which indicates that the predictors explain 90.84% of the variance in the "response variable- Highway tunnel condition rating." The ***P-value*** is **0.000** ($P\text{-value} < 0.001$), which means that the model is sound and significant at α -level of 0.001 and 0.01, and the null hypothesis has been

rejected. Table 4 shows the t-test results, where ***P-value*** for all predictors is almost zero, except for road width (0.042), but it was still significant ($0.042 < 0.05$). Consequently, the predictors were significantly related to the response (Highway tunnel condition).

6.2.2. Residual analysis

The error distribution is nearly normal, as shown in Figure S6, and a small shift from normality, which does not cause serious problems (Kutner et al. 2005). The outliers show this variance of the unusual observations and large residuals, as shown in Table 5. Figure S7 shows the residuals vs. the order of the data plot for the model. The result shows positive residuals at inner bands of X values, and the outer bands primarily consist of negative residuals, which is consistent with the outliers' observation.

7. Model Validation

To check the adequacy of the regression model, a comprehensive model validation procedure has been applied to all models. Approximately 20% of collected data is embedded in the regression model for comparing its results with the actual results. Average Invalidity Percent, Root Mean Square Error, and Mean Absolute Error have been used as validation measures (Zayed & Halpin 2005; Elwakil & Zayed 2012) as follows:

$$AIP = \frac{\sum_{i=1}^n |1 - (E_i/C_i)|}{n} \quad (2)$$

$$AVP = 1 - AIP \quad (3)$$

$$RMS = \frac{\sqrt{\sum_{i=1}^n (C_i - E_i)^2}}{n} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^n |C_i - E_i|}{n} \quad (5)$$

Where,

E_i : Predicted value for case i ;

C_i : Actual value; and

N : Number of bridges

The model validation is based on the randomly selected data for validation (80 Tunnel =20% of the collected data). Where it showed that, the AVP=83%, RMS=0.066<1, and MAE=0.0.466<1; as satisfying values to prove that the proposed models were able to forecast the highway tunnel condition. Figure S8 shows the actual vs. predicted output plot. The Figure shows that the predicted values are following the same pattern of real data. The significant factors on tunnel condition are identified, but their impacts are not specified as well as inherent uncertainty. The Monte Carlo simulation has been used to reduce uncertainty and simulate the real condition.

8. Stochastic Analysis

The Regression analysis has resulted in nine models, as shown in Table 3. The percentage of each model was calculated according to the distribution of 379 tunnels inside these models. It was found that four models have contained the majority of model data; Model 1 (35%), Model 4 (32%), Model 7 (13%), and Model 9 (12%), as represented in Figure S9.

The simulation has been performed separately for each regression model of four models; model 1, 4, 7, and 9; each contains more than 10% of modeling data. Furthermore, a sensitivity analysis was conducted to show the strength of each variable using Crystal Ball software © following the listed steps:

Step (1): Prepare the input variables for each model;

Step (2): Fit the data to a distribution: define the fit distribution for each continuous variable using chi-square as the goodness of fit;

Step (3): Define Forecast: assigning the mathematical regression formula for evaluating the tunnel condition;

Step (4): Run Simulation: Crystal Ball replaces the values in the "Assumption" cells with a random number drawn from the specified distribution. This process is repeated a predefined number of times/ iterations (Andersen & Brandstrup 2008; Kirkpatrick 2001); and

Step (5): Extracting the results of the simulation: Figure 3 shows the results. The tunnel condition ranges from 0 to 6. Figure (3-a) shows the probability of 99.97% for model 1. As well as the probability for the other three models was more than 95%, as represented in Figure (3-b), (3-c), and (3-d).

The proposed methodology guides the decision-maker to determine the uncertainty of predicted conditions to take appropriate maintenance prioritization action.

