Condition-based Maintenance for Long-Life Assets with Exposure to Operational and Environmental Risks

Abstract

This paper presents a new condition-based maintenance (CBM) model for long-life assets to address the potential risk caused by the decline of the operating environment. Two types of maintenance are formulated in the CBM model. Minor maintenance can mitigate the operational and environmental risk, and major maintenance can eliminate the accumulated damage within the asset. A continuous-time semi-Markov chain (CTSMC) is used for modeling the aging of the asset as well as the stochastic decline of the operating environment. To optimize the CBM policy in a mathematically tractable manner, we introduce a hypo-exponential approximation approach to match the first four moments of the sojourn time distribution of CTSMC. This approach guarantees a minimum representation of the CTSMC with non-fictitious surrogated Markov chain. The model provides both good mathematical tractability and sufficient generalizability. The practical impact of this research is demonstrated by applying it to a real industrial case of concrete bridge maintenance. It is observed that this approach results in a CBM plan with a lower asset lifecycle cost compared to current techniques.

Keywords: Maintenance, Operational and Environmental Risk, Stochastic Modelling, Moment Matching, Hypo-exponential Distribution.

1. Introduction

The positive impact of optimised maintenance strategies has been recognized in the literature (Pinjala, Pintelon, & Vereecke, 2006). Amongst all maintenance strategies, condition-based maintenance (CBM) has emerged as an increasingly important topic for both practitioners and academics. Many advanced CBM models have been developed to reduce operation and maintenance cost (Liu, Wu, Xie, & Kuo, 2017) and (Keizer, Teunter, Veldman, & Babai, 2018). Very often, the risk of the declining operating environment is not explicitly formulated in these models. Development of new materials and reliability innovations in new designs mean that assets nowadays tend to have a long lifetime. When assets operate over a long period of time, the decline (i.e., adverse change) of the operating environment during their lifetime may no longer be considered negligible. Therefore, developing lifecycle models to determine appropriate CBM policies for such types of assets under a declining operating environment is particularly important. In this paper, the decline of the operating environment is defined as a gradually increasing deviation of the operational environment from what the asset was designed and built for, which negatively affects the aging of the asset. In practice, the phenomena may manifest in three ways:
(1) Change in the macro-scale environment: (Stewart, Wang, & Nguyen, 2011) demonstrate the impact of climate change on carbonation-induced corrosion of concrete infrastructure caused by the increase in atmospheric CO₂ level and temperature. (Schweikert, Chinowsky, Kwiatkowski, & Espinet, 2014) explain the need to consider climate change in infrastructure planning.

(2) Accelerated deterioration after extreme events: An investigation has shown that the deterioration of Queensland roads increased rapidly after flooding events (Sultana, Chai, Chowdhury, & Martin, 2016).

(3) Consistent increase of load: Due to population increase and urbanization, traffic load on roads is likely to increase over time. (Rasul, Burrow, & Ghataora, 2016) show that the cumulative traffic load deteriorates the subgrade soil and may result in the loss of mechanical properties of road pavements.

The above examples clarify the importance of the decline of operating environment on the management of long-life assets. The risk of declining environment could result in a long-term detrimental effect on the asset’s aging and significantly reduce its lifetime if left unattended. Fortunately, some maintenance activities, such as anti-corrosive printing and repair of the supporting system, can effectively mitigate the risk. Hence, the explicit consideration of the long-term risk arising from such declining operating environment and its corresponding maintenance activities is essential to estimate lifecycle costs and risks, leading to better CBM policies. In this paper, we aim to capture this effect using stochastic modelling and explore its impact on the condition-based maintenance strategy to secure a resilient future for long-life assets.

Semi-Markovian model is a powerful approach for modeling the stochastic aging process. It relaxes the Markov assumption that is hardly met in practice by extending the sojourn time distribution from exponential distribution to general distributions. The semi-markovian model has been widely applied in multiple areas. Dimitrakos and Kyriakidis (Dimitrakos & Kyriakidis, 2008) designed an optimal preventive maintenance strategy for a production system with buffer capacity by using Semi-Markov decision processes. Salari and Makis developed a CBM for a multi-unit system using Semi-Markov decision process and applied it on wind farms (Salari & Makis, 2017). Lian et al. modeled an inventory replenishment policy using a semi-Markov kernel (Lian, Liu, & Zhao, 2009). Using such models to calculate the expected lifetime or evaluate the reliability normally requires the adoption of numerical solutions for the Kolmogorov equations (Lefebvre & Perotto, 2011; Kharoufeh, Cox, & Oxley, 2013) and discrete-event simulation (Manno, Chiacchio, Compagno, D’Urso, & Trapani, 2014). However, these methods are complicated and demand a high computational cost, especially for long-life assets with multiple deteriorating scenarios under different declining operating environments.
In this paper, we aim to develop an alternative approach to calculate the long-term expected cost of CBM policies for long life assets modeled by continuous time semi-Markov chains (CTSMC). Compared to the standard formulation of the semi-Markov decision process (Puterman, 1994), CTSMC can model the actions – such as maintenance – as an individual state. This provides a relatively straightforward method to model the duration of maintenance and the associated downtime penalty costs. Our approach contains two stages. In the first stage, we apply a hypo-exponential distribution to approximate all non-exponentially distributed sojourn times. After the approximation, the semi-Markovian model is then converted to a surrogate Markov model, which is easier to solve. The hypo-exponential approximation is a generalization of approximating using Erlang distribution (Thummler, Buchholz, & Telek, 2006). Also, unlike using Coxian distribution (C.-W. Park & Lee, 2014) and acyclic phase-type distribution (Bobbio, Horváth, & Telek, 2005) that will result in fictitious surrogate Markov chain that can hardly be associated with the actual condition of the asset (Yeh, 1997), the sub-states after hypo-exponential approximation can be deemed as a linear decomposition of the sojourn time. It implies that the sub-state can be easily calculated using the knowledge of the sojourn time in a condition state. In the extant literature in asset lifecycle and reliability modelling (Kharoufeh, Solo, & Ulukus, 2010; Guo, 2014) and other domains (Wang & Wilson, 1994; Arikan, Deshpande, & Sohoni, 2013; Ko & Pender, 2017), majority of the algorithms focus on matching the first two or three moments of the distribution. Because the fourth moment is directly related to the tail behavior of the sojourn time of distribution (Westfall, 2014), it is desirable to match the fourth moment for lifecycle modeling that requires formulation of long-term deteriorating characteristics. In our approach, we aim to match the first four moments as well as minimize the number of sub-states to avoid unnecessary calculation in the second stage. In the second stage, we use a three-layered CBM model that considers both minor maintenance and major maintenance. Since the surrogate Markov model is a continuous-time Markov model with good mathematical tractability, the long-term expected cost of the CBM policies can be derived analytically.

The rest of the paper is structured as follows. In section 2, we provide a description of the problem and an overview of the maintenance model. The outline of the overall approach is described step-by-step. Section 3 explains the hypo-exponential approximating approach for matching the first four moments and explores the analytical boundaries between the third and fourth moments. A guidance for selecting the minimum number of sub-states in surrogate Markov chain is also presented. This is essential for reducing unnecessary states in the surrogate Markov chain and provides a mathematically effective way for optimizing the CBM policy. A three-layered CBM model is formulated using the surrogate Markov chain. The steady-state probabilities are derived analytically to obtain a mathematical tractable expression for the lifecycle cost. Section 4 applies the overall approach to a practical case study of concrete bridge maintenance. The benefit of matching the additional fourth
2. System description and overview of the maintenance model

The designed service life of assets such as bridges is normally more than 30 years. Also, its aging process can be considered as a multi-state stochastic process. During its long lifetime, the operating environment may also decline stochastically, and this can accelerate the aging of the bridge. For instance, due to accumulated chloride ions, the surrounding environment can trigger chloride-induced deterioration. This could significantly reduce the lifetime of the bridge and cause unexpected disruption of service if left unattended. Apart from a self-revealing failure state, the condition state and operating environmental risks can only be assessed through periodic inspection. Based on the inspection outcome, two types of maintenance actions can be implemented. The first is minor maintenance, of which the purpose is to increase the level of protection against environmental hazards.

