Article

Optimizing Costs in a Reliability System under Markovian Arrival of Failures and Reposition by K-Policy Inspection

Delia Montoro-Cazorla 1 and Rafael Pérez-Ocón 2,*

1 Department of Statistics and Operational Research, University of Jaén, 23071 Jaén, Spain; dmontoro@ujaen.es
2 Department of Statistics and Operational Research, University of Granada, 18071 Granada, Spain
* Correspondence: rperez@ugr.es

Abstract: This paper presents an N warm standby system under shocks and inspections governed by Markovian arrival processes. The inspections detect the number of down units, and their replacement is carried out if there are a minimum K of failed units. This is a policy of the type (K, N) used in inventory theory. The study is performed via the up and down periods of the system (cycle); the distribution of these random times and the expected costs for each period comprising the cycle are determined on the basis of individual costs due to maintenance actions (per inspection and replacement of every unit) and others due to operation or inactivity of the system, per time unit. Intermediate addressed calculus are the distributions of the number of inspections by cycle and the expected cost involving every inspection, depending on the number of replaced units. The system is studied in transient and stationary regimes, and some reliability measures of interest and the cost rate are calculated. An optimization of these quantities is performed in terms of the number K in a numerical example. This general model extends to many others in the literature, and, by using the matrix-analytic method, compact and algorithmic expressions are achieved, facilitating its potential application.

Keywords: markovian arrival process; warm standby system; availability; costs; optimization

MSC: 90B25

1. Introduction

Systems with units in standby are common in industry. With units in reserve, the life of the machines enlarges. In reliability, this is a frequent structure. A warm standby is necessary in some cases in which access to the system is not immediate or difficult, as in meteorologic stations in remote locations. In these cases, an inspection allows for failed units to be detected and replaced. The alternative methods are to replace every unit when it fails or to replace all the units when they are nonoperational. In the first case, the number of replacements is greater than in the case we propose, and this increases the expected cost. In the second case, when the last operational unit fails, the system is nonoperational. This causes a stop in production until new units arrive. We propose an intermediate case in which the performance measures and the long-run cost rate can be optimized in terms of the number K.

Standby systems are common redundant structures, widely addressed in different domains. In reliability literature, they are often studied under several maintenance policies, repair modes, inspections and replacements. In [1], a redundant system performing phased missions exposed to random shocks was developed. Random inspections are considered in a redundant k-out-of-n: G system studied in [2], where potential loss of units is assumed, and [3] additionally incorporates multiple vacations on a cold standby system. Replacement policy K has been widely used as a useful tool to improve the availability of the system and the long-run cost rate. In [4,5], optimal repair–maintenance policies were applied to single systems. In the first one, repairable and nonrepairable failures are distinguished, the
system is replaced at the Nth repairable failure or at the arrival of nonrepairable failures, whichever occurs first, and two replacement models were compared. In the second one, the system is periodically inspected every T units of time, and two types of preventive replacements are provided, deterministic after M inspections and opportunity-based after N minor failures. The optimal policy is formulated as a function of these parameters. The Markov process has been applied in the study of reliability systems due to its versatility and the analytic expressions of the main reliability measures. The limitation of these processes is that the staying times in states are exponentially distributed. A variety of shock and wear systems with different repair and replacement criteria have been studied in [6–9]. Authors in [10] considered preventive maintenance of items under a Poisson random shock process, with double effect. A stochastic network was constructed for studying a warm standby n-system with dissimilar components using a Markov process, and the structure of minimal paths was performed in [11,12]. In [13] this technique was used, considering inspections at random times defining a renewal process. In [14], an expression for the joint interval reliability was obtained using the Markovian methodology for a general system that can be applicable to standby systems. This measure is of interest in the power transmission line reliability. The replacement policy K was applied in [15], where a system with a limited number of repairs before a replacement is presented. Phase-type distributions and Markov renewal processes are used more and more in the literature due to their generality in applications and the versatility of the calculations. The following papers are related to these methods [16–19], among others. In [16], a repairable system under the influence of environment conditions was modeled, in which replacement policy K was performed so that after the (K + 1)-th failure, the system is not repaired but replaced. The renewal process due to the replacements of the system was studied. In addition, under this methodology, a warm standby system submitted to shocks was studied in [17]. The replacement of the system occurred when it was down, and the number of replacements of the system with time was determined. Other papers deal with standby systems with repair under multiple vacations. For instance, in [18], a repairable warm standby system undergoing shocks is addressed, where a repairman takes multiple vacations based on the N-policy. Phase-type distributions and MAPs play an important role in that model, and performance measures of interest are obtained for the system in transient and stationary regimes. Other related work is [19], concerning a cold standby repairable system which incorporates working vacations and vacation interruptions. The existing literature on reliability and maintenance modeling is extensive, and many works related to this one in some respect can be found under different methodologies, approaches and application domains. For instance, in [20], a model is proposed to determine the inspection and opportunistic maintenance strategies of floating offshore wind turbines. In [21], multiple failures and simultaneous replacements after inspection are considered in a complex system, while authors in [22] focused on the maintenance of the server system. In this case, a single server stands Markovian arrivals and random shocks. As mentioned before, replacement policy K has demonstrated to play an important role when availability system and cost-optimal decision problems are investigated, as discussed in [23,24]. These papers determined the number of preventive maintenance before the system replacement or the planned replacement time to minimize the expected cost rate, respectively. Meanwhile, authors in [15] pointed out the optimal number of corrective repairs prior to the reposition of the system; other studies of effectiveness are [25–29], and so on.