9. Sensitivity Analysis

The sensitivity analysis is performed to test the significant variable's impact on the tunnel condition and quantify the considerable effect of each independent variable. Thus, a sensitivity analysis was conducted based on the developed regression models through two methods:

- i. Changing the values between minimum and maximum for one variable while keeping the other variables constant. Figure 4 shows three plots that expressed each significant variable against the predicted tunnel condition and described their trend.
- ii. Monte Carlo simulation has been applied to quantify the effect of each significant variable. The resulted sensitivity charts help to decide which variables affect the variation of tunnel condition primarily. This analysis allows maximizing

the positive impacts and reducing the negative ones. This sensitivity analysis has been represented in Figure 5.

As a result of the sensitivity analysis, three Figures (4-a), (4-b), and (4-c) have represented the trend of each variable. The annual average daily traffic, annual average daily traffic, and Roadway width have acted negatively versus predicted tunnel conditions. The tunnel deteriorates with increasing values for any of them. It is a logical relationship; growing values of these variables resulted in more loads and higher exposure to environmental impacts.

In addition to trend, identifying how much each parameter affects the tunnel condition helps understate their behavior. Therefore, further sensitivity analysis is conducted using the Crystal Ball © software to apply the Monte Carlo simulation. The results have been represented in Figure 5. It plotted four graphs based on four regression models that have contained the largest sample size. Figures from Figure (5-a) to Figure (5-d) have investigated the contribution to variance "impact value" for the essential parameters in each model. It concluded that annual average daily traffic and annual average daily truck traffic have the most significant impact on all models, keeping on the negative impact.

In comparison, Roadway width has the smallest positive contribution on the tunnel condition (7-10%) for all models. For models 1, 7, and 9, it was found that AADT has the highest contribution (48-52%), then followed by AADTT (38-44%). Except for model 4, both AADT and AADTT have switched their contribution, where the annual average daily truck traffic (AADTT) has the highest negative contribution than AADT. This result prioritizes maintenance to the tunnels with high annual average daily traffic and annual average daily truck traffic.

The sensitivity analysis has shown the same impacts and trends for AADT and AADTT and agreed with regression results. Except for the variable of road width, they agreed only on its relative contribution to variance. The sensitivity analysis using an excel sheet has shown a different correlation direction for the variable of roadway width than that appeared on the regression model. It could be attributed to the correlation trend depending on the population characteristic and changing in some predictable manner with another variable. While in regression and Monte Carlo simulation, the pattern is tested using the modeling data (observed data). In contrast, the excel sheet uses different roadway width values ranges from minimum to maximum, which is more realistic. However, the sensitivity analysis using excel showed more realistic results for the correlation trend for roadway width.

This result of significant impact as correlation direction and relative contribution to the variance for each parameter has agreed with the regression analysis results.

10. Results Analysis

The analysis processes have revealed the following summarized results:

- I. The proposed methodology allowed bridging the gap of disability of inspected tunnel conditions in the available NTI database. In addition to investigating the significant variables affecting the tunnel condition to formulate them in deterministic and stochastic modeling;
- II. The obtained results of the six groups of tunnel condition ratings are affected by the leading five variables through nine models; based on data in hand.
- III. Two-step cluster analysis has shown the ability to cluster the NTI database into six groups. Where groups 5 and 6 filled 32.5%, 26.6% respectively, contains the majority of the sample;

- IV. The assumed conditions of clustered groups depended on the characteristics of each of them, where group 6 has been assumed as *"excellent condition,"* contains the soil, highway service in the tunnel, and low values for AADT, AADTT, and RW as well as, group 5 "Very Good" contains rock, highway service in the tunnel, and lowest values of AADT, AADTT, and RW;
- V. On the other hand, group 1, which was classified as "poor condition," has filled 12% and contained mixed face as ground condition, and other services in the tunnel and medium traffic service for AADT, AADTT, and RW. While group 4 has proved the paper assumption, where it contained the highest traffic service for AADT, AADTT, and RW, but it contained good ground condition "soil" and "highway service in the tunnel," that enabled this group to survive to be in good conditions;
- VI. The analysis processes found that annual average daily traffic, annual average daily truck traffic, and road width are the most significant factors related to the clustered tunnel condition;
- VII. Significant negative relation with the higher impact between the tunnel condition and the significant independent variables have been tested and validated, except for road width that needs further research, where it gives logic negative trend using sensitivity analysis while the regression model indicated the positive one but with meager impact; and
- VIII. Further analysis of the significance of the other variables would bring a more realistic view for predicting tunnel conditions.

