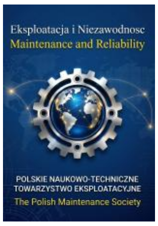


Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 2

https://ein.org.pl

The Polish Maintenance Society

Redundancy allocation and condition-based maintenance optimization in heterogeneous multi-state systems under imperfect monitoring and maintenance

Sebastián Miranda[1], Fredy Kristjanpoller[1,*], Pablo Viveros[1], Rodrigo Mena[1], Esteban Salgado[1]

[1] Industrial Engineering Department, Universidad Técnica Federico Santa Maria, Chile
* Corresponding Author: fredy.kristjanpoller@usm.cl

Received: 12 March 2026
Revised: 9 July 2026
Accepted: 12 September 2026
Online: 15 September 2026

Abstract

Existing redundancy allocation and maintenance optimization models commonly neglect imperfect monitoring and maintenance conditions and often rely on sequential decision approaches that may lead to suboptimal solutions. This study proposes a joint optimization framework for redundancy allocation and CBM in heterogeneous repairable multi-state weighted k-out-of-n systems. The objective is to minimize life-cycle cost while satisfying availability requirements under imperfect monitoring and maintenance conditions. Semi-Markov processes and universal generating functions are jointly used to model degradation, false alarms, missed detections, maintenance delays, and imperfect preventive actions. An adapted GA identifies high-quality feasible solutions under strict availability requirements. Results show that explicitly modeling imperfections reduces life-cycle cost by about 13%, while a sequential scheme increases cost by about 7% compared with the joint formulation. Sensitivity analyses reveal that demand and repair duration strongly affect performance and require joint adaptation of thresholds and redundancy.

Keywords: *multi-state systems, condition-based maintenance, imperfect monitoring, imperfect maintenance, semi-Markov process*

Article citation:
Miranda S, Kristjanpoller F, Viveros P, Mena R, Salgado E, Redundancy allocation and condition-based maintenance optimization in heterogeneous multi-state systems under imperfect monitoring and maintenance, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(2) http://doi.org/10.17531/ein/236719

Highlights

- Joint RAP–CBM optimization in heterogeneous multi-state systems.
- Imperfect monitoring and maintenance embedded in SMP modeling.
- GA with tailored penalties finds feasible solutions under strict requirements.
- Neglecting realistic conditions risks infeasibility; incorporating them reduces LCC by about 13%.
- Demand and repair time changes require joint, not separate, threshold–redundancy tuning.

1. Introduction

Complex engineering systems—such as power [1–3], aircraft actuation [4], pumping [5], and manufacturing systems [6] — operate under progressive degradation and multiple failure modes that affect performance. When components exhibit multiple performance levels rather than binary states, ensuring reliable operation while meeting demand becomes challenging.

Multi-state systems (MSS) provide a natural framework for modeling such behavior, allowing components to operate at different performance levels before failure [7]. Weighted k-out-of-n configurations are commonly used to represent redundant architectures in which system availability depends on whether the aggregated component performance exceeds a demand threshold [5,8]. In these systems, redundancy design directly determines the structural capacity to withstand degradation and failure.

At the operational level, PM policies—ranging from TBM to CBM—mitigate degradation and sustain performance. Redundancy and maintenance decisions are inherently interdependent: redundancy shapes feasible maintenance strategies, while maintenance performance influences availability, life-cycle cost (LCC), and the required redundancy level [9].

Despite this interdependence, redundancy allocation and maintenance planning are often optimized separately and typically assume perfect monitoring, instantaneous interventions, and complete restoration. In practice, false alarms, missed detections, imperfect repairs, and maintenance delays significantly affect system performance and cost. These limitations motivate the development of unified frameworks that jointly optimize redundancy design and maintenance policy under realistic conditions, enabling truly optimal solutions.

1.1. Redundancy allocation problem

Fyffe et al. [10] first introduced the RAP to maximize system reliability through optimal component allocation under cost and weight constraints. Since then, numerous studies have extended this approach under more realistic assumptions.

Recent work has focused on improving RAP flexibility and realism for MSS. Attar et al. [11] developed a simulation-based multi-objective optimization approach for redundancy allocation in free distributed repairable MSS, showing that NSGA-II outperforms SPEA2. Chen et al. [5] proposed a dynamic availability model combining CTMC and UGF for heterogeneous MSS, optimizing the availability-cost tradeoff via NSGA-II.

Extensions addressing non-exponential failures include Li et al. [12], who modeled phase-mission systems using SMP and improved GA, and Kim [13], who employed generalized phase-type distributions and parallel genetic algorithms to reduce computational effort. More recently, Li et al. [14] studied k-out-of-n:G MSS with random weight thresholds, showing that a customized UGF approach improves modeling flexibility. Sharifi and Taghipour [15] further extended RAP to series–parallel MSS with components exhibiting continuous performance levels, maximizing system availability through a UGF-based method.

1.2. PM and CBM optimization

Maintenance comprises activities aimed at preserving or restoring system functionality and preventing unexpected failures [16]. Strategies are generally classified as CM, performed after failure, and PM, which involves planned interventions to reduce breakdowns and improve reliability [17,18]. PM policies are commonly divided into TBM, which schedules interventions based on time or usage, and CBM, which relies on real-time monitoring and reliability models to anticipate failures and reduce unnecessary interventions [19–21].

Comparative studies highlight the advantages of CBM in multi-component systems. Andersen and Nielsen [22] show that CBM achieves lower costs than TBM in deteriorating series systems. Castro et al. [23] propose a semi-regenerative computational policy for systems with monitored and unmonitored components, combining corrective and just-in-time preventive actions. Oakley et al. [24] develop a CBM framework for systems with stochastic and economic dependencies, achieving cost reductions through penalty-based balancing.

However, many studies assume ideal monitoring and perfect restoration, assumptions that are rarely satisfied in practice. Measurement noise, sensor degradation, and operational constraints limit condition assessment accuracy, while maintenance actions often provide only partial restoration. Consequently, recent research increasingly considers imperfect monitoring and imperfect maintenance. Inspection errors are typically modeled as false negatives and false positives within probabilistic detection frameworks [25,26], while maintenance actions may be classified as perfect, imperfect, or minimal depending on the restoration level [27].

Recent studies have incorporated these aspects into reliability models. Azadeh et al. [28] compare CBM, CM, and PM under forecasting uncertainty using a Markov-based simulation framework, showing that accurate prognostics are critical to cost and reliability performance in CBM. Tang et al. [26] propose a three-stage degradation model incorporating imperfect inspection and repair through virtual age and improvement factors, modeling inspection errors as false negatives. Wu et al. [29] analyze wind turbine blade maintenance using Petri nets to evaluate the impact of monitoring reliability on maintenance decisions.

For multi-component systems, Zhang et al. [25] propose a reliability-centered hybrid PM policy that incorporates imperfect inspections and repairs while allocating penalty costs based on component criticality. Xu et al. [30] develop a selective maintenance framework under resource constraints using a non-homogeneous CTMC and evolutionary optimization. Kocer and Khaniyev [31] introduce a Markov-based imperfect preventive maintenance model incorporating age-reduction and failure-rate increase factors to optimize reliability and maintenance scheduling. Zhang et al. [32] later propose a hybrid condition-based opportunistic maintenance strategy considering economic dependence and imperfect maintenance.

Maintenance interventions are often delayed by resource and operational constraints–such as spare-part shortages, staffing limitations, or scheduling priorities–rendering immediate action infeasible even when an alarm signals an undesired state. Liu et al. [33] incorporate repair delays into a Monte Carlo-based fault tree model, improving availability estimation. Zhao et al. [34] show that delay distributions influence expected costs and alarm thresholds in CBM models. Li et al. [6] further analyze delay propagation across operational periods, highlighting that preventive decisions must account for both system health and accumulated delays.

1.3. Joint optimization of redundancy design and maintenance policy

As highlighted in the introduction, jointly optimizing redundancy design and maintenance policy is essential to achieving truly optimal solutions. Nevertheless, some studies address these decisions independently or sequentially, first determining the system configuration and subsequently defining the maintenance strategy. For example, Eryilmaz [35] optimizes the number of components in k-out-of-n systems before determining the preventive replacement time. Similarly, Hamdan et al. [36] propose a combined PM–CM policy for weighted k-out-of-n systems, triggered either by system age or by a threshold on the total working weight, while optimizing design and maintenance variables separately and suggesting their joint treatment as future research.

In contrast, other works adopt an integrated perspective. Liu et al. [37] jointly optimize redundancy allocation and imperfect maintenance in multi-state systems, accounting for cumulative deterioration after repairs. More recently, Hao and Zhu [38] develop a joint redundancy–maintenance model for load-sharing k-out-of-n systems under environmental uncertainty, showing that neglecting such interactions may lead to overestimated reliability and suboptimal decisions.

1.4. Gap analysis and paper contribution

Despite recent advances, existing joint optimization models rarely consider CBM strategies in MSS with heterogeneous components, limiting their applicability to realistic industrial systems in which degradation behavior differs across components and monitoring is essential [25,26,28]. In addition, practical factors–such as detection errors [30–32,34,37], repairable systems [5,12,13,15,34], imperfect or non-instantaneous repairs, and maintenance delays–are typically neglected [11,33,35,36,38]. Only studies [37,38] address the joint optimization of redundancy design and maintenance policy, but without incorporating CBM.

To address these limitations, this work proposes a joint optimization framework for redundancy allocation and CBM, incorporating degradation thresholds under imperfect monitoring, including false positives and false negatives, and maintenance actions with variable effectiveness. Both decisions are optimized simultaneously within a unified framework. Table 1 compares the proposed approach with previous studies. The main contributions are summarized below.

1. Integration of CBM under realistic monitoring and maintenance conditions in heterogeneous MSS with non-identical repairable components. False alarms, missed detections, maintenance delays, and imperfect PM outcomes, including perfect, imperfect, and minor repairs, are incorporated into the semi-Markov state-transition mechanism, enabling the assessment of their joint impact on system availability.
2. Joint RAP–CBM optimization model that simultaneously determines optimal component allocation and degradation thresholds. The model allows components to follow either RTF or CBM policies while minimizing LCC. The framework provides a unified tool applicable to both design and operational decision-making.
3. Adapted GA with problem-specific chromosome encoding and a tailored penalty scheme. The proposed algorithm enables efficient exploration of large combinatorial design spaces while prioritizing feasible solutions under strict availability constraints.

The remainder of this paper is organized as follows. Section 2 presents the problem formulation. Section 3 develops the availability model for the MSS, incorporating degradation thresholds, detection errors, imperfect maintenance, and delay times. Section 4 formulates the optimization problem and introduces the adapted GA. Section 5 presents a numerical application, validation against MCS, comparative analyses, and sensitivity results. Finally, Section 6 concludes the paper.

Table 1. Summary of differences between this study and existing literature.

Paper	Multi-component	Multistate	Heterogeneous components	Non-exponential distributions	Imperfect maintenance	Delay time	Imperfect monitoring	Redundancy design	PM	Joint optimization
Liu et al. [37]	X	X	X		X			X	X	X
Azadeh et al. [28]	X			X			X		X	
Attar et al. [11]	X	X	X	X	X			X		
Chen et al. [5]	X	X	X					X		
Eryilmaz [35]	X		X	X				X	X	
Zhao et al. [34]				X		X			X	
Li et al. [12]	X		X	X				X		
Liu et al. [33]	X		X	X		X				
Hamdan et al. [36]	X		X	X				X	X	
Kim [13]	X			X				X		
Li et al. [14]	X	X	X					X		
Zhang et al. [25]	X		X	X	X		X		X	
Hao and Zhu [38]	X			X				X	X	X
Tang et al. [26]		X		X	X		X		X	
Sharifi and Taghipour [15]	X	X	X					X		
Wu et al. [29]		X		X		X	X		X	
Xu et al. [30]	X	X		X	X				X	
Zhang et al. [32]	X			X	X				X	
Kocer and Khaniyev [31]	X			X	X				X	
This work	***X***	***X***	***X***	***X***	***X***	***X***	***X***	***X***	***X***	***X***

2. Problem formulation

The problem of jointly assigning equipment in a redundant system and defining a CBM policy is considered under imperfect monitoring and maintenance. Preventive interventions are triggered according to the detected degradation state and executed once predefined degradation thresholds are reached. These thresholds constitute the key decision variables of the maintenance policy.