1.1. Contributions

We present a standby system with multiple components submitted for shocks affecting the online component and the standby one in a different ways and maintained by inspection and replacement using policy K. The procedure for studying these systems is based on the matrix-analytic methods (MAMs) that have proven to be useful in complex systems and that allow the results to be presented in a well-structured form. Moreover, the staying times in states are not necessarily exponentially distributed. This technique based on the
state–space model seems to be appropriate for studying the reliability and maintenance of systems. In [30], different models are studied, discussed and compared. The arrival of the failures and inspections are governed by Markovian arrival processes (MAPs) that present two important properties: the interarrival times are not independent, and they extend many other arrival processes such as the Poisson process, the Markov-modulated Poisson process and the Phase-type renewal process. The replacement of the components occurs when an inspection arrives, and the number of down units reaches or surpasses a fixed number $K$. Extreme values in $K$, $K = 1$ and $K = N$ are of frequent use in maintenance modeling. The dynamic of the system establishes consecutive operational and nonoperational periods, the cycles. Explicit expressions for the expected costs associated to the cycles are calculated by considering the random number of inspections in each period, the mean number of replaced units in every inspection and the expected length of the operational and nonoperational periods. For the system, the availability, the rate of failure and the mean number of replacements are obtained in a transient regime. A study of the optimization of the long-run cost rate and the availability in terms of the values of $K$ and the mean time of inspections are performed in a numerical application.

This reliability system can also be applied to inventory control and risk analysis; failure, inspection, replacement and survival until time $t$ can be substituted by demand, revision, shortage, reinstatement and no shortage until time $t$ in inventory control and by claim, control of the number of claims, bankruptcy, and all claims until time $t$ in risk analysis. In inventory control, this model is assimilated to a $(K, N)$ one; there is a total of $N$ units to be sold, there is demand, and when the number in stock is less than $K$, an order is requested to the store to complete the stock to $S$ units; in any other case, there is no order. This model is useful when the size of the stock is necessarily limited by the capacity of the available space, as is the case for cars and large machines. Under this policy, the costs of the transport of the units from the store to the shop can be cheaper if the requests include several units. The standby structure is appropriate for material degrading with time. The failures (demands), the inspections (revisions), and the replacements (requests) are random, and we consider probabilistic structures with dependence among the interarrival time of the occurrences.

The general hypotheses established for the proposed model and the compact expressions achieved by using the MAM methodology expand its application possibilities. Many particular models previously addressed in the literature can be derived from it.

1.2. Organization

The paper is organized as follows. The model is constructed in Section 2. In Section 3, the stationary distribution and some performance measures are calculated first. Second, the up and down periods in the cycles and the involved costs are carried out. Finally, the performance measures of the system are calculated in a transient regime. In Section 4, a numerical example illustrates the optimization model in terms of $K$, some reliability measures and the long-run cost rate.

2. The Model

Before describing the model, we give the definition of elements that play an important role in the application of the matrix-analytic method: the Kronecker operators, the Phase-type distributions and the Markovian arrival processes.

2.1. Definitions

**Definition 1.** If $A$ and $B$ are rectangular matrices of orders $m_1 \times m_2$ and $n_1 \times n_2$, respectively, their Kronecker product $A \otimes B$ is the matrix of order $m_1 n_1 \times m_2 n_2$, written in compact form as $(a_{ij}B)$.

The Kronecker sum of the square matrices $C$ and $D$ of orders $p$ and $q$, respectively, is defined by $C \oplus D = C \otimes I_q + I_p \otimes D$, where $I_k$ denotes the identity matrix of order $k$. 
For more details about these operations see References [31,32].

**Definition 2.** The distribution function $H(\cdot)$ on $[0, \infty)$ of a phase-type distribution is
\begin{equation}
H(x) = 1 - \alpha \exp(Tx)e, \quad x \geq 0
\end{equation}

It is associated with a finite Markov process with one absorbent state. The initial vector of the process is $\alpha$. Matrix $T$ is the submatrix of the generator of the process restricted to the transient states; it is nonsingular. Vector $e$ is a column vector of 1’s. The absorption column vector is denoted by $T_0$, and it satisfies $-Te = T_0 \geq 0$. The order of the matrix and vectors involved are the same. It is said that the distribution has representation $(\alpha, T)$, and it is written $\text{PH}(\alpha, T)$. $H(x)$ is the distribution function of the first passage time of the Markov process for the absorbent state given the initial vector $\alpha$. The order of the distribution is the order of matrix $T$. A PH-renewal process is a renewal process whose distribution function is given by a PH-distribution.

**Definition 3.** Let $D$ be an irreducible infinitesimal generator of a Markov process. Let a sequence of matrices $D_k, k = 1, 2, ..., n$ be non-negative and the matrix $D_0$ with non-negative off-diagonal entries. The diagonal entries of $D_0$ are strictly negative, and it is nonsingular. All the matrices are square and have the same order. It is assumed that
\begin{equation}
D = D_0 + \sum_{k=1}^{n} D_k
\end{equation}

associated with this Markov process, there is a renewal Markov process performing an arrival process to real line operating as follows. Matrix $D_0$ governs the interarrival times, and matrix $D_k$ governs the arrival of type $k$. This is the MAP associated with the initial Markov process. The order of the MAP is the order of the involved matrices, and it is written $\text{MAP}(D_0, D_1, ... D_n)$.

### 2.2. Assumptions of the Model

Let a system be with $N$ units. Initially one unit is online, and the other $N - 1$ units are in warm standby. When the unit online fails, one of the units in standby, if any, becomes the online unit. Online and standby units are subject to shocks governed by two different and independent Markovian arrival processes. A shock cannot produce multiple failures to the standby units. The system is operational if at least one unit is up. The maintenance of the system is performed by inspection and replacement. The maintenance follows a $K$-policy: in every inspection, the number $j$ of failed units is observed, $0 \leq j \leq N$. If the number of failed units is $j \geq K$, all these units are replaced; in any other case, there is no action. Particular well-known cases are $K = 1$, in which all the failed units are replaced in every inspection, and $K = N$, in which all the units have failed and are replaced. Once the units are replaced, they return to standby if there is one operating online. In any other case, one of the new units goes online. The costs associated with the maintenance are: $c_I$: cost per inspection; $c_R$: cost per replacement of every unit; $c_D$: cost per unit of time that the system is nonoperational; $c_U$: cost per unit of time that the system is operational and $c_S$: cost for starting up the system after a failure. In advance, positive values in costs are interpreted as benefits, while negative values are interpreted as losses.

