A Markovian-Genetic Algorithm Model for Predicting Pavement Deterioration

Ala Sati
College of Engineering, University of Sharjah, Sharjah, United Arab Emirates
u20200159@sharjah.ac.ae

Saleh Abu Dabous
College of Engineering, University of Sharjah, Sharjah, United Arab Emirates
sabudabous@sharjah.ac.ae

Hissam Tawfik
College of Engineering, University of Sharjah, Sharjah, United Arab Emirates
htawfik@sharjah.ac.ae

Abstract

Pavement structures are constantly deteriorating due to many distresses, for instance cracks and rutting that are initiated and expanded. Deterioration models of pavement structures is an important component of pavement management systems (PMS). The deterioration of pavements has been extensively modeled using Markov chains. This paper aims at formulating a more efficient deterioration model to predict the condition of pavement sections. It is proposed to accomplish this by developing a Markovian deterioration model coupled with a meta-heuristic search optimization method, namely genetic algorithms (GA). An essential component of the Markov chain model is the transition probability matrix. In the proposed model, a standard percentage prediction method was used to calculate the transition probabilities. This is then calibrated by integrating the GA method with the Markov chain. The model is based on the historical international roughness index (IRI) data retrieved from the long-term pavement performance (LTPP) database. To test the validity of the method, a real-life case study is used and the performance of the developed model was assessed using both validation and testing data. For predicting pavement conditions, this study concluded that calibrating calculated transition probabilities using meta-heuristic optimization results in better performance than developing the transition probabilities using classical methods. The Markovian-GA model developed in the present study can be used to predict the future condition of pavement facilities in order to assist engineers in planning the optimum maintenance and rehabilitation (M&R) actions.

Keywords: Pavement condition; Infrastructure deterioration; PMS; LTPP; Genetic algorithm

1 Introduction

Civil infrastructure such as roads, bridges, airports, telephones, water networks, etc. are essential to society. Due to the importance of these facilities, infrastructure agencies are responsible for monitoring them to ensure that they remain in good condition to help the economy. Generally, there are many types of transportation infrastructure, including roads, streets, bridges, and parking lots. It is essential for researchers to pay attention to these since their deterioration can lead to serious consequences. Pavement structures that provide the surface of roads, streets, and bridges on which vehicles are especially essential due to their impact on vehicle serviceability and safety. Structural and surface characteristics both influence pavement performance in meeting design criteria throughout its service life.

Pavement deterioration refers to the worsening of pavement or the drop of pavement conditions.

The deterioration of the pavement state continues until the pavement fails and reaches the point where it can no longer serve its function. Deterioration has stochastic nature, and because sometimes a sudden necessity for immediate maintenance and rehabilitation (M&R) actions are needed, pavement management systems (PMS) have become an important part of the decision-making process. As an example, Abaza et al. (2004) integrated PMS with the Markovian model. The designed system becomes an effective decision-making tool that assists the engineers in planning and scheduling the M&R for pavement sections. The primary responsibilities of any PMS is to increase the effectiveness of the pavement decision-making process, broaden the decision-making scope, provide feedback on the decisions made, and maintain a consistent decision-making process at all organizational levels (Hudson et al., 1979). PMS can be used at local, country, state, and federal levels (Beckley, 2016) as illustrated in Golabi et al. (1982), which developed a PMS for the state of Arizona and Picado-Santos et al. (2004) that created a PMS for managing Lisbon pavements. Particularly, pavement management studies confirmed that PMS has proven to be an extremely useful tool in managing large state and metropolitan paving networks (Wolters et al., 2011). However, measuring and predicting the pavement performance is a critical part of any PMS.

In recent years, researchers have shown considerable interest in studying the mechanisms underlying infrastructure deterioration and developing models to forecast the state of various forms of infrastructure, both of which are crucial tasks. In order to make successful, advantageous M&R decisions, an accurate deterioration prediction method is necessary (Ahmed et al., 2020). In deterioration research, there is a great variety in using different data types and different methods to predict the condition of infrastructure facilities, as well as in each research, every developed model is based on a unique set of data. Models of deterioration are categorized into three categories.