The system consists of N_T independent repairable multi-state components belonging to N different types. For each type i, the discrete state space is $S^{(i)} = \{0,1,\dots,M^{(i)}\}$, where the states represent degradation and maintenance conditions. By construction, the maximum index $M^{(i)}$ is always even, since each degradation level is associated with two states: one operational state and one PM state.

The set of operating states is defined as $S_{OP}^{(i)} = \{2m | m = 1,\dots,M^{(i)}/2\}$, which corresponds to the degradation levels at which the component is functioning and producing output. As degradation progresses, the component transitions through these states in increasing numerical order, while its production capacity decreases monotonically, i.e., $g_2^{(i)} > g_4^{(i)} > g_6^{(i)} > \cdots > 0$. Thus, even-indexed states represent operational degradation levels, whereas the immediately preceding odd-indexed states correspond to the PM actions associated with those degradation levels. The set of preventive maintenance states is defined as $S_{PM}^{(i)} = \{2m-1 | m = 1,\dots,M^{(i)}/2\}$, which represents scheduled maintenance interventions during which the component is temporarily stopped to perform preventive actions. Finally, the corrective maintenance state is $S_{CM}^{(i)} = \{0\}$, which represents complete failure and requires corrective repair. In this state the component is unable to produce output, and therefore its production capacity is $g_0^{(i)} = 0$.

Under a weighted k-out-of-n structure, system performance depends on the aggregate capacity of its components. Each system state $s \in S_{system}$ corresponds to a configuration of component states $(j_1, j_2, \dots, j_{N_T})$, where $j_k \in S^{(i)}$. The associated performance level is denoted G_s and depends on the capacities provided by the components in that configuration. The system is operational when G_s meets or exceeds the required demand W.

CM restores components to an as-good-as-new condition. Preventive actions are performed before failure and may be perfect, imperfect, corresponding to partial restoration, or minor, in which case the degradation state is preserved while the component is probabilistically rejuvenated.

Monitoring uncertainty is modeled through detection errors. A Type I error corresponds to a premature alarm, whereas a Type II error represents failure to detect that the degradation threshold has been reached. If the threshold is surpassed without detection, an alarm eventually occurs with certainty. After an alarm, a maintenance delay time is considered, representing the interval between alert generation and intervention execution during which further degradation or failure may occur.

System performance is assessed through an LCC framework including acquisition, operation, repair, and inefficiency costs to capture the financial impact over the asset's useful life [32,39]. The objective is to minimize LCC subject to availability requirements and component-type constraints.

3. Modeling system and its components

3.1. Multi-state component modeling

Each component is modeled as an SMP, which generalizes continuous- and discrete-time Markov chains by allowing arbitrary sojourn-time distributions that depend on both the current state and the subsequent state. Let $\left(Z^{(i)}(t), t \in \mathbb{R}_+\right)$ denote the SMP associated with a component of type i, defined on a finite state space $S^{(i)}$ and representing its operational condition over time. The embedded Markov renewal process $\left(\left(J_n^{(i)}, T_n^{(i)}\right), n \geq 0\right)$ is introduced, where $T_0^{(i)} \leq T_1^{(i)} \leq \cdots$ denote the jump times and $J_n^{(i)}$ the successive states visited in $S^{(i)}$. These processes satisfy the Markov property [40]. Fig. 1 presents the state-transition diagram for a generic component.

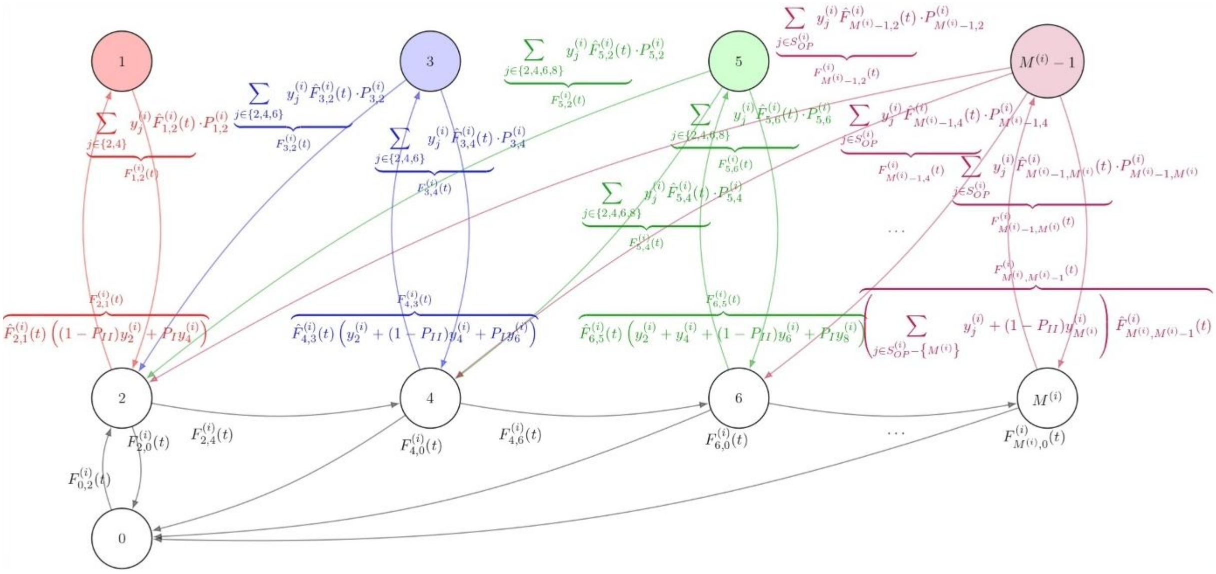


Figure 1. State diagram by component.

The SMP model is described by its kernel matrix $\boldsymbol{Q}^{(i)}(t)$, given by:

$$\boldsymbol{Q}^{(i)}(t) = \begin{bmatrix} Q_{0,0}^{(i)}(t) & Q_{0,1}^{(i)}(t) & \cdots & Q_{0,M^{(i)}}^{(i)}(t) \\ Q_{1,0}^{(i)}(t) & Q_{1,1}^{(i)}(t) & \cdots & Q_{1,M^{(i)}}^{(i)}(t) \\ \vdots & \vdots & \ddots & \vdots \\ Q_{M^{(i)},0}^{(i)}(t) & Q_{M^{(i)},1}^{(i)}(t) & \cdots & Q_{M^{(i)},M^{(i)}}^{(i)}(t) \end{bmatrix} \tag{1}$$

Each element of the matrix represents the probability that, given the component is currently in state j, the next transition occurs to state k and that this transition takes place within time t. Formally,

$$Q_{j,k}^{(i)}(t) = \mathbb{P}\left(J_{n+1}^{(i)} = k, T_{n+1}^{(i)} - T_n^{(i)} \leq t \mid J_n^{(i)} = j\right), \quad j,k \in S^{(i)}, t \geq 0, n \geq 0 \tag{2}$$

Imperfect condition monitoring is incorporated into the SMP by modifying the PM-triggering transitions, without altering the underlying physical degradation or failure processes. Let P_I and P_{II} denote the probabilities of Type I and Type II monitoring errors. A Type I error corresponds to a premature alarm, whereas a Type II error corresponds to a missed detection when the selected degradation threshold has been reached. For a component of type i, the CDF governing the transition from operational state $j \in S_{OP}^{(i)}$ to its corresponding PM state $j-1 \in S_{PM}^{(i)}$ is defined as follows:

$$F_{j,j-1}^{(i)}(t) = \hat{F}_{j,j-1}^{(i)}(t)\left(\sum_{x \in S_{OP}^{(i)}: x<j} y_x^{(i)} + (1-P_{II})y_j^{(i)} + P_I y_{j+2}^{(i)}\right) \tag{3}$$

where $y_x^{(i)}, y_j^{(i)}, y_{j+2}^{(i)}$ are binary variables indicating whether state j is selected as the degradation threshold for components of type i with $\sum_{j\in S_{OP}^{(i)}} y_j^{(i)} \leq 1$, and $\hat{F}^{(i)}_{j,j-1}(t)$ is the CDF of the transition time associated with the transition $j \to j-1$, which in this case corresponds to the delay time. The summation term captures situations in which the component has already exceeded the selected degradation threshold, in such cases, the PM transition is activated with certainty. The remaining terms account for correct detection and degradation overestimation, respectively. Note that the Type I error arises only when j is the state immediately preceding the selected degradation threshold, modifying the transition $j \to j-1$ through P_I.

To make the construction explicit, suppose that the selected degradation threshold for a component of type i is the operating state j_i^*, i.e., $y^{(i)}_{j_i^*} = 1$. The monitoring mechanism modifies only the distributions associated with transitions from operating states to their corresponding PM states. All physical degradation and failure transitions preserve their original transition-time distributions. For an operating state j, the PM-triggering distribution can be interpreted as follows:

1. If $j < j_i^*$, the component has not yet reached the selected threshold. Therefore, a transition to PM can only occur through a premature alarm. In the adopted formulation, this possibility is allowed for the operating state immediately preceding the selected threshold and is weighted by the Type I error probability P_I.
2. If $j = j_i^*$, the component has reached the selected threshold. A PM alarm is correctly triggered with probability $1-P_{II}$, while a Type II error prevents immediate detection with probability P_{II}.
3. If $j > j_i^*$, the component has already exceeded the selected threshold. In this case, the PM-triggering transition is enabled with probability one. The actual transition time, however, is still governed by the corresponding delay-time distribution and by the competing transitions from the same state.

Hence, once the threshold is fixed, equation (3) defines the modified PM-triggering transition-time distribution for each operating state. For example, if the degradation threshold is set at state 4, a PM alarm may be triggered prematurely from state 2 with probability P_I, correctly activated at state 4 with probability $1-P_{II}$, or missed at state 4 with probability P_{II}. Once the component degrades beyond the threshold, for instance to state 6, the PM-triggering transition becomes enabled with probability one.

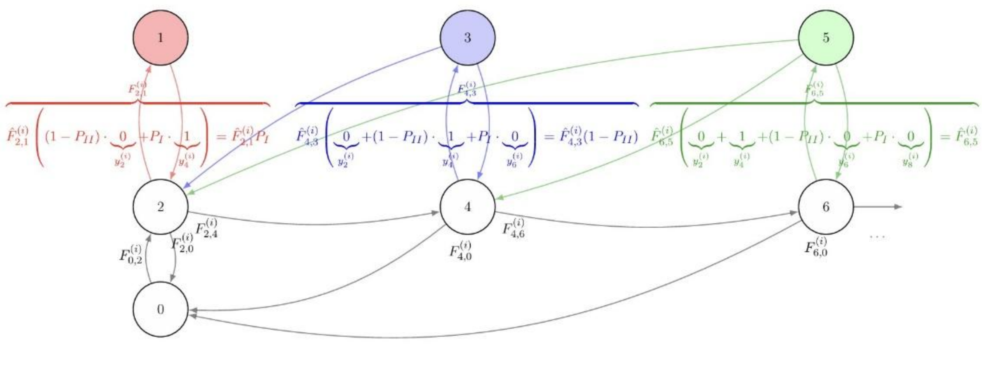


Figure 2. Illustration of PM-triggering transitions under imperfect monitoring.

