# A Probabilistic-Based Deterioration Model Using Ground Penetrating Radar

**Mohamed Marzouk[a], Eslam Mohammed Abdelkader[b], and Tarek Zayed[b]**

[a] Structural Engineering Department, Faculty of Engineering, Cairo University, Egypt
[b] Department of Building, Civil, and Environmental Engineering, Concordia University, Canada
E-mail: mm_marzouk@yahoo.com, eslam_ahmed1990@hotmail.com, zayed@encs.concordia.ca.

**Abstract – Infrastructure systems are very essential to every aspect of life on Earth. Existing Infrastructure is subjected to degradation while the demands are growing for a better infrastructure system in response to the high standards of safety, health, population growth, and environmental protection. Bridges play a crucial role in urban transportation networks. In addition to that, they are vulnerable to several deterioration agents such as the variable traffic loading, extreme weather conditions, deferred maintenance, etc. The development of Bridge Management Systems (BMSs) has become a fundamental imperative nowadays especially in the large transportation networks due to the huge variance between the need for maintenance actions, and the available funds to perform such actions. Condition assessment is regarded as one of the most critical and vital components of BMSs. Ground Penetrating Radar (GPR) is one of the non-destructive techniques (NDTs) that are used to evaluate the condition of bridge decks which are subjected to the rebar corrosion. This paper presents a corrosiveness index that is capable of evaluating the extent of severity of corrosion in concrete bridge decks. Different clustering algorithms are compared in order to select the most feasible clustering algorithm. A probabilistic deterioration model is constructed in order to model the future condition rating of bridge decks. Anderson darling test is used to select the most feasible probability density function, and the parameters of the probability density functions are obtained using maximum likelihood estimation. Finally, two case studies are presented to illustrate the capabilities of the proposed model.**

**Keywords –**
Bridge Management System; ground Penetrating Radar; bridge decks; corrosion; deterioration model; maximum likelihood estimation

## 1 Introduction

Bridges are vital links in transportation networks that should be safe, functional and serviceable during their service life to facilitate the mobility of people and transportation of goods which causes sustainable economic development. Concrete bridges are prone to high level of deterioration because of the variable traffic loading, extreme weather conditions, cycles of freeze and thaw, deferred maintenance, etc. More than 40% (700 bridges) of Egyptian bridge inventory out of 1,700 bridges have exceeded their maintenance limit and are at risk of failure.

American Association of State Highway and Transportation Officials (AASHTO) defined Bridge Management System (BMS) as "a system designed to optimize the use of available resources for inspection, maintenance, rehabilitation and replacement of bridges" [1]. AASHTO and Intermodal Surface Transportation Efficiency Act (ISTEA) defined five main components for BMS which are as follows [2]: 1) database for data storage, 2) condition rating model, 3) deterioration model, 4) cost model, and 5) optimization model for running system. The structure of BMS is shown in Figure 1 [2].

The database is the most essential component of the BMS where it is used to store information that is related to every bridge in the network. Bridge condition rating is based on field inspections to evaluate the condition of bridges. Deterioration model is used to predict the condition rating of different bridge elements over time. The cost model is divided into two main types which are: 1) agency costs, and 2) user costs. The agency costs ae based on the expenditure of maintenance, repair and rehabilitation activities (MR&R) to improve the condition rating of the bridges. The user costs are based on the impact of deterioration on road users as a function of the bridge condition [3]. The optimization model is implemented to determine the optimum maintenance, repair and rehabilitation activities for different bridge components. The proposed model focuses on two bridge management components which are: 1) the condition assessment model, and 2) the deterioration model. A corrosion index is developed based on the ground penetrating radar. Also, the weibull distribution is used to model the deterioration of bridge decks. Accurate condition rating and reliable deterioration models help in minimizing the risk of failure of aging infrastructures.