11. Limitations

This study has developed nine models for predicting tunnel conditions. The obtained results could push the wheel towards developing tunnel deterioration models from a

management perspective, not only from the structural view. However, two limitations should be considered to provide a reliable objective discussion: 1) One of the limitations of this study is the use of preliminary tunnel inventory data, which did not contain the inspected tunnel condition and not always have completed data; and 2) The developed models and rating conditions have depended on the generated groups of tunnel conditions as a result of cluster analysis, which could vary from the actual observations of inspected tunnels, so using a different approach in the future will be useful to able to generalize this result.

12. Conclusion

The proposed methodology and developed models have contributed to predicting the tunnel condition based on the NTI database. This research has proposed a condition prediction model for highway tunnels based on the Two-Step Clustering method as a based strategy that can automatically generate prior information from the tunnels inventory NTI database for further analysis. It was selected to characterize the profile of highway tunnels by tunnel condition. The applied methodology in this research has presented three-phased analysis processes; cluster analysis, regression modeling, and stochastic analysis to develop a model for predicting tunnel conditions.

The findings suggested six clusters by applying the suggested methodology; they exhibit distinct profile patterns of tunnel conditions based on the explanatory variables. Further regression and stochastic models were developed to explore the relationship between significant variables and generated cluster groups of tunnel conditions.

This study has investigated the impact of geometric, inspection, structure type, and material items on tunnel conditions. It concluded that ground condition, service in the tunnel, annual average daily traffic (AADT), annual average daily truck traffic (AADTT), and road width (RW) are the most significant variables.

As a result, nine models have been developed, categorized by two variables; ground condition and service in the tunnel. The regression output of tunnel condition prediction models has been used as a base to investigate the probability distribution and sensitivity analysis using Monte Carlo simulation. The statistical analysis has shown satisfying results for model significance and validation.

This study provided useful insights for academic researchers for developing prediction models of tunnel conditions and inspectors in highway authorities. For future studies, when the assessed condition will be available, it is recommended to use another approach to test the significance of other variables, such as age, tunnel shape, and the number of bores on tunnel condition. Moreover, it is recommended to collect a real condition rating instead of the cluster if available. This study recommends FHWA to add the following items to the current preliminary NTI, where the suggested items influence the tunnel deterioration. It will provide valuable data to tunnel deterioration modeling: The material types for structural elements and tunnel lining, the inspected defects, and the inspected tunnel condition. One of the limitations of this study is the use of preliminary tunnel inventory data, which did not contain the inspected tunnel condition and not always have completed data. The developed models and rating conditions have depended on the generated groups of tunnel conditions as a result of cluster analysis.

There is a persistent need for further efforts to apply the same methodology using the NTI database or similar database with different regions to identify the external regional factors such as; city size and weather conditions. Using different approaches, such as machine learning, to provide more reliable decisions is recommended. Moreover, the following techniques may be used; stochastic gradient descent, deep learning using the neural network to build the regression models. Furthermore, proposing a research track

to cover practitioners' and inspectors' concerns to have integrated prediction models of the theoretical condition and expert opinion in highway authorities.