Competing transitions from an operating state j are modeled through the semi-Markov kernel:

$$Q^{(i)}_{j,k}(t) = \int_0^t \prod_{m\in C_j^{(i)}\setminus\{k\}} \overline{F}^{(i)}_{j,m}(u)\, \mathrm{d}F^{(i)}_{j,k}(u) \qquad (4)$$

where $C_j^{(i)}$ denotes the set of states reachable from j, $F^{(i)}_{j,k}(t)$ is the CDF of the transition time associated with the transition $j \to k$, and $\overline{F}^{(i)}_{j,k}(t) = 1 - F^{(i)}_{j,k}(t)$ denotes the corresponding survival function. Equation (4) follows the competing risks formulation: the transition from j to k occurs at time u provided that none of the other admissible transitions from state j occurs before u. For PM-triggering transitions, the distributions $F^{(i)}_{j,k}(t)$ incorporated into equation (4) are modified according to equation (3). Thus, for a transition from an operating state j to its associated PM state, the corresponding semi-Markov kernel element is obtained as the probability that the PM-triggering event occurs at time t, while none of the other admissible transitions from state j occurs before t. For example, in the configuration illustrated in Fig. 2, where state 4 is selected as the degradation threshold, the transition $4 \to 3$ is governed by $F^{(i)}_{4,3}(t) = (1-P_{II})\hat{F}^{(i)}_{4,3}(t)$ and the corresponding semi-Markov kernel element is obtained directly from the competing-risks formulation as $Q^{(i)}_{4,3}(t) = \int_0^t \overline{F}^{(i)}_{4,0}(u)\, \overline{F}^{(i)}_{4,6}(u)\, d[(1-P_{II})\hat{F}^{(i)}_{4,3}(u)]$

Imperfect maintenance is modeled through probabilistic restoration levels combined with conditionally distributed maintenance durations. When the component is in a PM state $j \in S_{PM}^{(i)}$, the next state k must belong to the set of operable states $\mathcal{C}_j^{(i)} \subseteq S_{OP}^{(i)}$. In this case, the semi-Markov kernel simplifies to:

$$Q_{j,k}^{(i)}(t) = P_{j,k}^{(i)} F_{j,k}^{(i)}(t) = P_{j,k}^{(i)} \hat{F}_{j,k}^{(i)}(t) \sum_{x \in \mathcal{C}_j^{(i)} \cup \{j+3\}} y_x^{(i)} \quad (5)$$

where $\hat{F}_{j,k}^{(i)}(t)$ denotes the maintenance duration distribution and $P_{j,k}^{(i)}$ represents the probability that the maintenance action restores the component to operating state k, with $\sum_{k \in \mathcal{C}_j^{(i)}} P_{j,k}^{(i)} = 1$. The summation over $y_x^{(i)}$ includes the operating states from which the component can enter PM state j; the term $y_{j+3}^{(i)}$ captures cases where maintenance is triggered by an early alarm.

Under the adopted state indexing, a transition to $k = 2$ corresponds to a perfect repair, restoring the component to the best operating condition. Transitions to $2 < k < j$ correspond to imperfect repairs, while a transition to $k = j + 1$ represents a minor repair. Since operable states are indexed in increasing order of degradation, larger indices correspond to worse operating conditions, thereby defining a natural hierarchy of post-maintenance performance levels.

The restoration probabilities are parameterized through a single PM effectiveness parameter $q_{PM}^{(i)} \in (0,1)$, such that

$$P_{j,k}^{(i)} = \frac{\left(q_{PM}^{(i)}\right)^{k/2-1}}{\sum_{\ell=0}^{|\mathcal{C}_j^{(i)}|-1} \left(q_{PM}^{(i)}\right)^{\ell}} \quad (6)$$

which corresponds to a truncated geometric distribution over the feasible restoration levels. Using the geometric-series identity, this expression admits the closed form:

$$P_{j,k}^{(i)} = \frac{\left(1-q_{PM}^{(i)}\right)\left(q_{PM}^{(i)}\right)^{k/2-1}}{1-\left(q_{PM}^{(i)}\right)^{|\mathcal{C}_j^{(i)}|}} \quad (7)$$

Once maintenance is completed, the process transitions to the selected operating state k. The semi-Markov kernel elements can therefore be written according to the type of state j as:

$$Q_{j,k}^{(i)}(t) = \begin{cases} \int_0^t \prod_{m \in \mathcal{C}_j^{(i)} \setminus \{k\}} \overline{F}_{j,m}^{(i)}(u) \, \mathrm{d}F_{j,k}^{(i)}(u), \; j \in S_{OP}^{(i)} \cup S_{CM}^{(i)} \\ P_{j,k}^{(i)} \cdot F_{j,k}^{(i)}(t), \; j \in S_{PM}^{(i)} \end{cases} \quad (8)$$

The general formulation naturally covers the cases where the number of admissible transitions from state j is zero or one. When $|\mathcal{C}_j^{(i)}| = 0$, no further transitions are possible from state j, so the corresponding kernel elements are zero. When $|\mathcal{C}_j^{(i)}| = 1$, there is a single possible transition, and the competing-risks expression reduces to $Q_{j,k}^{(i)}(t) = F_{j,k}^{(i)}(t)$.

Remark 1. Although the proposed framework is developed for heterogeneous repairable multi-state components subject to condition-based maintenance, it also includes non-repairable components as a special case. This can be achieved by excluding the preventive-maintenance states $S_{PM}^{(i)}$ and the associated degradation-threshold variables $y_j^{(i)}$ for the corresponding component type. The component state space then contains only operating degradation states and the failure/replacement state ($S_{OP}^{(i)} \cup S_{CM}^{(i)}$). Under this interpretation, a transition to the failure state $S_{CM}^{(i)}$ represents the loss of the component, followed by replacement by a new unit when replacement is allowed. Therefore, the repairable setting considered in this work provides a more general modeling structure, while non-repairable behavior can be recovered through a simplified state-space definition.

Based on this semi-Markov formulation, the steady-state availability of a component is analytically evaluated using a two-stage approach [41]. First, the one-step transition probability matrix of the embedded Markov chain (EMC) associated with the SMP is constructed to compute the steady-state probabilities of the EMC. Subsequently, the sojourn time in each state is determined, and the steady-state probabilities of the SMP are obtained by combining the EMC steady-state distribution with the corresponding state sojourn times.

3.1.1. Stage 1: Steady-state EMC probabilities

The transition probability matrix for the EMC is $\boldsymbol{P}^{(i)}$. For each row of this matrix, it must satisfy $\sum_{k=1}^{M^{(i)}} p_{j,k}^{(i)} = 1, \forall j = 0, \ldots, M^{(i)}$. The one-step transition probability matrix is given as:

$$\boldsymbol{P}^{(i)} = \lim_{t \to \infty} \boldsymbol{Q}^{(i)}(t) = \begin{bmatrix} p_{0,0}^{(i)} & p_{0,1}^{(i)} & \cdots & p_{0,M^{(i)}}^{(i)} \\ p_{1,0}^{(i)} & p_{1,1}^{(i)} & \cdots & p_{1,M^{(i)}}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M^{(i)},0}^{(i)} & p_{M^{(i)},1}^{(i)} & \cdots & p_{M^{(i)},M^{(i)}}^{(i)} \end{bmatrix} \quad (9)$$

The following system of linear equations is formulated to find the steady-state probabilities vector of the EMC:

$$\vec{v}^{(i)} = \vec{v}^{(i)} \boldsymbol{P}^{(i)}, \sum_{j=1}^{M^{(i)}} v_j^{(i)} = 1, j \in S^{(i)} \quad (10)$$

3.1.2. Stage 2: Steady-state SMP probabilities

At this stage, the mean sojourn time, i.e., the time the process spends in each state, is determined to obtain the steady-state probability vector $\vec{\pi}^{(i)}$. The mean sojourn time is denoted by $\vec{\tau}^{(i)} = [\tau_0^{(i)}, \tau_1^{(i)}, \ldots, \tau_{M^{(i)}}^{(i)}]^T$, where each element is defined according to the structure of the outgoing transitions from the corresponding state. In particular, operational and failed states involve competing transition mechanisms, while PM states are governed by probabilistic restoration outcomes and non-competing maintenance durations, yielding a weighted expected sojourn time.

$$\tau_j^{(i)} = \begin{cases} \int_0^\infty \prod_{m\in C_j^{(i)}} \overline{F}_{j,m}^{(i)}(t)\,\mathrm{d}t, \; j \in S_{OP}^{(i)} \cup S_{CM}^{(i)} \\ \sum_{k\in C_j^{(i)}} P_{j,k}^{(i)} \cdot \int_0^\infty \overline{F}_{j,k}^{(i)}(t)\mathrm{d}t, \; j \in S_{PM}^{(i)} \end{cases} \tag{11}$$

Using the values of the steady-state probabilities of the EMC and the residence times in each state, the steady-state probability of state j for the SMP model is obtained as:

$$\pi_j^{(i)} = \frac{v_j^{(i)}\tau_j^{(i)}}{\sum_{k=0}^{M^{(i)}} v_k^{(i)}\tau_k^{(i)}}, j \in S^{(i)} \tag{12}$$

that is, the ratio of the time spent in state j to the total time spent.

Once the stationary distribution is obtained, the entrance rate to state j is defined as $\Lambda_j^{(i)}$, combining the stationary probability with the inverse of the mean sojourn time. This reflects both the proportion of time spent in the state and the average duration of each visit.

$$\Lambda_j^{(i)} = \frac{\pi_j^{(i)}}{\tau_j^{(i)}}, \quad j \in S^{(i)} \tag{13}$$

3.2. Multi-state system modeling

Among the main methodologies for evaluating MSS availability are the UGF technique and RA [3,42,43]. UGF, introduced by Ushakov [44], reduces the computational complexity of MSS by modeling discrete random variables. Its widespread use in reliability and availability studies stems from its ability to simplify modeling, accelerate computation, and ease numerical implementation [45].

3.2.1. UGF Method

The *u-function* of a component of type i relates the probability of each state to its respective performance level, and is defined as:

$$u^{(i)}(z) = \sum_{j=0}^{M^{(i)}} \pi_j^{(i)} z^{g_j^{(i)}} \tag{14}$$

The UGF associated with the MSS output performance distribution [46] is:

$$\begin{aligned} U(z) &= \Omega(u^{(1)}(z), u^{(2)}(z), \ldots, u^{(N)}(z)) \\ &= \sum_{j_1=0}^{M^{(1)}} \sum_{j_2=0}^{M^{(2)}} \cdots \sum_{j_N=0}^{M^{(N)}} \left(\prod_{i=1}^{N} \pi_{j_i}^{(i)} \, z^{\varphi(g_{j_1}^{(1)}, g_{j_2}^{(2)}, \ldots, g_{j_N}^{(N)})} \right) \end{aligned} \tag{15}$$

Where Ω is a composition operator and φ is the system structure function. Since the system under consideration is in a weighted k-out-of-n redundancy logical configuration, this is determined as:

$$\varphi(g_{j_1}^{(1)}, g_{j_2}^{(2)}, \ldots, g_{j_N}^{(N)}) = \sum_{i=1}^{N} g_{j_i}^{(i)} \tag{16}$$

Thus, the resulting system UGF becomes:

$$U(z) = \prod_{i\in I} (u^{(i)}(z))^{n_i} \tag{17}$$

Where n_i is the number of components of type i.