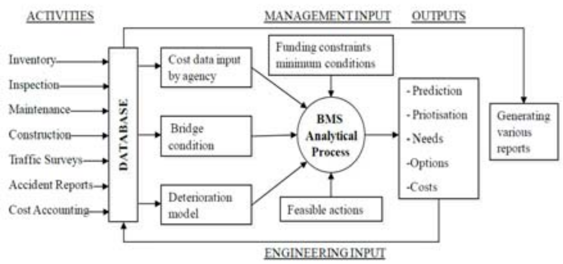


Figure 1: Typical structure of BMS [2]

## 2 Destructive and Non-Destructive Techniques

The condition assessment model heavily depends on the inspection type. The process of monitoring bridges should be cost effective, efficient, and fit for the purpose [4]. Thus, bridge inspectors should use techniques to assess accurately and effectively the corrosion of the bridge decks. Bridge inspection can be defined as “a process in which the defects on a bridge are identified, recorded to be used for assessing bridge condition”. Inspection techniques are divided into two main categories which are: 1) destructive techniques (DT), and 2) non-destructive techniques (NDT). Destructive Techniques (DT) provide accurate and direct results, but they cause damage to the element under investigation, and they are expensive and time-consuming. Non-Destructive Techniques (NDT) are inexpensive and quick, but they do not provide direct information about the element under inspection. NDTs gain popularity due to their various advantages including; providing a high level of safety for the labor staff, time saving, providing high rates of production in comparison to traditional methods.

Ground penetrating radar (GPR) is one of the non-destructive techniques that are used for field investigation in structural engineering. Ground Penetrating radar can determine the subsurface structure easily and accurately. Moreover, it has the capability of locating metallic and non-metallic objects. GPR transmits pulsed electromagnetic waves from the transmitting antenna which is located on the ground surface and signals are then received by the receiving antenna [5]. The proposed model utilizes GPR in order to evaluate the corrosion of the reinforcement rebars in the concrete bridge decks.

## 3 Model Development

The utilized GPR system incorporates two ground-coupled antennas which are: 2000 MHZ and 400 MHZ. The 2000 MHZ antenna is utilized because of its high-resolution capability. Thermal drift, electronic instability, cable length differences and variations in antenna airgap can cause ‘jumps’ in the air/ground wavelet first arrival time (usually referred to as the time-zero point). Therefore, traces require the adjusting to a common time-zero position before applying processing methods. This is usually achieved using some particular criteria (e.g., the air wave first break point or first negative peak of the trace) and is often done automatically by the processing software [6, 7, 8].

Several precautions are considered during the operation and processing processes. For the thermal drift, the scanning process was performed at night 2:00 am to overcome the internal heating of the equipment during the operation process. The antenna air gap is removed by moving the profile start time to the asphalt layer. For the dielectric constant of the material (concrete in our case), the wave length of the electromagnetic waves decreases as they encounter higher dielectric material. The utilized GPR is equipped with two antennas. The utilized GPR is equipped with two antennas which are: 2000 and 400 MHZ. Thus, in case the electromagnetic waves fail to reach the desired penetration depth, the lower frequency (400 MHZ) is utilized. The GRED HD software is used to extract the amplitudes of the top reinforcement rebars.

There are two main methods for GPR data analysis which are: 1) numerical amplitude method, and 2) image-based method. Numerical amplitude method depends on the value of amplitude of the reflected waves from the top layer of reinforcement. The higher the amplitude, the better the condition of the bars. This method main limitation is its lack of a clear value for the thresholds that define the different categories of corrosion. For example, the profiles of one bridge deck may have amplitude values from 10 dB to -5 dB, where 10 dB represents the best condition and -5 dB represents the worst for that bridge. Meanwhile, another of Bridges’ profiles may have amplitude values that range from -5 dB to -40 dB, where -5 dB represents the best condition and -40 dB represents the worst condition.

Therefore, a comparison between the clustering algorithms should be performed in order to select the most feasible clustering algorithm. After the selection of the most appropriate clustering algorithm, a corrosion map is developed and consequently a corrosion index is calculated. The second step is to construct a deterioration model. A stochastic model is constructed in order to overcome the vagueness and uncertainties associated with the deterministic models. The most suitable probability density function is defined based on the Anderson Darling statistic. The parameters of the probability density function are defined based on the maximum likelihood estimation algorithm.

## 3.1 Clustering Algorithms

The proposed model develops a corrosion map that is extensively based on the clustering algorithms. Clustering is the process of partitioning the dataset into a homogenous set of clusters without having any prior information about the clusters where the points within the same cluster share similar features. The selected clustering algorithms incorporate a combination of soft and hard clustering techniques. Hard clustering is the process of the assignment of data points to only one cluster such as K-means. On the other hand, soft clustering is the process of the assignment of data points to the clusters with different membership degrees such as Fuzzy C-means clustering (FCM).

