Life-Cycle of Structures and Infrastructure Systems – Biondini & Frangopol (Eds)

Life-cycle assessment of R.C. bridge components based on cluster analysis and stochastic process

F. Marsili & S. Keßler
Helmut-Schmidt-University/University of the Federal Armed Forces, Hamburg, Germany

F. Landi
University of Pisa, Pisa, Italy

ABSTRACT: This article proposes an approach to assess the life-cycle of reinforced concrete bridge components by applying a cluster algorithm and a stochastic model for damage evolution. The k-means algorithm is used to identify families of bridge components that deteriorate at similar rates. A measure of performance, i.e., silhouette width, supports the choice of the optimal number of clusters. Once the cluster model is defined, a gamma process is fitted to the data on the evolution of the conditions that belong to each family. By simulating the gamma process, the cumulative distribution function of time to failure is calculated for each cluster of components. The procedure applies to reinforced concrete bridge components in Switzerland, whose inspection and maintenance data is collected in the KUBA-DB database. This approach ensures that the expected service life of bridge components can be predicted with limited uncertainty.

1 INTRODUCTION

The management of existing infrastructure and its maintenance is an increasingly topical and relevant issue for the well-being of our societies. While, on the one hand, and as a result of a period strongly marked by new construction, physical infrastructure assets are aging and increasingly in need of maintenance, on the other hand, investment in maintenance is insufficient and even declining. Keeping infrastructures in good condition is therefore a challenge, requiring highly advanced approaches that can significantly improve maintenance efficiency. Remaining useful life should be understood as the time remaining before the infrastructure reaches a critical condition and maintenance is required to ensure safety, serviceability, or durability. However, and despite the fact that physical assets are ageing, only in few cases the condition has been judged critical, so that little information is available concerning the time when a critical condition is reached.

With the scope to increase maintenance efficiency, many infrastructure operators have undertaken a digitization process, which has led to the development of digital systems for managing infrastructure and its maintenance. Condition databases collect the results of periodic visual inspections and special tests, which are represented by digital data of different types, such as images, and qualitative and quantitative information on the progress of degradation. An important information is the qualitative judgment issued at the end of each inspection which regards the criticality of damage, as well as the condition of the component and the object as a whole, which is also linked to the urgency of the maintenance intervention.

A maintenance intervention is deemed necessary when the safety, durability or serviceability of the component and therefore possibly also of the object are jeopardized by the deterioration process. The judgment is expressed through an increasing condition index: the more significant the damage, the higher the index. The condition index can thus be interpreted as a measure of damage accumulation. The objects whose information is contained in these databases were built in different periods and therefore have different ages. As a result, not all the objects have already reached a critical condition. In addition to that, the objects have different characteristics, which affect their

DOI: 10.1201/9781003323020-303

robustness against aging in different ways. Since the information relating to objects in satisfactory condition are predominant over those in critical condition, the database is affected by the well-known problem of data imbalance, and information about the time in which the critical condition is reached cannot be directly derived from the database in most of the cases.

The evaluation of the time in which the damage reaches a critical condition is subject to uncertainties that can significantly influence the maintenance planning and therefore must not be neglected. One way to consider the uncertainties affecting the prediction of the remaining service life is to model damage evolution through probabilistic models such as Markov process or gamma process (Frangopol et al., 2004). Markov process is currently the most commonly used approach in bridge maintenance models. However, it has some drawbacks: first, it does not provide any reliability estimation, making it also difficult to link life-cycle analysis to structural reliability assessment; secondly, it is a condition model suitable for incorporating information from visual inspections, but not for predicting the attainment of conditions that have not already been observed. This represents a major problem in the case of imbalance data such as condition data. Gamma processes have so far been less applied, but do not have these drawbacks. The gamma process is suitable to model gradual damage monotonically accumulating over time, such as wear, fatigue, corrosion, crack growth, erosion, consumption, creep, swell, a degrading health index (van Noortwijk, 2009), as in the case of damage affecting civil infrastructures and reinforced concrete bridges. A gamma process can be fitted to available degradation data and used to assess the probability of reaching a critical condition as a function of time. Gamma processes have been already used for modelling fatigue damage evolution (Guida and Penta, 2015), deterioration of coating systems (Nicolai et al., 2007), and corrosion in reinforced concrete structures (Zhang et al., 2023). Modelling the deterioration as a gamma process is also suitable when visual inspections are involved and summarized into a condition index, provided that the condition index can be interpreted as a measure of damage accumulation (Edirisinghe et al., 2013).