In a weighted k-out-of-n system, its availability is defined as the probability that its performance level meets the required demand [7], and is expressed by the following formulation:

$$A_{system} = P\left(\varphi\left(g_{j_1}^{(1)}, g_{j_2}^{(2)}, \ldots, g_{j_N}^{(N)}\right) \geq W\right) = \sum_{s\in S_{system}} \pi_s \cdot I_{G_s} \tag{18}$$

Where W is the required demand and I_{G_s} is an indicator function defined as:

$$I_{G_s} = \begin{cases} 1, & G_s \geq W \\ 0, & G_s < W \end{cases}, \forall s \in S_{system} \tag{19}$$

Remark 2. The weighted k-out-of-n formulation adopted in this study is used because it naturally represents heterogeneous multi-state systems in which different component types may contribute different performance levels. Classical non-weighted binary and multi-state k-out-of-n systems are obtained as special cases by assigning identical capacities to components in equivalent operating states $g_j^{(i)} = 1$, $\forall i \in I$, $\forall j \in S_{OP}^{(i)}$. In that case, the system performance reduces to the number of functioning components satisfying the demand threshold. Thus, the weighted formulation does not restrict the applicability of the model; rather, it provides a more general representation that encompasses standard k-out-of-n structures.

4. Optimization problem

This section addresses the optimization of redundancy allocation and CBM policy design for MSS. The optimization model is first formulated, followed by a detailed description of the adapted metaheuristic developed to solve the problem.

4.1. Optimization model

The following joint Redundancy Allocation and Condition-Based Maintenance (RAP–CBM) optimization model is formulated.

$$\min \left\{CT_{inef} + CT_{rep} + CT_{op} + CT_{adq}\right\} \tag{20}$$

$$\text{s.t. } A_{system}\left(\{n_i, \vec{y}^{(i)}\}_{i\in I}\right) \geq A_{req} \tag{21}$$

$$\sum_{j\in S_{OP}^{(i)}} y_j^{(i)} \leq 1, \quad \forall i \in I \tag{22}$$

$$n_i \leq N_i, \quad \forall i \in I \tag{23}$$

$$n_i \geq 0, integers, \quad \forall i \in I \tag{24}$$

$$y_j^{(i)} \in \{0,1\}, \quad \forall i \in I, \forall j \in S_{OP}^{(i)} \tag{25}$$

whose LCC-based costs are broken down into:

- Total inefficiency costs:

$$CT_{inef} = c_{inef} \cdot \left(1 - A_{system}\right) \cdot H \cdot f_{up} \tag{26}$$

- Total costs of maintenance actions:

$$CT_{rep} = \sum_{i\in I} \sum_{j\in S_{PM}^{(i)} \cup S_{CM}^{(i)}} \left(c_{i,j}^{rep} \cdot \pi_j^{(i)}(\vec{y}^{(i)}) + c_{i,j}^{fix} \cdot \Lambda_j^{(i)}(\vec{y}^{(i)})\right) \cdot H \cdot f_{up} \cdot n_i \tag{27}$$

- Total operating costs:

$$CT_{op} = \sum_{i\in I} \sum_{j\in S_{OP}^{(i)}} c_{i,j}^{op} \cdot \pi_j^{(i)}(\vec{y}^{(i)}) \cdot H \cdot f_{up} \cdot n_i \tag{28}$$

- Total investment costs (equipment acquisition):

$$CT_{adq} = \sum_{i\in I} c_i^{adq} \cdot n_i \tag{29}$$

The system availability is obtained from equation (18) and implicitly captures the realistic operating conditions described in Section 3 through the stationary probabilities derived from the SMP formulation. Redundancy allocation and degradation-threshold decisions affect different levels of this availability evaluation. The redundancy variables determine the number and type of components entering the UGF formulation and, therefore, modify the system-level performance distribution. In contrast, the degradation-threshold variables determine when PM is triggered in the SMP model and, therefore, modify the stationary probabilities of operating, degraded, PM, and CM states for each component type. Since system availability is computed from both the component stationary probabilities and the redundancy structure, both sets of decisions jointly determine feasibility with respect to the availability requirement.

This coupling also affects the LCC objective. Redundancy decisions directly influence acquisition and operating costs, while threshold decisions influence maintenance, repair, and inefficiency costs through the frequency and duration of PM and CM states. As a result, higher redundancy may allow less conservative maintenance thresholds, whereas lower redundancy may require earlier preventive interventions to satisfy the availability constraint. This interaction motivates the joint optimization of redundancy allocation and CBM thresholds rather than treating them as independent or sequential decisions.

The variable $\pi_j^{(i)}(\vec{y}^{(i)})$ represents the probability that $\pi_j^{(i)}(\vec{y}^{(i)})$ i is in state j, as defined by the SMP. Additionally, $\vec{y}^{(i)}$ denotes the vector of variables $y_j^{(i)}$ associated with component i.

The decision variables in this model are the number of components of type i assigned to the system, denoted as n_i, and the binary variable $y_j^{(i)}$. The latter takes the value 1 if the degradation threshold is set to state j for components of type i, and 0 otherwise, where $j \in S_{OP}^{(i)}$ represents the operating states. When $\sum_{j\in S_{OP}^{(i)}} y_j^{(i)} = 0$, the component operates under an RTF policy. Regarding the constraints, equation (21) defines the availability requirements, equation (22) ensures that only one degradation threshold is assigned to components of type i, equation (23) limits the maximum number of components assignable for each type, and equations (24) and (25) describe the nature of the variables.

Remark 3. The model can be adapted to an operational planning context in which the system is already implemented, simply eliminate the acquisition cost CT_{adq} and subtract from the total cost the salvage cost CT_{sale}, which considers the resale or dismantling of unused components.

$$CT_{sale} = \sum_{i\in I} c_i^{sale} \left(N_i - n_i\right) \tag{30}$$

where c_i^{sale} denotes the unit salvage value of a component of type i, and N_i represents the number of components of type i initially available in the system. This modification shifts the optimization from system design to operational reconfiguration, without altering the structure of the proposed model.

4.2. Optimization technique

For a system with N component types and $M^{(i)}/2$ operable states per type $i = 1, \dots, N$, the solution space is $\prod_{i=1}^{N} N_i \cdot (M^{(i)}/2 + 1)$. Even simple series RAP systems with linear constraints are NP-hard [47], making exhaustive search infeasible. Consequently, metaheuristic approaches are adopted to obtain solutions within reasonable timescales. Universal metaheuristic algorithms such as GA [15,48], Particle Swarm Optimization (PSO) [49], and Artificial Bee Colony (ABC) [50], among others [11,14], are widely employed to explore the solution space efficiently [51]. Among these approaches, GA-based methods have shown strong performance in reliability optimization problems involving discrete decision variables and heterogeneous redundancy configurations [49].

4.2.1. Implementation of genetic algorithm

In this work, GA was adopted due to its flexibility for handling discrete decision variables and nonlinear availability constraints required by the proposed RAP–CBM formulation. Inspired by biological genetics, each solution is encoded as a finite-length string (chromosome), and a population of chromosomes is evolved through crossover, mutation, and selection. An elitism strategy is incorporated to preserve the best solution in each generation. The algorithm is summarized in Fig. 3, with details outlined below.

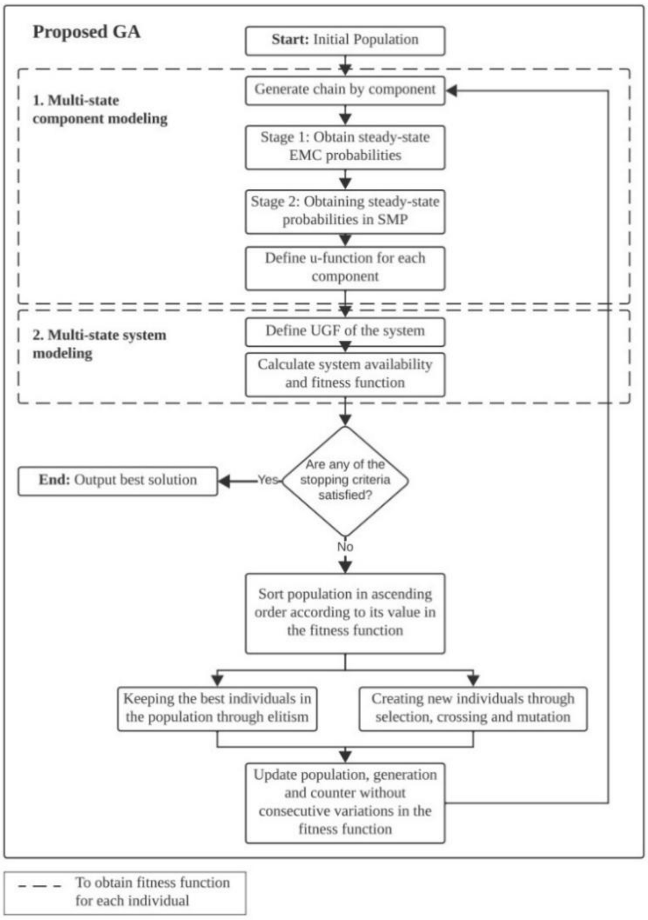


Figure 3. Proposed GA.

- **Coding solutions:** A chromosome is represented by an $N \times 2$ matrix, with one row per component type. The first column specifies the number of components of type i, and the second column identifies the state j where the binary variable $y_j^{(i)}$ equals 1. If this number is zero, all $y_j^{(i)}$ are zero, yielding an RTF policy for that component type. This structure provides precise representation within the algorithm.

$$\begin{array}{c c} & n_i \; j \text{ for } y_j^{(i)} \\ \begin{array}{c} \text{Component type 1} \\ \text{Component type 2} \\ \vdots \\ \text{Component type } N \end{array} & \begin{bmatrix} n_1 & j^{(1)} \\ n_2 & j^{(2)} \\ \vdots & \vdots \\ n_N & j^{(N)} \end{bmatrix} \end{array}$$

Figure 4. Chromosomal structure.

- **Initial population:** The maximum number of components of type i assignable to the system is n_i. Accordingly, the first entry in each row must be an integer in $[0, n_i]$, while the second entry must belong to $S_{OP}^{(i)} \cup \{0\}$.
- **Fitness function:** Since the proposed optimization model includes a constraint on the system's availability, a penalty is incorporated into the *fitness* function f of the adopted GA. Two parameter-free penalization schemes were evaluated: the first proposed in this work and the second adapted from [15]. The fitness function is defined as:

$$f = \left(CT_{inef} + CT_{rep} + CT_{op} + CT_{adq}\right) \cdot Penalty \quad (31)$$

with the following penalty formulations:

$$Penalty_1 = \left(1 + \frac{\max\{0, A_{req} - A_{system}\}}{1 - A_{req}}\right) \quad (32)$$

$$Penalty_2 = \left(\frac{1}{1 - \max\{0, A_{req} - A_{system}\}}\right) \quad (33)$$

Unlike conventional penalty methods that require user-defined coefficients, both penalization schemes considered here are parameter-free because they depend only on the availability violation. The proposed formulation further normalizes this violation with respect to the admissible unavailability margin $\left(1 - A_{req}\right)$. As a result, the penalty increases according to the relative severity of the violation: deviations that represent a small fraction of the admissible unavailability margin are penalized mildly, whereas deviations that substantially compromise the required availability receive stronger penalization. This structure avoids empirical calibration of penalty weights and makes the fitness function directly comparable across different availability requirements.

- **Chromosome selection:** To select the chromosomes needed for crossover and mutation operations, they are randomly sampled in groups of 3 and the one with the lowest f value is chosen.