RapidMiner 7.5 and KNIME 3.3.1 softwares [9, 10] are used as platforms to perform the clustering algorithms. The proposed methodology compares between five clustering algorithms which are: K-means, K-medoids, fuzzy C-means, expectation maximization, and X-means. K-means and k-medoids are two similar clustering algorithms. The proposed model assumes that the there are four categories for the condition of the bridge deck which are: “very severe”, “severe”, “medium”, and “good”. i.e., three thresholds

### 3.1.1 K-means algorithm

K-means clustering algorithm is based on minimizing the distance between the average squared Euclidean distance and the clusters’ centroids. The output of the clusters’ centroids is greatly influenced by the initial selection of the clusters’ centroids. The number of clusters in the K-means clustering algorithm must be defined initially. The clusters’ centroid is the mean of the data points within the cluster [11]. The output of the K-means is that the similarity between the data points within the same cluster is higher than the similarity between the data points in the different clusters [12]. The main distinct feature between K-means and k-medoids clustering algorithms is that one of the data points represents the centroid of the cluster in the case of k-medoids. K-means algorithm utilizes the mean of the data points.

The steps of K-means clustering are as follows [13]:

1- Select number of desired clusters $K$.
2- Select $K$ starting points randomly to be used as initial candidates for clusters’ centroids.
3- Calculate the distance between data points and cluster centroids.
4- Assign the data point to the cluster centroid which has the minimum distance between the data point and cluster centroids. The distance is simply the Euclidean distance.

$$d(x_i, C_j) = \sqrt{\sum_{d=1}^{n} (x_{id} - c_{id_i})^2} \quad (1)$$

5- Re-compute the new cluster centroids (centroid is the mean point of the cluster).
6- Repeat steps 3, 4, and 5 until convergence (centroid and data points no longer move).

X-means clustering algorithm is introduced in order to overcome a major drawback of the K-means clustering algorithm which is the necessity of knowing the number of clusters ( $K$ ) before processing. The X-means clustering algorithm tends to search the space to compute the clusters’ centroids based on the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). The expectation maximization clustering algorithm calculates the probabilities of cluster memberships based on one or more probability distribution. The number of the clusters is predetermined in the expectation maximization algorithm. Expectation maximization algorithm is based on maximizing the probability that the data point belongs to the clusters of the model.

### 3.1.2 Fuzzy C-means Algorithm

Fuzzy C-means (FCM) is an iterative clustering algorithm where each data point is assigned to one cluster or more based on the membership degrees. FCM was developed by Dunn in 1973 and improved by Bezdek in 1981. FCM is based on minimizing the following objective function [14].

$$J_w = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| \left(X_i - C_j\right)^2 \right\| \quad (2)$$

Where;
$m$ is a fuzzifier constant that is greater than one. $u_{ij}$ denotes the degree of membership of the $X_i$ in the cluster $j$ and it is between zero and one. $X_i$ is a $i-th$ data point in a d-dimensional space. $C_j$ represents the centroid of the $j-th$ cluster. $||*||$ is a norm distance that represents the similarity between the data point and the centroid of the cluster.

FCM starts by randomly initiating the cluster centroid. The second step is to construct the membership matrix. A membership matrix ($U_{(N\times C)}$) is composed of a group of membership degrees. The degree of membership ($u_{ij}$) can be calculated using Equation 3. The cluster centroids are then updated and can be calculated using Equation 4. The cluster centroids and the membership degrees are iteratively updated until the convergence criteria are satisfied. The convergence criteria is shown in Equation 5. The de-fuzzification process is performed using Equation 6 where the data point is assigned to the cluster that has the maximum degree of membership.

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|(x_i-c_j)\|}{\|(x_i-c_k)\|}\right)^{\frac{2}{m-1}}} \quad (3)$$

$$C_j = \frac{\sum_{i=1}^{N} u_{ij}^{m}\times x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \quad (4)$$

$$\max_{ij}\{|u_{ij}^{it+1} - u_{ij}^{it}|\} < \zeta \quad (5)$$

$$C_j = \arg_i\{max(u_{ij})\} \quad (6)$$

$$\sum_{j=1}^{c} u_{ij} = 1 \quad (7)$$

Where;

$\zeta$ is the termination constant between zero and one. $it$ refers to the number of iteration steps.

## 3.2 Clustering Validity Indices

Clustering is an un-supervised algorithm. Therefore, evaluating the output of the clustering algorithms is a matter of great importance. Assessing the clustering algorithms is much more difficult than the supervised algorithms because there is no “ground truth”, i.e., there are no-predefined classes for the domain problem. Moreover, it is very difficult to find the appropriate metrics to evaluate the quality of the generated clusters.