The assumptions of the system are the following:

1. The online unit is submitted to shocks governed by the $\text{MAP}(c, C_0, C_1)$, with initial vector $c$ and order $m$.
2. The standby units are submitted to shocks governed by the $\text{MAP}(d, D_0, D_1)$, with initial vector $d$ and order $s$.
3. The inspections occur following the $\text{MAP}(h, H_0, H_1)$, with initial vector $h$ and order $l$.
4. All the MAPs are independent.
2.3. Macro-States of the System

Macro-state \( i \) is the number of nonoperational units, \( 0 \leq i \leq N \). Three types of macro-states are distinguished: (1) \( i = 0, 1, ..., N - 2 \), the online unit and at least one of the standby units are operational if they are determined by the phases of the MAPs in the assumptions 1, 2, 3, in this order, and the associated vector is \((i_1, i_2, i_3)\), with \(i_1 = 1, ..., m, i_2 = 1, ..., s, i_3 = 1, ..., l\); (2) the only operational unit is the online one, \( i = N - 1 \); the associated vector is \((i_1, i_3)\), and (3) all the units are nonoperational; \( i = N \), the associated vector is \(i_3\).

To construct the generator, policy \( K \) allows for the grouping of the macro-states as follows: (1) \( i = 0, 1, ..., K - 1 \), when an inspection occurs and the system occupies any of these macro-states, no unit is replaced; in the rest of the macro-states, there is replacement, and two types of macro-states are differentiated, (2) \( i = K, K + 1, ..., N - 2 \), and (3) \( i = N - 1, N \), these are the border macro-states. The generator is constructed by blocks calculating the transition among these groups of macro-states.

2.3.1. Transition \( \{0, 1, ..., K - 1\} \rightarrow \{0, 1, ..., K - 1\} \)

Transition \( 0 \rightarrow 0 \) is governed by \( C_0 \oplus D_0 \oplus H = \tilde{A}_1 \); no shock arrives, and an inspection has no effect on the units. In Transition \( 0 \rightarrow 1 \), if a failure occurs to one unit (online or standby), the phase of the inspection cannot change; it is governed by \( (C_1 \oplus D_1) \otimes I = A_0 \). The rest of transitions \( i \rightarrow i \) and \( i \rightarrow i + 1 \) are identical, respectively, to \( 0 \rightarrow 0 \) and \( 0 \rightarrow 1 \). No other transitions are possible.

Transitions from \( \{0, 1, ..., K - 1\} \) to any other macro-state are null, except \( K - 1 \rightarrow K \) that occurs when a shock arrives and it is governed by \( A_0 \).

2.3.2. Transition \( \{K, K + 1, ..., N - 2\} \rightarrow \{K, K + 1, ..., N - 2\} \)

Transition \( K \rightarrow K \) is similar to \( 0 \rightarrow 0 \), but now an inspection cannot occur; it is governed by \( C_0 \oplus D_0 \oplus H_0 = A_1 \). The same occurs for the rest of diagonal blocks \( i \rightarrow i \), \( i = K + 1, ..., N - 2 \). Transition \( K \rightarrow K + 1 \) is like \( 0 \rightarrow 1 \), \( A_0 \); it is the same for \( i \rightarrow i + 1 \), \( i = K + 1, ..., N - 3 \). Transition \( N - 2 \) to \( N - 1 \) occurs if there are two operational units, the online and the standby ones, and a shock arrives; then one of the two units fails: (a) if the online fails (governed by \( C_1 \)), the MAP governing the arrival of shocks to the standby units can occupy any phase, and the standby unit becomes the online one; the inspection does not change (governed by \( I \)); and (b) similarly, the failure of the standby unit from any phase is governed by \( D_1 e \), the MAPs corresponding to the online unit and the inspection do not change. Taking into account the order of the macro-state vectors, the transition is governed by \( B_{N-2,N-1} = C_1 \otimes e \otimes I + I \otimes D_1 e \otimes I \). Other upper transitions cannot occur since there are no multiple shocks.

2.3.3. Transition \( \{K, K + 1, ..., N - 2\} \rightarrow \{0, 1, ..., K - 1\} \)

Transition \( i \rightarrow 0, i = K_1, ..., N - 2 \), occurs when an inspection arrives (governed by \( H_1 \)), and all the units are replaced. The MAPs related to the units do not change; it is governed by \( I \otimes I \otimes H_1 = A_2 \). No other transitions are possible.

2.3.4. Transition \( \{N - 1, N\} \rightarrow \{0, 1, ..., N\} \)

Transition \( N - 1 \rightarrow 0 \) occurs when an inspection arrives (governed by \( H_1 \)), and the MAP governing the arrival of shocks to the standby units is determined by \( I \otimes d \otimes H_1 = B_{N-1,0} \). Transition \( N - 1 \rightarrow N - 1 \) is governed by \( C_0 \oplus H_0 = B_{N-1,N-1} \). Transition \( N - 1 \rightarrow N \) is governed by \( C_1 e \otimes I \) since there is only one unit operating, and a failure occurs while the MAP governing the inspection does not change; the system fails. Transition \( N \rightarrow 0 \) occurs when an inspection arrives, and all the units are down; it is governed by \( c \otimes d \otimes H_1 = B_{N,0} \). Transition \( N \rightarrow N \) is governed by \( H_0 \). There are no more transitions.

The generator for \( 1 \leq K \leq N - 2, N \geq 3 \) is given by
\[ Q = Q_K = \begin{pmatrix}
\text{ine 0} & 0 & 1 & K - 1 & K & K + 1 & N - 2 & N - 1 & N \\
1 & \tilde{A}_1 & A_0 & \tilde{A}_1 & A_0 & & & & \\
\vdots & \vdots & & & & & & & \\
K - 1 & & & & & & & & \\
\text{ine K} & A_2 & \tilde{A}_1 & & & & & & \\
\vdots & \vdots & & & & & & & \\
K + 1 & & & & & & & & \\
\vdots & & & & & & & & \\
N - 2 & & & & & & & & \\
\text{ine N - 1} & A_2 & & & & & & & \\
N & & & & & & & & \\
B_{N-1,0} & & & & & & & & \\
B_{N,0} & & & & & & & & 
\end{pmatrix} \]  

For \( K = N - 1 \) and \( K = N \), we have, respectively,

\[ Q_{N-1} = \begin{pmatrix}
\text{ine 0} & 0 & 1 & N - 2 & N - 1 & N \\
1 & \tilde{A}_1 & A_0 & \tilde{A}_1 & A_0 & & & \\
N - 2 & & & & & & & & \\
\text{ine N - 1} & B_{N-1,0} & & & & & & & \\
N & B_{N,0} & & & & & & & 
\end{pmatrix} \]  

\[ Q_N = \begin{pmatrix}
\text{ine 0} & 0 & 1 & N - 2 & N - 1 & N \\
1 & \tilde{A}_1 & A_0 & \tilde{A}_1 & A_0 & & & \\
N - 2 & & & & & & & & \\
N - 1 & & & & & & & & \\
\text{ine N} & B_{N,0} & & & & & & & 
\end{pmatrix} \]  

with \( B_{N-1,N-1}^* = C_0 \oplus H \).