- **Crossover operator:** A one-point crossover is applied. Two parent chromosomes are selected, and a random breakpoint is defined by uniform variables over [0,2] (column) and $[0, N]$ (row). Segments before and after the breakpoint are exchanged, producing two offspring with combined information. Fig. 5 illustrates the case with breakpoint (2,1).

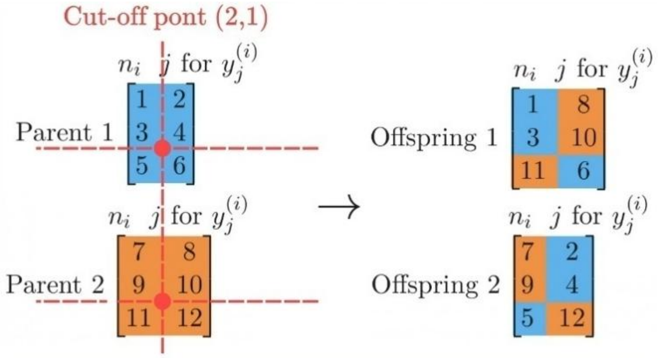


Figure 5. Single-point crossover operator.

- **Mutation operator:** To avoid premature convergence, mutation is introduced using a mask matrix with random values in $[0,1]$. Each gene with a mask value $\leq \mu_m$ mutates, generating a new value. Fig. 6 shows an example.

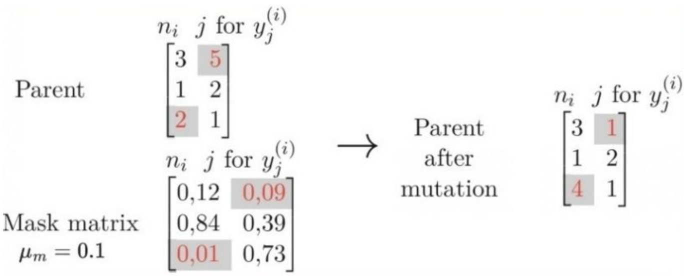


Figure 6. Mutation operator.

- **Elitism:** To prevent fitness degradation, the best solutions are passed directly to the next generation. The population is sorted by f values in ascending order, and the top $P_{elitism}$ percentile is retained.
- **Next generation:** The new population is formed as follows:
  - The top $P_{elitism}$ percentile of the current generation is preserved.
  - The remaining individuals are generated via crossover and mutation from the previous population.
- **Stopping criterion:** The algorithm halts after a fixed number of iterations or when the best solution shows no improvement over a given number of consecutive generations.

Remark 4. Although this work adopts a GA implementation, the proposed matrix-based solution encoding and penalized fitness formulation can also be adapted to a PSO framework. In this case, particle velocities may be updated as:

$$V^{l+1} = \omega V^l + c_1 r_1 \left(X^{\boldsymbol{best}} - X^l\right) + c_2 r_2 \left(GL^{\boldsymbol{best}} - X^l\right) \quad (34)$$

where X^l and V^l denote the particle position and velocity matrices at iteration l, ω is the inertia coefficient, c_1 and c_2 are the cognitive and social acceleration parameters, $r_1, r_1 \in U(0,1)$ are random coefficients, $X^{\boldsymbol{best}}$ represents the best position previously visited by the particle, and $GL^{\boldsymbol{best}}$ is the best global solution identified by the swarm. Since the RAP–CBM formulation involves discrete decision variables, the updated positions must be projected onto the feasible search space. For the redundancy variables, this can be expressed as

$$x_{n_i}^{l+1} = \max\{0, \min\{N_i, \mathrm{round}(x_{n_i}^l + v_{n_i}^{l+1})\}\} \quad (35)$$

while the degradation thresholds are updated by selecting the admissible operating state closest to the continuous PSO update:

$$x_{j_i}^{l+1} = \underset{j \in S_{OP}^{(i)}}{arg\ min} \left| j - \left(round\left(x_{j_i}^l + v_{j_i}^{l+1}\right)\right) \right| \quad (36)$$

thereby ensuring that threshold values remain restricted to feasible operating states.

5. Numerical examples

An illustrative power generation system is analyzed to demonstrate the applicability of the proposed framework under realistic operating and degradation conditions. Although the numerical data do not correspond to a specific industrial facility, they were constructed as a representative benchmark based on engineering judgment and publicly available information on power-generation systems. The objective is to provide a realistic heterogeneous multi-state environment for evaluating the proposed RAP–CBM framework, comparing alternative decision-making schemes, and conducting sensitivity analyses.

Four generator types with capacities ranging from 5 to 40 MW are considered. These represent distinct technological profiles, ranging from small, low-cost units with faster degradation dynamics to larger generators offering higher capacity and partial economies of scale, but exhibiting more demanding maintenance and aging characteristics. Acquisition, operation, and maintenance costs increase with capacity at different rates, creating nontrivial trade-offs between installing a few large units or several smaller ones. Consequently, no generator type is uniformly superior across all cost and performance dimensions, making the redundancy-allocation decision nontrivial. Table 2 summarizes the capacities and cost parameters of each generator type.

Each component is modeled using nine states: State 2 (fully functional), States 4, 6, and 8 (degraded with 5%, 10%, and 20% capacity loss, respectively), States 1, 3, 5, and 7 (stopped for PM), and State 0 (stopped for CM). Weibull distributions with shape parameters greater than one are used to represent progressive degradation, exponential distributions to model random failures, and lognormal distributions to capture the high variability typically associated with CM durations. The resulting transition curves are shown in Fig. 8 based on the state-transition structure presented in Fig. 7.

PM durations $tp_{j,k}^{(i)}$, $ti_{j,k}^{(i)}$, $tm_{j,k}^{(i)}$ are represented by Heaviside functions for perfect, imperfect, and minor maintenance actions, with probabilities $P_{j,k}^{(i)}$. Larger generators are assumed to have lower probabilities of complete restoration following PM actions, whereas maintenance durations increase with both equipment size and degradation severity. Table 3 summarizes the corresponding PM parameters.

Monitoring detects degradation with $(1 - P_{II}) = 85\%$ accuracy, $P_I = 10\%$ false alarms, and a fixed delay $t_{delay} = 2$ h, while a minimum system availability of 95% is imposed. Both the monitoring-error probabilities and the availability requirement are within ranges commonly adopted in the literature. The demand is $W = 50$ MW, the inefficiency cost is 1,000 \$/h, the planning horizon is $H = 3$ years under continuous operation, and the annual discount rate is 10%. Each generator type can be purchased up to four units.

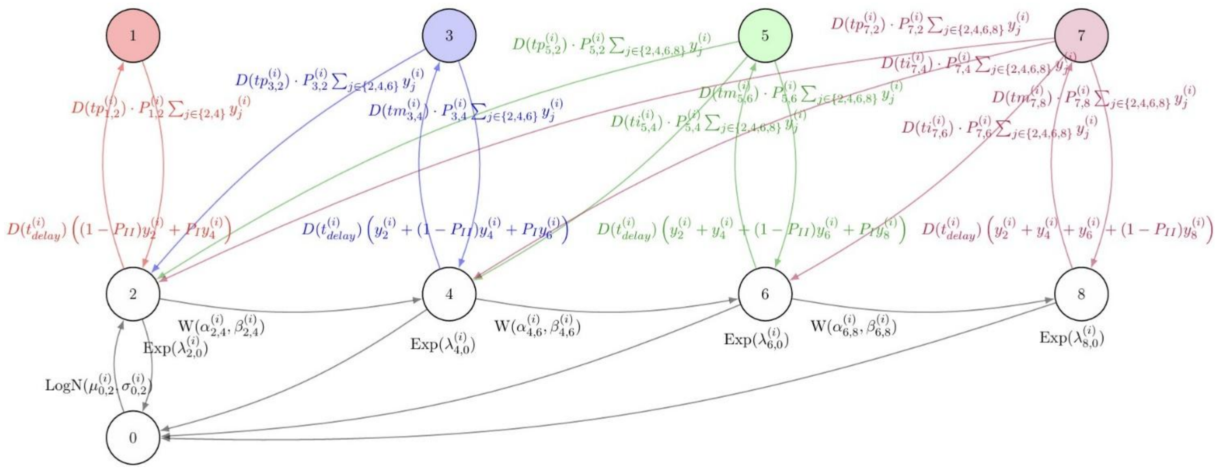
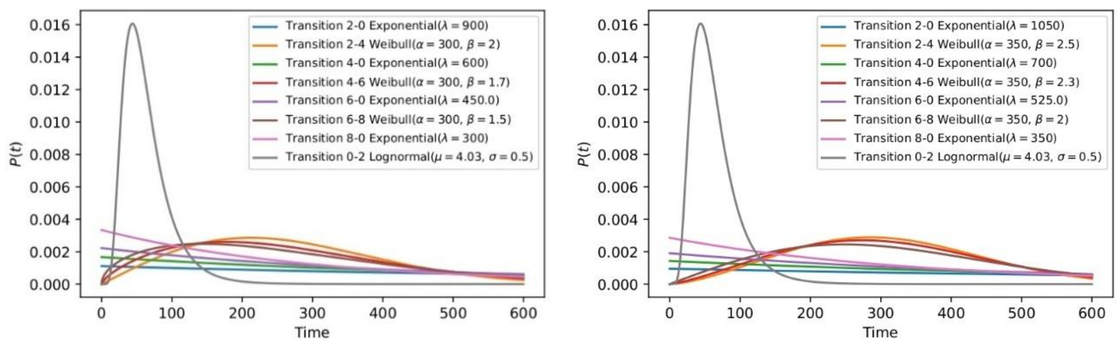


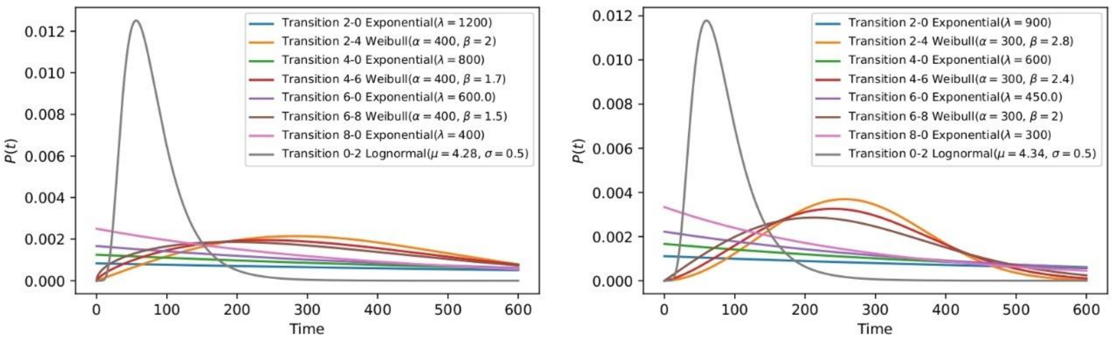
Figure 7. State diagram by generator.

Table 2. Data on capacity and unit acquisition, operation and labor costs by generator type.

Generator type	Capacity MW	Acquisition costo	Operating cost \$/hour	Labor cost \$/hour	Cost per PM event	Cost per CM event
1	5	\$100,800	12.5	25	\$1,000	\$12,000
2	10	\$140,000	20	35	\$2,500	\$30,000
3	20	\$420,000	50	60	\$3,000	\$36,000
4	40	\$680,000	65	100	\$5,000	\$60,000



(a) Generator type 1. **(b)** Generator type 2.



(c) Generator type 3. **(d)** Generator type 4.

Figure 8. Probability density functions for failure, degradation, and repair by generator type.

