Degradation-based preventive maintenance policy for railway transport systems

Fateme Dinmohammadi1*, B. Alkali1, M. Shafiee2
1Department of Mechanical Engineering, School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, United Kingdom, Fateme.Dinmohammadi@gcu.ac.uk
2Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

ABSTRACT

Railway transport maintenance plays an important role in delivering safe, reliable and competitive transport services. It also is one of the major costs for rail transport operations. According to several reports, the inspection and maintenance costs constitute a large portion of the life-cycle cost (LCC) for railway asset infrastructures (such as bridges, rail tracks, track beds and track equipment) and rolling stock components (e.g. chassis, bogies, wheels and wagons). In order to reduce the operating expenditure (OPEX) while maintaining high standards of safety, the asset managers must determine a planning period and find optimum preventive inspection policies for various railway systems, such that the total cost incurred over the life span is minimized and/or the rail network’s reliability is maximized. Common railway defects are caused by degradation processes such as rolling contact fatigue (RCF) or wear. The degradation of assets may result in substantial losses to the rail transport operators if it is not prevented in an efficient way. In this paper, we investigate an optimal age-dependent preventive inspection policy for railway assets subject to gradual degradation phenomena. The degradation processes initiate following Non-Homogenous Poisson Process (NHPP) and propagate according to gamma stochastic process. When the size of degradation reaches a critical level, the asset will unexpectedly fail and it has to undergo a corrective repair. This unexpected failure may also interrupt rail operations, cause passenger dissatisfaction or even some accidents like derailment or overturning. To avoid such undesired defects, the asset is preventively inspected at regular time intervals. The problem is to determine an optimal inspection time interval such that the long-run expected cost rate is minimized. The proposed model is applied to support maintenance decision-making for a railway asset on the Scottish rail network. The results show that the use of the proposed inspection policy allows a significant reduction of the maintenance cost compared to the strategy when only corrective repair is considered.

Keywords: Maintenance engineering technologies, Risk assessment, Modelling analysis and optimisation

Corresponding author: Fateme Dinmohammadi (email: Fateme.Dinmohammadi@gcu.ac.uk)

1. INTRODUCTION

The railway transport sector is a key enabler of Britain’s economy. The country’s rail network with a total length of approximately 16,209 kilometers track is the 18th largest network in the world [1]. The number of rail passengers as well as freight volumes have increased significantly in recent years. According to recent statistics published by the Office of Rail and Road (ORR), a total of 1.654 billion journeys were made in 2014–5, making the UK’s railway network the fifth most used in the world [2]. With the growing demand for rail services, the investment on railway has significantly increased during the last decade. Nevertheless, high operation and maintenance (O&M) costs act as a barrier to achieving a favourable financial performance of railway operations. According to several reports, the inspection and maintenance costs constitute a large portion of the life-cycle cost (LCC) for railway asset infrastructures (such as bridges, rail tracks, track beds and track equipment) and rolling stock components (e.g. chassis, bogies, wheels and wagons) [3, 4].

Nowadays, rail transport operators are under increasing pressure to reduce their O&M costs whilst maintaining reliability targets. Railway transport maintenance plays an important role in delivering safe, reliable and competitive transport services as it reduces the potential risk of defects and derailments. Generally, railway defects occur due to a number of specific causes that have been classified by many researchers. Olofsson and Nilsson [5] divided the defects of tracks into two types of surface-initiated and subsurface-initiated cracks. Cannon et al. [6]...
classified the rail track defects into three main groups: (i) defects originating from rail manufacture, (ii) defects originating from damage caused by inappropriate handling, installation and use, and (iii) defects caused by the exhaustion of the rail steel’s inherent resistance to fatigue damage. Dinmohammadi et al. [7] classified the modes of rolling stock defects into six groups, namely electrical faults, structural damages, functional failures, degradation, human errors, and natural (external) hazards. Also, some other classifications have been addressed in reference [8].

The majority of defects in the railway assets are caused by degradation processes such as rolling contact fatigue (RCF), wear, corrosion, erosion, etc. [9]. These degradation processes are very complex as they depend on various factors such as age, traffic density, axle load, asset material, track geometry, curvature, speed, and accumulated Million Gross Tones (MGT) [10]. Any of these forms or their combination can become a cause of a failure.