An example of this type of maintenance is anti-corrosion coating. This mitigates the environmental risk and restores the aging rate of the bridge back to normal. The second action is major maintenance, which can eliminate the accumulated damage of the bridge and bring the bridge back to as good as new state. Compared to minor maintenance, major maintenance may require a longer maintenance duration and higher cost. In practice, this type of maintenance includes overhaul and preventive replacement. If the bridge reaches the failure state, it will be replaced. All the maintenance actions and replacement incur a non-negligible duration, and the bridge will be closed to traffic and other users during maintenance, incurring a downtime penalty cost. The objective of this CBM model is to identify the optimal condition thresholds for minor and major maintenance actions such that the time-averaged lifecycle cost can be minimized.

Our CBM model has a three-layered structure. We use a layer index $l$ to signify different layers: $l = 0$ indicates the deterioration layer, $l = 1$ represents inspection layer, and $l = 2$ represents the maintenance layer. In the deterioration layer, we formulate the aging of the long-life asset under environmental risks as CTSMC. The state of the long-life asset is expressed as $(i,j,l)$, where $i \in \{0,1,\ldots,K\}$ indicates the deteriorating state of the asset. State $i = 0$ indicates the ‘as good as new’ state and state $i = K$ represents the failure of the asset and is an absorbing state. Assets may subject to a declined operating environment that results in an increase in the environmental risk and subsequently changes the aging mechanism. Our model formulates such type of environmental risk using an environmental index $j$. For the convenience of expression, we categorize the environmental index into three different levels denoted by $j \in \{0,1,2\}$ for normal-rated, risky, and severe environmental risk, respectively. Notice that this approach can cope with scenarios where $j$ has more than three levels. The sojourn time of states in the deterioration layer follows a General distribution. As the operating environment declines, the aging process may accelerate and exhibit different
characteristics. We represent the sojourn time in the normal-rated operating environment as $T_{n_i0}$ and the sojourn time in a declined operating environment as $T_{dl,j}$, where $j \in \{1,2\}$. The decline rates of the operating environment are denoted as $\lambda_{f_i,j}$. The state transition diagram of the long-life semi-Markovian deteriorating asset is illustrated in Figure 1.

Without loss of generality, we can combine the states $(K, j, 0), j \in \{0,1,2\}$ as the failure state $(K, 0, 0)$. Therefore, the overall number of states for the deterioration layer is $3K + 1$, and the state space $S = \{(i, j, 0) \cup (K, 0, 0)\mid i \in \{0, \ldots, K - 1\}, j \in \{0,1,2\}\}$.

To design and optimize the CBM policy in a mathematically tractable manner, we apply a hypo-exponential approximation to convert the CTSMC to a surrogate Markov chain. Hence the time-averaged lifecycle cost of the asset under CBM policies can be assessed analytically. In our approach, we aim to match the first four moments of the sojourn time distribution. It is important to consider the fourth moment at the moment matching step since it represents the tail behaviour of the sojourn time distribution that is associated with the long-term performance of assets and may in turn influence their lifecycle costs.

The overall approach developed in this paper can be summarized as follows:

**Step 1**: Identify the scale of applicability based on the analytical bounds of skewness and kurtosis of the hypo-exponential distribution and find the minimal representation for matching the first four moments.

**Step 2**: Approximate all the non-exponentially distributed sojourn times using the surrogate Markov chain.

**Step 3**: Construct the CBM model.
Step 4: Express the time-average lifecycle cost in term of state probabilities and derive the analytical form of the referencing state.

Step 5: Calculate the time-average lifecycle cost and optimize the maintenance thresholds for both minor maintenance and major maintenance.

This approach will now be described in more detail in the following sections.

3. Approximating approach

For each state with a non-exponential sojourn time distribution, the first four moments will be approximated by $n$ serially connected sub-states with different exponential distributions. After the conversion, the resulting continuous-time Markov chain is referred to as the surrogate Markov chain. We denote the surrogate Markov chain as $\{Y_t, t \geq 0\}$. Each condition is then decomposed into sub-states and renumbered from 0 to $k$. Thus, the state space of the surrogate Markov chain is $\tilde{S} = \{(i,j,0) \cup (k,0,0)| i \in \{0,1,...,k-1\}, j \in \{0,1,2\}\}$. It is worthwhile to mention that the number of sub-states in surrogate Markov $||\tilde{S}||$ is directly related to the number of equations to be solved for assessing CBM policies. Hence, it is important to keep $||\tilde{S}||$ to a minimum (a minimum representation) while matching the first four moments of the general sojourn time distribution.

3.1 Hypo-exponential approximation

To match high-order moments, we first calculate the moment generating function of a hypo-exponential distribution. Using the moment generating function, a given moment of the hypo-exponential distribution can be calculated by differentiation instead of integration. We denote a hypo-exponential distribution random variable as $Y$, which can be expressed as the sum of independent exponential random variables with different rates.

$$Y = X_1 + X_2 + \ldots + X_n$$

where $X_i$ is an exponential random variable and $n$ is a finite positive integer.

If $Y = \sum_{i=1}^{n} X_i$, where $X_i$ are independent exponential variables with non-identical rates $\lambda_i \neq \lambda_j, \forall i \neq j$, the Laplace transform of the probability density of $Y$ can be expressed as:

$$\mathcal{L}\{f_Y(t)\} = \prod_{i=1}^{n} \mathcal{L}\{f_{X_i}(t)\} = \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i + s}$$

To obtain the analytical expression of $f_Y(t)$, we need to apply inverse Laplace transform to the above equation:

$$f_Y(t) = \mathcal{L}^{-1}\{\mathcal{L}\{f_Y(t)\}\} = \mathcal{L}^{-1}\left(\prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i + s}\right)$$
By applying the Heaviside cover-up method (George Jr, 1996) on \( \prod_{i=1}^{n} \frac{1}{\lambda_i + s} \), it follows

\[
\mathcal{L}^{-1}\left\{ \prod_{i=1}^{n} \frac{\lambda_i}{\lambda_i + s} \right\} = \mathcal{L}^{-1}\left\{ \sum_{i=1}^{n} \lambda_i \prod_{j=1, j \neq i}^{n} \frac{\lambda_j}{\lambda_j - \lambda_i} \frac{1}{\lambda_i + s} \right\}
\]

For the convenience of expression, we denote \( l_i(0) = \prod_{j=1, j \neq i}^{n} \frac{\lambda_j}{\lambda_j - \lambda_i} \) which is also known as Lagrange basis polynomials of \( \lambda_i \). According to the linear property of the inverse Laplace transform, we have

\[
f_Y(t) = \sum_{i=1}^{n} \lambda_i l_i(0) \mathcal{L}^{-1}\left\{ \frac{1}{\lambda_i + s} \right\} = \sum_{i=1}^{n} \lambda_i l_i(0) e^{-\lambda_i t} I_{(0, \infty)}(t)
\]

The moment generating function of random variable \( Y \) is defined as \( M_Y(t) := \mathbb{E}[e^{tY}] = \int_{0}^{\infty} e^{tx} f_Y(x) dx \). It is a function of auxiliary mathematical variable of \( t \) that can be applied to find all the moments of the distribution.

\[
M_Y(t) = \int_{-\infty}^{\infty} e^{tx} \sum_{i=1}^{n} \lambda_i l_i(0) e^{-\lambda_i x} I_{(0, \infty)}(x) dx = \sum_{i=1}^{n} \lambda_i l_i(0) \int_{0}^{\infty} e^{tx} e^{-\lambda_i x} dx = \sum_{i=1}^{n} \frac{\lambda_i l_i(0)}{\lambda_i - t}
\]

