Joint Multi-Item Production and Condition-Based Maintenance Control of a System with Setup Times and Stochastic Demand

Alp Darendeliler[1][a], Dieter Claeys[1,2][b] and El-Houssaine Aghezzaf[1,2][c]

[1]Department of Industrial Systems Engineering and Product Design, Ghent University, Ghent, Belgium
[2]Industrial Systems Engineering (ISyE), Flanders Make, Kortrijk, Belgium

Keywords: Condition-Based Maintenance (CBM), Semi-Markov Decision Process (SMDP), Multi-Product, Reinforcement Learning, Stochastic Demand.

Abstract: In this paper, we address the joint production and Condition-based maintenance (CBM) planning problem for a deteriorating single-machine multi-product manufacturing system with uncertain product demands. The objective is to find an integrated production and maintenance policy such that the sum of expected setup, holding, lost sales, preventive and corrective maintenance costs is minimized. We formulate the problem as a semi-Markov decision process and propose a Q-learning algorithm for the problem. A numerical example is provided to illustrate the solution method.

1 INTRODUCTION

Preventive maintenance aims to reduce the likelihood of equipment failures that result in unexpected downtimes and production losses. Condition-based maintenance (CBM), as a preventive maintenance strategy, recommends performing maintenance activities based on the current equipment condition. If a CBM program is properly implemented, it can significantly save costs by reducing the unnecessary scheduled preventive maintenance operations (Jardine et. al, 2006).

In recent years, the integration of production lot-sizing and CBM has been studied by many researchers. The majority of the existing research studies the joint optimization of the economic production quantity (EPQ) and CBM for single-product systems under constant and deterministic demand rate. However, there are many cases in practice where multiple-items are produced by a single facility (machine) and the demand is uncertain.

In this study, we consider a stochastically degrading single-machine multi-product production/ inventory system under stochastic demand. Only a single product can be produced at a time. The produced items are stored in the inventory with limited capacity for each product, and holding costs are charged for the unconsumed products. Unsatisfied demand is lost, and a penalty cost is incurred for each. Switching production from one product to another or from the idle state to production takes a certain amount of time, and a setup cost is charged for this operation. The degradation level of the machine increases with production, and degradation behaviour changes with respect to the product being produced. Preventive or corrective maintenance brings the machine to "as good as new" state and they both take a deterministic time. The goal is to find a joint production/inventory and maintenance policy to minimize the long run expected average cost per unit time, which includes the holding, lost sales, setup, preventive and corrective maintenance costs.

We model the above-mentioned problem as a semi-Markov decision process (SMDP), in which the system state consists of the degradation level, the setup status of the machine and the stock levels of the products. Upon observing the state at a decision epoch, one of the following actions is taken: keeping the machine idle; producing a particular product that the machine has been set up; carrying out setup for a product in case the machine is idle or has been set up for another product; performing preventive

[a] https://orcid.org/0000-0002-5217-7444
[b] https://orcid.org/0000-0002-7666-2479
[c] https://orcid.org/0000-0003-3849-2218

Darendeliler, A., Claeys, D. and Aghezzaf, E.
Joint Multi-Item Production and Condition-Based Maintenance Control of a System with Setup Times and Stochastic Demand.
DOI: 10.5220/0011741300003396
In *Proceedings of the 12th International Conference on Operations Research and Enterprise Systems (ICORES 2023)*, pages 185-192
ISBN: 978-989-758-627-9; ISSN: 2184-4372

maintenance; or performing corrective maintenance (it is the only choice if the machine fails). To solve the underlying model, we adopt a Q-learning algorithm for SMDP.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature. The SMDP formulation is presented in Section 3. Then, in Section 4, we present the Q-learning algorithm adopted for our problem. In Section 5, a numerical example is given. Then, concluding remarks are provided in Section 6.

2 LITERATURE REVIEW

The integration of lot-sizing and time-based preventive maintenance has been extensively studied by many researchers (Aghezzaf et al., 2007; Ben Daya and Makhdoum, 1998; Ben-Daya, 2002; El-Ferik, 2008; Liao and Sheu, 2011; Suliman and Jawad, 2012; Shamsaei and Vyve, 2017). In recent years, integrated EPQ and CBM-based preventive maintenance models have been proposed with the aim of optimizing the lot-size and the degradation threshold, beyond which preventive maintenance is conducted. Jafari and Makis (2015) address the joint optimization of EPQ and preventive maintenance policy. The deterioration of the system is modelled by a proportional hazards model that considers the condition monitoring information and the age of the machine. Peng and van Houtum (2016) develop a joint optimization model of EPQ and CBM in which degradation is modelled as a continuous time and continuous state stochastic process. Khatab et al. (2019) investigate the problem of integrating production quality and CBM for a production system under periodic monitoring. Cheng et al. (2018) develop a model to optimize production, quality control and CBM policies for a system in which product quality depends on the degradation level. Jafari and Makis (2016) propose a model to jointly optimize EPQ and preventive maintenance policy for a partially observable two-unit system. Cheng et al. (2017) consider joint optimization of production lot-sizing and CBM for systems with multiple products. Preventive maintenance decision making depends on the predictive reliability and the structural importance measure of the components.

Fewer studies, however, develop integrated production and maintenance policies for systems with stochastic demand. To find the optimal policy for systems with stochastic demand, MDP and SMDP models are proposed (Iravani and Duenyas, 2002; Sloan, 2004; Jafari and Makis, 2019; Xiang et al., 2014). These studies assume that the system produces a single product type and the degradation is modelled by a Markov chain with a limited number of states. In this study, however, we propose a joint production and CBM policy for a multi-product production system with random product demands.