The degradation of assets may result in substantial losses to the rail transport operators if it is not prevented in an efficient way. Moreover, it may cause accidents [11], traffic disruption and ultimately passenger dissatisfaction. In order to control the rate of rail degradation, age-dependent preventive inspection policies have been extensively used by asset managers [12]. Under this policy, the railway asset is preventively inspected at fixed time intervals kT (k = 1, 2, …) after its installation. The asset is regarded as failed when the level of its degradation reaches an unacceptable size. In the event of asset failure between two consecutive preventive inspections, a corrective repair has to be undertaken. We assume that the costs for a preventive inspection and a corrective repair tasks are respectively C0 and C1, where C1 > C0 > 0. The main problem encountered in this policy is to determine the optimal inspection time interval T such that, under given physical/technical constraints, the railroad availability is maximized and/or O&M costs are minimized.

In this paper, we formulate an age-dependent preventive inspection policy for railway assets that are subjected to progressive degradation phenomenon. Degradation processes initiate following Non-Homogenous Poisson Process (NHPP) and propagate according to gamma stochastic process. If the size of degradation reaches a critical level, the asset will unexpectedly fail and it has to undergo a corrective repair. Otherwise, it will be preventively repaired at each inspection epoch. The explicit expression of the long-run expected cost function per unit time for the preventive inspection policy is derived and under certain conditions, the existence and uniqueness of the optimal solution are shown for the infinite-horizon case. The performance of the proposed policy in terms of cost is evaluated and compared to the case when only corrective repair is considered.

The rest of this paper is organized as follows. The assumptions and notation of the model are given in Section 2. In Section 3, we present the problem definition. The formulation of the optimization model and the properties of the optimal solution are discussed in Section 4. In Section 5, the model is applied to a real-life case study. Section 6 concludes this study.

2. ASSUMPTIONS AND NOTATION

In this section, we present the assumptions and notation used in our model formulation.

- The system starts functioning at time zero.
- The decision to repair the railway asset is made based on either its degradation level or its operational age.
- The degradation of the railway asset appears in the form of cracks, wear, corrosion, etc.
- A degradation process initiates following non-homogenous Poisson process (NHPP) and it accumulates gradually.
- The railway asset fails when its degradation size reaches a given threshold D. All defects are assumed to be instantly detected.
- The repair and inspection time is negligible. However, the planned preventive inspection is preferred to an unplanned corrective repair, because the potential passengers can be notified in advance and traffic disruption can be limited.
- The defect threshold D has a pre-specified value, which is determined by the original equipment manufacturer (OEM). The variable T is a decision variable and should be optimized.
\[ \{N(t) : t \geq 0\} \] number of degradation processes initiating in the interval \([0, t)\)

\[ m(t) [M(t)] \] intensity [mean value] function of \(N(t)\)

\[ P_j(t) \] the probability that exactly \(j\) degradation processes initiate in the interval \([0, t)\)

\[ T_j \] initiation time of the \(j\)th degradation process

\[ F_j(t) \] cumulative distribution function of \(T_j\)

\[ X_j(t) \] length/size/depth of the \(j\)th degradation process \(t\) units of time after its initiation

\[ D \] defect threshold (i.e. critical length/size/depth of degradation)

\[ U_j \] length of the interval between the initiation time of the \(j\)th degradation process to the time that it attains the critical threshold \(D\)

\[ g_{U_j}(\cdot) [G_{U_j}(\cdot)] \] probability density [cumulative distribution] function of \(U_j\)

\[ \Gamma(.) \{\gamma(-, -)\} \] gamma [incomplete gamma] function

\[ S_j \] time point that the length/size/depth of the \(j\)th degradation exceeds the critical threshold \(D\)

\[ S_{ij} \] time point that, for the first time, the length/size/depth of a degradation process exceeds the critical threshold \(D\)

\[ F_{S_{ij}(.)}(\cdot) [F_{S_{ij}(\cdot)}] \] cumulative distribution [survival] function of \(S_{ij}\)