After obtaining the moment generating function, the analytical expression of a given moment \( u \) of hypo-exponential distribution can be calculated by differentiating the moment generating function \( u \) times with respect to \( t \) and set \( t = 0 \) as:

\[
M_Y^{(u)}(0) = \sum_{i=1}^{n} \lambda_i l_i(0) \frac{d^u(\lambda_i - t)^{-1}}{dt^u} \bigg|_{t=0} = \sum_{i=1}^{n} \lambda_i l_i(0) \frac{u!}{(\lambda_i - t)^{u+1}} \bigg|_{t=0} = \sum_{i=1}^{n} u! \frac{l_i(0)}{\lambda_i^u}
\]

For hypo-exponential distributions, the first and second moments have been explored by (Magott & Skudlarski, 1993; Jang, Chung, Suh, & Rhee, 2006). We, therefore, focus on identifying the boundaries of the third and fourth moments. Skewness and kurtosis are standard ways to interpret the third and fourth moments (Johnson & Taaffe, 1989), as shown in equations (1) and (2) respectively.

\[
\gamma_Y = \frac{M_Y^{(3)}(0) - 3M_Y^{(1)}(0)M_Y^{(2)}(0) + 2M_Y^{(1)}(0)^3}{(M_Y^{(2)}(0) - M_Y^{(1)}(0)^2)^{3/2}} = \frac{2 \sum_{i=1}^{n} \lambda_i^{-3}}{(\sum_{i=1}^{n} \lambda_i^{-2})^{3/2}}
\]

\[
\kappa_Y = \frac{M_Y^{(4)}(0) - 4M_Y^{(1)}(0)M_Y^{(3)}(0) + 6M_Y^{(1)}(0)^2M_Y^{(2)}(0) - 3M_Y^{(1)}(0)^4}{(M_Y^{(2)}(0) - M_Y^{(1)}(0)^2)^2} = \frac{6 \sum_{i=1}^{n} \lambda_i^{-4}}{(\sum_{i=1}^{n} \lambda_i^{-2})^2} + 3
\]

where \( \gamma_Y \) and \( \kappa_Y \) expresses the skewness and the kurtosis of the hypo-exponential distribution respectively.

For a general distribution with known mean, variance, skewness, and kurtosis that are denoted as \( \mu_Z, \sigma_Z, \gamma_Z \) and \( \kappa_Z \) respectively, inspired by (Johnson & Taaffe, 1990) and (Öztürk & Dale, 1985), we
define the function $\mathcal{D}$ as the square distance between the first four standardized moments of the approximated hypo-exponential distribution and the underlying distribution.

$$
\mathcal{D} = \left( \sum_{i=1}^{n} \frac{\lambda_i^{-1}}{\sigma Z} \right)^2 + \left( \sum_{i=1}^{n} \frac{\lambda_i^{-2}}{\sigma Z} \right)^{1/2} + \left[ \frac{2 \sum_{i=1}^{n} \lambda_i^{-3}}{(\sum_{i=1}^{n} \lambda_i^{-2})^{3/2}} - \gamma_Z \right]^2 + \left[ \frac{6 \sum_{i=1}^{n} \lambda_i^{-4}}{(\sum_{i=1}^{n} \lambda_i^{-2})^2} + 3 - \kappa_Z \right]^2
$$

The objective of the hypo-exponential approximation is to minimize $\mathcal{D}$ by adjusting $\lambda_1, \ldots, \lambda_n$ while keeping $n$ as small as possible. If $n$ is known, $\arg_{\lambda_1, \ldots, \lambda_n} \{\min[\mathcal{D}]\}$ is a typical problem of minimization of nonlinear systems that can be solved by well-estimated approaches such as (Coleman & Li, 1996) and (Moré, 1978). Hence, the key step is to identify the smallest $n$. However, the dilemma is that if $n$ is too small, the hypo-exponential may no longer match the skewness and kurtosis, while if $n$ is too large, it may result in an unnecessary complication for calculating the lifecycle cost. As described by (Bobbio et al., 2005), one of the major difficulties of approximating with phase-type distribution lies in the unknown bounds of higher moments, which may result in unpredictable performance. To assist the approximation, we derive proposition 1 and 2 that express the interdependence between skewness and kurtosis and guide the selection of the minimum of $n$.

**Proposition 1**: For a hypo-exponential distribution, the lower bound of kurtosis is related to skewness as follows: $\kappa_Z \geq 3 + \gamma_Z^2$.

Proposition 1 is derived through proving the hypo-exponential distribution is infinite divisible, as shown in Appendix A.

**Proposition 2**: The minimum number of states $n$ to match the skewness $\gamma_Z$ and kurtosis $\kappa_Z$ of a given distribution using hypo-exponential distribution needs to satisfy the following necessary conditions.

$$
\begin{align*}
\left\{ \begin{array}{ll}
2 \leq \gamma_Z \leq \frac{2}{\sqrt{n-1}}, & n = 1 \\
\frac{6}{n+3} \leq \kappa_Z \leq \frac{6}{n-1} + 3, & n > 1, n \in \mathbb{N}
\end{array} \right.
\end{align*}
$$

The proof can be found in Appendix B, and the result can be visualized, as shown in Figure 2. Different pattern-fills are used to distinguish different regions to represent the third and fourth standardized moments with a minimum number of states. This result determines the feasibility of matching skewness and kurtosis by using hypo-exponential distribution, as well as recommends the minimum number of states in the surrogate Markov chain.
To demonstrate the performance of using the developed boundaries, an illustrative example is provided. In the example, we compare the performance of matching a Gamma distribution (benchmark) with the $n$ suggested by our derived boundary condition to the scenario with $n$ outside the suggested boundaries. The Gamma distribution has a shape parameter equals to 3.8, and a scale parameter equals to 1.316. The result is plotted in Figure 3.

Figure 3 demonstrates that with the same moment matching method, the one using the $n$ that suggested by our derived boundary has a better performance in matching the benchmark distribution.

We have so far demonstrated the feasibility and identified the boundaries for matching the third and fourth moments using a hypo-exponential approximation. By using this approximating approach, the CTSMC is transformed into a surrogate Markov chain whose states are capable of being interpreted as the sub-condition states of the asset. According to Proposition 2, the size of the surrogate Markov
chain can be minimized, thereby making it more practical for modeling the lifecycle of long-life assets by effectively avoiding unnecessary states.

3.2 Formulation and calculation of CBM policies

The results from section 3.1 allow us to transform a condition state into a serial connected sub-condition states with exponentially distributed sojourn time to match the first four moments of all non-exponentially distributed sojourn time in the CTSMC. In this section, we developed a three-layered CBM model on the resulting surrogate Markov chain. The model aims to find the optimal state to maintain the asset so that the time-averaged cost of the maintenance of the asset is minimized.

In general, the model contains three layers, namely deterioration layer \( l = 0 \), inspection layer \( l = 1 \), and maintenance layer \( l = 2 \). Both minor maintenance activities and major maintenance activities are considered within the model. The state transition diagram of the CBM model is illustrated in Figure 4.

![Figure 4: A three-layered CBM for assets under declining operating environment](image-url)
In Figure 4, the first deterioration layer is formulated by the surrogate Markov chain, which is interconnected with inspection and maintenance layers. In the inspection layer, we model the frequency of inspection as \( \lambda_{in} \) with a cost \( C_{in} \) and duration \( 1/\mu_{in} \). Based on the inspected condition and environmental index, we can decide the thresholds for minor maintenance and major maintenance. A minor maintenance activity is mainly used to increase the protection against environmental risk when the environmental index is ‘risky’ or ‘severe’. The minor maintenance has a duration \( 1/\mu_c \) and cost \( C_c \). A major maintenance activity restores the condition of an asset to as good as new state with a duration \( 1/\mu_M \) and a cost \( C_M \). If the asset deteriorates to the failure state \( (k,0,0) \), a replacement will be implemented with a duration \( 1/\mu_R \) and cost \( C_R \). We also consider a planned downtime cost as \( C_p \) caused by maintenance activities and an unplanned downtime cost as \( C_u \) caused by the replacement.