Although uncertainties are inevitable, it is necessary to assess the probability of reaching a critical condition and consequently to predict the remaining service life as precise as possible. In other words, it is necessary to identify groups of objects, or objects components, which show similarities in terms of condition evolution paths. Unsupervised learning techniques and cluster algorithms are data mining techniques suited for finding patterns in big data sets. A review of the application of these techniques to life-cycle assessment can be found in Ghoroghi et al. (2022). Cluster analyses have been already applied to identify spatial clusters of structurally deficient bridges (Amin et al., 2020), to identify damage patterns affecting bridges (Chang and Chi, 2019), to develop bridge deterioration models (Moscoso et al., 2022). Although a grouping of objects based on similar condition evolution paths has already been theorized and proposed by infrastructure operators (Marsili et al., 2018), this result has never been pursued in a completely empirical way, by applying cluster algorithms to condition databases for infrastructure management.

The objective of this paper is to assess the life-cycle of aging reinforced concrete (r.c.) bridge components by combining a cluster analysis with a stochastic process. The k-means algorithm is implemented to cluster bridge components with similar degradation paths. Once the families of bridge components with similar condition evolution have been identified, the gamma process is fitted to the data characterizing each cluster. This approach makes it possible to predict the service life of bridge component with higher precision and construct the cumulative distribution of time to reach an undesired condition, which can be referred to as time to failure, for each identified group of components. The paper is organized as follows: Section 2 gives some background information about the methods at the basis of the proposed approach, Section 3 presents the case study at which the procedure has been applied and Section 4 draws some conclusion.

2 METHODS

2.1 *Gamma process*

van Noortwijk (2009) presents a survey of gamma processes in maintenance. By following its presentation of gamma process, a random quantity X has a gamma distribution with shape parameter $v > 0$ and scale parameter $u > 0$ if its probability density function (PDF) is given by

$$\mathrm{Ga}(x|v,u) = \frac{u^v}{\Gamma(v)} x^{v-1} \exp(-ux) I_{0,\infty}(x) \tag{1}$$

where $I_A(x) = 1$ for $x \in A$ and $I_A(x) = 0$ for $x \notin A$ and $\Gamma(a) = \int_{z=0}^{\infty} z^{a-1} e^{-z} dz$ is the gamma function for $a > 0$. Furthermore $v(t)$ is a non-decreasing, right continuous, real valued function for $t \geq 0$ with $v(0) = 0$. The gamma process with shape function $v(t) > 0$ and scale parameter $u > 0$ is a continuous time stochastic process $X(t), t \geq 0$ with the following properties:

- $X(0) = 0$ with probability one;
- $X(\tau) - X(t) \sim \mathrm{Ga}(v(\tau) - v(t), u), t \in [0, \tau)$;
- $X(t)$ has independent increments.

$X(t)$ represents the deterioration at time t, $t \geq 0$. Its probability density function is given by

$$f_{X(t)}(x) = \mathrm{Ga}(x|v(t), u) \tag{2}$$

in which

$$E(X(t)) = \frac{v(t)}{u}, \tag{3}$$

$$\mathrm{Var}(X(t)) = \frac{v(t)}{u^2} \tag{4}$$

are the expectation and the variance, respectively.

Assuming H_0 as the initial value of the damage accumulation index, failure occurs when the damage accumulation index $H(t) = H_0 - X(t)$ reaches the critical level H_{crit}. The time at which failure occurs is especially called T_{crit}, which is also referred to as the first hitting time of the critical level H_{crit}. Then the cumulative distribution function (CDF) of time to failure can be written as

$$F(t) = \Pr\{T_{crit} \leq t\} = \Pr\{X(t) \geq H_{crit}\} = \int_{x=H_{crit}}^{\infty} f_{X(t)}(x)dx = \frac{\Gamma(v(t), uH_{crit})}{\Gamma(v(t))}, \tag{5}$$

where $\Gamma(a,x) = \int_{z=x}^{\infty} z^{a-1} e^{-z} dz$ is the incomplete gamma function for $x \geq 0$ and $a \geq 0$.

2.2 *Parameter estimation for the gamma process*

Let us assume to model the temporal variability in the deterioration with a gamma process. Empirical studies show that the expected deterioration at time *t* is often proportional to a power law

$$E(X(t)) = \frac{v(t)}{u} = \frac{ct^b}{u} \tag{6}$$

in which $c > 0$ and $b > 0$ are constant. The gamma process is called stationary if the expected deterioration is linear in time, i.e., when $b = 1$, and non-stationary when $b \neq 1$.