The first category is deterministic models, which are the models formulated by a mathematical or statistical equation. The second category is stochastic models, which rely on random variables which create uncertainty in the model. The last category is Artificial Intelligence Models, which are a type of model based on computer techniques that require human intelligence to be developed. In general, all the deterioration models, regardless of the method used in its development, describe how the facility will perform over time. Deterioration models have been extensively simulated using Markov chains (Yosri et al., 2021). For instance, Ranjith et al. (2013) developed a Markovian model to predict the future condition of a timber bridge. Moreover, Li et al. (2014) proposed a Markov chain to generate a deterioration model for urban bridges in Shanghai. For pavement, (Kobayashi et al., 2010; Surendrakumar et al., 2013; and Anyala et al., 2014) implemented the Markov chain method to predict the deterioration of pavement in different places. Among all the discussed Markovian deterioration models, the transition probabilities were the essential part of the Markov chain method.

In this study, the Markov chain approach was proposed to develop the pavement deterioration model. Transition probabilities were developed using a simple percentage prediction method and then calibrated using an optimization method in order to enhance the accuracy of the model prediction. To verify the accuracy of the suggested technique, the produced model’s performance was compared to results from previous research.

2 Materials and Methods

2.1 Data Extraction

In this study, historical pavement roughness data was utilized in developing the models. The data was retrieved from the Long-Term Pavement Performance (LTPP) database. LTPP is the largest pavement study ever conducted for the pavement sections of the United States (US) and Canada (Sati et al., 2020). This study was focused on pavement sections that had not been maintained or rehabilitated, which were 431 sections. The sections with the missing international roughness index (IRI) data were removed. The remaining sections used in this study were 362 sections. The characteristics of the pavement sections, which are the Age, Roadway Functional Class, Climatic Region, Freezing Index, Precipitation, Temperature, AADT, and AADTT were extracted, also from the LTPP database, to categorize the pavement sections.

2.2 Research Methodology

In this study, the methodology framework adopted, presented in Figure 1, consists of many stages. Many steps were executed in each stage. In the first stage, one stage has been performed, which is the collection and preparation of the roughness data. In the second stage, the collected data was converted into condition states utilizing the Federal Highway Administration (FHWA) ranges presented in Table 1. The data was categorized using cluster analysis. K-means cluster analysis is the most popular clustering method. In this method, the sum of the squared error (SSE), calculated by (1), between a cluster's empirical mean and the point within the cluster is minimized.

$$SSE = \sum_{k=1}^{K} \sum_{i=1}^{n} \|x_i - \mu_k\|^2 \quad (1)$$

Where n represents the number of observations within the cluster, K is the cluster number, x_i is the center of the observation, and μ_k is the cluster center.

The elbow method, which is a graphical method where the graph represents the relationship between the number of clusters and the sum of squared error, was proposed to find the optimum number of clusters. In the last stage, which is the core of this study, the model was first developed, then validated, and lastly tested. The following subsections will explain each in detail.

