



# Joint maintenance and just-in-time spare parts provisioning policy for a multi-unit production system

Nooshin Salari<sup>1</sup> · Viliam Makis<sup>1</sup>

Published online: 9 October 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

In this paper, two new joint maintenance and spare parts provisioning policies for a multi-unit production system are proposed. The production system consists of  $N$  identical, independent units, each subject to gradual deterioration. Production rates of the units depend on their operating states. Every maintenance action incurs a high set-up cost which includes the cost of sending a crew to the field, and it is therefore cost effective to maintain several units at the same time. The paper's main contribution is an analytical modeling of a multi-unit production system and the development of effective joint maintenance and spare parts ordering policies for such system. The states of the units are observable through regular inspections and the maintenance and spare part ordering decisions depend on the number of the failed units and the number of available spare parts. The process is formulated as a semi-Markov decision process with the optimality criterion being the minimization of the total long-run expected average cost per unit time. The objective is to determine the optimal levels of the number of failed units to place an order for the spare parts or to initiate group maintenance, and to find the optimal inspection interval. Numerical examples are provided to illustrate the proposed optimization model and to compare the two maintenance policies. Sensitivity analysis is conducted to analyze the effect of several cost components on the optimal levels and on the long run expected average cost rate.

**Keywords** Condition-based maintenance · Spare part provisioning · Multi-unit system · Production system · Semi-Markov decision process

## 1 Introduction

Most production systems are subject to deterioration with usage and age which leads to their failure. Failure of such systems incurs high replacement cost and the cost of lost demand due to their unavailability. Maintenance expenses are estimated to be in the range of 15–

---

✉ Viliam Makis  
makis@mie.utoronto.ca

Nooshin Salari  
nsalari@mie.utoronto.ca

<sup>1</sup> Mechanical and Industrial Engineering Department, University of Toronto, Toronto, ON, Canada

40% (in the study conducted by Maggard and Rhyne (1992)), or in the range of 15–70% (Bevilacqua and Braglia 2000) of the total production cost. So, taking proper maintenance actions is important to increase the availability and reduce the operating costs of the system. Corrective maintenance (CM) or preventive maintenance (PM) actions are performed on deteriorating production systems (Wang and Pham 1999) to keep the unavailability of the system down, and to decrease the cost of lost production. CM is performed after a system's failure, however PM is performed before a system's failure to prevent severe deterioration or failure of the system. PM activities generally consist of cleaning, lubrication, adjustment, alignment and replacement of sub-systems and components that are not functioning properly (Moghaddam and Usher 2011). Usually, PM is more effective than CM, as it can improve system performance by preventing or reducing unpredictable failure of the system resulting in its higher availability.

PM can be categorized into two classes: time-based preventive maintenance (TBPM) and condition-based preventive maintenance (CBM) (Wang et al. 2009). TBPM sets a periodic interval to perform PM regardless of the health status of a physical asset (Jardine et al. 2006). For example, Yeh et al. (2010) discussed an application of age-based maintenance policy for a leased equipment within the lease period, where PM decision was made considering a threshold for age. CBM suggests the required maintenance actions according to the actual condition of the system by means of degradation monitoring and failure prediction (Rahmati et al. 2018). Different CBM optimization models have been proposed in the literature, however, most of the reported CBM models are applicable to single unit systems (Dieulle et al. 2003; Wang 2000; Jiang et al. 2012; Niu et al. 2016).

Various maintenance models of multi-unit systems have been proposed for different system designs. For example, the problem of joint determination of a cost-optimal inspection and replacement policy for a deteriorating, complex multi-component manufacturing system is addressed in Ahmadi (2014). Authors assumed two possible states during operation: normal and degraded states. The system state is modeled using a proportional intensity model incorporating a damage process and a virtual age process generated by repair. The long-run average cost per unit time is derived by applying renewal reward process approach.

Group replacement policies can be implemented in a real production setting, where all units are as-good-as-new after maintenance (Barron 2015). A significant contribution has been made by Barron (2015), where the author presented three group replacement policies for a multi-component repairable, cold standby system. It is assumed that failure time has a phase-type (PH) distribution, and the long-run average cost rate was derived by applying renewal process.

A  $k$ -out-of- $n$  system consists of  $n$  identical units which can operate if at least  $k$  of the units operate (Asadi and Bayramoglu 2006). Due to the existence of economic dependencies among the units of a  $k$ -out-of- $n$  system, the combination of the optimal maintenance decisions for the units is not optimal for the whole system (Nicolai and Dekker 2008). There are several papers studying  $k$ -out-of- $n$  systems (Barron 2018; Eruguz et al. 2017; Bohlin and Warja 2015). Studying the availability and reliability of such systems has received a great deal of attention. Availability of  $R$ -out-of- $N$  repairable system is studied in Barron et al. (2004). The authors applied Markov renewal theory and semi-regenerative processes to formulate the availability of such a system. Recently, Barron and Yechiali (2017) developed another preventive maintenance policy for a 1-out-of- $N$  repairable system by applying dynamic programming, where the lifetimes of units follow a discrete phase-type distribution.

Two important special cases of  $k$ -out-of- $n$  systems are series and parallel systems corresponding to  $k = n$  ( $n$ -out-of- $n$ ) and  $k = 1$  (1-out-of- $n$ ), respectively. If a multi-unit system has a series configuration, the whole system stops operating whenever one of the units fails

(Zhou et al. 2009; Rao and Bhadury 2000; Moghaddam and Usher 2011). A parallel multi-unit system is a system that fails to operate if all the components stop operating (Tian et al. 2011; Keizer et al. 2018).

Most models of  $k$ -out-of- $n$  systems have focused on the reliability assessment of the system rather than on the development of maintenance policies for such systems. The multi-unit system considered in this paper is related to  $k$ -out-of- $n$  systems, but we are considering  $n$  independent units working in parallel (there are no stand-by units) and we propose maintenance policies for such a system. Applying proper maintenance policy is an effective way to keep the system in a good condition, but a common assumption in related, previously published papers is that spare parts are available at all times. This assumption reduces the complexity of maintenance problems, but it is not realistic for the majority of the real-world problems.

Assuming a system or its units to be either operating or failed is not applicable to many real world systems as the system or its units might be degraded but not failed (Kontoleon and Kontoleon 1974). Categorizing the system health state into three states has been considered in several papers (Kontoleon and Kontoleon 1974; Jia et al. 2016; Khaleghei and Makis 2015, 2016; Salari and Makis 2017b). For instance, Eryilmaz and Xie (2014) studied marginal and joint survival functions for the lifetime of two different three-state  $k$ -out-of- $n$  :  $G$  systems. Authors considered a system consisting of  $n$  components where each component could be in three different states: perfect functioning, partial performance and complete failure. In a research presented in Jia et al. (2016), authors considered an  $n$ -unit system, where each unit has three states: perfectly working, deteriorated and complete failure. Authors applied Markov theory and aggregated stochastic process theory to get reliability indices. We will show a detailed development for 3 states, that can be extended to a general case of  $n$  states (see “Appendix C”). We assume that deterioration process of each unit has two working states and a failure state which is absorbing. Assuming only two operational states is sufficient for CBM (e.g. Jafari et al. 2018) in most practical applications (see also the description in “Appendix C” on extending the model to a model with more than three states).

Most production systems are subject to high costs due to down time, and it is essential for these systems to have high availability. One efficient way of increasing availability and reducing the total cost is to jointly plan the maintenance actions and spare parts ordering. Spare parts are common inventory stock items which are required for maintenance of systems (Kennedy et al. 2002). Having a sufficient amount of spare parts at the right time when a maintenance decision is made is a crucial inventory control problem. It is therefore essential for a production manager to combine maintenance decisions with the spare parts ordering decisions.

Joint maintenance and inventory models have been introduced and studied recently by some researchers. Wang (2011) presented a joint spare part and maintenance inspection optimization model using the delay-time concept. A two-stage failure process characterized by the delay-time concept is used to model the failure and inspection processes. Optimal ordering quantity, ordering interval, and inspection interval are obtained by minimizing the long run expected average cost per unit time. An algorithm is proposed to determine the optimal decision process using a combination of analytical and enumeration approaches. Keizer et al. (2017) proposed a joint optimization of CBM and spares planning for multi-unit systems. A Markov decision process is applied to formulate the model. A decision is made at the beginning of each time unit whether or not to replace some units and to determine the number of spare parts to order.

Spare parts may be very expensive, and on the other hand, stocking is limited by space and cost. One effective strategy is the just-in-time provision of spare parts. The idea of just-in-time

provision of spare parts is to order and deliver spare parts whenever required instead of keeping them in stock permanently (Lanza et al. 2009). Just-in-time spare parts provisioning and maintenance optimization is studied in Lanza et al. (2009). Authors assumed that breakdowns of production can incur high cost when the required spare part is missing, so just-in-time strategy which depends on both reliability and economical aspects was applied in the model. A method was presented to calculate the optimal time to perform preventive maintenance and spare part provisioning using a stochastic optimization algorithm. This assumption of incurring high cost when spare parts are not available is a reasonable assumption for many production systems. In this paper, we consider just-in-time provision of spare parts as keeping spares in stock is not feasible and cost effective for many systems.

Based on our review of maintenance models for multi-unit production systems, no research papers have been published considering the combination of CBM, economic dependence, and spare parts provisioning, specially just-in-time concept for a multi-unit production system. There are papers studying the CBM and demand satisfaction for a production system. For example, Salari and Makis (2017a) proposed a CBM model for a multi-unit parallel system subject to deterioration. They considered production level of the system as a threshold for initiating preventive maintenance. They focused on demand satisfaction, however, they did not consider spare part provisioning in their model. We extend this model to a more complex model by considering two kinds of decisions, namely decision regarding spare parts ordering applying JIT approach, and also a maintenance decision. The other major difference is that the decision to perform maintenance in Salari and Makis (2017a) depends on the production rate of the system, whereas in our model the decision to perform maintenance depends on the number of failed units in the system. There are also papers studying the CBM and spare part provisioning for a multi-unit system subject to deterioration. For example, de Smidt-Destombes et al. (2006) considered a single  $k$ -out-of- $n$  system with deteriorating units and hot standby redundancy, where each unit had three possible observable states, namely as-good-as-new, degraded and failed. The conditions to initiate a maintenance, the stock level determination, the repair capacity and the repair priority settings are considered as the variables to control the availability. The authors presented two approximate methods to analyze the relation between the system availability and the control variables. Authors did not include production rates of the units and the demand satisfaction requirement in the model. They focused mainly on the availability of the system rather than studying the optimal maintenance policy for such system. In particular, although CBM and spare part provisioning have been considered for multi-unit systems, those studies did not consider the impact of deterioration level on the production rate of the units of the system and they assumed that the demand is always satisfied, which is not realistic.