Draft

13. References

- Ahluwalia, S. S. 2008. A Framework for Efficient Condition Assessment of the Building Infrastructure. Ph.D. Thesis, University of Waterloo, Waterloo.
- Al-Barqawi, H. & Zayed, T. 2006. Condition rating model for underground infrastructure sustainable water mains. *Journal of Performance of Constructed Facilities*, 20 (2, May), 126–135.
- Allen, C., Averso, M., Hargraves, C., & McNeil, S. 2015. Guide for the Preservation of Highway Tunnel Systems. Transportation Research Board.
- Andersen, N. J. H., & Brandstrup, J. 2008. Monte Carlo Simulation in Crystal Ball 7.3. Aarhus University, May.
- 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.
- Aoki, K., Yamamoto, K., & Kobayashi, K. 2007. Optimal inspection and replacement policy using a stochastic method for deterioration prediction. *The 11th World Conference on Transportation Research*, pp. 24-28.
- Antoniou, F., Konstantinidis, D., Aretoulis, G., & Xenidis, Y. 2018. Preliminary construction cost estimates for motorway underpass bridges. *International Journal of Construction Management*, 18(4), 321-330.
- 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.
- Beringer, J., & Hüllermeier, E. 2006. Online clustering of parallel data streams. *Journal of Data & Knowledge Engineering*, 58(2), 180-204.
- Berkhin, P. 2006. A survey of clustering data mining techniques. In *Grouping multidimensional data*. Springer, Berlin, Heidelberg, pp. 25-71.
- Chen, G., Jaradat, S. A., Banerjee, N., Tanaka, T. S., Ko, M. S., & Zhang, M. Q. 2002. Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data. *Institute of Statistical Science, Academia Sinica*, 241-262.
- Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. 2001. A robust and scalable clustering algorithm for mixed type attributes in a large database environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 263-268). ACM.
- Elhakeem A. 2005 An asset management framework for educational buildings with life-cycle cost analysis, Ph.D. Thesis, University of Waterloo, Waterloo.
- Elwakil, E. and Zayed, T., 2012. Data management for construction processes using fuzzy approach. In *Construction Research Congress 2012: Construction Challenges in a Flat World* (pp. 1222-1231).
- Elwakil, E. and Zayed, T., 2018. Construction productivity fuzzy knowledge base management system. *Canadian Journal of Civil Engineering*, 45(5), pp.329-338.
- English, G. 2016. Tunnel Operations, Maintenance, Inspection, and Evaluation Manual, 2015: Practical Implications for Fire Protection and Life Safety Systems. *Transportation Research Record: Journal of the Transportation Research Board* (2592), 162-168.

- Fares, H., Shahata, K., Elwakil, E., Eweda, A., Zayed, T., Abdelrahman, M., & Basha, I. 2012. Modeling 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.
- Kirkpatrick, T. J. 2001. Impact of specification changes on the chloride-induced corrosion service life of Virginia bridge decks (Doctoral dissertation, Virginia Tech).
- Kobayashi, K., Kaito, K., & Lethanh, N. 2010. Deterioration forecasting model with multistage Weibull hazard functions. *Journal of infrastructure systems*, 16(4), 282-291.
- Kogan, J., Nicholas, C., & Teboulle, M. 2006. Grouping multidimensional data. Springer-Verlag Berlin Heidelberg.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. 2005. Applied linear statistical models. 5th ed. USA: McGraw-Hill.
- Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. 2010. Understanding of internal clustering validation measures. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (pp. 911-916). IEEE.
- Markiz, N., & Jrade, A. 2019. Integrating an expert system with BrIMS, cost estimation, and linear scheduling at the conceptual design stage of bridge projects. *International Journal of Construction Management*, 1-16.
- Meguid, M. A., & Dang, H. K. 2009. The effect of erosion voids on existing tunnel linings. *Tunneling and Underground Space Technology*, 24(3), 278-286.
- Mirahadi, F., Elwakil, E., & Zayed, T. 2013. An Innovative Technique for Improving Productivity Forecasting Models. Conference: 4th Construction Specialty Conference, CSCE: Montreal, QC, Canada.
- Norušis, M. J. 2003. SPSS 12.0 Statistical Procedures Companion Prentice Hall. Upper Saddle River, NJ USA Google Scholar.
- Oregon Department of Transportation (ODOT). 2016. Bridge Condition Report & Tunnel Data. Available at: file:///U:/Personal/Downloads/Final_2016_bridge_tunnel_report_091316.pdf
- Richards, J. A. 1998. Inspection, maintenance, and repair of tunnels: international lessons and practice. *Tunneling and Underground Space Technology*, 13(4), 369-375.
- Safi, M., Sundquist, H., Karoumi, R., & Racutanu, G. 2013. Development of the Swedish bridge management system by upgrading and expanding the use of LCC. *Structure and infrastructure engineering*, 9(12), 1240-1250.
- Saitta, S., Raphael, B., & Smith, I. F. 2007. A bounded index for cluster validity. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 174-187). Springer, Berlin, Heidelberg.
- Salem, A., 2018. Automated Productivity Models for Earthmoving Operations (Doctoral dissertation, Concordia University).
- Salem, A. and Moselhi, O., 2020. AI-Based Cloud Computing Application for Smart Earthmoving Operations. *Canadian Journal of Civil Engineering*, (ja).