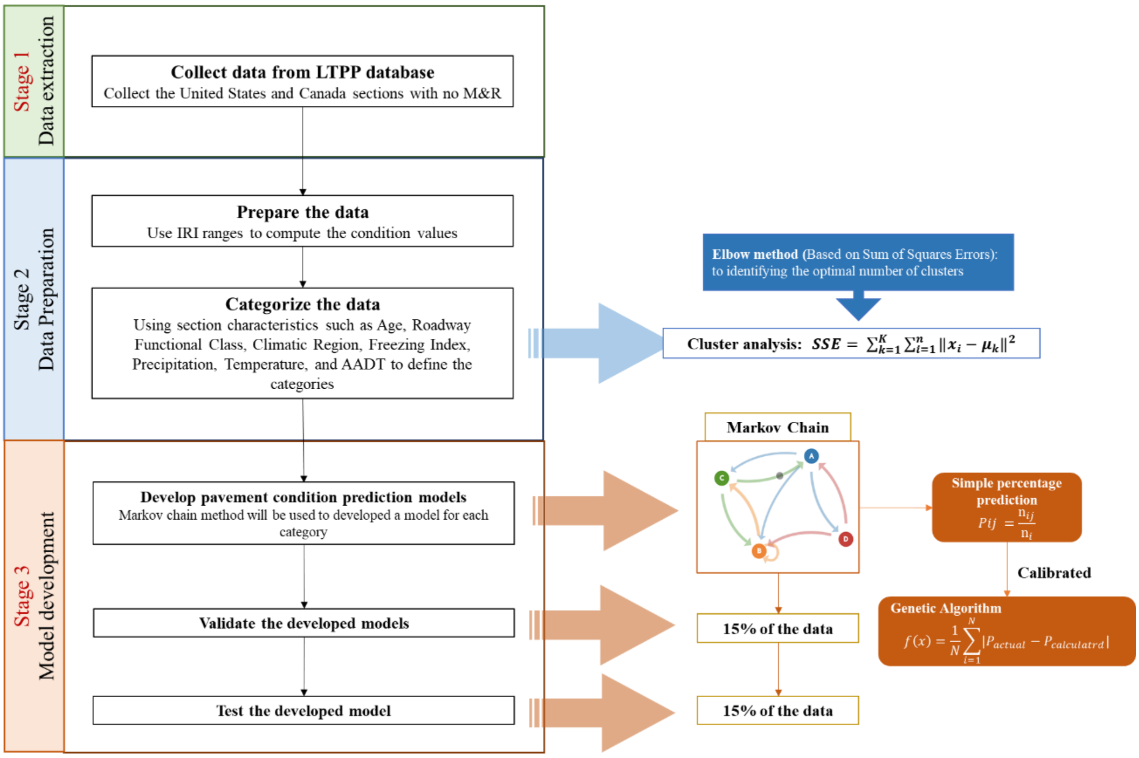

Fig. 1: Research methodology

Table 1: FHWA ranges

State	Category	IRI (m/km)	
		Roadway Functional Class	
		Interstate	Other
1	Very good	< 0.95	< 0.95
2	Good	0.95 – 1.49	0.95 – 1.49
3	Fair	1.50 – 1.89	1.50 – 2.69
4	Mediocre	1.90 – 2.69	2.70 – 3.48
5	Poor	2.69 <	3.48 <

2.3 Proposed Methods

2.3.1 Markov Chain

The Markov chain approach explains the sequence of potential events using a discrete number of events. The aim of the Markovian process is to base predictions purely on the current situation while ignoring the entire history. In a scientific sense, the probability of an event depends entirely on the current situation, regardless of the previous state. The following equation illustrates this process of independence, which is known as the Markov property:

$$\mathrm{P}(\mathrm{X}\,\mathrm{j}) = \mathrm{P}(\mathrm{X}\,\mathrm{t}+1 = \mathrm{i}\,\mathrm{t}+1 | \mathrm{X}\,\mathrm{t} = \mathrm{i}\,\mathrm{t}, \mathrm{X}\,\mathrm{t}-1 = \mathrm{i}\,\mathrm{t}-1, \ldots, \mathrm{X}\,1 = \mathrm{i}\,1, \mathrm{X}\,0 = \mathrm{i}\,0)) = \mathrm{P}(\mathrm{X}\,\mathrm{t}+1 = \mathrm{i}\,\mathrm{t}+1 | \mathrm{X}\,\mathrm{t} = \mathrm{i}\,\mathrm{t}) \tag{2}$$

Where $P(X_j)$, represents the probability of a future condition, t is the current state, while j is the state in the future. The odds of transitioning from state i to state j are known as the transition probabilities and are represented by P_{ij}. The transition probability matrix (n×n), or P, is depicted in (3) and contains all possible outcomes. The number of condition states is n, as well.

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1j} \\ P_{21} & P_{22} & \cdots & P_{2j} \\ \vdots & \vdots & \cdots & \vdots \\ P_{i1} & P_{i2} & \cdots & P_{ij} \end{bmatrix} \tag{3}$$

Using the Markovian process, the time step should be selected first. Then, the future condition vector (P_t) will be predicted by the multiplication of the transition probability matrix and the initial condition vector (P_0). This multiplication is raised to the power represented by the time step, as shown in (4). Lastly, the expected future condition will be calculated using (5).

$$P_j = P_0 \times P^t \tag{4}$$

$$Ev = \sum_{j=1}^{5} j \times P_j \tag{5}$$

In the simple percentage prediction method, the condition data was used to calculate the transitioning probability from state i to state j in terms of the simple mean. This probability is donated by P_{ij} and determined using the following equation.

$$Pij = \frac{\mathrm{n}_{ij}}{\mathrm{n}_i} \tag{6}$$

Where n_{ij} represents the transitions' number from state i to state j, and n_i represents the total number of elements in state i.

2.3.2 Genetic Algorithm Optimization

The most popular meta-heuristic optimization algorithm utilized recently due to its diversity of applications is the genetic algorithm (GA) (Vasuki, 2020). It computes successive generations of solutions based on a set of initial solutions and hypotheses.