In Section 3, performance measures in a transient regime will be calculated in terms of the transition probability functions; these depend on generator \( Q \).

3. Cycles of the System, Performance Measures and Costs

The system is first studied in a steady state. The stationary probability block vector is denoted by

\[ \pi = (\pi_0, \pi_1, ..., \pi_{K-1}, \pi_K, \pi_{K+1}, ..., \pi_{N-2}, \pi_{N-1}, \pi_N) \]  

The components \( \pi_i \) are vectors associated with the macro-states. This is calculated from the equations \( \pi Q = 0, \pi e = 1 \). We assume that \( 1 \leq K \leq N - 2, N \geq 3 \).

Operating in these previous expressions, we have

\[ \pi_i = \pi_0 (\tilde{R})^i, i = 1, ..., K - 1 \]
\[ \pi_i = \pi_0 (\tilde{R})^i R^{i+1}, i = K, ..., N - 2 \]
\[ \pi_{N-1} = \pi_0 (\tilde{R})^N R_{N-1} \]
\[ \pi_N = \pi_0 (\tilde{R})^N R_N \]  

with
\[
\tilde{R} = -A_0 \left[ \tilde{A}_1 \right]^{-1} \quad R = -A_0 [A_1]^{-1}
\]

\[
R_{N-1} = -B_{N-2,N-1} [B_{N-1,N-1}]^{-1} \quad R_N = -B_{N-1,N} [B_{N,N}]^{-1}
\]

The final expression for vector \( \pi_0 \) can be calculated from the following equations:

\[
\pi_0 \left[ \tilde{A}_1 + (\tilde{R})^{K-1} \left( \sum_{i=K}^{N-2} R_i^{K+1} A_2 + R_i^{N-K-1} R_{N-1} (B_{N-1,0} + R_N B_{N,0}) \right) \right] = 0
\]

\[
\pi_0 \left[ e + \sum_{i=1}^{K-1} (\tilde{R})^i e + R_i^{K-1} \left( \sum_{i=K}^{N-2} R_i^{K+1} e + R_i^{N-K-1} R_{N-1} (I + R_N) e \right) \right] = 1
\]

For \( K = N - 1 \), the equations and the solution are, respectively,

\[
\pi_i = \pi_0 (\tilde{R})^i, \quad i = 1, \ldots, N - 2
\]

\[
\pi_{N-1} = \pi_0 (\tilde{R})^{N-2} R_{N-1}
\]

\[
\pi_N = \pi_0 (\tilde{R})^{N-2} R_{N-1} R_N
\]

\[
\pi_0 \left[ \tilde{A}_1 + (\tilde{R})^{N-2} R_{N-1} (B_{N-1,0} + R_N B_{N,0}) \right] = 0
\]

\[
\pi_0 \left[ e + \sum_{i=1}^{N-2} (\tilde{R})^i e + (\tilde{R})^{N-2} R_{N-1} (I + R_N) e \right] = 1
\]

and for \( K = N \),

\[
\pi_i = \pi_0 (\tilde{R})^i, \quad i = 1, \ldots, N - 2
\]

\[
\pi_{N-1} = \pi_0 (\tilde{R})^{N-2} R_{N-1}
\]

\[
\pi_N = \pi_0 (\tilde{R})^{N-2} R_{N-1} R_N
\]

\[
\pi_0 \left[ \tilde{A}_1 + (\tilde{R})^{N-2} R_{N-1} R_N B_{N,0} \right] = 0
\]

\[
\pi_0 \left[ e + \sum_{i=1}^{N-2} (\tilde{R})^i e + (\tilde{R})^{N-2} R_{N-1} (I + R_N) e \right] = 1
\]

with \( R_{N-1}^* = -B_{N-2,N-1} \left[ B_{N-1,N-1}^* \right]^{-1} \).

Once this distribution is obtained, the following performance measures are calculated. The availability of the system is given by

\[
A = \sum_{i=0}^{N-1} \pi_i e
\]

The rate of occurrence of failures of the online unit is

\[
v_o = \sum_{i=0}^{N-2} \pi_i (C_i e \otimes e \otimes e) + \pi_{N-1} (C_e e \otimes e)
\]

The rate of occurrence of failures of the units in standby is

\[
v_{st} = \sum_{i=0}^{N-2} \pi_i (e \otimes D_i e \otimes e)
\]
For the system, the rate of failure has the expression
\[ v = \pi_{N-1}(C_I e \otimes e) \]  

(17)

3.1. Distribution of the up and Down Periods

We carry out a study of the system via the up and down periods. The system is operational while occupying the macro-states \( \{0, 1, ..., N-1\} \), and it is down in macro-state \( N \). The evolution of the system is an alternating sequence of up and down periods. A cycle is the time span of two consecutive up and down periods. The cycle initiates when the system occupies macro-state \( N \), an inspection occurs (all the units are replaced) and the system reinitiates. From this point, the system is in the up period until all the units fail. Then it occupies a down period that finishes when a new inspection arrives. The cycles are of industrial interest; in some cases, the lifetime of the machines is expressed in terms of the completed cycles. The aging and the costs increase with the number of cycles are of industrial interest; in some cases, the lifetime of the machines is expressed.