\[ h(.) \] failure rate function of \(S_{ij}\)

\[ a(x) \bullet b(x) \] convolution of two functions \(a(x)\) and \(b(x)\)

\[ T \] preventive inspection interval

\[ C_0 \] fixed cost of a preventive inspection

\[ C_1 \] fixed cost of a corrective repair

\[ \eta \] additional cost to the rail infrastructure owner resulting from a corrective replacement

\[ E[X_j] \] expected length of a maintenance cycle

\[ D(t) \] \(s\)-expected cost of the operating system over \([0, t)\)

\[ CR(T) \] long-run expected cost per unit time

3. PROBLEM DEFINITION

Railway assets are subject to various types of degradation modes, such as rolling contact fatigue (RCF), wear, corrosion, erosion, etc. The degradation process for these modes involves three following phases: (i) initiation, (ii) propagation (or growth), and (iii) the failure.

i. Suppose that the degradation processes initiate in the interval \([0, t)\) following a non-homogeneous Poisson process (NHPP), \(\{N(t) : t \geq 0\}\) with intensity function \(m(t)\) and mean value function \(M(t)\), i.e.,

\[ M(t) = \int_0^t m(x) \, dx, \quad t \geq 0. \quad (1) \]

where \(t\) is the age of the railway asset and \(M(t)\) is a non-decreasing function of \(t\) with \(M(0) = 0\). Then, the probability that exactly \(j (= 0, 1, 2, \ldots)\) degradation processes occur in the interval \([0, t)\), \(P_j(t)\) is given by

\[ P_j(t) = P\{N(t) = j\} = e^{-M(t)} \times \frac{(M(t))^j}{j!}. \quad (2) \]

Let \(T_j\) \((j = 0, 1, 2, \ldots)\) denote the initiation time of the \(j\)th degradation process in the railway asset, where \(T_0 = 0\). We assume that the asset degradation is detected by health monitoring techniques just when they arrive (for more see [13]). Then, the cumulative distribution function of the random variable \(T_j\) is given by

\[ F_j(t) = 1; F_j(t) = P\{T_j \leq t\} = \sum_{i=j}^{\infty} p_i(t), \quad j = 1, 2, \ldots. \quad (3) \]

ii. Propagation is the second phase of the degradation process which may be accelerated by adverse environmental conditions. Many models have been developed to study how various degradation processes in different railway assets propagate. For instance, Ringsberg [14] proposed a crack growth model for railway tracks in which the crack propagation life is divided into three stages: (i) shear stress driven initiation at the surface; (ii) transient crack growth behavior; and (iii) subsequent tensile and/or shear driven crack growth (see Figure 1).
In this paper, the degradation propagation is modeled using a stochastic gamma process, which represents the degradation length/size/depth evolution in time. The gamma process is a stochastic process with independent non-negative increments having a gamma distribution with identical scale parameter. The gamma process has been widely studied for different maintenance applications by several authors (see [15] for a thorough review on the use of gamma process in maintenance modeling). Also, it has been observed that the gamma process is satisfactorily fitted to data of different gradual degradation phenomena (such as wear and crack propagation) in railway industry [16]. Moreover, the existence of an explicit probability distribution function of gamma process permits feasible mathematical developments.

Let \( X_j(t) \) be the length/size/depth of the \( j \)th degradation process \( t \) units of time after its initiation. We assume that \( X_j(t) \) has a homogeneous gamma process with shape and scale parameters given by \( \alpha t \) and \( \beta \) respectively. Then, for \( t > 0 \), the density and the cumulative distribution function of the increment of the length/size/depth of the \( j \)th degradation process is given by [17]

\[
g_{\alpha t, \beta}(x) = \frac{\beta^{\alpha t} x^{\alpha t-1} e^{-\beta x}}{\Gamma(\alpha t)}, \quad x \geq 0; \quad \alpha, \beta > 0, \quad (4)
\]

and

\[
G_{\alpha t, \beta}(x) = \frac{\gamma(\alpha t, \beta x)}{\Gamma(\alpha t)}, \quad x \geq 0; \quad \alpha, \beta > 0, \quad (5)
\]

where \( \Gamma(\cdot) \) denotes the gamma function, i.e.,

\[
\Gamma(\nu) = \int_0^\infty z^{\nu-1} e^{-z} dz; \quad \gamma(\nu, u) = \int_u^\infty z^{\nu-1} e^{-z} dz, \quad \nu, u > 0.
\]

iii. Railway asset fails when the length/size/depth of a degradation process reaches a given threshold \( D \) (see Figure 2). In the event of asset failure, a corrective repair is performed and the system returns to an "as-good-as-new condition.