We classify the maintenance model into four phases. Phase I contains the states with a sub-condition index \( i < a \). In phase I, the asset under either the risky or severe environmental condition will be repaired by a minor activity. No major maintenance is required in this phase. Phase II contains the states with a condition index \( (a+1 \leq i \leq b) \). In phase II, the asset with a risky environmental index will be repaired by minor maintenance. However, if the asset is in a severely declined operating environment, major maintenance will be carried out. Phase III contains the states with a condition index \( (b \leq i < c) \). In phase III, major maintenance will be applied if the asset is operating in either a risky or severe environmental index. Phase IV contains the states with a sub-condition index \( (c+1 \leq i < k) \). In phase IV, all good, risky, and severe environmental indices will lead to major maintenance.

The objective of the maintenance model is to find the optimal values of phase thresholds as \( a, b, \) and \( c \) so that the lifecycle cost is minimized. The time-averaged cost of the maintenance of the asset can be expressed by equation (3):

\[
C = C_u \pi_{k,0,0} + C_p \left( \sum_{i=0}^{c} \sum_{j=1}^{2} \pi_{i,j,2} + \sum_{i=c+1}^{k-1} \sum_{j=0}^{2} \pi_{i,j,2} \right) + C_{in} \mu_{in} \left( \sum_{i=0}^{k-1} \sum_{j=0}^{2} \pi_{i,j,1} \right) + C_c \mu_c \left( \sum_{i=0}^{a} \sum_{j=1}^{2} \pi_{i,j,2} + \sum_{i=a+1}^{b} \pi_{i,1,2} \right) + C_M \mu_M \left( \sum_{i=a+1}^{b} \sum_{j=1}^{2} \pi_{i,j,2} + \sum_{i=b+1}^{k-1} \sum_{j=1}^{2} \pi_{i,j,2} \right) + C_R \mu_R \pi_{k,0,0} \]

To obtain the analytical expression for \( C \), it is critical to know the analytical expression of the required steady states probabilities, which requires to solve \( 9k - c + 1 \) equations, where \( k \) is the number of sub-conditions. Because the approximation is obtained by the hypo-exponential distribution, the configuration of resulting states follows a unique recursive pattern in each phase. We first express all the states as the ratio of referencing state \( \pi_{0,0,0} \). Because the sum of all steady-state
probabilities is 1, we can derive the analytical expression of the referencing state in a recursive form as Proposition 3.

**Proposition 3:** The referencing state can be expressed analytically with the recursive matrixes in phase $I$, $II$, $III$, and $IV$ as

\[
\pi_{0,0,0} = \left\{ \begin{array}{c} 1 + \frac{\lambda_{in}}{\mu_{in}} + 1 + \frac{\lambda_{in}}{\mu_{c}} + 1 + \frac{\lambda_{in}}{\mu_{a}} \sum_{i=0}^{\alpha} A^i \\
+ 1 + \frac{\lambda_{in}}{\mu_{in}} + 1 + \frac{\lambda_{in}}{\mu_{c}} + 1 + \frac{\lambda_{in}}{\mu_{a}} \sum_{i=0}^{b} A^a B^{b-i} \\
+ 1 + \frac{\lambda_{in}}{\mu_{in}} + 1 + \frac{\lambda_{in}}{\mu_{c}} + 1 + \frac{\lambda_{in}}{\mu_{a}} \sum_{i=0}^{c} A^a B^{b-a} C^{c-i} \\
+ 1 + \frac{\lambda_{in}}{\mu_{in}} + 1 + \frac{\lambda_{in}}{\mu_{c}} + 1 + \frac{\lambda_{in}}{\mu_{a}} \sum_{i=0}^{d} A^a B^{b-a} C^{c-b} D^{k-i-1} \\
+ \frac{\lambda_{nk-1,0}}{\mu_{R}} + \frac{\lambda_{dk-1,1}}{\mu_{R}} + \frac{\lambda_{dk-1,2}}{\mu_{R}} \left[ \begin{array}{c} 1 \\
\frac{\lambda_{f,0,0}}{\lambda_{f,0,0} + \lambda_{d,0,1} + \lambda_{in}} \\
\frac{\lambda_{f,0,0}}{\lambda_{f,0,0} + \lambda_{d,0,1} + \lambda_{in}} \end{array} \right]^{-1} \right\}
\]

where $A$, $B$, $C$ and $D$ are the recursive matrices in phase $I$, $II$, $III$, and $IV$ respectively. The proof of proposition 3 is shown in Appendix C. With the analytical expression of $\pi_{0,0,0}$, the rest of states can be expressed analytically by multiplying equation (4) with the associated ratios. Hence, by substituting of the analytical expressions of all the states to (3), the analytical expression for time-averaged cost can be achieved.

**4. Case study – CBM of concrete bridges**

In this section, we apply the overall approach to a case study of maintenance of concrete bridges. Bridges are important infrastructure assets that serve the public and drive economic growth. In practice, conditions of bridges are classified into 5 (0 to 4) discrete states. State 0 indicates that the bridge is in the ‘as good as new’ state, and state 4 indicates the functional failure of the bridge. Concrete bridge components may experience different exposure levels during their lifetimes, such as the risk of flooding, traffic load, aging environment, and condition of other components. The exposure levels for bridges are normally classified into three categories: mild, moderate, and severe. The exposure levels can be modelled using environmental indices described in section 4. Crack, creep, and
fatigue propagation are the major deterioration mechanisms for concrete bridges. In the literature, such types of deterioration are normally modeled by a Gamma process that is composed of independent increments of Gamma distribution (Van Noortwijk, 2009). With data and subjective estimates provided by the UK Bridges Board, the initial settings of sojourn times at each state are presented in Table 1.

Table 1: Initial setting of sojourn time distributions of each state

<table>
<thead>
<tr>
<th></th>
<th>$S_0$</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>$\mathcal{T} \sim \text{Exp}(0.0167)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.0633)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.0633)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.0633)$</td>
</tr>
<tr>
<td>(years)</td>
<td>$\mathbb{E}(\mathcal{T}) = 60$</td>
<td>$\mathbb{E}(\mathcal{T}) = 60$</td>
<td>$\mathbb{E}(\mathcal{T}) = 60$</td>
<td>$\mathbb{E}(\mathcal{T}) = 60$</td>
</tr>
<tr>
<td>Moderate</td>
<td>$\mathcal{T} \sim \text{Exp}(0.0167)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.095)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.19)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.38)$</td>
</tr>
<tr>
<td>(years)</td>
<td>$\mathbb{E}(\mathcal{T}) = 60$</td>
<td>$\mathbb{E}(\mathcal{T}) = 40$</td>
<td>$\mathbb{E}(\mathcal{T}) = 20$</td>
<td>$\mathbb{E}(\mathcal{T}) = 10$</td>
</tr>
<tr>
<td>Severe</td>
<td>$\mathcal{T} \sim \text{Exp}(0.05)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.2533)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.38)$</td>
<td>$\mathcal{T} \sim \Gamma(3.8, 0.76)$</td>
</tr>
<tr>
<td>(years)</td>
<td>$\mathbb{E}(\mathcal{T}) = 20$</td>
<td>$\mathbb{E}(\mathcal{T}) = 15$</td>
<td>$\mathbb{E}(\mathcal{T}) = 10$</td>
<td>$\mathbb{E}(\mathcal{T}) = 5$</td>
</tr>
</tbody>
</table>

In Table 1, $S_i$ indicates the $i^{th}$ state in the semi-Markov aging model. The sojourn times at state 0 are exponentially distributed, which do not require further treatment. The sojourn times of states 1 to 3 follow Gamma distribution. Based on the shape parameter of the Gamma distribution, the skewness and kurtosis are 1.0260 and 4.5789, respectively. According to Proposition 2, a minimum of $n = 4$ is required to match the third and fourth standardized moments of the Gamma distribution. Hence, the sojourn times at state $S_1$, $S_2$ and $S_3$ can be expressed as four serially connected sub-condition states, as shown in Table 2.