The costs involved in the system are: by inspection \((\alpha)\), by replacement \((\beta)\), by operation unit of time \((c_I)\), by nonoperation unit of time \((c_D)\), and by the starting up of the system after a failure \((c_S)\). We centered on the costs involved in the cycles. The expected costs during an up and down period are denoted by \(E[\Psi_I]\) and \(E[\Psi_D]\), respectively. The expected total cost per unit of time in a cycle is given by

\[ CT = \frac{E[\Psi_I] + E[\Psi_D]}{E[X_C]} \]  

(20)

The expected cost in a nonoperational period is \(E[\Psi_D] = c_D \mu_D\).
In advance, we also refer to $CT$ as the long-run cost rate. Note that a positive value implies a benefit, while a negative value means a loss.

For calculating the expected cost during an operational period, we must know the number of inspections that occurred in this period and the expected cost of every inspection that depends on the number of replaced units. Let $N_I$ be the number of inspections during an operational period. We calculate the distribution of this random variable considering how the inspections arrive to the system during this period. The events that must have been taken into account are the following:

(a) There is neither inspection nor failure of the system;
(b) There is an inspection, and the system does not fail: the operational period continues;
(c) There is no inspection, and the system fails: the up period finishes.

Every event of these has an associated submatrix derived from generator $Q$ collecting the transitions describing its occurrence. Given that the system is in an operational period, macro-state $N$ is not included.

Matrix $G_0$ is associated with event (a):

$$G_0 = \begin{pmatrix}
\begin{array}{c|cccc|cccc|c}
\text{ine } 0 & 0 & 1 & K - 1 & K & K + 1 & N - 2 & N - 1 \\
\text{ine } 1 & A_1 & A_0 & & A_1 & A_0 & & \\
\text{ine } K - 1 & & & \ddots & & & & \\
\text{ine } K & & & & \ddots & & & \\
\text{ine } K + 1 & & & & & \ddots & & \\
\text{ine } N - 2 & & & & & & \ddots & A_0 \\
\text{ine } N - 1 & & & & & & & A_1 \\
\end{array}
\end{pmatrix}
$$

(21)

Note that this matrix is obtained from $Q$ eliminating the macro-state $N$, and replacing the blocks $\tilde{A}_1$ by $A_1$, since inspections cannot occur in this event.

Matrix $G_1$ is associated with event (b):

$$G_1 = \begin{pmatrix}
\begin{array}{c|cccc|cccc|c}
\text{ine } 0 & 0 & 1 & K - 1 & K & K + 1 & N - 2 & N - 1 \\
\text{ine } 1 & A_2 & & & A_2 & \ddots & & \\
\text{ine } K - 1 & & & \ddots & & & & \\
\text{ine } K & & & & \ddots & A_2 & & \\
\text{ine } K + 1 & & & & & A_2 & \ddots & \\
\text{ine } N - 2 & & & & & & A_2 & B_{N - 1,0} \\
\text{ine } N - 1 & & & & & & & B_{N - 1,0} \\
\end{array}
\end{pmatrix}
$$

(22)

This matrix includes the blocks associated with transitions involved in the occurrence of inspections without failure of the system. Therefore, blocks $A_0$ are eliminated since there is no failure; blocks $\tilde{A}_1$ are replaced by $A_2$ for $0, 1, \ldots, K - 1$ (inspections without effect), maintained in column 0 and eliminated for $i = K, K + 1, \ldots, N - 1$ because an inspection occurs.
Matrix $G_2$ is associated with event (c):

$$
G_2 = \begin{pmatrix}
\text{ine 0} & N \\
1 & 0 \\
& \vdots \\
K - 1 & 0 \\
\text{ine K} & 0 \\
K + 1 & 0 \\
& \vdots \\
N - 2 & 0 \\
\text{ine N - 1} & B_{N-1,N}
\end{pmatrix}
$$

(23)

In this matrix, the transition blocks are collected, taking into account that there is no inspection, and the system fails. From these matrices are calculated the probability of the occurrence of inspections, matrix $F$, and the probability of failure of the system, matrix $F^0$, given by

$$F = (-G_0)^{-1}G_1$$

$$F^0 = (-G_0)^{-1}G_2e$$

(24)

These are, respectively, the matrices of the embedded Markov chain associated with the transitions due to inspections and failures of the system. The system fails from any phase, governed by $G_2e$. Note that transition $N - 1 \rightarrow N$ is governed by $B_{N-1,N} = C_i e \otimes I_i$; so the last block of $G_2e$ is $(C_i e \otimes I_i)e$ is a column vector of order $ml$ formed by the absorption rates to macro-state $N$ for the different phases of the inspection. Given that there are only inspections and failures of the system, we have $Fe + F^0 = e$. Matrix $F$ collects the probability of inspections and $F^0$ the probability of failures.

In macro-states $u = \{0, 1, ..., K - 1\}$, only inspections occur, and in macro-states $v = \{K, K + 1, ..., N - 1\}$, there are inspections and replacements. It is a partition of the set of operational macro-states. The costs associated with the macro-states are assigned in the following two matrices, where matrix $I_i$ denotes the identity matrix associated with macro-state $i, i = 0, 1, ..., K, K + 1, ..., N - 1$:

$$
\Psi_I = \begin{pmatrix}
I_0e_I \\
I_1e_I \\
\vdots \\
I_{K-1}e_I
\end{pmatrix}
$$

$$
\Psi_R = \begin{pmatrix}
I_K(e_I + Ke_R) \\
I_{K+1}(e_I + (K + 1)e_R) \\
\vdots \\
I_{N-1}(e_I + (N - 1)e_R)
\end{pmatrix}
$$

(25)

The new diagonal block matrix $\Psi$ summarizes the costs associated with the macro-states in which the inspections occur:

$$\tilde{\Psi} = \begin{pmatrix}
\Psi_I & 0 \\
\text{ine 0} & \Psi_R
\end{pmatrix}
$$

(26)

The up period initiates in macro-state 0 with initial probability vector $\phi_U$. The inspections occur in the macro-states with probabilities given by $\phi_U(-G_0)^{-1}G_1 = \phi_UF$. 