The main objective of this research is to develop a joint model for CBM and just-in-time spare parts provisioning for a multi-unit production system, focusing on satisfying of the demand for the system. We consider a multi-unit parallel system with deteriorating units, where deterioration process of each unit is a three-state continuous time homogeneous Markov chain with two working states and a failure state. Examples of such systems are energy systems such as solar panels or wind turbine farms, or production systems with a large number of machines. To model the effect of deterioration on the production rate of the system, we consider variable production rates of the units depending on their states. An example of application of the model considered in this paper is the maintenance of Active Phased Array Radar (APAR) described in de Smidt-Destombes et al. (2004). System under study in de Smidt-Destombes et al. (2004) is a radar which consists of several transmit and receive elements. Authors assumed that a certain percentage of the total number of elements is allowed to fail, without losing the function of the specific radar face. Set-up costs for

maintenance are considered to be high, so maintenance cannot be performed upon each element failure. This model is considered as a  $k$ -out-of- $N$  system, where system is operable when at least  $k$  units are operating. A special case  $k = 1$ , that is system is operable if at least 1 unit is operating, is similar to our proposed model.

Another application of the proposed model is in the area of fleet maintenance (El Moudani and Mora-Camino 2000), where the aim is to establish a fleet maintenance plan with maintenance contractors. Fleet of the vehicles is assumed as the system where each unit (bus, aircraft, etc.) operates independently and can deteriorate by usage and age.

The main contributions of this paper can be summarized as follows:

- We propose a new joint CBM and spare part provisioning model for a multi-unit parallel production system applying just-in-time ordering concept. No analytical research has been done on the joint maintenance and spare part provisioning modeling for multi-unit systems where units are working in parallel.
- Units of the system are subject to deterioration, and different production rates are considered in different working states. None of the previous studies in the area of maintenance modeling of multi-unit production systems presented an analytical development for CBM of multi-unit parallel production systems subject to deterioration.
- Demand satisfaction is another interesting contribution of this paper, which makes the modeling more difficult especially for the calculation of the expected cost.
- Applying semi-Markov decision process (SMDP) approach to formulate and solve jointly the maintenance and spare parts ordering decision problem for the multi-unit system described above using just-in-time concept, which is a new contribution to maintenance modeling and control research.

## 2 Problem assumptions

The system under study is defined by the following assumptions:

1. There is a production facility (e.g. wind farm), composed of several units (e.g. wind turbines) which are subject to deterioration.
2. The deterioration level of each unit and the spare parts inventory level become known through periodic inspections at discrete time epochs  $\{k\Delta : k = 1, 2, \dots\}$ , where  $\Delta$  is a fixed inspection interval.
3. An inspection reveals the exact state of the units. At each inspection time, an inspector is sent to the field to inspect the turbines, deterioration level of each unit is checked, and the number of units in each state and the inventory level are recorded.
4. Deterioration process  $\{X_t\}_{t \geq 0}$  of each unit is modeled as a three-state continuous time homogeneous Markov chain with two working states  $O = \{0, 1\}$  and a failure state  $F$  which is absorbing,  $\Omega = \{0, 1, F\}$ . State 0 means that the unit is as-good-as-new, state 1 means that the unit is in a warning state, and state  $F$  is a failure state.
5. In our model, we assume that when the unit leaves state  $i$ , it next enters state  $j$  with a known transition probability function  $P_{ij}(t)$ . Transition from state  $i$  to  $j$  is meaningful if  $j \geq i$  since deterioration is increasing with time.
6. System state space description keeps track of the number of units in each state and the inventory level of the spare parts at each inspection time. The state space for the whole system can be defined as  $W = \{(n_F, n_1, S) | n_1 + n_F \leq N, n_1, n_F, S \geq 0\}$ , where  $n_F, n_1$  represent the number of units in the failure state and warning state, respectively, and  $S$  represents the on-hand inventory.

7. Units produce with different production rates in states 0 and 1.
8. We consider known demand and production rates depending on the states. Units are capable of producing at different production rates, and we study the effect of deterioration on the total production rate in the model.
9. Unsatisfied demand incurs high cost of lost demand and excess production in an interval can be sold at a market price.
10. At each inspection time, there are several possible actions. A decision is made whether or not to initiate maintenance and on the ordering of spare parts. Maintenance is initiated if the number of failed units in the system is found to be greater than or equal to  $R$  units. More specifically, maintenance time of the system is defined as the first inspection time when the number of failed units exceeds the maintenance threshold  $R$  ( $R > 0$ ). Hence, the action space is defined as:  $A = \{M, O\}$ .  $M$  and  $O$  can be 0 or 1, where  $M = 1$  represents initiation of maintenance, and  $M = 0$  represents no maintenance.  $O = 1$  means that an order is placed, and  $O = 0$  means that no order is placed. Two maintenance policies are considered in this paper:

- First maintenance policy prescribes replacement of the failed units when at an inspection time there are at least  $R$  failed units in the system, where  $R$  is a decision variable. We assume that there is a set-up time to prepare and send a crew to perform maintenance. We also assume that sufficient number of repairmen is available so that all units can be maintained simultaneously. However the maintenance cost for both policies depends on the number of units maintained.
- Second maintenance policy prescribes replacement of failed units and preventive maintenance of units in state 1, when at an inspection time there are at least  $R$  failed units in the system. If the decision is to perform maintenance, failed units are replaced and there is an opportunity to perform preventive maintenance on the units in the warning state. After maintenance is performed, all the units are working in state 0.

Thus, the maintenance action is decided by comparing the number of failed units with the threshold  $R$ , and maintenance is initiated if at a decision epoch ( $k\Delta$ ),  $n_F \geq R$ , where  $(n_F, n_1, S)$  is the state of the system at this epoch.

11. We assume that the total time to perform corrective or preventive maintenance,  $T_R$  (or  $T_P$ ) have known density functions  $f_{T_R}(t)$  (or  $f_{T_P}(t)$ ). We assume that  $T_P \geq_{st} T_R$  so that  $F_{T_R}(t) \geq F_{T_P}(t)$ .

**Remark 1** We say that the random variable  $X$  is stochastically larger than the random variable  $Y$ , written as  $X \geq_{st} Y$ , if, for all  $t$ :

$$P(X > t) \geq P(Y > t)$$

12. Both preventive and corrective maintenance actions are perfect i.e. they will restore the units to the as-good-as-new condition.
13. To perform replacement, spare parts are required and all units share the same pool of spares. Spare parts ordering depends on the number of failed units and the number of available spare parts in the system at each inspection time. We define a threshold to order spare parts as  $R - x$ ,  $x \geq 1$  where both  $R$  and  $x$  are decision variables and  $R$  is the threshold to perform maintenance. Assume that the state of the system at an inspection time is  $(n_F, n_1, S)$ . If  $R > n_F \geq R - x$ ,  $x \geq 1$ , an order of size  $R - S$  for spares is placed and the total number of spares is equal to  $R$  after the delivery. If  $n_F \geq R$ , maintenance is initiated, and the missing spares are obtained through an emergency order.

14. We assume that the space and the budget for holding the spare parts is limited, so JIT ordering strategy is applied in the model. The system starts operating and there is no spare part available in the system. When the number of failed units reaches  $R - x$ , an order for  $R$  spare parts is placed. When the number of failed units reaches or exceeds  $R$ , maintenance is initiated. In this case, the system requires at least  $R$  spare parts (which are available if the order was placed before), and if more than  $R$  spare parts are required, an emergency order is placed.

**Remark 2** The number of available spare parts in the system at each inspection time can be 0 or  $R$  ( $S \in \{0, R\}$ ), due to just-in-time ordering.

15. There are two types of ordering for spare parts: a regular order with a fixed lead time and an emergency order with a negligible lead time and a higher ordering cost compared to regular ordering.

16. The following cost components are considered in the model:

- $C_I$ : Inspection cost per unit at each inspection time
- $C_F$ : Failure replacement cost of each unit
- $C_P$ : Preventive maintenance cost of each unit
- $C_D$ : Cost of lost production per unit time, when the total production rate of the system is below the demand rate  $D$
- $C_E$ : Profit rate from the excess production, when the total production rate of the system is higher than the demand rate  $D$
- $C_{KM}$ : Set-up cost of performing maintenance
- $C_{KI}$ : Set-up cost of performing inspection
- $C_H$ : Holding cost rate per spare part
- $C_O$ : Regular ordering cost per unit
- $C_{es}$ : Emergency ordering cost per unit

Figure 1 shows the decision at each inspection time depending on the number of failed units and the number of available spare parts in the system.

Two types of ordering of spares is considered in the model. A regular order with a fixed lead time  $T_o$  and an emergency order with a negligible lead time and a higher ordering cost per spare part. We assume that the regular order lead time  $T_o$  is smaller than a reasonable inspection interval  $\Delta$  so that the regular order will arrive by the next inspection time.

If the number of failed units at an inspection time exceeds  $R$  units (maintenance is required), maintenance crew is sent to perform maintenance. If there is not enough spares available in the system, emergency order is placed to cover the required number of spares.

The decision variables in the joint maintenance and inventory optimization model include the number of failed units to initiate maintenance ( $R$ ), ordering time, which is the first inspection time when the number of failed units exceeds  $R - x$ ,  $x \geq 1$ , and inspection interval ( $\Delta$ ).

Our objective is to minimize the long-run expected average cost per unit time.

### 3 State definition

It is assumed that the sojourn times in states 0 and 1 are exponentially distributed. Transition probability from state  $i$  to  $j$  for each unit is equal to  $P_{ij}(t)$  for  $t \geq 0$  which is obtained by solving the Kolmogorov backward differential equations. Details can be found in “Appendix A”.

**Fig. 1** Decision made at each inspection time depending on the number of failed units and the number of available spare parts

		S	
		0	R
$n_F$	0	Do nothing	Do nothing
	...		
	$R - x - 1$	Place a regular order	Do nothing
	$R - x$		
	...		
$R - 1$	Perform maintenance	Perform maintenance	
R			
...			
N			

States of the system are defined as follows:

1. State (0, 0, 0): initial state when all units are in state 0 and there is no spare part available in the system.
2. State ( $n_F, n_1, S$ ): there are  $n_F$  units in the failure state,  $n_1$  units in state 1,  $n_0 = N - n_1 - n_F$  units in state 0, and  $S$  spare parts are available in the system.