The first step in GA is the random creation of the initial population of the solution. The next step is to evaluate each chromosome in the population using the objective function. Then, GA operations will be utilized to create new chromosomes. The new chromosomes will be evaluated as well. The old population will be replaced with new ones. Then, these procedures will be repeated until a near-optimal solution is found. GA has three important operations, reproduction, crossover, and mutation. In addition, there are other operations that can be added to the algorithm based on the need, such as Elitism.

1. Reproduction: is a GA operation that first arranges the chromosomes according to their objective function values, then selects the members from the current population randomly and copies them to the next population. The most commonly used method for selection is the roulette wheel.
2. Crossover: is an operator that crosses over the genes of randomly selected parents to produce a new offspring. It may be one point, two points, or multipoint.
3. Mutation: is a GA operator that is implemented to ensure diversity in the population. In this operation, one or more genes in the parent chromosome string are altered to produce totally new offspring.
4. Elitism: is an operation that can be added before the main GA operations to ensure the survival of the best solutions. This operation copies the elite solutions to the next generation without any changes. If the predefined number of iterations has been reached, an elitist chromosome does not change significantly from one iteration to the next, or when an absolute number of generations is reached, the production of the new generation will cease.

In this study, GA was chosen because of its capability to search for a population of solutions on a global scale instead of a single-based search solution. In this study, GA was used as a calibration process of the transition probabilities generated by the percentage prediction method.

2.3.3 Genetic Algorithm-Based Markovian model

In this case, GA was used to minimize the mean absolute error (MAE) between the transition probabilities based on historical data, which are actual, and the estimated transition probabilities, as shown in (7).

$$Min\ MAE = \frac{1}{N}\sum_{i=1}^{N} |P_{actual} - P_{calculated}| \tag{7}$$

Where N is the number of probabilities in the raw, and P_x is the transition probability. Since the problem is using data for the sections with no M&R, the transition probabilities that were optimized are:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} \\ 0 & P_{22} & P_{23} & P_{24} & P_{25} \\ 0 & 0 & P_{33} & P_{34} & P_{35} \\ 0 & 0 & 0 & P_{44} & P_{45} \\ 0 & 0 & 0 & 0 & P_{55} \end{bmatrix}$$

These probabilities are subject to many constraints. The first one is the summation of each row, which is equal to one.

- $\sum_{j=1}^{5} P_{ij} = 1$, i= 1,2,3,4, & 5

The second one is the probability of staying in the same condition rather than transitioning to the next condition for one time step that is always higher. Following is the mathematical formulation of this constraint:

- $P_{11} > P_{12}, P_{12} > P_{13}, P_{13} > P_{14}, P_{14} > P_{15}$
- $P_{22} > P_{23}, P_{23} > P_{24}, P_{24} > P_{25}$
- $P_{33} > P_{34}, P_{34} > P_{35}$
- $P_{44} > P_{45}$

The last constraint is that the probabilities are always between zero and one. Rarely can they be exactly zero or one.

- $0 \leq P_{ij} \leq 1$

Based on the nature of the problem in this study, each row is optimized separately since the row's probabilities are related to each other, while the probabilities of different rows are not related. MATLAB software was used to implement the GA calculations. Before the main GA operation, 0.05 of the population was counted as elite solutions. In addition, the rate of the crossover was 0.8, which means that 80% of the population was crossed over. Moreover, the default mutation option in MATLAB, which is mutation Gaussian, was used. In this option, each entry of the parent vector is given a random number according to a Gaussian distribution.

3 Results and Discussions

3.1 Cluster analysis

After performing K-means cluster analysis and in order to identify the number of clusters, the Elbow method was applied. The results were based on the sum of squared errors (SSE). Table 2 represents the SSE results that vary with the clusters' number. By graphically representing the results (Figure 2) the cluster's optimum number is equal to three. In addition, the number of sections in each cluster and the division of the data in each cluster are represented in Figure 3.

Table 2: SSE Results

No of clusters	SSE
2	2,305.33
3	1,731.75
4	1,658.42
5	1,466.28
6	1,200.77
7	1,181.78
8	1,119.79
9	994.86
10	949.14

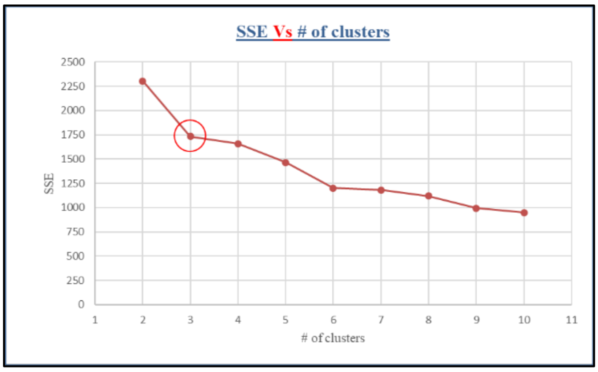

Fig. 2: SSE vs. the number of clusters

Fig. 3: Data division for each cluster

3.2 Markov Chain Transition Probabilities

The following matrices represent the results of calculating the transition probabilities using the traditional method, percentage prediction, then calibrating these probabilities using a Genetic algorithm.