The expected cost due to the inspections and replacements during an up period in the different macro-states is given by matrix $\phi_U(-G_0)^{-1}\Psi G_1$, and the total expected cost is

$$CMIU = \phi_U(-G_0)^{-1}\Psi G_1 e$$  \hspace{1cm} (27)

To clarify this result, we write matrices $(-G_0)^{-1}$ and $G_1$ in terms of the blocks of the partition $\{u,v\}$ of the set of macro-states; the arrival of an inspection without replacement occurs in the macro-states in $u$, and the ones including replacement after an inspection are in $v$. We can write

$$(-G_0)^{-1} = \begin{pmatrix} V_0(u,u) & V_0(u,v) \\ V_0(v,u) & V_0(v,v) \end{pmatrix}$$ \hspace{1cm} (28)

$$G_1 = \begin{pmatrix} V_1(u,u) & V_1(u,v) \\ V_1(v,u) & V_1(v,v) \end{pmatrix} = \begin{pmatrix} V_1(u,u) & 0 \\ V_1(v,u) & 0 \end{pmatrix}$$ \hspace{1cm} (29)

These blocks of matrix $(-G_0)^{-1}$ can be calculated by computational methods in practical cases. Note that $V_1(u,u) = I \otimes \text{diag}(A_2)$ and $V(v,u)$ is a rectangular matrix with all the blocks null except the ones including replacement after an inspection, as can be seen above.

We have, formally,

$$(-G_0)^{-1}\Psi G_1 = \begin{pmatrix} V_0(u,u)\Psi_1 V_1(u,u) + V_0(u,v)\Psi_R V_1(v,u) & 0 \\ V_0(v,u)\Psi_1 V_1(u,u) + V_0(v,v)\Psi_R V_1(v,u) & 0 \end{pmatrix} \hspace{1cm} (30)$$

We interpret the block

$$V_0(u,u)\Psi_1 V_1(u,u) + V_0(u,v)\Psi_R V_1(v,u)$$  \hspace{1cm} (31)

The factors of the first summand indicate that the up period initiates in an element of $u$ (really is 0), and an inspection occurs. The associated cost is given by $\Psi_1$, and after, the inspection returns to a macro-state in $u$. The second summand indicates that the up period initiates in an element of $u$, a failure occurs and there is a transition to $v$. An inspection arrives later with associated cost given by $\Psi_R$ and returns again to set $u$ (really enters 0). The other block can be interpreted in a similar way.

Taking into account the initial conditions $\phi_U = (a_U, 0)$, the other block of the previous matrix is not used, and the mean cost per inspection in the up period is given by

$$CMIU = \phi_U(G_0)^{-1}\Psi G_1 e = a_U \left[ V_0(u,u)\Psi_1 V_1(u,u) + V_0(u,v)\Psi_R V_1(v,u) \right] e \hspace{1cm} (32)$$

Once the cost per inspection is calculated, we must calculate the number of inspections occurring in an up period, denoted by $N_i$. Previously matrices $F$ and $F^0$ have been defined. In terms of them, we have

$$P(N_i = 0) = \phi_U F^0$$
$$P(N_i = 1) = \phi_U F F^0$$
$$\vdots$$
$$P(N_i = n) = \phi_U F^n F^0$$

This is the expression of a discrete $PH_d(f,F)$-distribution with $f = \phi_U F$, being

$$P(N_i = n) = f F^{n-1} F^0 \quad \text{and} \quad E[N_i] = f(I - F)^{-1} e$$ \hspace{1cm} (34)
The cost in the up period is denoted by $\Psi_U$, and the total mean cost in this period is

$$E[\Psi_U] = c_s + CMIU \cdot E[N_I] + c_U \mu_U = c_s + CMIU \cdot f(I - F)^{-1} e + c_U \mu_U$$  \hspace{1cm} (35)

Taking into account the mean cost of the down period by cycle, the long-run cost rate is given by

$$CT = \frac{E[\Psi_U] + E[\Psi_D]}{E[X_C]} = \frac{c_s + CMI \cdot f(I - F)^{-1} e + c_U \mu_U + c_D \mu_D}{-\phi_D T_D}$$  \hspace{1cm} (36)

### 3.3. Performance Measures

The system is studied in a stationary regime, but from matrix $Q$ it is possible to calculate the transition probability matrix $P(t) = P_{ij}(t)$, with $P_{ij}(t) = P\{X(t) = j|X(0) = i\}$ and $0 \leq i, j \leq N$ macro-states, from the matrix equation $P'(t) = QP(t)$ under the initial condition $P(0) = I$. It is well-known that $P(t) = \exp(Qt)$, and this expression is calculated using computational methods. The performance measures of the system in a transient regime can be deduced from this expression. Particularly, we calculate the availability using computational methods. The performance measures of the system in a transient condition is given by

The availability has the expression

$$A(t) = a_0 \sum_{i=0}^{N-1} P_{0i}(t) e$$  \hspace{1cm} (38)

with $a_0 = c \otimes d \otimes h$ and $a = (a_0, 0, ..., 0)$.

The failure rate of the system is

$$r(t) = \frac{a \exp(Q^t e)(-Q^te)}{a \exp(Q^t e)}$$  \hspace{1cm} (39)

The mean number of replacements to the system in $(0, t]$ is calculated introducing new matrices derived from the ones of the MAPs and taking into account the arrivals of the event of interest. The transitions can be classified in two: those not producing a renewal of the system, governed by $M_0(K)$, and those producing a renewal of the system, denoted by $M_1(K)$. The expressions of these matrices for $1 \leq K \leq N - 2$, $N \geq 3$, are

$$M_0(K) = \begin{pmatrix}
\begin{array}{cccccccc}
0 & 1 & K - 1 & 0 & 1 & K - 1 & 0 & 1 & K - 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
A_1 & A_0 & A_0 & \ddots & \ddots & \ddots & \ddots & \ddots & B_{N-1,0} \\
A_2 & A_1 & A_1 & \ddots & \ddots & \ddots & \ddots & \ddots & H_0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
A_{N-2} & A_{N-1} & A_{N-1} & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
B_{N-1,0} & B_{N-1,1} & B_{N-1,1} & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\end{array}
\end{pmatrix}$$  \hspace{1cm} (40)
\[
M_1(K) = \begin{pmatrix}
0 & 1 & K - 1 & K & K + 1 & N - 2 & N - 1 & N \\
\vdots & & \vdots & & \vdots & \vdots & \vdots & \vdots \\
K - 1 & \text{ine } K & K + 1 & \text{ine } N - 1 & \text{ine } N & \text{ine } 0 & \text{ine } 1 & B_{N,0}
\end{pmatrix}
\]

It can be seen that \( M_0(K) + M_1(K) = M(K) \) is a generator for any value of \( K \). The mean number of replacements in \((0, t]\) is

\[
E(N(t)) = \lambda t + a \exp(M(K)t - I)(M(K) - e\theta)^{-1}M_1(K)e, t \geq 0
\]

\( \theta \) being the stationary vector associated with \( M(K) \) and \( \lambda = \theta M_1(K)e \).