Consider a system whose state is observed at inspection epochs and at each epoch a decision is made and costs are incurred as a consequence of the decision made. This controlled dynamic system is called an SMDP when the following Markovian properties are satisfied: if at a decision epoch the action  $a$  is chosen in state  $i$ , then the expected time and expected cost until the next decision epoch and the state at the next epoch depend only on the present state  $i$  and the subsequently chosen action  $a$ . So, sojourn time, expected cost, and the state at the next epoch are independent of the past history of the system (Tijms 1986).

SMDP is determined by the following quantities:

- $P_{ij}$  = the probability that the system will be in state  $j \in W$  at the next decision epoch given the current state is  $i \in W$ . Note that  $P_{ij}$  depends on the time till the next decision epoch.
- $\tau_i$  = the expected time until the next decision epoch given the current state is  $i \in W$ .
- $C_i$  = the expected cost incurred until the next decision epoch given the current state is  $i \in W$ .

From the theory of SMDP for given control limits ( $R, x$ ) and inspection interval  $\Delta$ , the long-run expected average cost per unit time  $g(\Delta, R, x)$  can be obtained by solving the following linear equations (Tijms 1986):

$$V_m = C_m - g(\Delta, R, x) \cdot \tau_m + \sum_{k \in W} P_{m,k} \cdot V_k$$

$$V_j = 0 \text{ for an arbitrarily selected single state } j \in W, \tag{1}$$

where the quantities  $V_m$  are the relative values dependent on the control limits when starting in state  $m \in W$ . The optimal values of the control limits and inspection interval can be obtained as follows:

$$(\Delta^*, R^*, x^*) = \underset{\substack{\Delta > 0 \\ R > 0 \\ x > 0}}{\text{arg inf}} g(\Delta, R, x)$$



These optimal control parameters minimize the long run expected average cost  $g(\Delta^*, R^*, x^*)$  which is obtained by solving the system of linear equations (1) considering different combinations of  $\Delta$ ,  $R$ , and  $x$ .

Next, we derive the transition probabilities, the expected costs and sojourn times for each state of the system.

### 3.1 Computing the transition probabilities for the system

In this section, we derive the transition probability matrix for a system consisting of  $N$  units. Transition probabilities for each state depend on the decision made in that state.

If the decision at an inspection time is to have no order and no maintenance, the system can transit from state  $(n_F, n_1, S)$  to state  $(n'_F, n'_1, S')$ .

**Remark 3** Assume that the state of the system at the current inspection time is  $(n_F, n_1, S)$ , and the state of the system at the next inspection time is  $(n'_F, n'_1, S')$ . If the decision is to do nothing at the current inspection time, we can write:

$$n'_F \geq n_F, n'_0 = N - n'_F - n'_1, n_0 = N - n_F - n_1, n'_0 \leq n_0, S = S'.$$

**Theorem 1** *The transition probability function when the decision is to continue operation (no maintenance or ordering is initiated) is given by the following equations:*

$$P_{(n_F, n_1, S)(n_F+i, n'_1, S)}(\Delta) = \begin{cases} \sum_{j=n_1-n'_1}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^{n'_1-n_1+j} & n'_1 < n_1 \\ \quad \times P_{0F}(\Delta)^{i-j} P_{11}(\Delta)^{n_1-j} P_{1F}(\Delta)^j & \\ \sum_{j=0}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{j} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^j & n'_1 = n_1 \\ \quad \times P_{0F}(\Delta)^{i-j} P_{11}(\Delta)^{n_1-j} P_{1F}(\Delta)^j & \\ \sum_{j=0}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^{n'_1-n_1+j} & n'_1 > n_1, \\ \quad \times P_{0F}(\Delta)^{i-j} P_{11}(\Delta)^{n_1-j} P_{1F}(\Delta)^j & \end{cases} \quad (2)$$

where  $i = n'_F - n_F$ ,  $i \geq 0$  is the total number of units which failed in the system during the next time interval  $\Delta$ , and  $j$  is the number of units which failed from state 1 during that time interval.

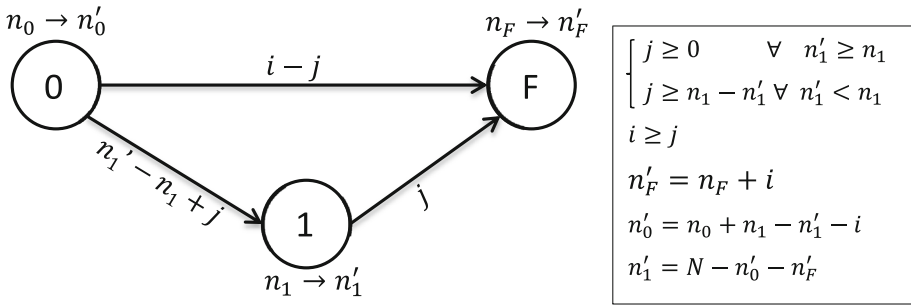


Fig. 2 Number of units changing state in a time interval of length  $\Delta$

**Proof** Transition from state  $(n_F, n_1, S)$  to state  $(n_F + i, n_1', S)$  occurs when  $i$  units fail in the next time interval of length  $\Delta$ . These failed units can be from state 0, or state 1, or both, depending on  $n_1'$ .

Figure 2 shows the number of units which change their state in the next time interval of length  $\Delta$ .

Define  $j$  as the number of units which fail from state 1 in the time interval of length  $\Delta$ , where  $j \leq i$ .

From the total  $i$  units which fail during the next time interval of length  $\Delta$ , we assume that  $j$  units have failed from state 1 and the remaining  $i - j$  units have failed from state 0.

If  $j$  units fail from state 1,  $n_1' - n_1 + j$  units must deteriorate to state 1, and the other units do not deteriorate or fail (stay in the same state) during time  $\Delta$ . The values of  $n_1'$ ,  $n_1$ , and  $i$  are known based on the current state  $(n_F, n_1, S)$  and the next state  $(n_F + i, n_1, S)$  of the system at the next inspection interval, but the values of  $j$  can vary.

Two different cases can occur which are summarized as follows:

- Consider the case  $n_1' < n_1$  where at least  $n_1 - n_1'$  units can fail from state 1 in the time interval of length  $\Delta$ , in this case  $\min\{j\} = n_1 - n_1'$ . The maximum number of units which can fail from state 1 depends on the value of  $i$ . As  $j \leq i$ , the number of units which can fail from state 1 cannot exceed  $i$ . On the other hand, the maximum number of units available in state 1 which can fail is equal to  $n_1$ , so that  $\max\{j\} = \min\{n_1, i\}$ .
- If  $n_1' \geq n_1$ , the minimum number of units which can fail from state 1 is equal to 0, so the minimum value of  $j$  is equal to 0 and the maximum value of  $j$  is the same as in the previous case.

□

If at an inspection time the number of failed units exceeds  $R$ , a decision is made to perform maintenance. If the number of failed units is greater than the number of available spares, an emergency order is placed to have the required spares on hand.

**Theorem 2** *The transition probability when the decision is to perform maintenance is given by the following formulas:*

– For the first maintenance policy:

$$P_{(n_F, n_1, S)(0, n'_1, 0)}(T_R) = \begin{cases} \sum_{i=n_1-n'_1}^{N-n_F-n'_1} \sum_{j=n_1-n'_1}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \left( \int_0^\infty P_{00}(u)^{n'_0} \right. & n'_1 < n_1 \\ \quad \times P_{01}(u)^{n'_1-n_1+j} P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j \times f_{T_R}(u) du & \\ \sum_{i=0}^{N-n_F-n'_1} \sum_{j=0}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{j} \left( \int_0^\infty P_{00}(u)^{n'_0} P_{01}(u)^j & n'_1 = n_1 \\ \quad \times P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j \times f_{T_R}(u) du & \\ \sum_{i=0}^{N-n_F-n'_1} \sum_{j=0}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \left( \int_0^\infty P_{00}(u)^{n'_0} & n'_1 > n_1 \\ \quad \times P_{01}(u)^{n'_1-n_1+j} P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j \times f_{T_R}(u) du & \end{cases} \quad (3)$$

Transition probability in this case depends on the total number of units which failed in the system ( $i$ ), and the number of units which failed from state 1 ( $j$ ) during  $T_R$ .

– For the second maintenance policy:

$$P_{(n_F, n_1, S)(0, 0, 0)}(T_P) = 1 \quad (4)$$

**Proof** • First maintenance policy prescribes corrective replacement of failed units and the units which fail during  $T_R$ . After performing maintenance, all failed units are replaced, there are no spare parts available in the system, and the state of the system after maintenance is  $(0, n'_1, 0)$ , where  $0 \leq n'_1 \leq N - n_F$ .

During  $T_R$ , which has a known density function  $f_{T_R}(t)$ , units in states 0 and 1 are still working and their state may change. The transition probability for the first maintenance policy is the same as that in Eq. 2, however lead time is a random variable and the number of units which transit to failure state in the system ( $i$ ) during  $T_R$  is not fixed. Proof is the same as the proof of Theorem 2, but we need to condition on the random time  $T_R$ .

It is known that  $i \geq j$ , so the minimum number of units that can fail in the system is equal to the minimum number of units which fail from state 1, so  $\min\{i\} = \min\{j\}$  for all the cases.

There are  $N - n_F$  operational units in the system in state  $(n_F, n_1, S)$  which can fail by the end of  $T_R$ . We know that state of the system after  $T_R$  is  $(0, n'_1, 0)$  which means that from  $N - n_F$  operational units in the system,  $n'_1$  units are working in state 1 after  $T_R$ . So the maximum number of units which can fail in the system during  $T_R$  is equal to  $N - n_F - n'_1$ , that is  $\max\{i\} = N - n_F - n'_1$ .

- For the second maintenance policy, all the failed units are replaced correctively and units in state 1 are maintained preventively by the end of  $T_P$ . After  $T_P$ , all the units are in state 0, and there are no spares available in the system. Therefore, transition from state  $(n_F, n_1, S)$  to state  $(0, 0, 0)$  occurs with probability one.

□

If at an inspection time the state of the system is  $(n_F, n_1, S)$ , where  $R - x \leq n_F < R$  and  $S < R$ , a decision is made to place an order (but not to perform maintenance). The order of size  $R - S$  is placed which will arrive before the next inspection time ( $T_o < \Delta$ ).