Table 2: Sojourn times of sub-states after hypo-exponential decomposition for matching the first four standardized moments

<table>
<thead>
<tr>
<th>Condition state</th>
<th>$S_0$</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-state</td>
<td>$S_{0,0}$</td>
<td>$S_{1,0}$</td>
<td>$S_{1,1}$</td>
<td>$S_{1,2}$</td>
</tr>
<tr>
<td>Mild (years)</td>
<td>60</td>
<td>15.4</td>
<td>13.2</td>
<td>15.4</td>
</tr>
<tr>
<td>Moderate (years)</td>
<td>60</td>
<td>8.82</td>
<td>10.3</td>
<td>10.3</td>
</tr>
<tr>
<td>Severe (years)</td>
<td>20</td>
<td>4.19</td>
<td>4.01</td>
<td>3.15</td>
</tr>
</tbody>
</table>
In Table 2, $\bar{S}_i^j$ denotes the $j^{th}$ sub-state in the surrogate Markov chain for approximating the $i^{th}$ condition state in the semi-Markov chain. After the approximation, the four aging conditions in the semi-Markov chain are expanded to 13 sub-condition states in the surrogate Markov chain. Because the surrogate Markov chain is a continuous time Markov chain with exponentially distributed sojourn time, the deterioration rate can be calculated by inverting the sojourn time.

We consider the environmental decline as a stochastic process. The decline rate from mild exposure level to moderate exposure level is every 10 years. The decline rate from the moderate exposure level to severe exposure level is every 2 years. In the UK, the condition of bridges is investigated by general inspection every 2 years. We assume that the average duration of inspection is 4 hours and costs £100. A minor maintenance activity such as patching takes 2 days and cost £6000. Major maintenance that eliminates the accumulated damage of bridges takes one week and cost £200,000. All maintenance activities on the bridge will cause planned traffic interruption. The traffic management cost, in this case, is £5000 per day. A replacement of the concrete bridge can last as long as a month and cost up to £300,000. More importantly, it can also cause a safety issue and unplanned traffic interruption, which will cost £20,000 per day. The parameter setting of the declining operating environment, inspection, and maintenance are provided in Table 3.

Table 3: Parameter setting of the declining operating environment, inspection, and maintenance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{f,0}$</td>
<td>0.1 year$^{-1}$</td>
</tr>
<tr>
<td>$\lambda_{f,1}$</td>
<td>0.5 year$^{-1}$</td>
</tr>
<tr>
<td>$\lambda_{in}$</td>
<td>0.5 year$^{-1}$</td>
</tr>
<tr>
<td>$\mu_{in}$</td>
<td>6 day$^{-1}$</td>
</tr>
<tr>
<td>$C_{in}$</td>
<td>£100</td>
</tr>
<tr>
<td>$\mu_c$</td>
<td>0.5 day$^{-1}$</td>
</tr>
<tr>
<td>$C_c$</td>
<td>£6000</td>
</tr>
<tr>
<td>$\mu_M$</td>
<td>0.1429 day$^{-1}$</td>
</tr>
<tr>
<td>$C_M$</td>
<td>£200,000</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>0.033 day$^{-1}$</td>
</tr>
<tr>
<td>$C_R$</td>
<td>£300,000</td>
</tr>
<tr>
<td>$C_p$</td>
<td>£5000 day$^{-1}$</td>
</tr>
<tr>
<td>$C_u$</td>
<td>£20000 day$^{-1}$</td>
</tr>
</tbody>
</table>

In this case, we aim to find the optimal maintenance thresholds of phases I, II, III, and IV to minimize the long-term maintenance cost of concrete bridges. By using the three-layered CBM model, the time-
averaged long-term cost can be calculated by equation (4). The cost is minimized when $a = 4, b = 6$, and $c = 6$. The minimum long-term cost $C_{\text{min}}$ is £4667.40 per year. The time-averaged long-term costs at the combination of different phase thresholds for $0 \leq a, b, c \leq 12, a, b, c \in \mathbb{N}$ are illustrated in Figure 5. The size of the markers $s \propto (C - C_{\text{min}} + \epsilon)^{-1}$ where $\epsilon$ is a significantly small number to avoid the undefined value.

![Figure 5: Time-averaged long-term costs at different phase thresholds](image)

For performance comparison, we have applied a similar approach where we match the first three moments of the sojourn time of condition states. The resulting sub-condition states are tabulated in Table 4.

**Table 4: Sojourn times of sub-states after hypo-exponential decomposition for matching the first three moments**

<table>
<thead>
<tr>
<th>Condition state</th>
<th>$S_0$ (Years)</th>
<th>$S_1$ (Years)</th>
<th>$S_2$ (Years)</th>
<th>$S_3$ (Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-state</td>
<td>$S_{0,0}$</td>
<td>$S_{1,0}$</td>
<td>$S_{1,1}$</td>
<td>$S_{1,2}$</td>
</tr>
<tr>
<td>Mild</td>
<td>60</td>
<td>14.1</td>
<td>14.1</td>
<td>18.6</td>
</tr>
<tr>
<td>Moderate</td>
<td>60</td>
<td>11.0</td>
<td>11.0</td>
<td>11.0</td>
</tr>
<tr>
<td>Severe</td>
<td>20</td>
<td>4.13</td>
<td>4.13</td>
<td>4.13</td>
</tr>
</tbody>
</table>

We repeat the same approach to optimize the maintenance thresholds with the first three matched moments. In this case, the optimized maintenance thresholds are $a = b = c = 6$. Figure 6 compares
the time-averaged long-term cost with the first four matched moments and the first three matched moments.

![Figure 6: Time-averaged long-term cost when b = c = 6](image)

By matching the first three moments, the minimum time averaged cost is £4706.40 per year when $a = 6$. With four matching moments, the minimum time-averaged cost is reduced to £4667.40 per year when $a = 4$, which lead to a 0.84% annual saving of the maintenance budget. Although the resulting difference seems small in the first instance, similar benefits can be rolled out to similar bridges or other concrete infrastructure when planning regional maintenance. For instance, according to (Parlikad & Catton, 2018), the average revenue budget for local authorities to maintain bridges is about £2.3 million. In the cases with a similar scale, a saving of £19,320 can be expected for local authorities. Moreover, for bridge maintenance, many sensors such as vibration sensors, moisture sensors, and fiber optic sensors can be implemented to further improve maintenance performance. However, it may require a large initial investment in sensors, ICT, and data centres. Our developed approach is formulated using the practically available inspection data and can be used until investment in such condition-monitoring technologies is made. Moreover, our model can lead to maintenance practices that can potentially avoid the excessive aging of the assets caused by the decline of the operating environment. In practice, it promotes less costly maintenance by employing minor maintenance to mitigate excessive aging and reducing the chance of costly major maintenance and replacement.

**Concluding Remarks**

The paper presents a new technique to optimise the CBM policy for long-life assets. We presented a mathematically tractable way for optimizing the maintenance policy through a hypo-exponential distribution. This approach addresses one of the shortcomings of current models in the literature that matches the first three moments resulting in fictitious sub-states that are undesirable for designing a CBM policy. Our approach transforms the semi-Markovian aging process using a hypo-exponential
decomposition so that the surrogate Markov chain could be deemed as the sub-states of the asset. More importantly, the tail behavior of sojourn time distribution could be preserved with minimal representation.

This research emphasises the importance of matching the fourth moment and provides fundamental analysis on the boundaries of skewness and kurtosis, which contributes to the selection of the minimum number of states in the surrogate Markov chain. The resulting three-layered CBM model helps determine optimal maintenance decisions based on the combined information of aging and operating environment risk and minimises the time-averaged operation cost. The analytical expression of the reference state makes the model mathematical tractable. Finally, the practical case study highlights the benefit of the approach on annual maintenance budget planning for regional authorities.

In the future, we aim to capture and formulate the transitive influence from the operating environment with the assistance of sensory information.