The proposed model utilizes two clustering validity approaches to assess the quality of the generated clusters and to identify the optimal partition of clusters which are: 1) Davies-Bouldin index, and 2) Dunn index. The objective of the clustering validity approaches is to select the most feasible thresholds that ensure that the clusters are compact and well-separated, i.e., maximize the inter-cluster distance (distance between the clusters), and minimize the intra-cluster distance (distance between data points within the same cluster) (see Figure 2).

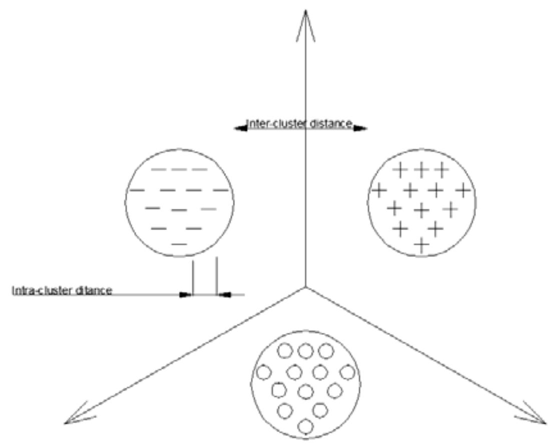


Figure 2: Overview of the clustering process

### 3.2.1 Davies-Bouldin Index

Davies-Bouldin index is a ratio between the sum of intra-cluster scatter to the inter-cluster separation. The Davies-Bouldin index can be calculated using the following Equation [15].

$$DBI = (\frac{1}{k}\sum_{w,v=1}^{k} \max_{v\neq w}(\frac{D_w+D_v}{d(c_w,c_v)}) \quad (8)$$

Where;

$D$ represents the intra-cluster distance. $d$ represents the inter-cluster distance.

The intra-cluster distance ($D$), and the inter-cluster distance ($d$) can be calculated using Equations 9, and 10, respectively based on the Euclidean distance principle. The intra-cluster distance is the average distance between the data points and the cluster centroid. The inter-cluster distance is the distance between the centroid of the two clusters.

$$D = \left(\frac{\sum_{h=1}^{t}||X_a - C_w||}{N_w}\right) \quad (9)$$

$$d = dist(C_w, C_v) = \sqrt{(C_{w1}-C_{v1})^2 + (C_{w2}-C_{v2})^2 \ldots \left(C_{wq}-C_{vq}\right)^2} \quad (10)$$

Where;

$X_a$ is an arbitrary data point that belongs to a cluster $w$. $C_w$, and $C_v$ represent the centroid of clusters $w$, and $v$, respectively. $N_w$ represents number of data points in the cluster $w$. The smaller Davies-Bouldin index denotes that the clusters are compact, and the centers of the clusters are far away from each other (Sahani and Bhuyan 2014). $t$ is the number of the data points in the cluster. $q$ is the number of the dimensions of the model.

### 3.2.2 Dunn Index

The Dunn index is used to assess the quality of the clusters and it can be calculated using Equation 11 [16].

$$\mathrm{DUI} = \max_{1\le w\le k}\left(\min_{1\le v<K, v\ne w}\left(\min_{i\le K\le K}\left(\frac{d(x_w,x_v)}{\max(D(x_k))}\right)\right)\right) \quad (11)$$

Where;

D, and d are defined as above. The larger the Dunn index indicates that the clusters are compact and well-separated.

The clustering index is performed in order to compare between the five clustering algorithms. The clustering index can be calculated using the following equation.

$$CLU = \frac{\mathrm{DUI} - DBI}{2} \quad (12)$$

Where;

DUI represents the Dunn index while $DBI$ represents the Davies-Bouldin index. $CLU$ denotes the clustering index.

Surfer 12 is a plotting and mapping software that is utilized to develop the corrosion map for the concrete bridge decks. Finally, a Corrosion Index (CI) can be calculated as follows.

$$CI = \frac{\sum_{i=1}^{4} Q_i \times W_i}{\sum_{i=1}^{4} Q_i} \quad (13)$$

Where; $Q_i$ represents the quantity of a bridge element in category $i$. $W_i$ represents the weighting factor for a bridge element in category $i$. The weighting factors for the "good", "medium", "severe", and "very severe" categories are assumed 100%, 70%, 50%, and 20%, respectively.