Let \( U_j \) be the length of the interval between the initiation time of the \( j \)th degradation process to the time that it attains the critical threshold \( D \), i.e.,

\[
U_j = \inf \{ t \geq 0 : X_j(t) \geq D \}, \quad j = 1, 2, ..., \quad (6)
\]

![Figure 1. Cr...](image)

![Figure 2. Rail asset fails at degradation level D](image)
\[ G_U(t) \equiv G_U(t) = \frac{\Gamma(\alpha t, \beta D)}{\Gamma(\alpha t)}, \quad t \geq 0; \quad \alpha, \beta > 0. \quad (8) \]

We denote by \( S_j \) the time point that the length/size/depth of the \( j \)th degradation process exceeds the critical threshold \( D \). Then, \( S_j = T_j + U_j, \quad j = 1, 2, \ldots \). \quad (9)

**Lemma.** Let \( I_\alpha(\cdot) \) denote the indicator function that is defined as \( I_\alpha(x) = 1 \) for \( x \in \mathbb{A} \), and 0 otherwise. Let \( \{ N_S(t) : t \geq 0 \} \) be the counting process associated with the random variables \( S_j \) \( (j = 1, 2, \ldots) \), that is,
\[
N_S(t) = \sum_{j=1}^{\infty} I_{[0,t]}(S_j), \quad (10)
\]

Then, having in mind that the convolution of any functions \( a(.) \) and \( b(.) \) is given by
\[
a(x) \ast b(x) = \int_0^x a(x-t) \, db(t),
\]
\[
\{N_S(t) : t \geq 0\}
\]

is an NHPP with intensity function,
\[
h(t) = m(t) \ast g_U(t), \quad (11)
\]

where \( g_U(t) \) is given by Eq. (7) \cite{18}.

### 4. MORE FORMULATION AND ANALYSIS

The railway Asset is preventively inspected and repaired when its operational age attains a value of \( T (> 0) \). The cost of a preventive repair is \( C_0 \), whereas the cost of a corrective repair is \( C_1 \). Let \( \eta \geq 0 \) represent the cost parameter referring to an additional cost resulting from an unexpected failure, i.e., \( \eta = C_1 - C_0 \).

Let \( X_j \) denote a maintenance cycle defined by the time interval between maintenance actions (either corrective or preventive). Under the assumptions of the model, we have
\[
X_j = \min \{ S_{[1]}, T_j \}, \quad (12)
\]

where \( S_{[1]} \) denotes the time that, for the first time, a degradation process exceeds the critical threshold \( D \), i.e.,
\[
S_{[1]} = \min \{ S_j, \quad j = 1, 2, \ldots \}, \quad (13)
\]

Then, by using lemma, the survival function of \( S_{[1]} \) is given by
\[
\bar{F}_{S_{[1]}(t)} = P\{ S_{[1]} > t \} = P\{ N_S(t) = 0 \}
\]
\[
= \exp \left\{ - \int_0^t h(x) \, dx \right\}, \quad (14)
\]

where \( h(.) \) is the failure rate function of \( S_{[1]} \), and is given by Eq. (11). Then, the expected length of a maintenance cycle \( E[X_j] \), is given by
\[
E[X_j] = \int_0^T \bar{F}_{S_{[1]}(t)} \, dt, \quad T > 0. \quad (15)
\]