**Theorem 3** *The transition probability function when the decision is to place a regular order for spare parts is given as follows ( $\Delta > T_o$ ):*

$$P_{(n_F, n_1, S)(n_F+i, n'_1, R)}(\Delta) = \begin{cases} \sum_{j=n_1-n'_1}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^{n'_1-n_1+j} & n'_1 < n_1 \\ \quad \times P_{0F}(\Delta)^{i-j} P_{11}(\Delta)^{n_1-j} P_{1F}(\Delta)^j \\ \sum_{j=0}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{j} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^j & n'_1 = n_1 \\ \quad \times P_{0F}(\Delta)^{i-j} P_{11}(\Delta)^{n_1-j} P_{1F}(\Delta)^j \\ \sum_{j=0}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^{n'_1-n_1+j} & n'_1 > n_1 \\ \quad \times P_{0F}(\Delta)^{i-j} P_{11}(\Delta)^{n_1-j} P_{1F}(\Delta)^j \end{cases} \tag{5}$$

**Proof** Ordering spare parts does not affect the deterioration process of the units. The regular order arrives before the next inspection time, so that the state of the system at that time is  $(n_F + i, n'_1, R)$ , where the number of spare parts increases to  $R$ . The transition probability from state  $(n_F, n_1, S)$  to state  $(n'_F, n'_1, R)$  in the next time interval of length  $\Delta$  is the same as the transition probability from state  $(n_F, n_1, S)$  to state  $(n'_F, n'_1, S)$  in Eq. 2, but the number of spares increased to  $R$  by the time  $\Delta$ . □

### 3.2 Derivation of the expected costs

The expected cost at each inspection time is given by the sum of the inspection cost, maintenance cost (if maintenance is performed), ordering cost (if an order is placed), profit from excess production, cost of lost demand, and holding cost (for on hand spare parts).

We define  $PR = p_0 \times n_0 + p_1 \times n_1$  ( $n_0 = N - n_F - n_1$ ) as the total production rate of the system at each time which depends on the number of units in states 0 and 1. If the total production rate of the system ( $PR$ ) drops below the demand rate ( $D$ ) at time  $t$ , shortage occurs with a cost of lost demand depending on the amount of shortage and sojourn time in that state.

Let us define  $C_{rt}(n_F, n_1, S)$  as the cost rate (cost rate of lost demand or cost rate of excess production) of the system that is charged whenever the system is in state  $(n_F, n_1, S)$ , which is written as follows:

$$C_{rt}(n_F, n_1, S) = C_D \times \text{Max}\{0, D - ((N - n_F - n_1)p_0 + n_1p_1)\} - C_E \times \text{Max}\{0, ((N - n_F - n_1)p_0 + n_1p_1) - D\} \tag{6}$$

We define  $CR(t)$  as the cost of lost demand (or profit from excess production) during a time interval of length  $t$ . If the decision is not to perform maintenance,  $t = \Delta$ , and if the decision is to perform maintenance,  $t$  is a random variable with a density function  $f_{T_R}$  (or  $f_{T_P}$ ).

**Theorem 4** *The expected cost of lost demand (or profit from excess production) during the time interval  $t$  given the current system state is given as follows:*

$$E[CR(t)|(n_F, n_1, S)] = \sum_{i=x}^{N-n_F-n_1} \sum_{n'_1=\max\{n_1-i, 0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i, n_1\}} C_{rt}(n_F + i, n'_1, S')$$

$$\begin{aligned} & \times \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \int_0^t P_{00}(u)^{n'_0} P_{01}(u)^{n'_1-n_1+j} \\ & \times P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j du \end{aligned} \tag{7}$$

where  $x = 0$  for  $n'_1 \geq n_1$ , and  $x = n_1 - n'_1$  for  $n'_1 < n_1$ .

**Proof** Suppose that for the continuous-time Markov chain  $\{X_t\}$  the rates  $v_i = \sum_{j \neq i} q_{ij}$  (of the units of the system) are bounded in  $i \in \Omega$ . Then:

$$\begin{aligned} E[CR(t)|(n_F, n_1, S)] &= \sum_{(n_F+i, n'_1, S)} C_{rt}(n_F+i, n'_1, S') E_{(n_F, n_1, S)(n_F+i, n'_1, S')}(t) \\ &= \sum_{(n_F+i, n'_1, S)} C_{rt}(n_F+i, n'_1, S') \times \int_0^t P_{(n_F, n_1, S)(n_F+i, n'_1, S')}(u) du, \end{aligned}$$

where  $E_{(n_F, n_1, S)(n_F+i, n'_1, S')}(t)$  is the expected amount of time that the process is in state  $(n_F+i, n'_1, S)$ , and the summation is over all the reachable states from the current state  $(n_F, n_1, S)$ . Cost rate of  $C_{rt}(n_F+i, n'_1, S')$  is given by the following formula:

$$\begin{aligned} C_{rt}(n_F+i, n'_1, S') &= C_D \times \text{Max}\{0, D - ((N - (n_F+i) - n'_1)p_0 + n'_1 p_1)\} \\ &\quad - C_E \times \text{Max}\{0, ((N - (n_F+i) - n'_1)p_0 + n'_1 p_1) - D\} \end{aligned}$$

□

Expected cost at each state can be derived based on the decision made in that state. To decide about the required action, we compare the number of failed units in each state with the critical levels  $R$ , and  $R - x$ .

The total expected cost of the system in state  $(n_F, n_1, S)$  when the decision is not to perform maintenance and not to place an order is equal to:

$$C_{(n_F, n_1, S)}(\Delta) = C_{KI} + C_I(N - n_F) + C_H \cdot S \cdot \Delta + E(CR(\Delta|(n_F, 1, S))) \tag{8}$$

The total expected cost of the system in state  $(n_F, n_1, S)$  when the decision is not to perform maintenance but to place a regular order is equal to:

$$\begin{aligned} C_{(n_F, n_1, S)}(\Delta) &= C_{KI} + C_I(N - n_F) + C_O(R - S) \\ &\quad + C_H(S \cdot T_o + R(\Delta - T_o)) + E(CR(\Delta|(n_F, 1, S))) \end{aligned} \tag{9}$$

If at an inspection time, decision is made to perform maintenance, maintenance team is sent to the field to initiate maintenance with the total maintenance time equal to  $T_R$  (or  $T_P$ ). If there is a need for more spare parts (number of failed units is greater than the number of available spare parts) an emergency order is placed. Expected cost of emergency ordering and maintenance cost can be derived by conditioning on the number of units in states  $F$  and  $1$  at the end of  $T_R$  (or  $T_P$ ).

Note that if the decision is to perform maintenance in state  $(n_F, n_1, S)$ , after  $T_R$  (or  $T_P$ ) there are no available spare parts in the system ( $S = 0$ ). Based on the idea of just-in-time ordering  $S = 0$  or  $R$  at each inspection time. If at an inspection time  $n_F \geq R$  maintenance is initiated, and during  $T_R$  the number of failed units can increase to  $n_F + i \geq R, i \geq 0$ . The maximum number of available spare parts in the system is equal to  $R$  that is less than or equal to the number of failed units in the system ( $S \leq R \leq n_F + i$ ). All the failed units are replaced correctively by the end of  $T_R$  (or  $T_P$ ) and at least  $n_F + i \geq S$  spare parts are required in the system, so all the spare parts will be used and  $S = 0$  by the end of  $T_R$  (or  $T_P$ ).

For the first maintenance policy when the number of failed units exceeds  $R$  at an inspection time, failed units are correctively replaced by the end of  $T_R$ . For the second maintenance policy, when the total number of failed units at an inspection time exceeds  $R$ , failed units are correctively replaced and units in state 1 are preventively maintained by the end of  $T_P$ .

**Theorem 5** *The total expected cost of performing maintenance incurred in state  $(n_F, n_1, S)$ , when the decision is to initiate maintenance is as follows:*

– *First maintenance policy:*

$$\begin{aligned}
 E(C|T_R, (n_F, n_1, S)) &= C_{KM} + C_I(N - n_F) + \int_0^\infty C_H \cdot S \cdot f_{T_R}(u) du \\
 &+ \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i, 0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \\
 &\times \left[ \left( C_{es}(n_F+i-S) + C_F(n_F+i) \right) \int_0^\infty P_{00}(u)^{n'_0} P_{01}(u)^{n'_1-n_1+j} \right. \\
 &\times P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j f_{T_R}(u) du \\
 &+ C_{ri}(n_F+i, n'_1, S') \int_0^\infty \int_0^u P_{00}(t)^{n'_0} P_{01}(t)^{n'_1-n_1+j} \\
 &\left. \times P_{0F}(t)^{i-j} P_{11}(t)^{n_1-j} P_{1F}(t)^j f_{T_R}(u) dt du \right] \tag{10}
 \end{aligned}$$

where  $x = 0$  for  $n'_1 \geq n_1$ , and  $x = n_1 - n'_1$  for  $n'_1 < n_1$ .

– *Second maintenance policy:*

$$\begin{aligned}
 E(C|T_P, (n_F, n_1, S)) &= C_{KM} + C_I(N - n_F) + E(C_H|T_P) \\
 &+ E(C_R|T_P) + E(C_{EM}|T_P) + E(C_{CM} + C_{PM}|T_P) \\
 &= C_{KM} + C_I(N - n_F) + \int_0^\infty C_H \cdot S \cdot f_{T_P}(u) du \\
 &+ \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i, 0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i, n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \\
 &\times \left[ \left( C_{es}(n_F+i-S) + C_F(n_F+i) + C_P(n'_1) \right) \int_0^\infty P_{00}(u)^{n'_0} \right. \\
 &\times \left( P_{01}(u)^{n'_1-n_1+j} P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j f_{T_P}(u) du \right) \\
 &+ C_{ri}(n_F+i, n'_1, S') \times \int_0^\infty \int_0^u P_{00}(t)^{n'_0} P_{01}(t)^{n'_1-n_1+j} \\
 &\left. \times P_{0F}(t)^{i-j} P_{11}(t)^{n_1-j} P_{1F}(t)^j f_{T_P}(u) dt du \right] \tag{11}
 \end{aligned}$$

where  $x = 0$  for  $n'_1 \geq n_1$ , and  $x = n_1 - n'_1$  for  $n'_1 < n_1$ .