## 4 Deteriroation Models

Deterioration model is the most crucial component of the BMS because it enables the transportation authorities to predict the future bridge condition ratings. Planning of maintenance, repair and rehabilitation activities (MR&R) of bridges is based on calculating accurate future bridge condition ratings. A high-quality deterioration model enables infrastructure managers to optimize MR&R activities and minimize un-planned maintenance activities. The deterioration model constructs a relationship between the facility condition rating and group of explanatory variables such as age, traffic volume, weather conditions, percentage of commercial vehicles, etc.

Goodness of fit is a statistical measure that is used to determine the compatibility of fitting set of data to probability distributions. Goodness of fit is a means to evaluate different probability density functions and to determine discrepancies between observed values and expected values in a certain statistical model. There are several goodness o fit tests such as Kolmogorov-Smirnov test, Anderson Darling and chi-squared test. The proposed model utilizes Anderson Darling statistic.

Anderson Darling test ($A^2$) is used to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. Anderson darling provides more weight to the distributions' tail than the Kolmogorov-Smirnov test. Anderson Darling Statistic can be calculated as follows [17].

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2\mathrm{i}-1) \times (\ln F(x_i) + \ln(1 - F(X_{n-i+1}))) \quad (14)$$

There are different methods for parameter estimation of probability density function such as maximum likelihood estimation and least squares, method of moments estimation (MME), and median rank regression (MRR). The proposed model utilizes maximum likelihood estimation (MLE) algorithm as a parameter estimation method. MLE is based on finding the parameters that maximizes the likelihood function where the observations are assumed to be independent. The MLE is characterized by being is asymptotically efficient where larger the sample size the more likely the parameters converge to precise values.

MLE is based on finding the unknown parameter θ that maximizes the $\log L(\theta/y)$ because it is often easier to maximize the log-likelihood function than the likelihood function itself [18].

$$\log L(\theta/y) = \sum_{i=1}^{n} \log(f(y_i|\theta)) \quad (15)$$

Where;

$Y_i$ represents a set of independent variables. $f(y_i|\theta)$ denotes the probability density function of the random variable $y_i$

## 5 Model Implementation

There are two case studies to validate the proposed model. The length and the width of the scanned bridge deck of El-Kobba bridge are 20 meters, and 3 meters, respectively. The length and the width of the scanned bridge deck of Torra bridge are 21 meters, and 6 meters. The data collection of El-Kobba bridge deck is shown in Figure 3.

Figure 3: GPR data collection for El-Kobba bridge deck

The grids spacing in both the longitudinal and transversal directions were taken to be 1.0 m. Consequently, the number of scans in the longitudinal and transversal directions in El-Kobba bridge are 3, and 21, respectively. The number of scans in the longitudinal and transversal directions in Torra bridge are 6, and 20, respectively. A plan of the scans of El-Koba bridge deck is shown in Figure 4.

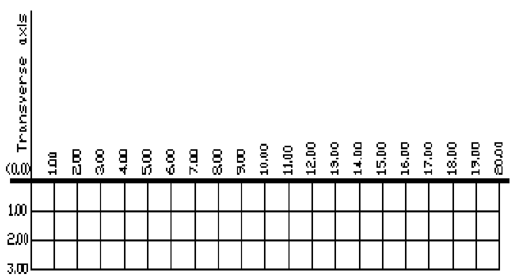


Figure 4: A plan of the scanned paths in El-Kobba bridge deck

After the data collection process, the GPR data profiles are transferred to a PC for processing. The software IDS GRED HD is used to process the B-scans. IDS GRED HD is used to compute the travel time, and the amplitude of each reinforcing rebar. The profile obtained from the GPR is depicted in Figure 5.

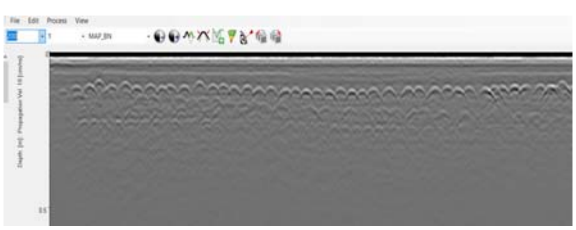

Figure 5: GPR profile of Torra bridge deck

Rebar locations were clear in the raw data as well as obvious areas of deterioration where rebar reflections were weak. Areas of the bridge deck having weak reflection amplitude values are typically indicative of the deterioration. These weaker reflections can be due to several factors, including high chloride content, concrete deterioration or corrosion of the embedded steel rebar, which all attenuate the radar signal.