Appendix A. Proof of Proposition 1

For a distribution to be infinitely divisible, it requires that the characteristic function \( \phi(t) \) can be expressed as

\[
\ln \phi_{\chi}(t) = itc + \int_0^\infty \left( e^{itx} - 1 - \frac{itx}{1 + x^2} \right) h(x) \, dx
\]  

(A1)

where \( c \) is real and \( xh(x)/1 + x^2 \) is integrable between 0 and \( \infty \). For example, the characteristic function of exponential distribution with rate \( \lambda \) is \( \lambda/\lambda - it \). In this case, the equation holds when

\[
c = \int_0^\infty e^{-\lambda x} \, dx \text{ and } h(x) = x^{-1}e^{-\lambda x}
\]

By substituting the \( c \) and \( h(x) \) into equation (A3) we have:

\[
\ln \frac{\lambda}{\lambda - it} = \int_0^\infty (e^{itx} - 1)x^{-1}e^{-\lambda x} \, dx
\]

Based on the property of the characteristic function \( \phi_Y(t) = \phi_{X_1}(t)\phi_{X_2}(t) \ldots \phi_{X_n}(t) \), we have

\[
\ln \phi_Y(t) = \ln \left( \prod_{j=1}^n \frac{\lambda_j}{\lambda_j - it} \right) = \sum_{j=1}^n \ln \left( \frac{\lambda_j}{\lambda_j - it} \right)
\]

Let \( c = \int_0^\infty \sum_{j=1}^n \frac{e^{-\lambda_j x}}{1 + x^2} \, dx \) and \( h(x) = \sum_{j=1}^n \frac{e^{-\lambda_j x}}{x} \). Then by substituting \( c \) and \( h(x) \) into equation (A1), it follows:

\[
\ln \phi_Y(t) = it \int_0^\infty \sum_{j=1}^n \frac{e^{-\lambda_j x}}{1 + x^2} \, dx + \int_0^\infty (e^{itx} - 1)x^{-1} \sum_{j=1}^n e^{-\lambda_j x} \, dx - it \int_0^\infty \frac{\sum_{j=1}^n e^{-\lambda_j x}}{1 + x^2} \, dx
\]
The result shows that the hypo-exponential distribution is infinite divisible. As (Rohatgi & Székely, 1989) proved, \( \kappa \geq 1 + \gamma^2 \) for any given distribution. The inequalities can be further sharpened to \( \kappa \geq 3 + \gamma^2 \) if the distribution is infinite divisible. Therefore, the interdependence between skewness and kurtosis for hypo-exponential distribution is constrained by the sharp boundary as expressed by equation (A2).

\[ \kappa \gamma \geq 3 + \gamma^2 \]  

(A2)

Appendix B. Proof of Proposition 2

According to equation (1)

\[ \left( \frac{\gamma^3}{2} \right)^{1/3} = \left( \frac{\sum_{i=1}^{n} \lambda_i^{-3}}{n} \right)^{1/2} \]

By using the inequality of Hölder mean

\[ \left( \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{-3} \right)^{1/3} \geq \left( \frac{1}{n} \sum_{i=1}^{n} \lambda_i^{-2} \right)^{1/2} \]

The equality holds when \( \lambda_i \) is identical. Hence,

\[ \left( \frac{\gamma^3}{2} \right)^{1/3} \geq \left( \frac{1}{n} \right)^{1/6} \]

Therefore,

\[ \gamma \geq \frac{2}{\sqrt{n}} \]

(A3)

By using Cauchy-Schwarz inequality

\[ \left( \sum_{i=1}^{n} 1 \cdot \lambda_i^{-2} \right)^2 \leq n \sum_{i=1}^{n} \lambda_i^{-4} \]

The equality holds when \( \lambda_i^{-2} \) are identical. Therefore, the
\[ \kappa_Y = \frac{6 \sum_{i=1}^{n} \lambda_i^{-4}}{(\sum_{i=1}^{n} \lambda_i^{-2})^2} + 3 \geq \frac{6}{n} + 3 \]  

(A4)

Based on Proposition 1, the necessary boundaries for matching the third and fourth moments are expressed in equation (A2), (A3) and (A4). Equations (A3) and (A4) are related to the number of states \( n \). Equation (A2) is independent of \( n \). When \( n = 1 \), it is equivalent to use an exponential distribution for approximation. In this case, the boundaries for skewness and kurtosis can be calculated using equation (A3) and (A4) as \( \gamma_Z \geq 2 \) and \( \kappa_Z \geq 9 \). By substituting \( n \in \{2, 3, 4, \ldots \} \) into equation (A3) and (A4), we can calculate the boundaries for matching the third and fourth moments for a given \( n \). Then, the feasible region of phase-type approximation using hypo-exponential distribution can be visualized as shown in Figure 2. Figure 2 shows that when \( n \) increases the region expands. When \( n \) approaches infinity, the region is the largest and identical to all the positive region above the curve boundary \( \kappa_Y = 3 + \gamma_Y^2 \). To match a given distribution with skewness \( \gamma_Z \) and kurtosis \( \kappa_Z \), the minimum number of state \( n \) can be identified when the boundaries conditions satisfy at \( n \) and not satisfy at \( n - 1 \) as shown below:

\[
\left\{ \begin{array}{l}
\frac{2}{\sqrt{n}} \leq \gamma_Z \leq \frac{2}{\sqrt{n} - 1} \\
\frac{6}{n} + 3 \leq \kappa_Z \leq \frac{6}{n - 1} + 3
\end{array} \right.
\]

Appendix C. Proof of Proposition 3

Equations based on the input and output equilibrium

For \( i = 0 \)

\[
\begin{align*}
\pi_{0,0,0}(\lambda_{n0,0} + \lambda_{f0,0} + \lambda_{in}) &= \mu_{n0} \pi_{0,0,1} + \mu_c (\pi_{0,1,2} + \pi_{0,2,2}) \\
+ \mu_M (\sum_{a<i \leq b} \pi_{i,2,2} + \sum_{b<i \leq c} \sum_{j<s \leq 2} \pi_{i,j,2} + \sum_{c<i \leq k} \sum_{j<s \leq 2} \pi_{i,j,2}) + \mu_R \pi_{k,0,0}, & j = 0, l = 0 \\
\pi_{0,1,0}(\lambda_{f0,1} + \lambda_{d0,1} + \lambda_{in}) &= \lambda_{f0,0} \pi_{0,0,0} & j = 1, l = 0 \\
\pi_{0,2,0}(\lambda_{d0,2} + \lambda_{in}) &= \lambda_{f0,1} \pi_{0,1,0}, & j = 2, l = 0 \\
\pi_{0,j,1} \mu_{in} &= \lambda_{in} \pi_{0,0,0}, & 0 \leq j \leq 2, l = 1 \\
\pi_{0,j,2} \mu_c &= \mu_{in} \pi_{0,j,1}, & 1 \leq j \leq 2, l = 2
\end{align*}
\]

For \( 0 < i \leq a \)

\[
\begin{align*}
\pi_{i,0,0}(\lambda_{ni,0} + \lambda_{fi,0} + \lambda_{in}) &= \lambda_{ni-1,0} \pi_{i-1,0,0} + \mu_{in} \pi_{i,0,1} + \mu_c (\pi_{i,1,2} + \pi_{i,2,2}), & j = 0, l = 0 \\
\pi_{i,1,0}(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) &= \lambda_{fi,0} \pi_{i,0,0} + \lambda_{di-1,1,0} \pi_{i-1,1,0}, & j = 1, l = 0 \\
\pi_{i,2,0}(\lambda_{di,2} + \lambda_{in}) &= \lambda_{fi,1} \pi_{i,1,0} + \lambda_{di-1,2,0} \pi_{i-1,2,0}, & j = 2, l = 0 \\
\pi_{i,j,1} \mu_{in} &= \lambda_{in} \pi_{i,0,0}, & 0 \leq j \leq 2, l = 1 \\
\pi_{i,j,2} \mu_c &= \mu_{in} \pi_{i,j,1}, & 1 \leq j \leq 2, l = 2
\end{align*}
\]
For $a < i \leq b$

\[
\begin{align*}
\pi_{i,0,0}(\lambda_{ni,0} + \lambda_{fl,0} + \lambda_{in}) &= \lambda_{ni-1,0}\pi_{i-1,0,0} + \mu_{in}\pi_{i,0,1} + \mu_{c}\pi_{i,1,2}, & j = 0, l = 0 \\
\pi_{i,1,0}(\lambda_{fl,1} + \lambda_{di,1} + \lambda_{in}) &= \lambda_{fl,0}\pi_{i,0,0} + \lambda_{di-1,1,0}\pi_{i-1,1,0}, & j = 1, l = 0 \\
\pi_{i,2,0}(\lambda_{di,2} + \lambda_{in}) &= \lambda_{fl,1}\pi_{i,1,0} + \lambda_{di-1,2,0}\pi_{i-2,0}, & j = 2, l = 0 \\
\pi_{i,j,1}\mu_{in} &= \lambda_{in}\pi_{i,0,0}, & 0 \leq j \leq 2, l = 1 \\
\pi_{i,j,2}\mu_{c} &= \mu_{in}\pi_{i,j,1}, & j = 1, l = 2 \\
\pi_{i,j,2}\mu_{M} &= \mu_{in}\pi_{i,j,1}, & j = 2, l = 2 
\end{align*}
\]