Let \( D(t) \) be the \( s \)-expected cost of operating the system for the time interval \([0, t)\). From the renewal reward theorem (see \cite[p. 52]{19}), the expected cost rate, denoted by \( CR(t) \), is the expected operational cost incurred in a maintenance cycle divided by the expected cycle length, i.e.,
\[
CR(T) = \lim_{\alpha \to \infty} \frac{D(t)}{t} = \frac{(C_0 + \eta) \times F_{S_{[1]}(T)} + C_1 \times \bar{F}_{S_{[1]}(T)}}{\int_0^T \bar{F}_{S_{[1]}(t)} \, dt}, \quad (16)
\]

where \( F_{S_{[1]}(\cdot)} [\bar{F}_{S_{[1]}(\cdot)}] \) is the cumulative distribution [survival] function of \( S_{[1]} \). The problem is to find the optimal value of \( T^* \) that minimizes the objective function \( CR(T) \), given in Eq. (16). Therefore, the proposed optimization model can be formulated as follows:
\[
\text{minimise} \quad CR(T) = \frac{C_0 + \int_0^T \eta h(t) \, \bar{F}_{S_{[1]}(t)} \, dt}{\int_0^T \bar{F}_{S_{[1]}(t)} \, dt}, 0 < T \leq T_{\text{max}}, \quad (17)
\]
The constraint $T \leq T_{\text{max}}$ implies that for safety requirements or due to the presence of physical/technical constraints (such as technology obsolescence and design modifications), the inspection time interval should not exceed some finite upper limit. The following theorem solves this problem.

**Theorem.** If $h(T)$ is strictly increasing in $t$, and $\eta h(T_{\text{max}}) > CR(T_{\text{max}})$, there exists an unique and finite minimum $T^* \in (0,T_{\text{max}})$ that verifies the following equation:

$$
\bar{F}_{\delta_{\tau}}(T^*) + h(T^*) \times \int_0^{T^*} \bar{F}_{\delta_{\tau}}(t) \, dt = \frac{C_0}{\eta} + 1 \tag{18}
$$

whereas, if $h(T)$ is non-decreasing in $t$, and $\eta h(T_{\text{max}}) \leq CR(T_{\text{max}})$, then $T^* = T_{\text{max}}$ (implying maximum preventive replacement interval).

**Proof.** The single-variable optimization model in Eq. (17) is a special case of the framework studied by Aven [20, pp. 151–152]. The optimal $T^*$ can be obtained by differentiating $CR(T)$ with respect to $T$ and setting it equal to zero, if it is an interior point of the feasible region. If none of the solutions of Eq. (17) is within the feasible region, we need to investigate the behavior of $CR(T)$ over the feasible region. If $CR(T)$ is a decreasing function of $T$, then the optimal preventive inspection interval should be set to $T_{\text{max}}$.

**Remark.** Suppose that the degradation processes initiate following a homogeneous Poisson process with constant rate $m (>0)$.

Then, from Eq. (11), we have

$$
h(T) = mG_{\delta_{\tau}}(T) \tag{19}
$$

Also, suppose that there is neither physical nor technical constraints on the preventive inspection interval time, i.e., $T_{\text{max}} \to \infty$. Now, if

$$
\int_0^{T^*} \bar{F}_{\delta_{\tau}}(t) \, dt > \frac{C_0 + \eta}{m\eta}
$$

there exists a unique and finite minimum optimal age $T^*$ that minimizes the function $CR(T)$, whereas, if

$$
\int_0^{T^*} \bar{F}_{\delta_{\tau}}(t) \, dt \leq \frac{C_0 + \eta}{m\eta},
$$

the optimal maintenance policy will be repairing the failed railway asset at its critical degradation level $D$.

5. **A CASE STUDY**

In this section, we present an application of the proposed inspection policy to the Europe’s only heavy haul line, Iron Ore Line (Malmbanan). The Malmbanan is a 473km/294mi railway freight line in northern Sweden that runs from Luleå via Gällivare and Kiruna to Narvik in Norway (Figure 3).

![Figure 3. Iron Ore Line (Malmbanan)](image)

The data has been collected from the literature [10, 21-25]. We assume that the arrival of cracks to the rail track follows a Poisson process with rate $\hat{m} = 0.144$ per month (i.e., the mean-time-to-initiate a crack is around seven months). The length of the cracks follows a gamma process with parameters $\hat{\alpha} = 0.576$ and $\hat{\beta} = 1.50$ ($\alpha/\beta$ is 0.384mm per month). The railway track breaks when the length of the crack exceeds the rail web thickness, i.e., $D=16.5$mm.