**Proof** is given in ‘‘Appendix B’’.

□

### 3.3 Calculation of expected sojourn times

The expected sojourn time is the expected time until the next decision epoch if action  $a$  is chosen in the present state. So, the expected sojourn time in state  $(n_F, n_1, S)$ , when the decision is not to perform maintenance, is equal to the inspection interval  $\Delta$ , as the system state is revealed after  $\Delta$  time units (the next decision epoch is after the next  $\Delta$  time units). A transition to a different state may or may not occur during the next  $\Delta$  time units.

The expected sojourn time in state  $(n_F, n_1, S)$ , when the decision is to perform maintenance is equal to  $E(T_R)$  or  $E(T_P)$  for the first and the second maintenance policy, respectively.

## 4 Numerical example

In this section, we present the results of a numerical example to illustrate performance of the proposed maintenance policies. We obtain the optimal level to initiate maintenance, optimal level to place an order for spares and the optimal inspection interval for different values of the profit rate from excess production, and lost demand cost rate.

Salari and Makis (2017a) considered a wind farm consisting of 10 wind turbines, and they assumed that gearbox is subject to condition monitoring. We consider the same wind farm with 10 wind turbines working in parallel.

Deterioration process of each gearbox is described by a continuous time homogeneous Markov chain with two operating states  $\{0, 1\}$ , and a failure state  $\{F\}$  which is absorbing,  $\Omega = \{0, 1, F\}$ . The sojourn time in healthy state (state 0) has an exponential distribution with parameter  $\nu_0 = q_{01} + q_{02}$ , and the sojourn time in the warning state (state 1) has an exponential distribution with parameter  $\nu_1 = q_{12}$ .

We consider the weekly-based deterioration process of each gearbox given by Byon and Ding (2010).

$$P = \begin{bmatrix} 0.93 & 0.04 & 0.03 \\ 0 & 0.95 & 0.05 \\ 0 & 0 & 1 \end{bmatrix}$$

Using Eq. 13 we can derive the parameters of the exponential distributions corresponding to the sojourn time in healthy and warning states, respectively.

The unit of measurement of production rate of a wind turbine is  $kWh$ , so we readjust the parameters by changing the time unit from 1 week to 1 h (to be consistent with our example) and we obtain the following parameters:

$$q_{01} = 0.2619 \times 10^{-3}, \quad q_{02} = 0.1726 \times 10^{-3}, \quad q_{12} = 0.3036 \times 10^{-3}$$

We assume that each turbine is working in the wind farm with a rated capacity of 1.5 MW and capacity factors of 30% and 22% in states 0 and 1, respectively (Salari and Makis 2017a). These capacity factors result in the output of  $p_0 = 450$  kWh for the units in healthy state, and the output of  $p_1 = 330$  kWh for the units in warning state, respectively. We assume that the demand rate of the wind farm is  $D = 2000$  kWh, which is constant.

The maintenance cost parameters are given by: (Salari and Makis 2017a; Byon and Ding 2010)

$$C_F = \$78368, \quad C_P = \$8182, \quad C_I = \$50, \quad C_{KI} = 2000, \quad C_{KM} = \$6000$$

The parameters related to the spare part ordering are:

$$C_O = \$500, T_o = 150 \text{ (h)}, C_{es} = \$1000, C_h = \$2 \text{ per hour per spare part}$$

Set-up time to perform maintenance follows a gamma distribution with parameters  $\alpha$  and  $\beta$ :

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$$

We consider  $\alpha = 6, \beta = 0.1$  for the first maintenance policy (mean value of 60 h) and  $\alpha = 6, \beta = 0.08$  for the second maintenance policy (mean value of 75 h).

The average price of electricity to ultimate customers in residential sector in U.S. ranges from 9.56 to 29.39 cents per kWh (EIA 2019). Five different cost rates of lost demand are considered: 0.10, 0.2, 0.3, 0.4, and 0.5 \$/kWh (10, 20, 30, 40 and 50 cents per kWh). Profit rate from the excess production ( $C_E$ ) can vary depending on the ability to sell the excess production in the market, so we will find the optimal production rate to do maintenance for different values of  $C_E$ .

If at an inspection time the total number of failed units exceeds the optimal level ( $R^* - x^*$ ), a regular order of size  $R^*$  is placed and if the total number of failed units exceeds  $R^*$ , maintenance is initiated. Tables 1 and 2 present the optimal values and the expected average costs for different cost parameters of the first and the second maintenance policies, respectively (note that  $\Delta$  increased in 100h).

Results in Table 1 indicate that when  $C_D = 0.1$  and  $C_E = 0$  the optimal inspection interval is  $\Delta^* = 1600$ h. Optimal maintenance level is  $R^* = 6$ , which means that maintenance is initiated when the total number of failed units at an inspection time is greater than or equal

**Table 1** Optimal values for Policy 1 (corrective maintenance)

Lost demand cost rate	Profit rate	Optimal values			Expected average cost
		Inspection interval	# of failed units to do maintenance	# of failed units to order spares	
0.1	0	1600	6	5	140.36
	0.02	1100	6	5	125.70
	0.04	2500	2	1	106.97
0.2	0	900	5	4	149.59
	0.02	700	5	4	130.94
	0.04	1700	2	1	109.51
0.3	0	600	5	4	152.97
	0.02	600	5	4	133.67
	0.04	1500	2	1	110.67
0.4	0	500	5	4	155.34
	0.02	700	4	3	135.14
	0.04	1300	2	1	111.4
0.5	0	400	5	4	157.34
	0.02	700	4	3	136.10
	0.04	1200	2	1	111.93



**Table 2** Optimal values for Policy 2 (corrective and preventive maintenance)

Lost demand cost rate	Profit rate	Optimal values			Expected average cost
		Inspection interval	# of failed units to do maintenance	# of failed units to order spares	
0.1	0	1400	6	5	135.74
	0.02	3000	2	1	114.22
	0.04	1200	2	1	85.93
0.2	0	900	5	4	142.40
	0.02	2200	2	1	116.99
	0.04	1100	2	1	86.11
0.3	0	700	5	4	145.16
	0.02	1900	2	1	118.34
	0.04	1000	2	1	86.25
0.4	0	500	5	4	147.17
	0.02	1700	2	1	119.23
	0.04	900	2	1	86.30
0.5	0	800	4	3	148.94
	0.02	1600	2	1	119.86
	0.04	900	2	1	86.45

to 6 units. If at an inspection time  $5 \leq n_F < 6$  a regular order for  $6 - S$  spare parts (if  $S = 0$  order for 6 spare parts and if  $S = R^* = 6$  order for 0 spare parts) is placed ( $x^* = 1$ ,  $(R^* - x^*) = 5$ ). If at an inspection time  $n_F \geq 6$  (maintenance is initiated) and there are not enough spare parts available in the system, an emergency order is placed for the required spare parts. The corresponding expected average cost rate for the whole wind farm is equal to \$140.36 per hour. This high level for initiating maintenance is due to the low cost rate of lost demand and profit rate. When the lost demand cost rate is low and profit rate is low (in this case  $C_E = 0$ ), it is not profitable to perform maintenance regularly, so maintenance level is high as we expect. To analyze the effect of the profit rate on the optimal values, we obtain the optimal values for  $C_E = 0.02$  and  $C_D = 0.04$ .

For  $C_D \in \{0.1, 0.2\}$  by increasing profit rate from 0 to 0.02, the optimal inspection interval decreases but  $R^*$ , and  $x^*$  remain constant. When profit rate increases from 0.02 to 0.04, inspection interval increases but  $R^*$  decreases to a lower level of  $R^* = 2$ , and  $x^*$  remains constant ( $x^* = 1$ ).

When  $C_D = 0.3$ , optimal values do not change for the increase of  $C_E$  from 0 to 0.02. Next by the increase of  $C_E$  from 0.02 to 0.04,  $\Delta^*$  increases but  $R^*$  decreases to the lower value of  $R^* = 2$ , and  $x^*$  remains constant ( $x^* = 1$ ), which is the same behavior as in the previous cases.

When  $C_E \in \{0.4, 0.5\}$ , by the increase of profit rate  $\Delta^*$  increases,  $R^*$  decreases, and  $x^*$  remains constant ( $x^* = 1$ ).

Furthermore, results indicate that for a given lost demand cost rate by the increase of profit rate we can observe two behaviors:

1. Optimal inspection interval decreases (or remains constant),  $R^*$  and  $x^*$  remain constant, or

2. Optimal inspection interval increases (less frequent inspection is required) and this effect is balanced by the decrease in  $R^*$ , and  $x^*$  remains constant.

We note that  $x^*$  is not affected by  $C_E$  and  $C_D$ , as it depends mostly on the holding cost rate ( $C_H = 2$ ) that is assumed to be the same for all the cases.

We can also see from Table 1 that for a given lost demand cost rate by the increase of the profit rate, the long-run expected average cost rate of the wind farm decreases. However, for a given profit rate by the increase of the lost demand cost rate, the long-run expected average cost rate of the wind farm increases (as we expect).

Second maintenance policy assumes corrective replacement of the failed gearboxes and preventive maintenance of gearboxes in the warning state, when the total number of failed units at an inspection time is greater than or equal to an optimal level  $R^*$ . A regular order of size  $R^* - S$  for spare parts is placed when the number of failed units is greater than or equal to the optimal level  $R^* - x^*$ . We compute the optimal inspection interval, optimal level to initiate maintenance, and optimal level to place an order. The results are shown in Table 2.

Results of Table 2 indicate that for a given  $C_D$  by the increase of  $C_E$  from 0 to 0.02, optimal inspection interval increases and  $R^*$  decreases to  $R^* = 2$ . Next, by the increase of  $C_E$  from 0.02 to 0.04, optimal inspection interval decreases and  $R^*$  does not change ( $R^* = 2$ ). For all of these cases,  $x^*$  is not affected ( $x^* = 1$ ).

Compared to the previous policy, the decrease in the optimal level  $R^*$  is more drastic for Policy 2.