The variance is

\[
\text{Var}(\lambda N(t)) = (\lambda - 2\lambda^2 - 2\theta M_1(K)(M(K) - e\theta)^{-1}M_1(K)e)t + 2\theta M_1(K)(M(K) - e\theta)^{-1}(\exp(M(K)t - I)(M(K) - e\theta)^{-1}
\]

For \( K = N - 1 \) and \( K = N \), the calculations are similar.

Some other measures can be calculated following the present procedure, including measures referred to the cycles defined before.

4. Numerical Example

A numerical and graphical study of the optimization of the maintenance (policy \( K \)) is applied to a system with \( N = 5 \). The unit time is denoted by u.t. The matrix parameters of the MAPs governing the arrival of shocks are

| Arrival of shocks to online unit | \( c = (1, 0) \) | \( C_0 = \begin{pmatrix} -4 & 1 \\ 2 & -7 \end{pmatrix} \) | \( C_1 = \begin{pmatrix} 0 & 3 \\ 2 & 3 \end{pmatrix} \) |
| Arrival of shocks to the standby units | \( d = (1, 0) \) | \( D_0 = \begin{pmatrix} -9 & 0 \\ 1 & -1 \end{pmatrix} \) | \( D_1 = \begin{pmatrix} 8 & 1 \\ 0 & 0 \end{pmatrix} \) |
| Occurrence of an inspection | \( h = (1, 0) \) | \( H_0 = \begin{pmatrix} -2.4 & 0 \\ 2.4 & -6 \end{pmatrix} \) | \( H_1 = \begin{pmatrix} 2.4 & 0 \\ 2.4 & 1.2 \end{pmatrix} \) |

The assigned costs are

\( c_U = 1.5 \quad c_D = -1.58 \quad c_I = -0.05 \quad c_R = -0.10 \quad c_S = -0.07 \)

The study of this system comprises: (1) the performance reliability measures and the long-run cost rate, including values and graphics for the system with \( K = 3 \); (2) a comparison of the measures for different values of \( K \), tables and graphics; (3) a selection of the optimal policy given the performance measures and costs; and (4) an analysis of the frequency of the inspection in the measures for different values of \( K \) in the exponential case.

• Performance measures for \( K = 3 \)

For \( K = 3 \), the values of the performance measures previously defined in a transient regime are given in Table 1.
Table 1. Performance measures in transient regime for $K = 3$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A(t)$</td>
<td>1</td>
<td>0.9974</td>
<td>0.9688</td>
<td>0.9073</td>
<td>0.8388</td>
<td>0.7843</td>
<td>0.7487</td>
<td>0.7281</td>
<td>0.7171</td>
<td>0.7113</td>
<td>0.7084</td>
</tr>
<tr>
<td>$r(t)$</td>
<td>0</td>
<td>0.1107</td>
<td>0.5745</td>
<td>1.0341</td>
<td>1.2833</td>
<td>1.3605</td>
<td>1.3556</td>
<td>1.3282</td>
<td>1.3016</td>
<td>1.2804</td>
<td>1.2636</td>
</tr>
<tr>
<td>$N(t)$</td>
<td>0</td>
<td>0</td>
<td>0.0034</td>
<td>0.0178</td>
<td>0.0484</td>
<td>0.0940</td>
<td>0.1504</td>
<td>0.2134</td>
<td>0.2802</td>
<td>0.3488</td>
<td>0.4185</td>
</tr>
</tbody>
</table>

The transient measures are plotted in Figure 1.

Figure 1 shows that the point $t = 0.5$ is a singular point. The availability reaches the equilibrium at $t = 0.5$ with a value close to 0.7. The failure rate of the system increases until $t = 0.5$ and then decreases slowly. The number of replacements increases, as expected, and the rate seems to be linear.

In Table 2 the performance measures and costs by cycle are given.

Table 2. Performance measures and costs by cycle for $K = 3$.

<table>
<thead>
<tr>
<th>Policy</th>
<th>$A$</th>
<th>$\mu_U$</th>
<th>$\mu_C$</th>
<th>$\rho_C$</th>
<th>$CT$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = 3$</td>
<td>0.7055</td>
<td>0.9983</td>
<td>1.415</td>
<td>0.7055</td>
<td>0.3120</td>
</tr>
</tbody>
</table>

The system is operational 70.55% of the time of the cycle, the operational mean time by cycle is 0.9983, and the mean time of the down period is 0.4167. Therefore, the expected length of the cycle is 1.415. The long-run cost rate is 0.3120.

- Comparing performance measures

In Figure 2, the availability of the system for all the values of $K$ is plotted. The way in which the availability decreases with $K$ is shown; the point in which all the curves approach to a constant is between 0.5 and 1, nearer to 1. There is a significant difference between $K = 5$ and the others.
Figure 2. Availability function for systems $K = 1, 2, 3, 4, 5$.

In Figure 3 the failure rate of the system for all the values of $K$ is plotted. Up to $t = 0.5$, the curves increases then, tends to be constant between 1 (exponential) and 1.5 for $K = 1, 2, 3, 4$, but for $K = 5$, the rate increases slowly and takes values greater than 3.

Figure 3. Rate of system failure function for systems $K = 1, 2, 3, 4, 5$.

Figure 4 represents the mean number of replacements in the system for different values of $K$. As expected, all the plots increase with time and are approximately linear for $t > 1.5$ for systems $K = 1, 2, 3, 4$. The system $K = 5$ has an expected number significantly different to the others.
In Table 3, the stationary distribution $A$, the mean operational time $\mu_U$, the mean time of the cycle $\mu_C$, the proportion of the operational time $\rho_C$, and the long-run cost rate $CT$ are given for all the values of $K$.