Let us consider a typical data set of inspection times t_i, $i = 1, \ldots, n$, where $t_0 < t_1 < \ldots < t_n$, and corresponding observations of the cumulative amounts of deterioration x_i, $i = 1, \ldots, n$, where $0 = x_0 \leq x_1 \leq \ldots \leq x_n$. The parameter b can be determined based on engineering experience or estimated from data, according to a least square method (Hu et al., 2022)

$$b = \frac{n \sum_{i=1}^{n} (\log t_i)(\log x_i) - \left(\sum_{i=1}^{n} \log t_i\right)\left(\sum_{i=1}^{n} \log x_i\right)}{n \sum_{i=1}^{n} (\log t_i)^2 - \left(\sum_{i=1}^{n} \log t_i\right)^2}. \tag{7}$$

The parameters c and u can be estimated according to different approaches such as method of maximum likelihood, method of moments, method of Bayesian statics. According to the method of moments (Hu et al., 2022)

$$c = \frac{x_n^2\left[1 - \sum_{i=1}^{n}\left(\frac{\Delta t_i}{t_n}\right)^2\right]}{t_n S_Y^2}, \tag{8}$$

$$u = \frac{x_n\left[1 - \sum_{i=1}^{n}\left(\frac{\Delta t_i}{t_n}\right)^2\right]}{S_Y^2} \tag{9}$$

in which S_Y^2 is given by

$$S_Y^2 = \sum_{i=1}^{n}(\Delta x_i - \bar{Y}\Delta t_i)^2, \tag{10}$$

and

$$\bar{Y} = \frac{\sum_{i=1}^{n}\Delta x_i}{\sum_{i=1}^{n}\Delta t_i} = \frac{x_n}{t_n}. \tag{11}$$

2.3 *The k-means algorithm*

The k-means is one of the simplest and most efficient as well as most widely used partitional clustering algorithms. The algorithm starts by choosing K representative points as the initial centroids. Each point is then assigned to the closest centroid based on a proximity measure, usually the Euclidean distance metric. Once the clusters are formed, the centroids for each cluster are updated. The algorithm then iteratively repeats these two steps until a convergence criterion is met and the centroids do not change anymore. In particular, the objective function employed by k-means is the Sum of Squared Errors (SSE).

Given a set of m observation $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m$ in which each observation is a d-dimensional real vector, k-means clustering aims to partition the m observations into $k \leq m$ sets $\mathbf{D} = D_1, D_2, \cdots, D_k$ so as to minimize the SSE (Aggarwal and Reddy, 2014)

$$\underset{\mathbf{D}}{\operatorname{argmin}} = \sum_{k=1}^{K}\sum_{\mathbf{z}_j \in D_k} ||\mathbf{z}_j - \mathbf{d}_k||^2, \tag{12}$$

$$\mathbf{d}_k = \frac{\sum_{\mathbf{z}_j \in D_k}\mathbf{z}_j}{|D_k|}, \tag{13}$$

in which $\mathbf{d}_k$ is the centroid of cluster D_k and $|D_k|$ is the size of D_k.

The major factors that can impact the performance of the k-means algorithm are the following: 1) Choice of the initial centroids; 2) Estimation of the number of clusters K. Several methods are proposed in the literature to tackle these factors. In this work, the k-means++ algorithm is applied to select the initial centroids. According to this technique, the first cluster center is chosen uniformly at random from the data set. The next centroid is chosen randomly from the remaining data points with probability proportional to its distance from the point's closest existing cluster center. The problem of estimating the correct number of clusters is addressed by calculating the silhouette coefficient. This performance measure is based on the calculation of the the intra- and inter-cluster distances. For a given point $\mathbf{z}_j$, first the average of the distances to all points in the same cluster is calculated. This value is set to e_j. Then for each cluster that does not contain $\mathbf{z}_j$, the average distance of z_j to all the data points in each cluster is computed. This value is set to f_j. Using these two values the silhouette coefficient of a point is estimated. The average of all the silhouettes in the dataset is called the average silhouette width

$$P = \frac{\sum_{j=1}^{m}\frac{f_j - e_j}{\max(f_j, e_j)}}{m}, \tag{14}$$

and the larger its value, the higher the quality of clustering.