Average length of the rail replacement after a break is $L = 8$ meters. The cost of 60E1 railway track (including neutralization) per meter is 2,250 SEK. Average labour cost per hour (including the track worker cost, track welder cost, and inspection personnel cost) is 625 SEK. The hourly rate of hiring the welding equipment or service vessels for maintenance,
replace or inspection of the railway track is 80 SEK. The mean time required to perform a maintenance (either corrective or preventive) is 4 hours. However, the corrective type may cause traffic disruption that incurs an additional cost \( \eta \) to the route operator. The physical lifetime of 60E1 railway track is considered to be equal to six years (72 months). We wrote a MATLAB program for the minimization of the expected cost rate, as given in Eq. (17). The pictorial representation of the expected cost rate as a function of the inspection interval \( T \) for three different values of \( \eta = 0, \eta = 5,000, \) and \( \eta = 10,000 \) SEK is shown in Figure 4.

![Figure 4. Expected cost rate for different values of \( \eta \).](image)

From Figure 4, it is found that the use of optimal age-dependent preventive inspection policy allows a significant reduction of the maintenance cost compared to the strategy when only corrective repair is considered (the corresponding cost is the asymptote of the path, when \( T \) tends to \( T_{\text{max}} \)). The percentage reduction of the maintenance cost achieved through applying the optimal age-dependent preventive inspection policy is obtained as

\[
r = \frac{CR(T_{\text{max}}) - CR(T^*)}{CR(T_{\text{max}})} \times 100.
\]

The optimal value of \( T^* \) and the corresponding expected cost rate, \( CR(T^*) \), the expected cost rate for corrective maintenance policy, \( CR(T_{\text{max}}) \), and the percentage reduction of the maintenance cost, \( r \) are presented in Table 1. It can be seen that as the cost parameter \( \eta \) increases, the optimal inspection interval \( T^* \) becomes shorter, however the expected cost rate, \( CR(T^*) \) increases. Also, when a large additional cost is likely to be incurred by the infrastructure owner in corrective maintenance case, applying the age-dependent preventive inspection policy will be more efficient than corrective repair and has a huge potential to reduce the maintenance cost. For instance, when the cost parameter \( \eta \) is 10,000, the age-dependent policy allows for approximately \( 10\% \) reduction of the maintenance cost compared to the corrective repair policy.

<table>
<thead>
<tr>
<th>Unit</th>
<th>( T^* ) month</th>
<th>( \eta = 0 )</th>
<th>( \eta = 5,000 )</th>
<th>( \eta = 10,000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CR(T^*) ) SEK/month</td>
<td>448.38</td>
<td>546.40</td>
<td>594.81</td>
<td></td>
</tr>
<tr>
<td>( CR(T_{\text{max}}) ) SEK/month</td>
<td>448.38</td>
<td>554.94</td>
<td>661.49</td>
<td></td>
</tr>
<tr>
<td>( r ) %</td>
<td>0</td>
<td>1.54</td>
<td>10.08</td>
<td></td>
</tr>
</tbody>
</table>

6. CONCLUSIÓN

In this paper, an optimal age-dependent preventive inspection policy is presented for railway assets subject to progressive degradation. Under this policy, the railway asset is either preventively repaired at fixed time intervals \( kT \) \((k = 1, 2, \ldots)\) or it undergoes a corrective repair action at an unacceptable degradation level \( D \). This study can be extended in many directions to make it more practical in maintenance management of railway industry. Some of the possible extensions are:

(a) In this paper, we assumed that the cost discount rate (the time value of money) is zero. More work is needed to investigate the optimal solution for a discounted case with a positive cost discount rate \( \alpha \); (b) Formulating and analyzing the model when the degradation can be only detected if their length/size/depth reaches a detection threshold \( c (>0) \); and finally,
(c) Providing a cost comparison of the proposed age-dependent inspection policy with other common strategies such as reliability-centred maintenance (RCM).

We have worked on some of these extensions and our findings will be reported in the near future.

REFERENCES