- Sarstedt, M., & Mooi, E. 2014. Cluster analysis. 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. A 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 the repair of concrete tunnels. In *Eco-Efficient Repair and Rehabilitation of Concrete Infrastructures* (pp. 637-672). Woodhead Publishing Series in Civil and Structural Engineering.
- Yamany, M.S. and Elwakil, E., 2020. Modelling of cast-in-place concrete tunnel liners condition. *Structure and Infrastructure Engineering*, 16(8), pp.1147-1160.
- Zayed, T. M., & Halpin, D. W. 2005. Pile construction productivity assessment. *Journal of Construction Engineering and Management*, 131(6), 705-714.

Figures Legends

- Figure 1. Research Methodology
- Figure 2. Theoretical Deterioration Model for Highway Tunnels
- Figure 3. Simulation Result for Predicted Tunnel Condition
- Figure 4. Sensitivity Analysis of Significant Variables for Predicting Tunnel Condition
- Figure 5. Sensitivity Analysis Using Monte Carlo Simulation

Tables Legends

- Table 1. Variables Description (FHWA 2015)
- Table 2. Summary Statistics of the Significant Variables Affecting Highway Tunnel Condition
- Table 3. Developed Regression Models for Predicting Highway Tunnel Condition
- Table 4. Regression Output
- Table 5. Unusual Observations

Figure 1. Research Methodology

Figure 2. Theoretical Deterioration Model for Highway Tunnels

Figure 3-a. Simulation result for Predicted tunnel condition MODEL(1)

Figure 3-b. Simulation result for Predicted tunnel condition MODEL(4)

Figure 3-c. Simulation result for Predicted tunnel condition MODEL(7)

Figure 3-d. Simulation result for Predicted tunnel condition MODEL(9)

Figure 3. Simulation Result for Predicted Tunnel Condition

Figure 4-a. Sensitivity Analysis of AADT

Figure 4-b. Sensitivity Analysis of AADTT

Figure 4-c. Sensitivity Analysis of RW

Figure 4. Sensitivity Analysis of Significant Variables for Predicting Tunnel Condition

Figure 5-a. Sensitivity Analysis for Model (1)

Figure 5-b. Sensitivity Analysis for Model (4)

Figure 5-c. Sensitivity Analysis for Model (9)

Figure 5-d. Sensitivity Analysis for Model (7)

Figure 5. Sensitivity Analysis Using Monte Carlo Simulation

Table 1. Variables Description (FHWA 2015)

	Suggested Variables	Code	Description
Service items	Age	AGE	The age of the bridge 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 data	Tunnel Length (Feet)	TL	Length of the tunnel to the nearest foot.
	Road Width(Feet)	RW	The most restrictive minimum distance between curbs or rails.
Structure type and	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) <i>Soil</i> : consisting primarily of clay, silt, sand, gravel, or a mixture, 2) <i>Rock</i> , 3) <i>Mixed Face</i>