3 CASE STUDY

3.1 *KUBA-DB*

The Federal Roads Office (ASTRA) is the Swiss authority for road infrastructure management. In the context of the maintenance of ASTRA's engineering structures, inspection is of primary importance in order to detect damage at an early stage and to assess the current condition of the structure or the individual parts of the structure. Three different types of inspection can be carried out: primary, intermediate and special. The findings of primary and intermediate inspections are damage processes and information collected during the inspection is collected in the database KUBA-DB. A set of homogeneous damage processes within the same segment of the structural member and having the same effects on the functionality of the component forms a damage group. Each damage group is assigned to a condition class, which describes the condition of a relevant area of the structural element. Five condition classes have been defined, from "good condition" class to "alarming condition" class, with which a condition index from 1 to 5 is associated. Then, an assessment of the condition of the whole structural component is made, based on the damage groups and their effect on the safety and on the functionality of the component. This information is than aggregated at higher level to determine the condition of the whole object.

3.2 *Description of the procedure*

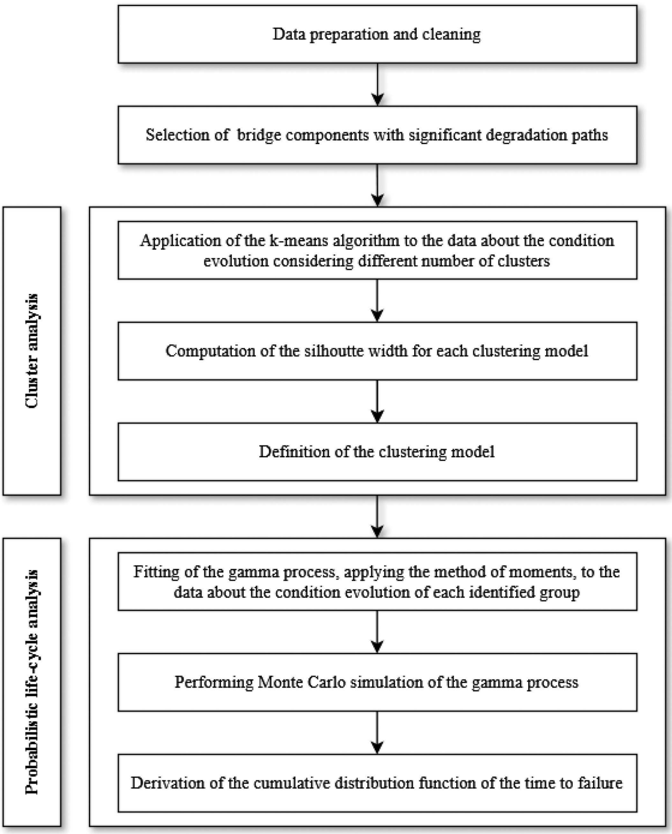
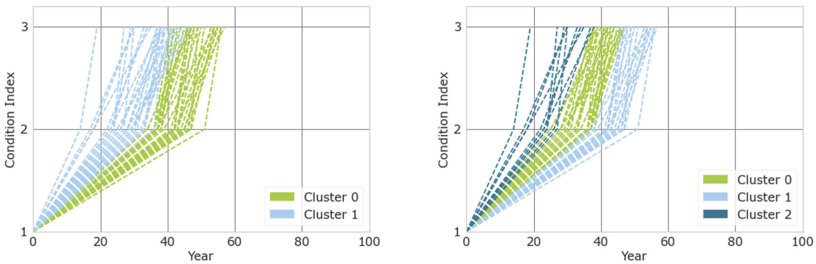


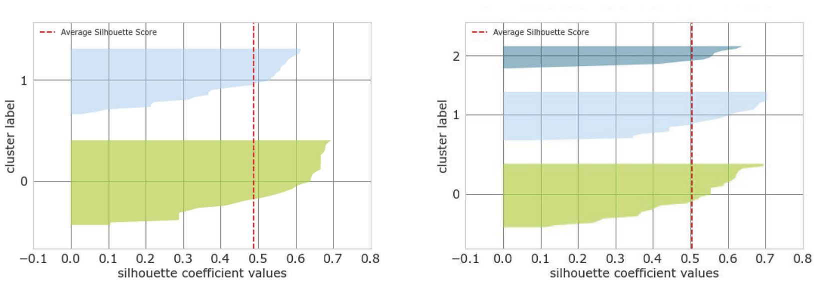
Figure 1. Flow chart summarizing the applied procedure.