Policy 2 gives lower expected average cost rates compared to Policy 1. When Policy 2 is applied, all the components are operating in the healthy state after maintenance, however, when Policy 1 is applied, only units in the failure state are replaced, so there is a smaller number of units working in the healthy state after maintenance compared to Policy 2. As a result, the total production rate of the wind farm is higher after maintenance for Policy 2. This is due to the higher production rate of the units in the healthy state ( $p_0 = 450$  kWh) compared to the units in the warning state ( $p_1 = 330$ ). So, the cost of lost demand is lower and the profit from selling extra production is higher for Policy 2 compared to Policy 1. Results of

**Table 3** Expected cost savings obtained by applying Policy 2 instead of Policy 1

$C_D$	$C_E$	Expected cost saving (%)
0.1	0	3.3
	0.02	9.1
	0.04	19.7
0.2	0	4.8
	0.02	10.7
	0.04	21.4
0.3	0	5.1
	0.02	11.5
	0.04	22.1
0.4	0	5.3
	0.02	11.8
	0.04	22.5
0.5	0	5.3
	0.02	11.9
	0.04	22.8

both tables indicate that for a given  $C_D$  the average cost decreases when  $C_E$  increases. Tables 1 and 2 confirm that the second policy outperforms the first policy.

Table 3 presents the expected cost savings associated with applying Policy 2 instead of Policy 1. For a given  $C_E$  with the increase of  $C_D$ , expected cost savings associated with applying Policy 2 increases. Furthermore, for a given  $C_D$  by the increase of profit rate, expected cost savings associated with applying Policy 2 increases.

We can observe that expected cost saving by applying Policy 2 instead of Policy 1 for the lowest values of  $C_D$  and  $C_E$  ( $C_D = 0.1$ ,  $C_E = 0$ ) is equal to %3.3, however for the highest values of  $C_D$  and  $C_E$  ( $C_D = 0.5$ ,  $C_E = 0.04$ ) the expected cost saving increases to %22.8. Therefore, performance of Policy 2 is better for the higher values of lost demand cost rate and the higher values of excess production profit rate.

## 5 Conclusions and future research

In this paper, we have developed two new joint CBM and just-in-time spare parts provisioning policies for a multi-unit production system. We have considered a multi-unit parallel production system consisting of  $N$  identical, independent units, each subject to gradual deterioration. Deterioration process of each unit is assumed to be a three-state continuous time homogeneous Markov chain with two working states and a failure state. Production rate of units depends on the working state, indicating the effect of deterioration of units on the production rate. Units are subject to inspection and maintenance and each inspection or maintenance action entails a fixed cost of sending an inspector or crew to the field but it is charged only once, when actions on the several units of the system are performed. The proposed joint maintenance and spare part provisioning policies belongs to the class of the control-limit policies, where decisions regarding maintenance and spare parts ordering are made by comparing the number of failed units with the critical thresholds. Both maintenance policies prescribe corrective replacement of failed units when the total number of failed units exceeds a critical level. Second maintenance policy prescribes preventive maintenance of units in the warning state as well. We have applied SMDP framework to formulate the joint maintenance and spare parts ordering decision problem with the optimality criterion being the minimization of the long-run expected average cost per unit time. The expected long run cost rate model has been developed for a general multi-unit system to determine the optimal level to initiate maintenance, the optimal level to order spare parts, and the optimal inspection interval.

A numerical example has been developed to illustrate the proposed joint maintenance and spare part provisioning model and to compare both policies. The results have shown that Policy 2 gives lower expected average cost rates compared with Policy 1. For both policies for a given profit rate, when the lost demand cost rate increases, the optimal level to perform maintenance decreases or remains constant. For both policies, the decrease in the optimal level for performing maintenance is balanced by the increase in the optimal inspection interval. We also noticed that performance of Policy 2 depends on the lost production cost rate and excess production profit rate. By the increase of these rates, performance of Policy 2 increases (compared to Policy 1).

There are a variety of interesting extensions and topics for future research. One direction would be to have stochastic demand instead of constant demand rate in the model. The stochastic demand can depend on the off-peak and on-peak periods for energy systems, so that maintenance can be performed in the off-peak periods to decrease the unavailability

of the system. Another interesting topic for future research would be to investigate the case where deterioration process of each unit has more than three states. We have considered a three state deterioration process for each unit which is a reasonable assumption for many real world applications, but this model can be extended to a general  $N$  state deterioration process. Such an extension would lead to both interesting theoretical and practical challenges, especially in the calculation of the transition probabilities and expected costs. Another possible future research topic would be to consider the general extension of the exponential distribution which is the phase-type distribution (PH). The proposed mathematical formulation becomes so complicated which makes it intractable, and by applying PH distribution it is possible to overcome this barrier.

**Acknowledgements** The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for the financial support under Grant No. RGPIN 121384-11, and the two anonymous referees for their constructive comments.

### Appendix A: Calculation of transition probabilities for each unit

The instantaneous deterioration transition rates  $q_{ij}$ ,  $i, j \in \Omega$  for each unit are defined by:

$$q_{ij} = \lim_{u \rightarrow 0} \frac{P(X_{t+u} = j | X_t = i)}{u} < +\infty, \quad j \neq i$$

$$\text{and } q_{ii} = - \sum_{j \neq i} q_{ij} \tag{12}$$

where  $i, j \in \{0, 1, F\}$ , and the state transition matrix  $P(t)$  can be written as follows:

$$P(t) = \begin{bmatrix} e^{-v_0 t} & \frac{q_{01}(e^{-v_1 t} - e^{-v_0 t})}{v_0 - v_1} & 1 - e^{-v_0 t} - \frac{q_{01}(e^{-v_1 t} - e^{-v_0 t})}{v_0 - v_1} \\ 0 & e^{-v_1 t} & 1 - e^{-v_1 t} \\ 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

where the exponential distribution parameters of the sojourn times are denoted by  $v_0 = q_{01} + q_{0F}$  and  $v_1 = q_{1F}$ .

We assume that the state process is non-decreasing with probability 1, i.e.,  $q_{ij} = 0$  for all  $j < i$  and failure state is absorbing. The Kolmogorov’s backward differential equations for given transition rates can be written as follows:

$$\begin{aligned} P'_{00}(t) &= -v_0 P_{00}(t) \\ P'_{01}(t) &= q_{01} P_{11}(t) - v_0 P_{01}(t) \\ P'_{02}(t) &= q_{01} P_{12}(t) + q_{02}(t) P_{22}(t) - v_0 P_{02}(t) \\ P'_{11}(t) &= -v_1 P_{11}(t) \\ P'_{12}(t) &= v_1 P_{22}(t) - v_1 P_{12}(t) \end{aligned}$$

Extensive details on the Laplace transform can be found in Debnath and Bhatta (2014). Taking Laplace transform of the above equations, the following equations are obtained:

$$\begin{aligned} s\bar{P}_{00}(s) - 1 &= -v_0 \bar{P}_{00}(s) \\ s\bar{P}_{01}(s) &= q_{01} \bar{P}_{11}(s) - v_0 \bar{P}_{01}(s) \\ s\bar{P}_{02}(s) &= q_{01} \bar{P}_{12}(s) + \frac{q_{02}}{s} - v_0 \bar{P}_{02}(s) \end{aligned}$$

$$\begin{aligned}
 s\bar{P}_{11}(s) - 1 &= -\nu_1\bar{P}_{11}(s) \\
 s\bar{P}_{12}(s) &= \frac{\nu_1}{s} - \nu_1\bar{P}_{12}(s),
 \end{aligned}$$

where  $\bar{P}_{ij}(s)$  is the Laplace transform of the transition probability  $P_{ij}(t)$ , which is written as:  $\bar{P}_{ij}(s) = \int_0^\infty e^{-st} P_{ij}(t)dt$ . By solving the above equations we get:

$$\begin{aligned}
 \bar{P}_{00}(s) &= \frac{1}{s + \nu_0}, \quad \bar{P}_{01}(s) = \frac{q_{01}}{(s + \nu_0)(s + \nu_1)}, \quad \bar{P}_{02}(s) = \frac{\nu_0\nu_1 + q_{02}s}{s(s + \nu_0)(s + \nu_1)} \\
 \bar{P}_{11}(s) &= \frac{1}{s + \nu_1}, \quad \bar{P}_{12}(s) = \frac{\nu_1}{s(s + \nu_1)}.
 \end{aligned}$$

By taking the inverse Laplace transform of the above equations, we derive the transition probabilities as follows:

$$\begin{aligned}
 P_{00}(t) &= e^{-\nu_0 t}, \quad P_{01}(t) = \frac{q_{01}(e^{-\nu_1 t} - e^{-\nu_0 t})}{\nu_0 - \nu_1}, \\
 P_{02}(t) &= 1 - P_{00}(t) - P_{01}(t) = 1 - e^{-\nu_0 t} - \frac{q_{01}(e^{-\nu_1 t} - e^{-\nu_0 t})}{\nu_0 - \nu_1} \\
 P_{11}(t) &= e^{-\nu_1 t}, \quad P_{12}(t) = 1 - e^{-\nu_1 t}.
 \end{aligned}$$

### Appendix B: Proof of Theorem 5

Assume that the state of the system is  $(n_F, n_1, S)$  and the decision is to perform maintenance. If the state of the system changes to  $(n_F + i, n'_1, S)$ ,  $i \geq 0$  during  $T_R$  (or  $T_P$ ) an emergency order of size  $n_F + i - S$  is placed. The expected cost of emergency ordering in state  $(n_F, n_1, S)$  for different values of  $i$  is as follows:

$$\begin{aligned}
 E(C_{EM}|T_R, (n_F, n_1, S)) &= \sum_{(n_F+i, n'_1, S)} E(C_{EM}|(n_F + i, n'_1, S)) \\
 &\times \int_0^\infty P_{(n_F, n_1, S)(n_F+i, n'_1, S)}(u) f_{T_R}(u) du \\
 &= \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i, 0\}}^{n_1+n_0-i} C_{es}(n_F + i - S) \\
 &\times \int_0^\infty P_{(n_F, n_1, S)(n_F+i, n'_1, S)}(u) f_{T_R}(u) du \\
 &= \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i, 0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i, n_1\}} C_{es}(n_F + i - S) \\
 &\times \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \int_0^\infty P_{00}(u)^{n'_0} P_{01}(u)^{n'_1-n_1+j} \\
 &\times P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j \cdot f_{T_R}(u) du \tag{14}
 \end{aligned}$$

where  $x = 0$ , for  $n'_1 \geq n_1$ , and  $x = n_1 - n'_1$ , for  $n'_1 < n_1$ .