Draft

Table 2. Summary Statistics of the Significant Variables Affecting Highway Tunnel Condition

Variables	Mean	Standard deviation	Skewness
Annual ADT	30670	46275	2.47
Annual ADTT	1687	2792	3.38
Service in Tunnel (ST)	1.431	1.2204	2.54
Road Width (RW)	32.39	24.05	3.93
Ground Conditions (GC)	1.8384	0.7716	0.29

Draft

Table 3. Developed Regression Models for Predicting Highway Tunnel Condition

Model number	Ground Conditions	Service in Tunnel	Condition Tunnel prediction for Equation Regression
1 Model	Soil	Highway	$5.7142 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 2	Soil	Highway and Pedestrian	$3.216 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 3	Soil	Other	$3.571 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 4	Rock	Highway	$5.0617 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 5	Rock	Highway and Pedestrian	$2.563 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 6	Rock	Other	$2.918 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 7	Mixed face	Highway	$3.3338 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 8	Mixed face	Highway and Pedestrian	$0.835 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$
Model 9	Mixed face	Other	$1.190 - 0.000004 \text{ AADT} - 0.000066 \text{ AADTT} + 0.00258 \text{ RW}$

Draft

Table 4: Regression Output

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	628.856	89.8365	409.59	0.000
Annual ADT	1	5.159	5.1587	23.52	0.000
Annual ADTT	1	4.370	4.3695	19.92	0.000
Road Width	1	0.912	0.9119	4.16	0.042
Ground Conditions	2	154.167	77.0836	351.45	0.000
Service in Tunnel	2	86.858	43.4289	198.01	0.000

DF: The total degrees of freedom (DF) are the amount of information in the data. **Adj SS:** Adjusted sums of squares are measures of variation for different components of the model. **Adj MS:** Adjusted mean squares measure is how much variation a term or a model explains, assuming that all other terms are in the model, regardless of the order they were entered. **F-Value:** The F-value is used to determine whether the term is associated with the response. **P-Value:** The p-value is a probability that measures the evidence against the null hypothesis.

Draft

Table 5. Unusual Observations

Obs	Cluster group	Fit	Resid	Std Resid	Large residual	Unusual
	6.000	6.224	-0.224	-0.55		X
6	6.000	6.231	-0.231	-0.56		X
7	1.000	-0.171	1.171	2.64	R	X
26	4.000	5.264	-1.264	-2.73	R	
27	4.000	5.255	-1.255	-2.71	R	
33	4.000	5.167	-1.167	-2.51	R	
34	4.000	5.161	-1.161	-2.50	R	
35	4.000	5.191	-1.191	-2.56	R	
39	4.000	5.271	-1.271	-2.73	R	
41	4.000	5.216	-1.216	-2.61	R	
43	4.000	5.216	-1.216	-2.61	R	
59	4.000	4.371	-0.371	-0.88		X
80	4.000	2.237	1.763	3.90	R	
88	4.000	3.446	0.554	1.36		X
102	2.000	3.271	-1.271	-3.16	R	X
106	4.000	5.393	-1.393	-3.03	R	
112	4.000	5.238	-1.238	-2.67	R	
117	4.000	4.942	-0.942	-2.03	R	
146	4.000	5.067	-1.067	-2.29	R	
148	4.000	1.887	2.113	4.71	R	X
152	2.000	3.013	-1.013	-2.52	R	X
169	4.000	3.835	0.165	0.37		X
175	4.000	4.927	-0.927	-2.02	R	
177	4.000	5.238	-1.238	-2.67	R	
223	1.000	0.554	0.446	1.00		X
226	4.000	2.549	1.451	3.16	R	
243	1.000	0.362	0.638	1.43		X
244	1.000	0.359	0.641	1.43		X
245	1.000	0.361	0.639	1.43		X
249	2.000	0.858	1.142	2.83	R	X
250	2.000	0.858	1.142	2.83	R	X