For an optimal policy $K$ in terms of the operational time and costs, the values of $\rho_C$ and $CT$ in terms of $K$ are plotted in Figure 5. In both cases, the values decrease with $K$ in a similar way.
Table 3. Performance measure and the long-run cost rate for systems $K = 1, 2, 3, 4, 5$.

<table>
<thead>
<tr>
<th>Policy</th>
<th>$A$</th>
<th>$\mu_U$</th>
<th>$\mu_C$</th>
<th>$\rho_C$</th>
<th>$CT$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = 1$</td>
<td>0.7395</td>
<td>1.1826</td>
<td>1.5993</td>
<td>0.7395</td>
<td>0.3404</td>
</tr>
<tr>
<td>$K = 2$</td>
<td>0.7274</td>
<td>1.1123</td>
<td>1.529</td>
<td>0.7275</td>
<td>0.3326</td>
</tr>
<tr>
<td>$K = 3$</td>
<td>0.7055</td>
<td>0.9983</td>
<td>1.415</td>
<td>0.7055</td>
<td>0.3120</td>
</tr>
<tr>
<td>$K = 4$</td>
<td>0.6731</td>
<td>0.8579</td>
<td>1.2746</td>
<td>0.6731</td>
<td>0.2677</td>
</tr>
<tr>
<td>$K = 5$</td>
<td>0.6057</td>
<td>0.6401</td>
<td>1.0568</td>
<td>0.6057</td>
<td>0.1672</td>
</tr>
</tbody>
</table>

- Availability and the long-run cost rate in terms of the inspection times

We consider the mean time of inspections a parameter and determine the performance measures in terms of it. This is a quantity that can be established by the researcher, and it allows us to optimize some performance measures. It is assumed that the inspection times are exponentially distributed. In Figure 6, the long-run cost rates in terms of the different values of $K = 1, 2, 3, 4, 5$ are given for the inspection mean time (IMT) $\mu_I = 0.10, \mu_I = 0.30, \mu_I = 0.50$. For $K = 1$, the benefits increase when the mean time between inspections increases. For $K > 1$, the behavior is different. For $K = 2$, the long-run cost rates for $\mu_I = 0.10$ and $\mu_I = 0.50$ are identical and less than for $\mu_I = 0.30$. For $\mu_I = 0.10$, the optimal long-run cost rate is reached for $K = 4$. Until this value, the long-run cost rate is notably increasing; then it decreases shortly for $K = 5$.

![Figure 6](image)

**Figure 6.** The long-run cost rate versus $K$ for $\mu_I = 0.10, \mu_I = 0.30, \mu_I = 0.50$.

In Figure 7, the long-run cost rates in terms of the mean time of inspections for different values of $K$ are plotted. It is observed that the behavior of the graphics are very similar. It is optimum when the mean time between inspections is 0.10 and $K = 5$. 
5. Conclusions

A major novelty in the present paper is the inclusion of the MAMs for the study of complex systems. The generality of the assumptions extends the application of the model to other systems. We can see that by handling the matrix calculations, it is possible to express quantities algorithmically since the operations in the formulae involving matrices are presented in a well-structured form. This is highlighted in the study of the costs and in the calculation of the performance measures. These measures have been calculated for the
cycles and for the total system, since the aging of a system can be measured by the time or by the number of cycles. Applying the methods to the study of a system and varying the parameters depending on the researcher reveals that complete information on the quantities affecting the performance of the system can be obtained. By adding new assumptions to the system, such as lifetime of the units, repairs and others, the study can be extended to more complex systems. Moreover, this methodology is suitable for the extension of the model in other ways. For instance, considering an appropriate MAP, simultaneous failures in units can be incorporated, and a natural continuation takes place when units in the system are K-grouped, following a modular structure. Many industrial systems have this structure; it can allow time and cost reductions when analysis of failures and maintenance, such as replacements, are set by blocks.

**Author Contributions:** Formal analysis, D.M.-C. and R.P.-O.; Investigation, R.P.-O.; Methodology, D.M.-C. and R.P.-O.; Software, D.M.-C.; Supervision, R.P.-O.; Writing—original draft, D.M.-C.; Writing—review & editing, R.P.-O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Notation**

The following abbreviations are used in this manuscript:

- \( MAP(c, C_0, C_1) \) shock arrival process for online unit
- \( MAP(d, D_0, D_1) \) shock arrival process for standby units
- \( MAP(h, H_0, H_1) \) inspection arrival process
- \( I_r \) identity matrix of order \( r \)
- \( e_r \) column vector of 1’s of order \( r \)
- \( Q_K, K = 1, 2, ..., N \) generator of the process under policy \( K \)
- \( \pi_i \) steady state vector associated with the macro-state \( i \)
- \( \pi \) steady state vector
- \( A \) steady state availability
- \( v_o \) rate of occurrence of failures for online unit
- \( v_{st} \) rate of occurrence of failures for standby units
- \( v \) rate of occurrence of failures for the system
- \( PH(\phi_U, T_U) \) distribution of the up period of the system, \( (X_U) \)
- \( PH(\phi_D, T_D) \) distribution of the down period of the system, \( (X_D) \)
- \( X_C \) length of a cycle of the system
- \( \mu_U \) expected length of the up period of the system
- \( \mu_D \) expected length of the down period of the system
- \( \mu_C \) expected length of a cycle of the system
- \( c_I \) cost per inspection
- \( c_R \) cost per every unit replacement
- \( c_U \) cost by operational unit time
- \( c_D \) cost by nonoperational unit time
- \( c_S \) cost due to the starting-up of the system
- \( E[\Psi_U] \) expected cost during an up period
- \( E[\Psi_D] \) expected cost during a down period
- \( CMIU \) expected cost per inspection, depending on the number of replacements, in the up period
- \( E[N_I] \) expected number of inspections in the up period
- \( CT \) long-run cost rate
- \( A(t) \) availability of the system in time \( t \)
- \( R(t) \) reliability of the system in time \( t \)
- \( N(t) \) mean number of replacements of the system \((0, t] \)
- \( r(t) \) failure rate of the system in time \( t \)