The RapidMiner software is used in order to perform the K-means clustering, expectation maximization cluster, X-means, and K-medoids. KNIME platform is used in order to perform the fuzzy C-means clustering. The developed clustering model using the RapidMiner platform is shown in Figure 6. The developed clustering model using the KNIME platform is shown in Figure 7.

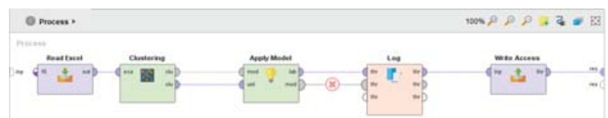


Figure 6: Interface of the RapidMiner platform


Figure 7: Interface of the KNIME platform

The clustering indices of the five clustering algorithms in Tora and El-Kobba bridges are shown in Table 1. As shown in Table 1, K-medoids has the largest clustering index and therefore K-medoids is the most feasible clustering algorithm. Moreover, K-medoids has the largest clustering algorithm in El-Kobba bridge which means that K-medoids is the most feasible clustering algorithm

Table 1: Clustering indices of the two bridges

| Clustering algorithm | Clustering index (EL-Kobba) | Clustering index (Torra) |
|---|---|---|
| K-means | 1.4118 | 1.1566 |
| K-medoids | 1.8954 | 2.8291 |

| | | |
|---|---|---|
| X-means | 1.3948 | 1.2322 |
| Expectation maximization | 1.5585 | 1.3896 |
| Fuzzy C-means | 1.5118 | 2.3124 |

The corrosion map for El-Kobba bridge is depicted in Figure 8 where 18.67% of the bridge is a “very severe” condition, 43.98% of the bridge is in a “severe” condition, 30.71% is in a “medium” condition, and 6.64% is in a “good” condition. Based on Equation (13), the Corrosion Index is 70.76% which means that the bridge deck is in the “medium” category. The amplitude thresholds are expressed in volts. The amplitude thresholds are 0.106, 0.202, and 0.325. The corrosion map for Torra bridge is depicted in Figure 9 where 14.8% of the bridge is a “very severe” condition, 22.37% of the bridge is in a “severe” condition, 26.64 is in a “medium” condition, and 36.18% is in a “good” condition. The Corrosion Index is 68.98% which means that the bridge deck is in the “medium” category. The amplitude thresholds are -0.477, -0.3605, and -0.1999.

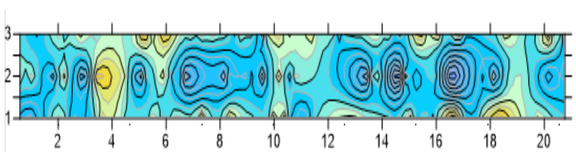

Figure 8: Corrosion map of El-Kobba bridge

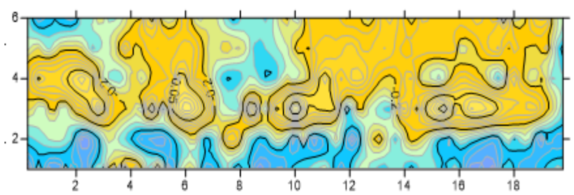

Figure 9: Corrosion map of Torra bridge

The attributes of the bridges are stored in shapefile format for documentation purposes as shown in Figure 10. The shapefile format is a well-known geospatial vector data format for geographic information system (GIS) softwares.

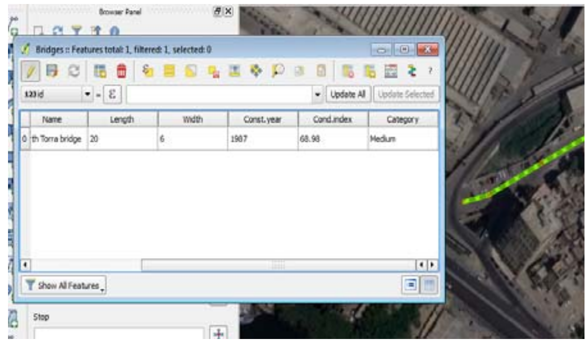


Figure 10: Storing of the bridges’ attributes

Based on the Anderson Darling test, weibull distribution is the best-fit distribution followed by the lognormal distribution, and finally the exponential distribution, whereas the Anderson Darling statistics of the previous distributions are 0.٤٤٨٨, 0.٤٧٠٥, and 0.٦٥٣٩, respectively. The parameters of the weibull distributions are obtained based on the maximum likelihood estimation, whereas the shape factor and scale factor are ١,٩١٧٤, and ٥٠,٩٣٥٤, respectively. The deterioration model based on the weibull distribution is shown in Figure 11.