For $b < i \leq c$

\[
\begin{align*}
\pi_{i,0,0}(\lambda_{ni,0} + \lambda_{fl,0} + \lambda_{in}) &= \lambda_{ni-1,0}\pi_{i-1,0,0} + \mu_{in}\pi_{i,0,1}, & j = 0, l = 0 \\
\pi_{i,1,0}(\lambda_{fl,1} + \lambda_{di,1} + \lambda_{in}) &= \lambda_{fl,0}\pi_{i,0,0} + \lambda_{di-1,1,0}\pi_{i-1,1,0}, & j = 1, l = 0 \\
\pi_{i,2,0}(\lambda_{di,2} + \lambda_{in}) &= \lambda_{fl,1}\pi_{i,1,0} + \lambda_{di-1,2,0}\pi_{i-2,0}, & j = 2, l = 0 \\
\pi_{i,j,1}\mu_{in} &= \lambda_{in}\pi_{i,0,0}, & 0 \leq j \leq 2, l = 1 \\
\pi_{i,j,2}\mu_{M} &= \mu_{in}\pi_{i,j,1}, & 1 \leq j \leq 2, l = 2 
\end{align*}
\]

For $c < i < k$

\[
\begin{align*}
\pi_{i,0,0}(\lambda_{ni,0} + \lambda_{fl,0} + \lambda_{in}) &= \lambda_{ni-1,0}\pi_{i-1,0,1}, & j = 0, l = 0 \\
\pi_{i,1,0}(\lambda_{fl,1} + \lambda_{di,1} + \lambda_{in}) &= \lambda_{fl,0}\pi_{i,0,0} + \lambda_{di-1,1,0}\pi_{i-1,1,0}, & j = 1, l = 0 \\
\pi_{i,2,0}(\lambda_{di,2} + \lambda_{in}) &= \lambda_{fl,1}\pi_{i,1,0} + \lambda_{di-1,2,0}\pi_{i-2,0}, & j = 2, l = 0 \\
\pi_{i,j,1}\mu_{in} &= \lambda_{in}\pi_{i,0,0}, & 0 \leq j \leq 2, l = 1 \\
\pi_{i,j,2}\mu_{M} &= \mu_{in}\pi_{i,j,1}, & 1 \leq j \leq 2, l = 2 
\end{align*}
\]

For $i = k$

\[
\pi_{k,0,0}\mu_{R} = \pi_{k-1,0,0}(\lambda_{nk-1,0} + \lambda_{dk-1,1} + \lambda_{dk-1,2})
\]

From the list of equations, only $\pi_{i,0,0}$ are dependent on the probabilities of $\pi_{i-1,j,0}$, $\pi_{i,j,1}$ and $\pi_{i,j,2}$ can be expressed by only $\pi_{i,0,0}$. If we use $\pi_{0,0,0}$ as the reference state, the $\pi_{0,1,0}$ and $\pi_{0,2,0}$ can be expressed as shown in the following equations:

\[
\begin{align*}
\pi_{0,1,0} &= \frac{\lambda_{f0,0}}{(\lambda_{f0,1} + \lambda_{d0,1} + \lambda_{in})}\pi_{0,0,0} \\
\pi_{0,2,0} &= \frac{\lambda_{f0,1}}{(\lambda_{d0,2} + \lambda_{in})(\lambda_{f0,1} + \lambda_{d0,1} + \lambda_{in})}\pi_{0,0,0}
\end{align*}
\]

$\pi_{i,j,0}$ when $0 < i \leq a$ are dependent on $\pi_{i,1,0}$, $\pi_{i,2,0}$, $\pi_{i,0,1}$, $\pi_{i,1,2}$ and $\pi_{i,2,2}$, which can be expressed as:

\[
\begin{align*}
\pi_{i,1,0} &= \frac{\lambda_{fi,0}}{(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in})}\pi_{i,0,0} + \frac{\lambda_{di-1,1,0}}{(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in})}\pi_{i-1,1,0}
\end{align*}
\]
By substituting the expressions, we have

\[ \pi_{i,0} = a_{0,0}\pi_{i-1,0,0} + a_{0,1}\pi_{i-1,1,0} + a_{0,2}\pi_{i-1,2,0} \]

where

\[ a_{0,0} = \frac{\lambda_{ni-1,0}(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]

\[ a_{0,1} = \frac{\lambda_{in}\lambda_{di-1,1,0}(\lambda_{fi,1} + \lambda_{di,2} + \lambda_{in})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]

\[ a_{0,2} = \frac{\lambda_{in}\lambda_{di-1,2,0}(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]

\[ \pi_{i,1,0} = c_{1,0}\pi_{i-1,0,0} + c_{1,1}\pi_{i-1,1,0} + c_{1,2}\pi_{i-1,2,0} \]

\[ a_{1,0} = \frac{\lambda_{fi,0}\lambda_{ni-1,0}(\lambda_{di,2} + \lambda_{in})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]

\[ a_{1,1} = \frac{\lambda_{di-1,1,0}(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]

\[ a_{1,2} = \frac{\lambda_{fi,0}\lambda_{in}\lambda_{di-1,2,0}}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]

\[ \pi_{i,2,0} = c_{2,0}\pi_{i-1,0,0} + c_{2,1}\pi_{i-1,1,0} + c_{2,2}\pi_{i-1,2,0} \]

\[ a_{2,0} = \frac{\lambda_{fi,1}\lambda_{fi,0}\lambda_{ni-1,0}}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{di,2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{di,2} + \lambda_{in} + \lambda_{fi,1})} \]
To make the equations more compact, we can transfer the equations into matrix format.

\[
\begin{align*}
a_{2,1} &= \frac{\lambda_{f1}\lambda_{d_{i-1,0}}(\lambda_{ni,0} + \lambda_{fi,0})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{d2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{d2} + \lambda_{in} + \lambda_{fi,1})} \\
a_{2,2} &= \frac{\lambda_{d_{i-1,0}}(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}\lambda_{d_{i-1,0}}}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{d2} + \lambda_{in})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}(\lambda_{d2} + \lambda_{in} + \lambda_{fi,1})}
\end{align*}
\]

In phase I, we denote the recursive matrix as \( A \). Therefore, we can analytically express any deterioration state \( i \) within phase I.

\[
\begin{pmatrix}
\pi_{i,0,0} \\
\pi_{i,1,0} \\
\pi_{i,2,0}
\end{pmatrix} = A^i \begin{pmatrix}
\pi_{0,0,0} \\
\pi_{0,1,0} \\
\pi_{0,2,0}
\end{pmatrix} = A^i \begin{pmatrix}
1 \\
\lambda_{f0,0} + \lambda_{d0,1} + \lambda_{in} \\
\lambda_{d0,2} + \lambda_{in}
\end{pmatrix}, \forall 0 \leq i \leq a \quad \text{(A5)}
\]