Table 3. PM parameters by generator type and operating condition ($q_{PM}^{(1)} = 0.45$, $q_{PM}^{(2)} = 0.50$, $q_{PM}^{(3)} = 0.54$, $q_{PM}^{(4)} = 0.58$).

Component	Transition	Time hours	Probability
1	1 → 2	2.2	1
	3 → 2	3.08	0.69
	3 → 4	4.4	0.31
	5 → 2	3.52	0.61
	5 → 4	4.4	0.27
	5 → 6	5.72	0.12
	7 → 2	3.96	0.57
	7 → 4	4.84	0.26
	7 → 6	6.16	0.12
	7 → 8	6.6	0.05
2	1 → 2	2.2	1
	3 → 2	3.08	0.67
	3 → 4	4.4	0.33
	5 → 2	3.52	0.57
	5 → 4	4.84	0.29
	5 → 6	5.72	0.14
	7 → 2	4.4	0.53
	7 → 4	5.28	0.27
	7 → 6	6.16	0.13
	7 → 8	7.92	0.07
3	1 → 2	2.8	1
	3 → 2	3.92	0.65
	3 → 4	5.6	0.35
	5 → 2	4.48	0.55
	5 → 4	5.6	0.29
	5 → 6	7.28	0.16
	7 → 2	5.04	0.5
	7 → 4	6.16	0.27
	7 → 6	7.28	0.15
	7 → 8	8.4	0.08
4	1 → 2	9	1
	3 → 2	12	0.63
	3 → 4	7.2	0.37
	5 → 2	6	0.52
	5 → 4	7.2	0.3
	5 → 6	7.8	0.18
	7 → 2	6	0.47
	7 → 4	7.2	0.27
	7 → 6	9	0.16
	7 → 8	10.8	0.09

5.1. MCS model

Simulation methods, particularly Monte Carlo approaches, are crucial for modeling complex systems where traditional analytical methods fall short. Widely used in reliability analysis, risk assessment, and performance evaluation, MCS generates random samples from probability distributions to evaluate key system parameters through iterative simulations [11,52].

To validate the proposed approach, its availability estimates were compared against those obtained via MCS under the modeling assumptions adopted in this work. The comparison was conducted for a system configuration $(n_1 = 1, \sum_{j \in S_{OP}^{(1)}} y_j^{(1)} = 0, n_3 = 2, y_6^{(3)} = 1, n_4 = 1, y_4^{(4)} = 1)$. Over a three-year horizon, the proposed method yields a system availability of 86.463%, while MCS produces 86.496% based on 100,000 simulation runs, resulting in a relative difference below 0.05%. Although a substantial reduction in computation time is observed (99.1%), this result is expected given the analytical nature of the proposed formulation and is reported solely to support its practical viability.

5.2. Hyperparameters and analysis of fitness function penalization

The stagnation limit was evaluated using a population size of 50, a maximum of 300 generations, and 15% elitism and mutation rates. Stall values of 50, 75, and 100 generations were tested over 10 independent runs. As shown in Table 4, larger stagnation limits produced negligible differences in average LCC, but substantially increased the average number of generations required for convergence. In most runs, a stagnation limit of 50 reached the same high-quality solutions with lower computational effort. Since no run exceeded 100 generations under this limit, the final GA configuration adopted a stagnation limit of 50 and a maximum of 100 generations.

Table 4. Effect of stagnation limit on GA convergence and solution quality.

Stagnation limit	Average LCC	Average generations	Δ Average LCC	Δ Average generations
50	27,486,247.90	63.0	-	-
75	27,480,359.63	86.7	0.021423%	37.46%
100	27,480,359.63	123.6	0.021423%	96.19%

The penalty formulations defined in (32) and (33) were evaluated within the fitness function given by (31). For moderate availability requirements, both penalty schemes yield identical solutions. However, for $A_{req} \geq 0.97$, the traditional penalty converges to infeasible solutions, as the fitness remains more favorable than the additional redundancy or maintenance cost required to satisfy the availability constraint. In contrast, the proposed scheme penalizes deviations proportionally to the availability shortfall, guiding the GA toward feasible or near-feasible solutions and preventing convergence to infeasible local optima.

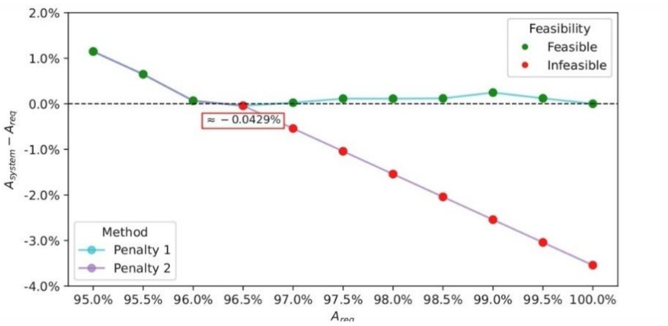


Figure 9. Comparison of feasibility of GA solutions by penalty method (Evaluated to A_req=99.999%).

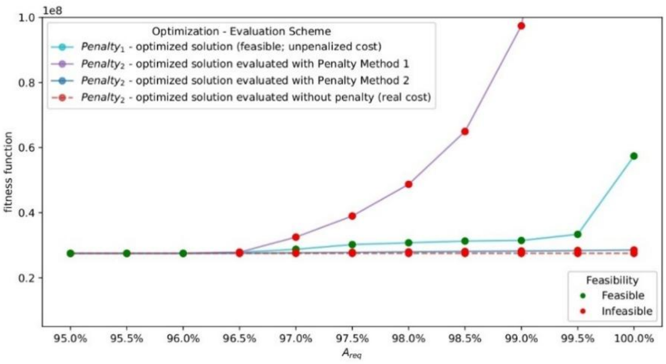


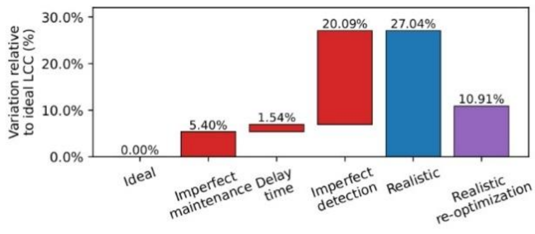
Figure 10. Effect of penalty methods on GA optimization and cost evaluation.

5.3. Optimization model results

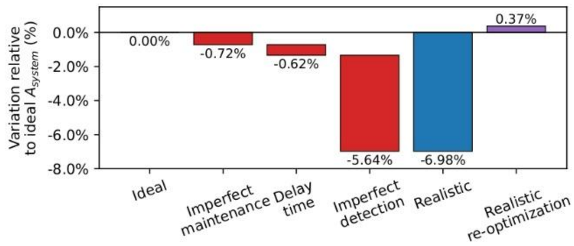
The best solution obtained from the adapted GA, when decoded, yields the number of generators that should be assigned to the system and the degradation thresholds. The optimal configuration includes four type 3 generators, which are preventively intervened when the continuous monitoring system detects that they are operating at 90% of their capacity (state 6). The system has an availability of 96.15%, and its LCC amounts to $27,477,415. The algorithm took 8 minutes 59 seconds to find the solution, which was reached at generation 64, as the solution remained unchanged for 50 consecutive iterations.

5.4. Impact of realistic conditions and proactive policy

The model is first optimized under idealized assumptions, namely perfect maintenance, error-free condition monitoring, and instantaneous maintenance execution, yielding a baseline solution denoted as X. This solution is subsequently evaluated under progressively more realistic assumptions, while keeping the decision variables fixed. Fig. 11 illustrates the resulting variations in LCC and system availability using a waterfall chart, highlighting the marginal contribution of each realistic factor. The results show that detection errors have the dominant impact on cost and availability degradation, whereas maintenance delays produce a comparatively limited effect.



(a) Cumulative impact of non-ideal operational factors on LCC.



(b) Cumulative impact of non-ideal operational factors on system availability.

Figure 11. Impact of non-ideal operational conditions on system performance: (a) LCC and (b) system availability.

When the model is optimized directly under realistic conditions, the optimal solution achieves 96.07% availability with an LCC of $27,506,857. In contrast, the solution obtained under ideal assumptions becomes infeasible when realistically evaluated, with availability dropping to 89.11% and LCC increasing to \$31,473,056, thus violating the 95% requirement. Explicitly incorporating operational imperfections during optimization yields a 12.7% cost reduction, highlighting the economic relevance of the proposed framework.

To highlight the economic benefits of proposed CBM policy, its performance was compared with an RTF strategy. The optimal RTF configuration yields an LCC of \$32,239,300, whereas the CBM policy reduces costs by 14.77%. This advantage becomes even more pronounced under conditions of higher inefficiency costs or longer repair durations.

5.5. Sequential vs joint optimization

The proposed joint RAP–CBM framework was compared with a sequential strategy in which redundancy allocation was first optimized assuming a RTF policy, and the degradation thresholds were subsequently adjusted with fixed redundancy. This order reflects the traditional distinction between system design and operational decision stages. The sequential optimization was solved through exhaustive enumeration, ensuring identification of the best feasible solution at each stage.

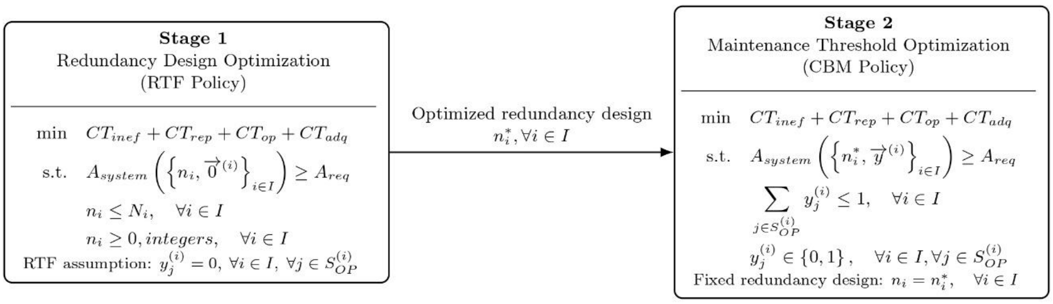


Figure 12. Two-stage sequential optimization strategy adopted for comparison with the proposed joint RAP–CBM framework.

The sequential approach resulted in a configuration with one component of type 3 and two components of type 4, where both component types were preventively intervened when operating at 90% of their nominal capacity (state 6). The obtained LCC was \$29,398,361, representing an increase of approximately 7% compared with the proposed joint optimization framework, thereby empirically illustrating the coupling between redundancy allocation and maintenance-threshold decisions for the studied instance.

5.6. Sensitivity analysis

A one-dimensional sensitivity analysis was conducted to evaluate the impact of key parameters. Figs. 13 and 14 show the percentage variation in LCC relative to the base case. The average CM duration and required demand exhibit the strongest influence on cost: longer repairs substantially increase LCC and shift degradation thresholds toward healthier states, while higher demand increases redundancy. In contrast, system inefficiency cost mainly scales the objective value without altering the optimal configuration.

The minimum availability requirement and component limits affect the solution only under restrictive settings. Regarding monitoring imperfections, eliminating Type I and Type II errors would reduce costs by 18.98% and 7.81%, respectively. Although sensitivity graphs suggest limited variation in LCC even for error probabilities of 40%, this behavior is largely driven by the parameterization of the base case, where differences in cost, duration of preventive actions, and performance across degradation states are relatively moderate. Detection errors primarily induce an implicit shift in the effective intervention threshold: higher Type I error probabilities anticipate interventions, whereas higher Type II probabilities delay them. Consequently, their economic impact depends on how strongly degraded-state operation is penalized. Maintenance delays have a relatively marginal impact on costs. The observed change in the LCC trend is explained by the fact that, for delays exceeding 5 h, the redundancy configuration remains unchanged while the optimal degradation thresholds shift toward earlier preventive interventions. Preventive maintenance effectiveness exhibits a similarly limited influence on costs.