Figure 1 summarizes the applied procedure for the analysis of the data. The first step is to prepare the data for analysis. In this step, data affected by inconsistencies due to human errors in their recording are removed, such as typing error. The data are cleaned and transformed into an appropriate format to perform the analysis. In a second step, the data for each structural component are selected, focusing on those structural components that show significant deterioration in their condition, i.e., those for which the condition index has experienced at least an increase from 1 to 2 and from 2 to 3. Such components are able to provide relevant information on the development of their condition, which will allow the fitting of a gamma process. Then, cluster analysis is performed, considering a variable number of clusters and calculating for each cluster model the performance measure. In general, it is possible to identify between two and three families of components that show similar condition development. More families can be identified only if a larger data set is available. The performance measure, i.e., silhouette width, supports the choice of the optimal number of clusters and is supplemented with a visual inspection of the identified clusters. Once the cluster model has been defined, a gamma process is fitted to the data belonging to each identified cluster. The parameters of the gamma process are determined based on Equations 7, 8, 9. Their definition allows a Monte Carlo simulation of the gamma process and the derivation of the mean time to failure as well as its CDF.



(a) K = 2 (b) K = 3

Figure 2. Results of k-means clustering for two and three centers.



(a) K = 2 (b) K = 3

Figure 3. Silhouette plot of k-means clustering for 64 samples and two and three centers.

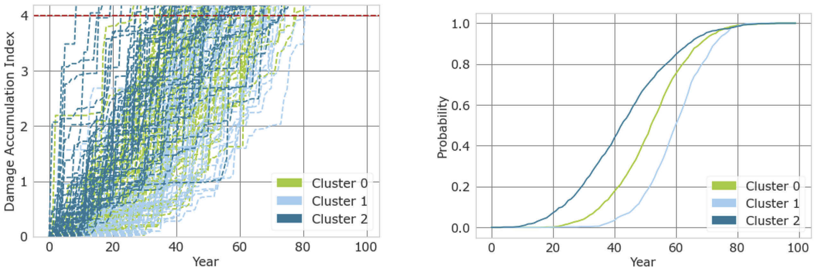
3.3 *Results*

For illustrative purposes, the results for a single component, namely the deck slab of reinforced concrete bridges, are reported. In a total of $m = 64$ cases, the condition of the deck slab deteriorate significantly over time, and the condition index increases from 1 up to 3.

The cluster analysis with the k-means algorithm is performed considering the evolution of the condition index over time of the selected elements. Figure 2 shows the clusters identified in the case where $K = 2$ and $K = 3$. For each cluster model the average silhouette width is calculated (Figure 3). Results reveal that the silhouette width of the model characterized by three clusters ($P(K = 3) = 0.505$) is higher than the silhouette width of the model characterized by two clusters (P(K = 2) = 0.488). For this reason, the optimal number of clusters is three, and this model is assumed and considered in the subsequent analysis. The cluster analysis identifies three families of bridge deck slab that show markedly different condition evolution, which can be referred to as slow (cluster 1), normal (cluster 0) and fast (cluster 2). The next step is to fit the gamma process to the data characterizing each cluster, that is, to statistically estimate the parameters b, c, and u for each family of bridge deck slab. The results of the inspection, namely the condition index, must be first adjusted to obtain an index representing the accumulation of deterioration over time. Hence, the condition index is transformed into a damage accumulation index through the following equation

Table 1. Estimated parameters of the gamma process and average time to failure resulting from its simulation for each cluster.

	b	u	c	Average time to failure (year)
Cluster 0	1.7027	1.3333	0.0069	52.4
Cluster 1	2.3245	1.3333	0.0004	60.2
Cluster 2	1.0930	1.3333	0.0935	43.6



(a) Simulation of the gamma processes (50 samples for each cluster)

(b) Cumulative distribution functions of time to failure (1000 samples for each cluster)

Figure 4. Probabilistic models for life-cycle assessment.

$$\text{Damage accumulation index} = \text{Condition index} - 1, \tag{15}$$

which is considered to fit the gamma process.

It is assumed that the critical damage accumulation level is 4, which corresponds to the condition index value 5, which is the critical value at which the component no longer fulfills a function (safety, serviceability, durability).

After that the parameters of the gamma process have been estimated (Table 1), a Monte Carlo simulation is performed (Figure 4). The Monte Carlo simulation allows the derivation of the average time to failure (Table 1) and the time to failure CDF for each identified cluster, which is shown in Figure 4. The CDF describes the trend of the failure probability as a function of time, which is directly related in statistical terms to the service life of the component. The average time to failure especially represents the expected service life of the component. Based on this result, it can be concluded that the group of components having a fast rate of deterioration has an expected service life of 43.6 years, the normal one of 52.4 years, and the slow one of 60.2 years.