Cluster 1

$$\begin{bmatrix} 0.9577 & 0.0276 & 0.0110 & 0.0047 & 0.0000 \\ 0 & 0.9411 & 0.0493 & 0.0104 & 0.0001 \\ 0 & 0 & 0.9508 & 0.0467 & 0.0025 \\ 0 & 0 & 0 & 0.9902 & 0.0098 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Cluster2

$$\begin{bmatrix} 0.8705 & 0.0787 & 0.0364 & 0.0147 & 0.0007 \\ 0 & 0.9521 & 0.0332 & 0.0137 & 0.0010 \\ 0 & 0 & 0.9762 & 0.0206 & 0.0038 \\ 0 & 0 & 0 & 0.9730 & 0.0270 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Cluster 3

$$\begin{bmatrix} 0.9367 & 0.0363 & 0.0182 & 0.0089 & 0.0002 \\ 0 & 0.9596 & 0.0325 & 0.0078 & 0.0002 \\ 0 & 0 & 0.9875 & 0.0110 & 0.0016 \\ 0 & 0 & 0 & 0.9167 & 0.0833 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

3.3 Performance Results

Table 3 represents the root mean squared error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) results for both validation and testing data. In addition, the performance of different studies was collected and presented in Table 4 to compare it with the performance of the proposed model in this study.

Table 3: RMSE results

	Validation		Testing		
Cluster No.	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**R^2**
1	0.447	0.070	0.316	0.025	0.8334
2	0.277	0.051	0.555	0.115	0.7505
3	0.408	0.065	0.236	0.014	0.8622

Table 4: The models' performance in different studies

The study	Method	RMSE	R^2
Yang et al. (2006)	ANN	0.379	-
	Markov chain	0.346	-
Attoh-Okine et al. (2009)	MARS	-	0.76
Bianchini & Bandini, P. (2010)	Multiple linear regression	0.2272 - 0.3790	0.8022 - 0.9196

In summary, after developing the models, the performance of each one was calculated. Focusing on the testing results, the RMSE was found to be 0.316, 0.555, and 0.236 for cluster 1, cluster 2, and cluster 3, respectively. Moreover, the MAPE values are 0,025 for the first cluster, 0,115 for the second one, and 0,014 for the third cluster. The R^2 ranges between 0.07505 and 0.08622. The results of this study were compared with previous studies that used different methods (Table 4); the results demonstrated that the developed models have acceptable performance. In addition, it is clear from the results that the model of Cluster 3 has the highest performance while Cluster 2 has the lowest performance. To summarize, this study found that the Markov chain performed very well in predicting pavement conditions when calibrating the transition probabilities using GA.

4 Conclusion

This study was performed to enhance the accuracy of the Markov chain deterioration model for pavement sections. Data on pavement roughness was collected from the LTPP database. IRI values were converted into condition states using the ranges specified by FHWA. In order to classify the data into categories, pavement characteristics such as age, roadway functional class, climatic region, freezing index, precipitation, temperature, AADT, and AADTT were used. K-means cluster analysis was integrated with the elbow method to define the number of clusters. According to the results, three clusters were found to be the optimum partitioning. Each cluster's models were developed using the Markov chain method. A simple percentage prediction approach was presented to determine the transition probabilities, which are an essential part of the Markov chain.

Then, the developed transition probabilities were calibrated using genetic algorithms. The data in each cluster was divided into training, validation, and testing sets. The validation and testing data were used to test the model's performance by utilizing RMSE, MAPE, and R^2. The ranges of RMSE, MAPE, and R^2 were 0.236 – 0.555, 0.014 – 0.115, and 0.7505 – 0.8622, respectively. The developed models in this research were compared to other models utilized in previous studies and were found to outperform previous models. This research concluded that the Markov chain models have better accuracy in predicting pavement conditions rather than other models. Furthermore, using the genetic algorithm to calibrate the transition probabilities improves the model's accuracy.