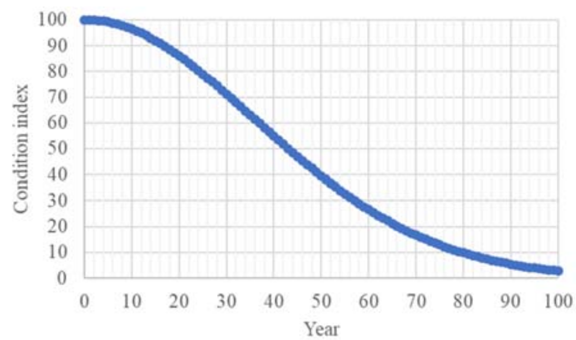


Figure 11: Deterioration model based on weibull distribution

## 6 Acknowledgment

This project was funded by the Academy of Scientific Research and Technology (ASRT), Egypt, JESOR-Development Program - Project ID: 40

## 7 References

[1] AASHTO Highways subcommittee on Bridges and Structures. *The Manual for Bridge Evaluation*, 2011, doi:10.1017/CBO9781107415324.004.

[2] Elbehairy, H. *Bridge Management System with Integrated Life Cycle Cost Optimization*. PH.D

thesis, Waterloo University, Canada, 2007.

[3] Morcous, G. Comparing the use of artificial neural networks and case-based reasoning in modeling bridge deterioration. *Annual Conference of the Canadian Society for Civil Engineering*, 1–9, 2002.

[4] Alani, Amir M., Morteza A., and Gokhan K. Applications of Ground Penetrating Radar (GPR) in Bridge Deck Monitoring and Assessment. *Journal of Applied Geophysics*. 97, 45–54, 2013, doi:10.1016/j.jappgeo.2013.04.009.

[5] Furuta, K., Ishikawa, J. *Anti-personnel Landmine Detection for Humanitarian Demining*, Springer-Verlag London, 2009.

[6] Olhoeft, G.R., and Smith III, S., S. Automatic processing and modeling of GPR data for pavement thickness and properties. *8th International Conference on Ground Penetrating Radar*, 188-193, 2000.

[7] Nobes, D. C. Geophysical surveys of burial sites: a case study of the Oaro urupa, *Geophysics*, 64(2): 357-367, 1999.

[8] Cassidy, N. J. Ground penetrating radar data processing, modelling and analysis. *Ground penetrating radar: theory and applications*, 141-176, 2009.

[9] RapidMiner, Inc. RapidMiner 7.5. On-line: https://rapidminer.com/products/studio/ (accessed 8.6.2016), Boston, USA.

[10] KNIME. KNIME 3.3.1. On-line: https://www.knime.org/ (accessed 8.6.2016), Zürich, Switzerland.

[11] Poteras, C. M., Mihaescu, M. C., and Mocanu, M. An Optimized Version of the K-Means Clustering Algorithm. Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, 2, 695–699, 2014.

[12] Atabay, H. A., Sheikhzadeh, M. J., and Torshizi, M. A Clustering Algorithm Based on Integration of K-means and PSO. *1st Conference on Swarm Intelligence and Evolutionary Computation*, 59–63, 2016.

[13] Sawant, K. B. Efficient Determination of Clusters in K-Mean Algorithm Using Neighborhood Distance. International *Journal of Emerging Engineering Research and Technology*, 3(1), 22–27, 2015.

[14] Keskin, G. A. Using integrated fuzzy DEMATEL and fuzzy C: means algorithm for supplier evaluation and selection. *International Journal of Production Research*, 2014.

[15] Davies, D. L., and Bouldin, D. W. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227, 1979.

[16] Dunn, J. C. Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, 4(1), 95–104, 1974.

[17] Love, P. E. D., Wang, X., Sing, C., and Tiong, R. L. K. Determining the Probability of Cost Overruns. *Journal of Construction Engineering and Management*, 139, 321–330, 2013.

[18] Nielsen, M. A. *Parameter estimation for the two-parameter weibull distribution*. M.Sc. thesis, Brigham Young University, United States of America, 2011.