4 CONCLUSIONS

The assessment of the remaining service life, understood as the service life before a degradation process reaches a critical level for which a maintenance intervention is deemed urgent, is an important step in improving the management of physical assets of infrastructure. Conscious of this, this article proposes an approach to assess the life-cycle of reinforced concrete bridge components by applying an unsupervised learning technique and a stochastic process. A cluster analysis based on the k-means algorithm is applied to identify families of bridge components that degrade with similar rates. A gamma process is fitted to the data concerning the evolution of the degradation that characterizes each family. By simulating the gamma process, the average expected service life of each family of components can be estimated. The simulation also allows the estimation of the time-dependent probability of damage exceeding a threshold level, called the probability of failure. An application of this approach to the case of reinforced concrete bridges in Switzerland, whose inspection results are contained in the KUBA-DB database, is developed. In particular, condition data related to the deck slab of reinforced concrete bridges is analyzed with the described procedure. Results reveal that three families of components whose condition evolves in a similar way can be identified. A performance measure, namely the silhouette width, supports the choice of the optimal number of clusters. Finally, it can be said that through this approach the service life of components can be predicted with reduced uncertainties.

ACKNOWLEDGMENT

This research study is funded by the FFG project "ENDURE - Estimation of the remaining service life of bridges through the development and testing of hybrid models", whose aim is to improve bridge life-cycle assessment analysing the infrastructure management databases of the D-A-CH countries.

REFERENCES

Aggarwal, C. C. and Reddy, C. K. (2014), *Data Clustering, Algorithms and Applications*, Taylor Francis Group.

Amin, R., Kocak, S., Sevil, H., Peterson, G., Hamilton, J., Rivera, B. and Barraco, T. (2020), 'A spatio-temporal cluster analysis of structurally deficient bridges in the contiguous usa', *Developments in the Built Environment* **4**, 100034.

Chang, K. and Chi, S. (2019), 'Bridge clustering for systematic recognition of damage patterns on bridge elements', *Journal of Computing in Civil Engineering* **33**(5), 04019028.

Edirisinghe, R., Setunge, S. and Zhang, G. (2013), 'Application of gamma process for building deterioration prediction', *Journal of Performance of Constructed Facilities* **27**(6), 763–773.

Frangopol, D. M., Kallen, M.-J. and Noortwijk, J. M. V. (2004), 'Probabilistic models for life-cycle performance of deteriorating structures: review and future directions', *Progress in Structural Engineering and Materials* **6**(4), 197–212.

Ghoroghi, A., Rezgui, Y., Petri, I. and Beach, T. (2022), 'Advances in application of machine learning to life cycle assessment: a literature review', *The International Journal of Life Cycle Assessment* **27**(3), 433–456.

Guida, M. and Penta, F. (2015), 'A gamma process model for the analysis of fatigue crack growth data', *Engineering Fracture Mechanics* **142**, 21–49.

Hu, C., Fan, H. and Wang, Z. (2022), *Gamma Process-Based Degradation Modeling and Residual Life Prediction*, Springer Singapore, Singapore, pp. 77–97.

Marsili, F., Bödefeld, J., Croce, P. and Landi, F. (2018), Bayesian approaches to lifetime pre- diction, *in* 'Safety and reliability - safe societies in a changing world, proceedings of the 28th International European Safety and Reliability Conference (ESREL 2018)', Trondheim, Norway.

Moscoso, Y. F. M., Rincón, L. F., Leiva-Maldonado, S. L. and e Matos, J. A. S. C. C. (2022), 'Bridge deterioration models for different superstructure types using markov chains and two-step cluster analysis', *Structure and Infrastructure Engineering* **0**(0), 1–11

Nicolai, R. P., Dekker, R. and van Noortwijk, J. M. (2007), 'A comparison of models for measurable deterioration: An application to coatings on steel structures', *Reliability Engineering System Safety* **92** (12), 1635–1650. Special Issue on ESREL 2005.

van Noortwijk, J. (2009), 'A survey of the application of gamma processes in maintenance', *Reliability Engineering System Safety* **94**(1), 2–21. Maintenance Modeling and Application.

Zhang, Z., Ren, X., Niu, Q., Zhang, Y. and Zhao, B. (2023), 'Durability degradation simulation of rc structure based on gamma process considering two-dimensional chloride diffusion and life probabilistic prediction', *Structures* **48**, 159–171.