This study was limited to sections with no maintenance and rehabilitation. For future work, other statistical methods such as Poisson distribution can be used to calculate the transition probabilities and then calibrated using a genetic algorithm or other optimization methods. Additionally, data from other infrastructure systems can be used to implement the proposed model to ensure its applicability.

References

Abaza, K. A., Ashur, S. A. & Al-Khatib, I. A. (2004). “Integrated pavement management system with a Markovian prediction model”. *Journal of Transportation Engineering, 130*(1), 24-33. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:1(24)

Ahmed, R., Zayed, T. & Nasiri, F. (2020). “A hybrid genetic algorithm-based fuzzy Markovian model for the deterioration modeling of healthcare facilities”. *Algorithms, 13*(9), 210. https://doi.org/10.3390/a13090210

Anyala, M., Odoki, J. B., & Baker, C. J. (2014). “Hierarchical asphalt pavement deterioration model for climate impact studies”. *International Journal of Pavement Engineering, 15*(3), 251-266. https://doi.org/10.1080/10298436.2012.687105

Attoh-Okine, N. O., Cooger, K., & Mensah, S. (2009). “Multivariate adaptive regression (MARS) and hinged hyperplanes (HHP) for doweled pavement performance modeling”. *Construction and Building Materials, 23*(9), 3020-3023. https://doi.org/10.1016/j.conbuildmat.2009.04.010

Beckley M. E. (2016), “Pavement Deterioration Modeling Using Historical Roughness Data,” *Arizona State University*.

Bianchini, A. & Bandini, P. (2010). ″Prediction of pavement performance through neuro‐fuzzy reasoning″. *Computer‐Aided Civil and Infrastructure Engineering, 25*(1), 39-54. https://doi.org/10.1111/j.1467-8667.2009.00615.x

Golabi, K., Kulkarni, R. B. & Way, G. B. (1982). “A statewide pavement management system.” *Interfaces, 12*(6), 5-21. https://doi.org/10.1287/inte.12.6.5

Hudson, W. R., Haas, R. & Pedigo, R. D. (1979). Pavement management system development. *NCHRP report*.

Kobayashi, K., Do, M. & Han, D. (2010). “Estimation of Markovian transition probabilities for pavement deterioration forecasting”. *KSCE Journal of Civil Engineering*, *14*(3), 343-351. https://doi.org/10.1007/s12205-010-0343-x

Li, L., Sun, L. & Ning, G. (2014). “Deterioration prediction of urban bridges on network level using Markov-chain model”. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2014/728107

Santos, et al. (2004). “Pavement management system for Lisbon”. In *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, 157(3), 157-165. Thomas Telford Ltd. https://doi.org/10.1680/muen.2004.157.3.157

Ranjith, et al. (2013). “Deterioration prediction of timber bridge elements using the Markov chain.” *Journal of Performance of Constructed Facilities*, *27*(3), 319-325. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000311

Sati, A. S., Dabous, S. A. & Zeiada, W. (2020). “Pavement deterioration model using Markov chain and international roughness index”. In IOP Conference Series: *Materials Science and Engineering*, *812*(1), 012012). IOP Publishing. DOI: 10.1088/1757-899X/812/1/012012

Surendrakumar, K., Prashant, N. & Mayuresh, P. (2013). “Application of Markovian probabilistic process to develop a decision-support system for pavement maintenance management”. *International Journal of Scientific & Technology Research*, *2*(8), 295-303.

Vasuki, A. (2020). *Nature-inspired optimization algorithms*. CRC Press.

Wolters, et al. (2011). “Implementing Pavement Management Systems for Local Agencies: Implementation Guide” (No. ICT-11-094-1).

Yang, et al. (2006). “Modeling crack deterioration of flexible pavements: Comparison of recurrent Markov chains and artificial neural networks”. *Transportation Research Record*, *1974*(1), 18-25.

Yosri, et al. (2021). “Genetic algorithm-Markovian model for predictive bridge asset management”. *Journal of Bridge Engineering*, *26*(8), 04021052. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001752

Cite as: Sati A., Abu Dabous S. & Tawfik H., “A Markovian-Genetic Algorithm Model for Predicting Pavement Deterioration”, *The 2nd International Conference on Civil Infrastructure and Construction (CIC 2023)*, Doha, Qatar, 5-8 February 2023, DOI: https://doi.org/10.29117/cic.2023.0089