For the first maintenance policy, the expected cost of corrective replacement in state  $(n_F, n_1, S)$  is equal to:

$$\begin{aligned}
 E(C|T_R, (n_F, n_1, S)) &= C_{KM} + C_I(N - n_F) + E(C_H|T_R) \\
 &\quad + E(CR|T_R) + E(C_{EM}|T_R) + E(C_{CM}|T_R) \\
 &= C_{KM} + C_I(N - n_F) + \int_0^\infty C_H \cdot S \cdot f_{T_R}(u)du \\
 E(\text{Cost (profit) from production}) &\left\{ \begin{aligned} &+ \sum_{i=x}^{N-n_F-n_1} \sum_{n'_1=\max\{n_1-i,0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i,n_1\}} C_{rI}(n_F + i, n'_1, S') \\ &\times \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \int_0^\infty \int_0^u P_{00}(t)^{n'_0} \\ &\times P_{01}(t)^{n'_1-n_1+j} P_{0F}(t)^{i-j} P_{11}(t)^{n_1-j} P_{1F}(t)^j f_{T_R}(u) dt du \end{aligned} \right. \\
 E(\text{Cost of emergency ordering}) &\left\{ \begin{aligned} &+ \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i,0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i,n_1\}} C_{es}(n_F + i - S) \\ &\times \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \int_0^\infty P_{00}(u)^{n'_0} \\ &\times P_{01}(u)^{n'_1-n_1+j} P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j f_{T_R}(u) du \end{aligned} \right. \\
 E(\text{Cost of failure replacement}) &\left\{ \begin{aligned} &+ \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i,0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i,n_1\}} C_F(n_F + i) \\ &\times \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \int_0^\infty P_{00}(u)^{n'_0} \\ &\times P_{01}(u)^{n'_1-n_1+j} P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j f_{T_R}(u) du \end{aligned} \right. \\
 &= C_{KM} + C_I(N - n_F) + \int_0^\infty C_H \cdot S \cdot f_{T_R}(u)du \\
 &\quad + \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i,0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i,n_1\}} \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \\
 &\quad \times \left[ \left( C_{es}(n_F + i - S) + C_F(n_F + i) \right) \int_0^\infty P_{00}(u)^{n'_0} P_{01}(u)^{n'_1-n_1+j} \right. \\
 &\quad \times P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j f_{T_R}(u) du \\
 &\quad + C_{rI}(n_F + i, n'_1, S') \int_0^\infty \int_0^u P_{00}(t)^{n'_0} P_{01}(t)^{n'_1-n_1+j} \\
 &\quad \left. \times P_{0F}(t)^{i-j} P_{11}(t)^{n_1-j} P_{1F}(t)^j f_{T_R}(u) dt du \right] \tag{15}
 \end{aligned}$$

where  $x = 0$  for  $n'_1 \geq n_1$ , and  $x = n_1 - n'_1$  for  $n'_1 < n_1$ .

The expected cost of corrective replacement and preventive maintenance for the second policy in state  $(n_F, n_1, S)$  is equal to:

$$\begin{aligned}
 E(C_{CM} + C_{PM}|T_P, (n_F, n_1, S)) &= \sum_{(n_F+i, n'_1, S)} E(C_{CM} + C_{PM})(n_F + i, n'_1, S) \\
 &\quad \times \int_0^\infty P_{(n_F, n_1, S)(n_F+i, n'_1, S)}(u) f_{T_P}(u) du \\
 &= \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i,0\}}^{n_1+n_0-i} (C_F(n_F + i) + C_P(n'_1)) \\
 &\quad \times \int_0^\infty P_{(n_F, n_1, S)(n_F+i, n'_1, S)}(u) f_{T_P}(u) du
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=x}^{N-n_F-n'_1} \sum_{n'_1=\max\{n_1-i,0\}}^{n_1+n_0-i} \sum_{j=x}^{\min\{i,n_1\}} (C_F(n_F+i) + C_P(n'_1)) \\
 &\times \binom{n_0}{i-j} \binom{n_1}{j} \binom{n_0-i+j}{n'_1-n_1+j} \int_0^\infty P_{00}(u)^{n'_0} P_{01}(u)^{n'_1-n_1+j} \\
 &\times P_{0F}(u)^{i-j} P_{11}(u)^{n_1-j} P_{1F}(u)^j \times f_{T_P}(u) du \tag{16}
 \end{aligned}$$

where  $x = 0$  for  $n'_1 \geq n_1$ , and  $x = n_1 - n'_1$  for  $n'_1 < n_1$ .

### Appendix C: Extension to more than 3 state deterioration process

First, we will show how to extend the model to a more complex case of a 4 state deterioration process (we consider 4 states for each unit of the system). This change affects the number of system states, transition probabilities, production rate and the expected cost. In this section we provide the formula to calculate the transition probabilities. The calculation of the expected cost is similar to to the previous case (3 state deterioration process).

Consider a system consisting of  $N$  units where deterioration process of each unit is modeled as a 4-state continuous time homogeneous Markov chain with three working states and a failure state  $F$  which is absorbing,  $\Omega = \{0, 1, 2, F\}$ . The state space for the whole system can be defined as  $W = \{(n_F, n_2, n_1, S) \mid n_1 + n_2 + n_F \leq N, n_1, n_2, n_F, S \geq 0\}$ , where  $n_F$ ,  $n_2$ , and  $n_1$  represent the number of units in the failure state, state 2, and state 1, respectively, and  $S$  represents the on-hand inventory. Based on Remark 2, the number of available spare parts in the system at each inspection time can be 0 or  $R$ , so the number of states for this system is equal to  $\frac{(N+1)(N+2)(N+3) \times 2}{3!}$ .

The transition probability function when the decision is to continue operation is as follows:

$$\begin{aligned}
 &P_{(n_F, n_2, n_1, S)(n_F+i, n'_2, n'_1, S)}(\Delta) = \\
 &\sum_{k=\max\{0, n_2-n'_2\}}^{\min\{i, n_2\}} \sum_{j=\max\{0, i-k-n_0\}}^{\min\{i-k, n_1\}} \sum_{m=\max\{0, n_1-n'_1-j\}}^{\min\{n_1-j, n'_2-n_2+k\}} \binom{n_0}{i-j-k} \binom{n_0-i+k+j}{n'_1-n_1+m+j} \\
 &\binom{n_0-i+k-n'_1+n_1-m}{n'_2-n_2+k-m} \binom{n_1}{j} \binom{n_1-j}{m} \binom{n_2}{k} P_{00}(\Delta)^{n'_0} P_{01}(\Delta)^{n'_1-n_1+j+m} \\
 &P_{02}(\Delta)^{n'_2-n_2+k-m} P_{0F}(\Delta)^{i-j-k} P_{11}(\Delta)^{n_1-m-j} P_{12}(\Delta)^m P_{1F}(\Delta)^j P_{22}(\Delta)^{n_2-k} P_{2F}(\Delta)^k \tag{17}
 \end{aligned}$$

Figure 3 shows the number of units which change their state in the next time interval of length  $\Delta$  when deterioration process of units has 4 states.

The transition probability when the decision is to perform maintenance is given by the following formulas:

- For the first maintenance policy,

$$\begin{aligned}
 &P_{(n_F, n_2, n_1, S)(0, n'_2, n'_1, 0)}(T_R) = \\
 &\sum_{i=\max\{0, n_1+n_2-n'_1-n'_2\}}^{N-n_F-n'_1-n'_2} \sum_{k=\max\{0, n_2-n'_2\}}^{\min\{i, n_2\}} \sum_{j=\max\{0, i-k-n_0\}}^{\min\{i-k, n_1\}} \sum_{m=\max\{0, n_1-n'_1-j\}}^{\min\{n_1-j, n'_2-n_2+k\}} \binom{n_0}{i-j-k}
 \end{aligned}$$

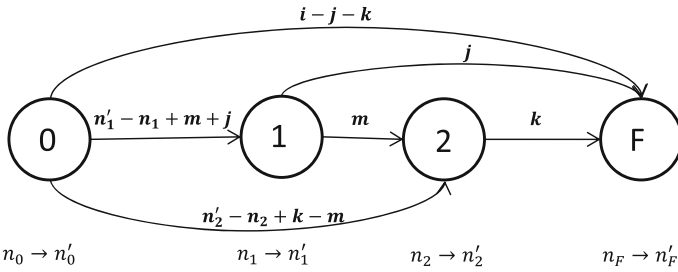


Fig. 3 Number of units changing state in a time interval of length  $\Delta$

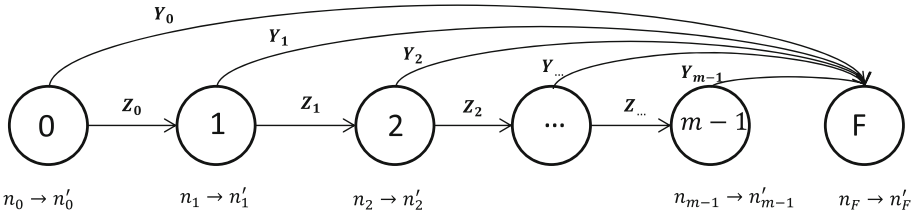


Fig. 4 Number of units changing state in a time interval of length  $\Delta$

$$\begin{aligned}
 & \binom{n_0 - i + k + j}{n'_1 - n_1 + m + j} \binom{n_0 - i + k - n'_1 + n_1 - m}{n'_2 - n_2 + k - m} \binom{n_1}{j} \binom{n_1 - j}{m} \binom{n_2}{k} \\
 & \left( \int_0^\infty P_{00}(u)^{n'_0} P_{01}(u)^{n'_1 - n_1 + j + m} P_{02}(u)^{n'_2 - n_2 + k - m} P_{0F}(u)^{i - j - k} P_{11}(u)^{n_1 - m - j} \right. \\
 & \left. P_{12}(u)^m P_{1F}(u)^j P_{22}(u)^{n_2 - k} P_{2F}(u)^k \times f_{TR}(u) du \right) \tag{18}
 \end{aligned}$$

– For the second maintenance policy,

$$P_{(n_F, n_2, n_1, S)(0,0,0,0)}(T_P) = 1$$

As can be seen, the number of system states increases dramatically and the calculation of transition probability becomes quite complicated by adding new states. We assume that a transition to a different state may or may not occur during the next  $\Delta$  time units, which makes it difficult to calculate transition probabilities for the system. Some simplification is possible for reliable systems, for which transitions to some states can be eliminated. So, we can consider a system consisting of  $N$  units where deterioration process of each unit is modeled as an  $m$ -state continuous time homogeneous Markov chain with  $m - 1$  working states and a failure state  $F$  which is absorbing,  $\Omega = \{0, 1, 2, \dots, m - 1, F\}$ . The multi-state structure is considered where transitions are allowed either to the neighbor state or to the failure state. For highly reliable systems, we can omit any intermediate transition between states during each inspection interval (see Fig. 4).