The analysis indicates that shortening corrective intervention times—through improved diagnostics, spare part availability, and skilled teams—is the most effective lever for reducing LCC, while additional standby capacity helps stabilize availability during demand peaks. Detection errors primarily induce an implicit shift in the effective intervention threshold: higher Type I probabilities anticipate interventions, whereas higher Type II probabilities delay them. Their economic impact therefore depends on how strongly operation in degraded states is penalized within the cost structure. The proposed framework, along with the monitoring of key performance indicators such as repair time and available capacity, facilitates the dynamic adjustment of maintenance decisions as operating conditions evolve.

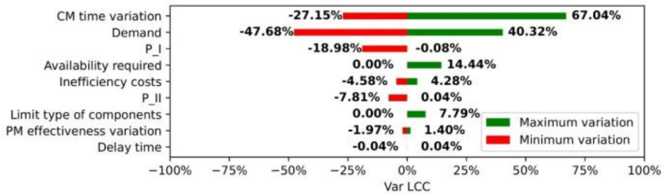
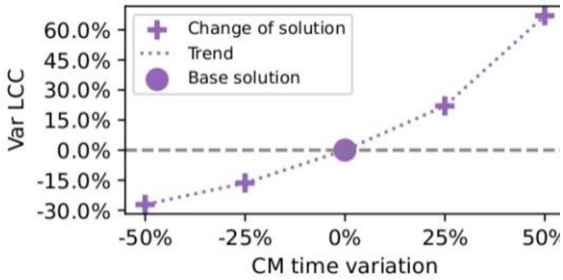
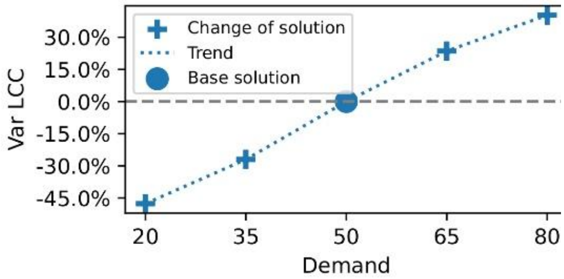


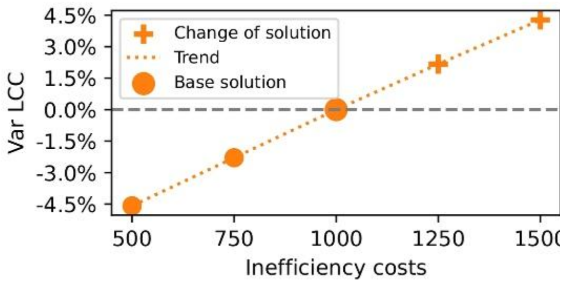
Figure 13. Percentage variation of the LCC by sensitized parameter.



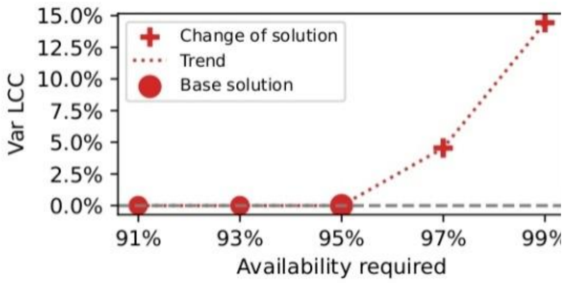
(a) CM time variation.



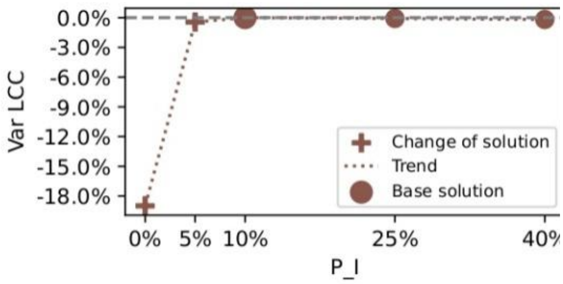
(b) Demand.



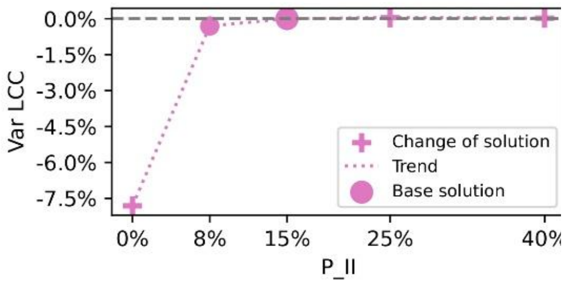
(c) Inefficiency cost.



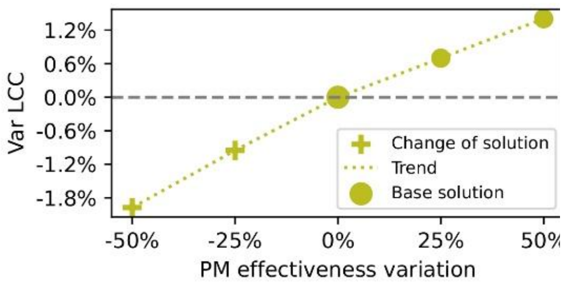
(d) Availability required.



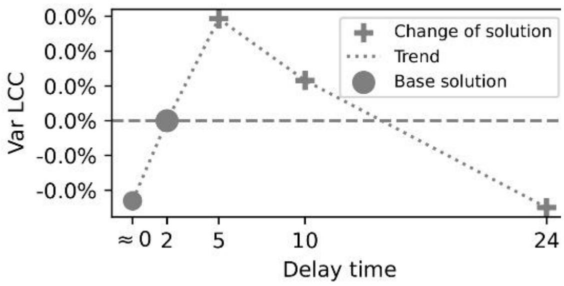
(e) Type I error probability.



(f) Type II error probability.



(g) PM effectiveness.



(h) Delay time.

Figure 14. Sensitivity of LCC to model parameters.

6. Conclusions and future work

This study addressed a joint RAP–CBM problem for weighted

k-out-of-n systems with repairable, non-identical components and multiple degradation states from an LCC perspective. The proposed framework integrates degradation-based intervention thresholds and explicitly incorporates realistic operational features, including monitoring errors, maintenance delays, non-instant repairs, and heterogeneous PM outcomes.

The combined SMP–UGF approach accurately captures the multi-state stochastic behavior of the system, matching MCS results with deviations below 0.05% while reducing computational time by approximately 99%. The resulting optimization problem was efficiently solved using an adapted GA, where the proposed availability penalty scheme proved essential for identifying feasible solutions under stringent constraints.

Results demonstrate the economic relevance of modeling operational imperfections. Detection errors exhibited the strongest influence on system performance, primarily by inducing implicit shifts in intervention thresholds, whereas maintenance delays had a comparatively minor impact. Optimization under idealized assumptions resulted in solutions that failed to satisfy availability requirements when assessed under realistic conditions, with a marked increase in LCC. Incorporating operational imperfections directly into the optimization process ensured feasibility and achieved substantial cost savings. Comparisons with a RTF policy showed that the proposed CBM strategy achieves an LCC reduction of approximately 15%. Sequential optimization yields about 7% higher LCC than the joint approach, highlighting the benefit of integrated optimization. Sensitivity analyses further confirmed that demand levels and corrective maintenance duration are the main cost drivers. True optimality requires the joint adaptation of degradation thresholds and redundancy to parameter variations, as adjusting only one dimension leads to suboptimal configurations.

Overall, the proposed framework provides a computationally efficient tool for jointly supporting design and maintenance planning decisions in heterogeneous multi-state systems under realistic operating conditions.

Future research may extend the model by incorporating stochastic optimization to account for demand fluctuations and improve robustness. CBM policies based on periodic or aperiodic inspections could also be explored as cost-effective alternatives when continuous monitoring is not feasible. Finally, the use of more advanced metaheuristics may further improve solution quality and computational efficiency.

Acknowledgment

The first author work was funded by the National Agency for Research and Development (ANID) / Scholarship Program / MAGISTER NACIONAL / 2024 - 22242071.

References

1. Huang S, Lei B, Gao K, Wu Z, Wang Z. Multi-state system reliability evaluation and component allocation optimization under multi-level performance sharing. *IEEE Access* 2021; 9: 88820-34. https://doi.org/10.1109/ACCESS.2021.3090405.
2. Eryilmaz S. Reliability analysis of multi-state system with three-state components and its application to wind energy. *Reliability Engineering & System Safety* 2018; 172: 58-63. https://doi.org/10.1016/J.RESS.2017.12.008.
3. Huang T, Xiahou T, Li Y-F, Qian H-M, Liu Y, Huang H-Z. Reliability assessment of wind turbine generators by fuzzy universal generating function. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2021; 23: 308-14. https://doi.org/10.17531/ein.2021.2.10.
4. Cui X, Li T, Wang S, Shi J, Ma Z. Reliability modeling based on power transfer efficiency and its application to aircraft actuation system. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2020; 22: 282-96. https://doi.org/10.17531/ein.2020.2.11.
5. Chen Z, Zhao T, Jiao J, Ren F. Availability analysis and optimal design of multistate weighted k -out-of- n systems with component performance requirements. *IEEE Access* 2018; 6: 51547-55. https://doi.org/10.1109/ACCESS.2018.2865933.
6. Li S, Yang Z, Fang C, Luo H, He J, Li J, et al. Optimal maintenance strategy for imperfect manufacturing system considering operational availability under delay propagation. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2026; 28(1): 207148. https://doi.org/10.17531/ein/207148.
7. Lisnianski A, Levitin G. Multi-state system reliability: Assessment, optimization and applications. World scientific; 2003.
8. Sharifi M, Taghipour S. Inspection interval optimization of a weighted-k-out-of-n system with identical multi-state load-sharing