Similarly, we can express the phase II \((a + 1 \leq i \leq b)\) deterioration states as a recursive equation A2:

\[
\begin{pmatrix}
\pi_{n,0,0} \\
\pi_{n,1,0} \\
\pi_{n,2,0}
\end{pmatrix} = A^a B^{i-a} \begin{pmatrix}
\pi_{0,0,0} \\
\pi_{0,1,0} \\
\pi_{0,2,0}
\end{pmatrix} = A^a B^{i-a} \begin{pmatrix}
1 \\
\lambda_{f0,0} + \lambda_{d0,1} + \lambda_{in} \\
\lambda_{d0,2} + \lambda_{in}
\end{pmatrix}, \forall 0 \leq i \leq a \quad \text{(A6)}
\]

where \( B \) is a \( 3 \times 3 \) matrix

\[
B = \begin{pmatrix}
b_{0,0} & b_{0,1} & 0 \\
b_{1,0} & b_{1,1} & 0 \\
b_{2,0} & b_{2,1} & b_{2,2}
\end{pmatrix}
\]

\[
\begin{align*}
b_{0,0} &= \frac{\lambda_{ni-1,0}(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}} \\
b_{0,1} &= \frac{\lambda_{in}\lambda_{d_{i-1,0}}}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}} \\
b_{1,0} &= \frac{\lambda_{fi,0}\lambda_{ni-1,0}}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}} \\
b_{1,1} &= \frac{\lambda_{d_{i-1,0}}(\lambda_{ni,0} + \lambda_{fi,0})}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{di,1} + \lambda_{in}) - \lambda_{in}\lambda_{fi,0}}
\end{align*}
\]
\[ b_{2,0} = \frac{\lambda_{f,1}\lambda_{f,0}\lambda_{i-1,0}}{(\lambda_{n,0} + \lambda_{f,0})(\lambda_{f,1} + \lambda_{d,1} + \lambda_{i})} \left( \lambda_{d,2} + \lambda_{i} \right) - \lambda_{i} \lambda_{f,0} \left( \lambda_{d,2} + \lambda_{i} \right) \]

\[ b_{2,1} = \frac{\lambda_{f,1}\lambda_{d,i-1,0}(\lambda_{n,0} + \lambda_{f,0})}{(\lambda_{n,0} + \lambda_{f,0})(\lambda_{f,1} + \lambda_{d,1} + \lambda_{i})} \left( \lambda_{d,2} + \lambda_{i} \right) - \lambda_{i} \lambda_{f,0} \left( \lambda_{d,2} + \lambda_{i} \right) \]

\[ b_{2,2} = \frac{\lambda_{d,i-1,0}}{\lambda_{d,2} + \lambda_{i}} \]

For phase \( \text{III} \) \((b + 1 \leq i \leq c)\), we have

\[ \begin{bmatrix} \pi_{n,0,0} \\ \pi_{n,1,0} \\ \pi_{n,2,0} \end{bmatrix} = A^a B^{b-a} C^{i-b} \begin{bmatrix} \pi_{0,0,0} \\ \pi_{0,1,0} \\ \pi_{0,2,0} \end{bmatrix} = A^a B^{b-a} C^{i-b} \begin{bmatrix} 1 \\ \frac{\lambda_{f,0}}{(\lambda_{f,0} + \lambda_{d,0,1} + \lambda_{i})} \\ \frac{\lambda_{f,0} \lambda_{f,1}}{(\lambda_{d,0,2} + \lambda_{i})(\lambda_{f,0} + \lambda_{d,0,1} + \lambda_{i})} \end{bmatrix} \]  

(A7)

Where \( C \) is a \( 3 \times 3 \) matrix

\[ C = \begin{bmatrix} c_{0,0} & 0 & 0 \\ c_{1,0} & c_{1,1} & 0 \\ c_{2,0} & c_{2,1} & c_{2,2} \end{bmatrix} \]

\[ c_{0,0} = \frac{\lambda_{ni-1,0}}{\lambda_{ni,0} + \lambda_{fi,0}} \]

\[ c_{1,0} = \frac{\lambda_{ni-1,0}\lambda_{fi,0}}{(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{d,1} + \lambda_{i})} \]

\[ c_{1,1} = \frac{\lambda_{di-1,0}}{\lambda_{f,1} + \lambda_{d,1} + \lambda_{i}} \]

\[ c_{2,0} = \frac{\lambda_{ni-1,0}\lambda_{fi,0}\lambda_{fi,1}}{(\lambda_{d,2} + \lambda_{i})(\lambda_{ni,0} + \lambda_{fi,0})(\lambda_{fi,1} + \lambda_{d,1} + \lambda_{i})} \]

\[ c_{2,1} = \frac{\lambda_{di-1,0}\lambda_{fi,1}}{(\lambda_{d,2} + \lambda_{i})(\lambda_{fi,1} + \lambda_{d,1} + \lambda_{i})} \]

\[ c_{2,2} = \frac{\lambda_{di-1,2,0}}{\lambda_{d,2} + \lambda_{i}} \]

For phase \( \text{IV} \) \((c + 1 \leq i \leq k)\), we have
\[
\begin{bmatrix}
\pi_{n,0,0} \\
\pi_{n,1,0} \\
\pi_{n,2,0}
\end{bmatrix}
= A^a B^{b-a} C^{c-b} D^{i-c} \begin{bmatrix}
\pi_{0,0,0} \\
\pi_{0,1,0} \\
\pi_{0,2,0}
\end{bmatrix}
\]
\[
= A^a B^{b-a} C^{c-b} D^{i-c} \begin{bmatrix}
1 \\
\frac{1}{\lambda_{f,0,0}} \\
\frac{1}{\lambda_{f,0,1}}
\end{bmatrix}
\frac{\lambda_{f,0,0} \lambda_{f,0,1}}{(\lambda_{d,0,2} + \lambda_{in})(\lambda_{f,0,1} + \lambda_{d,0,1} + \lambda_{in})}
\]

(A8)

Where \( D \) is a \( 3 \times 3 \) matrix

\[
D = \begin{bmatrix}
d_{0,0} & 0 & 0 \\
d_{1,0} & d_{1,1} & 0 \\
d_{2,0} & d_{2,1} & d_{2,2}
\end{bmatrix}
\]

\[
d_{0,0} = \frac{\lambda_{n-i,0}}{\lambda_{n,0} + \lambda_{f,0} + \lambda_{in}}
\]

\[
d_{1,0} = \frac{\lambda_{n-i,0} \lambda_{f,0}}{(\lambda_{n,0} + \lambda_{f,0} + \lambda_{in})(\lambda_{f,1} + \lambda_{d,1} + \lambda_{in})}
\]

\[
d_{1,1} = \frac{\lambda_{d,1}}{\lambda_{f,1} + \lambda_{d,1} + \lambda_{in}}
\]

\[
d_{2,0} = \frac{\lambda_{n-i,0} \lambda_{f,0} \lambda_{f,1}}{(\lambda_{d,2} + \lambda_{in})(\lambda_{n,0} + \lambda_{f,0} + \lambda_{in})(\lambda_{f,1} + \lambda_{d,1} + \lambda_{in})}
\]

\[
d_{2,1} = \frac{\lambda_{d,1,0} \lambda_{f,1}}{(\lambda_{d,2} + \lambda_{in})(\lambda_{f,1} + \lambda_{d,1} + \lambda_{in})}
\]

\[
d_{2,2} = \frac{\lambda_{d,1,2,0}}{\lambda_{d,2} + \lambda_{in}}
\]

Based on the list of equations, states \( \pi_{i,j,1} \) and \( \pi_{i,j,2} \) can be expressed by \( \pi_{i,j,0} \). By recalling the sum of all states probability is equal to 1, therefore by substituting A5 to A8, the analytical expression of referencing state of \( \pi_{0,0,0} \) can be derived.

\[\blacksquare\]

**Acknowledgement**

The research presented in this paper was supported by the Innovation and Knowledge Centre for Smart Infrastructure and Construction funded by the Engineering and Physical Sciences Research Council (EP/N021614/1) and Innovate UK.

**References**