The state space for the whole system can be defined as  $W = \{(n_F, n_{m-1}, \dots, n_1, S) \mid n_1 + \dots + n_F \leq N, n_1, \dots, n_F, S \geq 0\}$ , where  $n_F, n_i \ i \in \{0, 1, \dots, m - 1\}$  represent the number of units in the failure state, and state  $i$ , respectively, and  $S$  represents the on-hand inventory.

The transition probability function when the decision is to continue operation is given by:



$$\begin{aligned}
 &P_{(n_F, n_{m-1}, \dots, n_1, S)(n_F+i, n'_{m-1}, \dots, n'_1, S)}(\Delta) = \\
 &\min\{i, n_{m-1}, n_{m-1}-n'_{m-1}+n_{m-2}\} \min\{i-Y_{m-1}, n_{m-2}, n_{m-2}-n'_{m-2}+n_{m-3}\} \dots \min\{i-Y_{m-1}-\dots-Y_2, n_1, n_1-n'_1+n_0\} \\
 &\sum_{Y_{n-1}=\max\{0, n_{m-1}-n'_{m-1}\}} \sum_{Y_{m-2}=\max\{0, n_{m-2}-n'_{m-2}\}} \dots \sum_{Y_1=\max\{0, n_1-n'_1\}} \\
 &\binom{n_0}{Z_0} \binom{n_0 - Z_0}{i - Y_1 - Y_2 - \dots - Y_{m-1}} \binom{n_1}{Z_1} \binom{n_1 - Z_1}{Y_1} \dots \binom{n_{m-1}}{Y_{m-1}} \\
 &P_{0,0}^{n'_0} P_{(0,1)}^{Z_0} P_{(0,F)}^{Y_0} P_{1,1}^{n_1-Y_1-Z_1} P_{(1,2)}^{Z_1} P_{(1,F)}^{Y_1} \dots P_{m-1,m-1}^{m-1-Y_{m-1}} P_{(m-1,F)}^{Y_{m-1}}
 \end{aligned}$$

where

$$\begin{aligned}
 Z_{m-2} &= n'_{m-1} - n_{m-1} + Y_{m-1} \\
 Z_{m-3} &= n'_{m-2} - n_{m-2} + Y_{m-2} + Z_{m-2} \\
 &\dots \\
 Z_1 &= n'_2 - n_2 + Y_2 + Z_2 \\
 Z_0 &= n'_1 - n_1 + Y_1 + Z_1.
 \end{aligned}$$

## References

Ahmadi, R. (2014). Optimal maintenance scheduling for a complex manufacturing system subject to deterioration. *Annals of Operations Research*, 217(1), 1–29.

Asadi, M., & Bayramoglu, I. (2006). The mean residual life function of a k-out-of-n structure at the system level. *IEEE Transactions on Reliability*, 55(2), 314–318.

Barron, Y. (2015). Group replacement policies for a repairable cold standby system with fixed lead times. *IIE Transactions*, 47(10), 1139–1151.

Barron, Y. (2018). Group maintenance policies for an R-out-of-N system with phase-type distribution. *Annals of Operations Research*, 261(1–2), 79–105.

Barron, Y., Frostig, E., & Levikson, B. (2004). Analysis of r-out-of-n repairable systems: The case of phase-type distributions. *Advances in Applied Probability*, 36(1), 116–138.

Barron, Y., & Yechiali, U. (2017). Generalized control-limit preventive repair policies for deteriorating cold and warm standby markovian systems. *IIEE Transactions*, 49(11), 1031–1049.

Bevilacqua, M., & Braglia, M. (2000). The analytic hierarchy process applied to maintenance strategy selection. *Reliability Engineering & System Safety*, 70(1), 71–83.

Bohlin, M., & Warja, M. (2015). Maintenance optimization with duration-dependent costs. *Annals of Operations Research*, 224(1), 1–23.

Byon, E., & Ding, Y. (2010). Season-dependent condition-based maintenance for a wind turbine using a partially observed markov decision process. *IEEE Transactions on Power Systems*, 25(4), 1823–1834.

de Smidt-Destombes, K. S., van der Heijden, M. C., & van Harten, A. (2004). On the availability of a k-out-of-n system given limited spares and repair capacity under a condition based maintenance strategy. *Reliability Engineering & System Safety*, 83(3), 287–300.

de Smidt-Destombes, K. S., van der Heijden, M. C., & Van Harten, A. (2006). On the interaction between maintenance, spare part inventories and repair capacity for a k-out-of-n system with wear-out. *European Journal of Operational Research*, 174(1), 182–200.

Debnath, L., & Bhatta, D. (2014). *Integral transforms and their applications*. Boca Raton: CRC Press.

Dieulle, L., Berenguer, C., Grall, A., & Roussignol, M. (2003). Sequential condition-based maintenance scheduling for a deteriorating system. *European Journal of operational research*, 150(2), 451–461.

EIA (2019). Average price of electricity to ultimate customers by end-use sector. <https://www.eia.gov>.

El Moudani, W., & Mora-Camino, F. (2000). A dynamic approach for aircraft assignment and maintenance scheduling by airlines. *Journal of Air Transport Management*, 6(4), 233–237.

Eruguz, A. S., Tan, T., & van Houtum, G. J. (2017). Optimizing usage and maintenance decisions for k-out-of-n systems of moving assets. *Naval Research Logistics (NRL)*, 64(5), 418–434.

Eryilmaz, S., & Xie, M. (2014). Dynamic modeling of general three-state k-out-of-n: G systems: Permanent-based computational results. *Journal of Computational and Applied Mathematics*, 272, 97–106.

- Jafari, L., Naderkhani, F., & Makis, V. (2018). Joint optimization of maintenance policy and inspection interval for a multi-unit series system using proportional hazards model. *Journal of the Operational Research Society*, 69(1), 36–48.
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483–1510.
- Jia, X., Shen, J., & Xing, R. (2016). Reliability analysis for repairable multistate two-unit series systems when repair time can be neglected. *IEEE Transactions on Reliability*, 65(1), 208–216.
- Jiang, R., Kim, M. J., & Makis, V. (2012). A bayesian model and numerical algorithm for cbm availability maximization. *Annals of Operations Research*, 196(1), 333–348.
- Keizer, M. C. O., Teunter, R. H., & Veldman, J. (2017). Joint condition-based maintenance and inventory optimization for systems with multiple components. *European Journal of Operational Research*, 257(1), 209–222.
- Keizer, M. C. O., Teunter, R. H., Veldman, J., & Babai, M. Z. (2018). Condition-based maintenance for systems with economic dependence and load sharing. *International Journal of Production Economics*, 195, 319–327.
- Kennedy, W., Patterson, J. W., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of production economics*, 76(2), 201–215.
- Khaleghi, A., & Makis, V. (2015). Model parameter estimation and residual life prediction for a partially observable failing system. *Naval Research Logistics (NRL)*, 62(3), 190–205.
- Khaleghi, A., & Makis, V. (2016). Reliability estimation of a system subject to condition monitoring with two dependent failure modes. *IIE Transactions*, 48(11), 1058–1071.
- Kontoleon, J., & Kontoleon, N. (1974). Reliability analysis of a system subject to partial and catastrophic failures. *IEEE Transactions on Reliability*, 23(4), 277–278.
- Lanza, G., Niggeschmidt, S., & Werner, P. (2009). Optimization of preventive maintenance and spare part provision for machine tools based on variable operational conditions. *CIRP Annals-Manufacturing Technology*, 58(1), 429–432.
- Maggard, B. N., & Rhyne, D. M. (1992). Total productive maintenance: A timely integration of production and maintenance. *Production and Inventory Management Journal*, 33(4), 6.
- Moghaddam, K. S., & Usher, J. S. (2011). Preventive maintenance and replacement scheduling for repairable and maintainable systems using dynamic programming. *Computers & Industrial Engineering*, 60(4), 654–665.
- Nicolai, R. P., & Dekker, R. (2008). Optimal maintenance of multi-component systems: A review. In *Complex system maintenance handbook* (pp. 263–286). London: Springer.
- Niu, C., Liang, X., Ge, B., Tian, X., & Chen, Y. (2016). Optimal replacement policy for a repairable system with deterioration based on a renewal-geometric process. *Annals of Operations Research*, 244(1), 49–66.
- Rahmati, S. H. A., Ahmadi, A., & Govindan, K. (2018). A novel integrated condition-based maintenance and stochastic flexible job shop scheduling problem: Simulation-based optimization approach. *Annals of Operations Research*, 269(1–2), 583–621.
- Rao, A. N., & Bhadury, B. (2000). Opportunistic maintenance of multi-equipment system: A case study. *Quality and Reliability Engineering International*, 16(6), 487–500.
- Salari, N., & Makis, V. (2017a). Comparison of two maintenance policies for a multi-unit system considering production and demand rates. *International Journal of Production Economics*, 193, 381–391.
- Salari, N., & Makis, V. (2017b). Optimal preventive and opportunistic maintenance policy for a two-unit system. *The International Journal of Advanced Manufacturing Technology*, 89(1–4), 665–673.
- Tian, Z., Jin, T., Wu, B., & Ding, F. (2011). Condition based maintenance optimization for wind power generation systems under continuous monitoring. *Renewable Energy*, 36(5), 1502–1509.
- Tijms, H. C. (1986). *Stochastic modelling and analysis: A computational approach*. Hoboken: Wiley.
- Wang, H., & Pham, H. (1999). Some maintenance models and availability withimperfect maintenance in production systems. *Annals of Operations Research*, 91, 305–318.
- Wang, L., Chu, J., & Mao, W. (2009). A condition-based replacement and spare provisioning policy for deteriorating systems with uncertain deterioration to failure. *European Journal of Operational Research*, 194(1), 184–205.
- Wang, W. (2011). A joint spare part and maintenance inspection optimisation model using the delay-time concept. *Reliability Engineering & System Safety*, 96(11), 1535–1541.
- Wang, W. B. (2000). A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance. *International Journal of Production Research*, 38(6), 1425–1436.
- Yeh, R. H., Chang, W. L., & Lo, H. C. (2010). Optimal threshold values of age and two-phase maintenance policy for leased equipments using age reduction method. *Annals of Operations Research*, 181(1), 171–183.

Zhou, X., Xi, L., & Lee, J. (2009). Opportunistic preventive maintenance scheduling for a multi-unit series system based on dynamic programming. *International Journal of Production Economics*, 118(2), 361–366.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.