components. *Reliability Engineering & System Safety* 2023; 238: 109412. https://doi.org/10.1016/J.RESS.2023.109412.
9. Tian Z, Levitin G, Zuo MJ. A joint reliability–redundancy optimization approach for multi-state series–parallel systems. *Reliability Engineering & System Safety* 2009; 94: 1568-76. https://doi.org/10.1016/J.RESS.2009.02.021.
10. Fyffe DE, Hines WW, Lee NK. System reliability allocation and a computational algorithm. *IEEE Transactions on Reliability* 1968; R-17: 64-9. https://doi.org/10.1109/TR.1968.5217517.
11. Attar A, Raissi S, Khalili-Damghani K. A simulation-based optimization approach for free distributed repairable multi-state availability-redundancy allocation problems. *Reliability Engineering & System Safety* 2017; 157: 177-91. https://doi.org/10.1016/J.RESS.2016.09.006.
12. Li XY, Li YF, Huang HZ. Redundancy allocation problem of phased-mission system with non-exponential components and mixed redundancy strategy. *Reliability Engineering & System Safety* 2020; 199: 106903. https://doi.org/10.1016/J.RESS.2020.106903.
13. Kim H. Markov-based reliability model for a mixed redundant system and parallel genetic algorithm with knowledge archives for a redundancy allocation problem. *Reliability Engineering & System Safety* 2023; 240: 109585. https://doi.org/10.1016/J.RESS.2023.109585.
14. Li J, Xue L, Wang G, Zhou H. Availability evaluation and design optimization of multi-state k-out-of-n: G systems with random performance requirements. *IEEE Access* 2023; 11: 119106-17. https://doi.org/10.1109/ACCESS.2023.3327431.
15. Sharifi M, Taghipour S. Redundancy allocation problem with a mix of components for a multi-state system and continuous performance level components. *Reliability Engineering & System Safety* 2024; 241: 109632. https://doi.org/10.1016/J.RESS.2023.109632.
16. Mobley RK. An introduction to predictive maintenance. Elsevier; 2002.
17. Guiras Z, Turki S, Rezg N, Dolgui A. Optimal maintenance plan for two-level assembly system and risk study of machine failure. *International Journal of Production Research* 2019; 57: 2446-63. https://doi.org/10.1080/00207543.2018.1521017.
18. Toshio N. Maintenance theory of reliability. Springer-Verlag; 2005. https://doi.org/10.1007/1-84628-221-7.
19. Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 2006; 20: 1483-510. https://doi.org/10.1016/J.YMSSP.2005.09.012.
20. Ahmad R, Kamaruddin S. An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering* 2012; 63: 135-49. https://doi.org/10.1016/J.CIE.2012.02.002.
21. Alaswad S, Xiang Y. A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety* 2017; 157: 54-63. https://doi.org/10.1016/J.RESS.2016.08.009.
22. Andersen JF, Nielsen BF. A comparative study of time-based maintenance and condition-based maintenance for multi-component systems. *Reliability Engineering & System Safety* 2025; 256: 110759. https://doi.org/10.1016/J.RESS.2024.110759.
23. Castro IT, Basten RJI, Houtum GJ van. Maintenance cost evaluation for heterogeneous complex systems under continuous monitoring. *Reliability Engineering & System Safety* 2020; 200: 106745. https://doi.org/10.1016/J.RESS.2019.106745.
24. Oakley JL, Wilson KJ, Philipson P. A condition-based maintenance policy for continuously monitored multi-component systems with economic and stochastic dependence. *Reliability Engineering & System Safety* 2022; 222: 108321. https://doi.org/10.1016/J.RESS.2022.108321.
25. Zhang F, Liao H, Shen J, Ma Y. Optimal maintenance of a system with multiple deteriorating components served by dedicated teams. *IEEE Transactions on Reliability* 2023; 72: 900-15. https://doi.org/10.1109/TR.2022.3190274.
26. Tang X, Xiao H, Kou G, Peng R. Optimal inspection policy for a three-stage system with imperfect inspection and repair. *IEEE Transactions on Reliability* 2024; 73: 1669-83. https://doi.org/10.1109/TR.2024.3353755.
27. Pham H, Wang H. Imperfect maintenance. *European Journal of Operational Research* 1996; 94: 425-38. https://doi.org/10.1016/S0377-2217(96)00099-9.
28. Azadeh A, Asadzadeh SM, Salehi N, Firoozi M. Condition-based maintenance effectiveness for series–parallel power generation system—a combined markovian simulation model. *Reliability Engineering & System Safety* 2015; 142: 357-68. https://doi.org/10.1016/J.RESS.2015.04.009.
29. Wu W, Prescott D, Remenyte-Prescott R, Saleh A, Ruano MC. An asset management modelling framework for wind turbine blades considering monitoring system reliability. *Reliability Engineering & System Safety* 2024; 252: 110478. https://doi.org/10.1016/J.RESS.2024.110478.

30. Xu Y, Pi D, Yang S, Zio E. Knowledge transfer-based multifactorial evolutionary algorithm for selective maintenance optimization of multistate complex systems. *IEEE Transactions on Reliability* 2024; 73: 1341-52. https://doi.org/10.1109/TR.2023.3324701.
31. Kocer Ozturk M, Khaniyev T. Optimal maintenance policy for a markov deteriorating system under reliability limit. *Eksploatacja i Niezawodność – Maintenance and Reliability* 2024; 26: 190865. https://doi.org/10.17531/ein/190865.
32. Zhang W, Zhang X, He S, Zhao X, He Z. Optimal condition-based maintenance policy for multi-component repairable systems with economic dependence in a finite-horizon. *Reliability Engineering & System Safety* 2024; 241: 109612. https://doi.org/10.1016/J.RESS.2023.109612.
33. Liu L, Zhao G, Sun Y, Hu W. Reliability analysis method of a repairable system with delay and correlation maintenance. *2020 Prognostics and Health Management Conference (PHM-Besançon)*, IEEE; 2020, p. 132–6. https://doi.org/10.1109/PHM-Besancon49106.2020.00028.
34. Zhao X, He S, He Z, Xie M. Optimal condition-based maintenance policy with delay for systems subject to competing failures under continuous monitoring. *Computers & Industrial Engineering* 2018; 124: 535-44. https://doi.org/10.1016/j.cie.2018.08.006.
35. Eryilmaz S. The number of failed components in a k-out-of-n system consisting of multiple types of components. *Reliability Engineering & System Safety* 2018; 175: 246-50. https://doi.org/10.1016/J.RESS.2018.03.027.
36. Hamdan K, Tavangar M, Asadi M. Optimal preventive maintenance for repairable weighted k-out-of-n systems. *Reliability Engineering & System Safety* 2021; 205: 107267. https://doi.org/10.1016/J.RESS.2020.107267.
37. Liu Y, Huang H-Z, Wang Z, Li Y, Yang Y. A joint redundancy and imperfect maintenance strategy optimization for multi-state systems. *IEEE Transactions on Reliability* 2013; 62: 368-78. https://doi.org/10.1109/TR.2013.2259193.
38. Hao Y, Zhu X. Redundancy design and preventive maintenance for a load-sharing multiasset system considering uncertain environmental conditions. *IEEE Transactions on Industrial Informatics* 2024; 20: 9308-19. https://doi.org/10.1109/TII.2024.3383492.
39. Kristjanpoller F, Cárdenas-Pantoja N, Viveros P, Pascual R. Wind farm life cycle cost modelling based on oversizing capacity under load sharing configuration. *Reliability Engineering & System Safety* 2023; 236: 109307. https://doi.org/10.1016/J.RESS.2023.109307.
40. Limnios N. Reliability measures of semi-markov systems with general state space. *Methodology and Computing in Applied Probability* 2012; 14: 895-917. https://doi.org/10.1007/s11009-011-9211-5.
41. Distefano S, Trivedi KS. Non‐markovian state‐space models in dependability evaluation. *Quality and Reliability Engineering International* 2013; 29: 225-39. https://doi.org/10.1002/qre.1305.
42. Bisht V, Singh SB. Lz‐transform approach to evaluate reliability indices of multi‐state repairable weighted k‐out‐of‐n systems. *Quality and Reliability Engineering International* 2023; 39: 1043-57. https://doi.org/10.1002/qre.3279.
43. Gu L, Wang G, Zhou Y, Peng R. Reliability optimization of multi-state systems with two performance sharing groups. *Reliability Engineering & System Safety* 2024; 241: 109580. https://doi.org/10.1016/J.RESS.2023.109580
44. Ushakov IA. A universal generating function. *Soviet Journal of Computer and Systems Sciences* 1986; 24: 37.
45. Li YF, Zio E. A multi-state model for the reliability assessment of a distributed generation system via universal generating function. *Reliability Engineering & System Safety* 2012; 106: 28-36. https://doi.org/10.1016/J.RESS.2012.04.008.
46. Li W, Zuo MJ. Reliability evaluation of multi-state weighted k-out-of-n systems. *Reliability Engineering & System Safety* 2008; 93: 160-7. https://doi.org/10.1016/J.RESS.2006.11.009.
47. Chern MS. On the computational complexity of reliability redundancy allocation in a series system. *Operations Research Letters* 1992; 11: 309-15. https://doi.org/10.1016/0167-6377(92)90008-Q.
48. Peiravi A, Nourelfath M, Zanjani MK. Redundancy strategies assessment and optimization of k-out-of-n systems based on markov chains and genetic algorithms. *Reliability Engineering & System Safety* 2022; 221: 108277. https://doi.org/10.1016/J.RESS.2021.108277.
49. Gholinezhad H. A new model for reliability redundancy allocation problem with component mixing. *Reliability Engineering & System Safety* 2024; 242: 109815. https://doi.org/10.1016/J.RESS.2023.109815.
50. Zhang J, Li L, Chen Z. Strength–redundancy allocation problem using artificial bee colony algorithm for multi-state systems. *Reliability Engineering & System Safety* 2021; 209: 107494. https://doi.org/10.1016/J.RESS.2021.107494.
51. Kristjanpoller F, Michell K, Kristjanpoller W, Crespo A. Fleet optimization considering overcapacity and load sharing restrictions using genetic algorithms and ant colony optimization. *AI EDAM* 2020; 34: 104-13. https://doi.org/10.1017/S0890060419000428.
52. Oszczypała M, Konwerski J, Ziółkowski J, Małachowski J. Reliability analysis and redundancy optimization of k-out-of-n systems with

random variable k using continuous time markov chain and monte carlo simulation. *Reliability Engineering & System Safety* 2024; 242: 109780. https://doi.org/10.1016/J.RESS.2023.109780.

ACRONYMS	
PM	Preventive Maintenance.
CM	Corrective Maintenance
CBM	Condition-based Maintenance
SMP	Semi-Markov Process
UGF	Universal Generating Function
GA	Genetic Algorithm
MCS	Monte Carlo Simulation
RAP	Redundancy Allocation Problem
CTMC	Continuous-Time Markov Chain
EMC	Embedded Markov Chain
MSS	Multi-state System
CDF	Cumulative Distribution Function
LCC	Life Cycle Cost
RTF	Run-to-Failure

NOTATIONS	
i	Index of component type
j, k	Indices of states
N	Number of different component types
$g_j^{(i)}$	Capacity of components of type i in state j
W	Required demand
$S_{OP}^{(i)}$	Set of operable states for components of type i
$S_{PM}^{(i)}$	Set of states in preventive maintenance for components of type i
$S_{CM}^{(i)}$	Set of states in corrective maintenance for components of type i
$\mathcal{C}_j$	Set of all states reachable from state j
P_I	False positive probability in state detection
P_{II}	False negative probability in state detection
t_{delay}	Delay time between alert emission and start of intervention
$y_j^{(i)}$	Binary variable, 1 if the degradation threshold is set to state j in components of type i, 0 otherwise, $j \in S_{OP}^{(i)}$
$F_{j,k}^{(i)}(t), \hat{F}_{j,k}^{(i)}(t)$	Free cumulative distribution function to transition from state j to k in components of type i
$P_{j,k}^{(i)}$	Probability of performing preventive maintenance that leads from state j to state k on components of type i
$q_{PM}^{(i)}$	Preventive maintenance effectiveness parameter governing the probabilistic restoration outcome after preventive maintenance for components of type i
$Q^{(i)}(t)$	Kernel matrix for components of type i
$P^{(i)}$	One-step transition probabilities EMC matrix for components of type i
$\vec{\tau}^{(i)}$	Vector of mean sojourn times for components of type i
$\vec{v}^{(i)}$	Steady-state probability vector of the EMC for components of type i
$\vec{\pi}^{(i)}$	Steady-state probability vector of the SMP for components of type i
$u^{(i)}$	u-function for components of type i
$U(z)$	Universal generating function of the system
A_{system}	System availability
n_i	Number of components of type i used in the system
c_{inef}	System inefficiency cost per hour
$c_{i,j}^{op}$	Hourly operating cost of components of type i in state j
$c_{i,j}^{rep}$	Hourly repair cost of components of type i in state j
$c_{i,j}^{fix}$	Fixed cost associated with a maintenance event for components of type i in state j
c_i^{adq}	Acquisition cost of components of type i
f_{up}	Upgrade factor
H	Planned operating time in hou
A_{req}	Minimum availability required
N_i	Number of components of type i available