Darendeliler et al. (2022) has recently studied joint optimal production/inventory and CBM control for a multi-product manufacturing system under stochastic product demands. It is assumed that the system is reviewed at equidistant time points, so the durations of producing a lot and maintenance are assumed to be equal. The present paper relaxes this assumption and extends the work by modelling the problem as a SMDP, in which the system is reviewed at the completion of a unit production, setup and maintenance. Also, the previous work does not take the production setup times into account, while they are incorporated in the present model.

In literature, the problem of planning the lot-size and sequence of several products on a single machine with random product demands is known as the stochastic economic lot scheduling problem (SELSP). In the SELSP, the objective is to find a policy that proposes whether to continue the production of the current item, whether to switch to another product or whether to keep the machine idle so as to minimize the total expected average cost. Obtaining such a policy, which dynamically distributes the finite production capacity among the products to be reactive to the stochastic demands, processing and setup times, is a challenging problem (Sox et al., 1999). Winands et al. (2011) categorize SELSPs based on their sequencing and lot-sizing strategies. Our model's production policy could be considered in the category of dynamic sequence and global lot-sizing, in which there is no predetermined production sequence, and the quantity of the lot-size depends on the stock levels of all products and the machine status rather than depending only on the stock level of the product currently setup. The majority of the SELSP models do not consider the effect of equipment deterioration and maintenance on the production policies. However, in this study, we incorporate CBM policy in the SELSP problem.

There are few studies that consider dynamic sequencing and global lot-sizing for the SELSP. Qiu and Loulou (1995) model the SELSP as a SMDP and solve limited-size problems by the successive approximation method. Wang et al. (2012) apply two reinforcement learning algorithms to the SELSP with the random demand and processing times. Löhndorf and Minner (2013) propose an approximate value iteration method and compare its performance with the global search for parameters of simple control policies.

3 PROBLEM FORMULATION

We consider a single machine production system producing N items one at a time. The processing time of product n takes ρ_n time units, and the machine must be set up for product n if it is going to be produced. The setup for item n takes ST_n time units and incurs a cost cs_n. Finished products are stored in an inventory with limited storage capacity $\bar{x}_n$ for each product $n \in \{1, \dots, N\}$. A holding cost of ch_n is charged per product per unit time for product n. Demand for each product follows a Poisson process with rate λ_n. The unsatisfied demand is lost and a lost sales cost cl_n is incurred per item. The machine has multiple degradation states $\{1, \dots, F\}$, where 1 is the as-new state, and F is the failure state. If the machine degradation level is observed to be F, then corrective maintenance is immediately performed which takes CT time unit and costs cc. Conducting preventive maintenance in an operational state takes PT time units and costs cp. Both preventive and corrective maintenance bring the machine to the as good as new state (state "1"). Note that the system cannot be interrupted during production, setup and maintenance operations. The objective is to find the optimal production/inventory and maintenance policy that minimizes the total long-run average cost per time unit.

We formulate the problem as a semi-Markov decision process (SMDP), where the state is described as $s = (i, k, x_1, \dots, x_N)$ with degradation level $i \in \{1, \dots, F\}$, machine status $k \in \{0, \dots, N\}$ and the product inventories $x_1, \dots, x_N$. When the machine status is $k = n$, the machine has been set up for product $n \in \{1, \dots, N\}$. The system is reviewed at decision epochs, which are epochs at which demand for any product has just arrived in case the machine is idle, or setup or production for an item has just been completed, or preventive or corrective maintenance has just been performed. At a decision epoch, based on state s, an action $a \in A(s)$ is chosen. The eligible actions are: (1) keeping the machine idle $(a = 0)$; (2) producing item n $(a = n)$; (3) carrying out preventive maintenance $(a = -1)$; (4) conducting corrective maintenance $(a = -2)$ if and only if failure occurs. If product n is decided to be produced and the machine is set up for another product or is in the idle state, then the machine is going to be set up for item n.

For state $s = (i, k, x_1, \dots, x_N)$ and action a, the transition probabilities for the degradation can be expressed as

$$P_{ij}^{k,a} = \begin{cases} 1 & if\ a = 0, i = j \\ 1 & if\ a = -2, i = F, j = 1 \\ 1 & if\ a = -1, j = 1 \\ p_{ij}(n) & if\ a = n = k, j \in \{i, \dots, F\} \\ 1 & if\ a = n \neq k, j = i \\ 0 & otherwise, \end{cases} \tag{1}$$

where, the degradation remains at the same level until the next decision epoch if no production occurs, and it will be at state j in case of production.

The expected time until the next decision epoch is given by

$$\tau(s, a) = \begin{cases} 1/\sum_{n=1}^{N} \lambda_n & for\ a = 0 \\ CT & for\ a = -2 \\ PT & for\ a = -1 \\ \rho_n & for\ a = n = k \\ ST_n & for\ a = n \neq k. \end{cases} \tag{2}$$

For state $s = (i, k, x_1, \dots, x_N)$, the transition probabilities of the stock level of product $n = 1, \dots, N$, and the machine status k under the action a are as follows:

$$T_n(k', x_n' | k, x_n, a) = \begin{cases} P\left(D_n(t) = x_n - x_n^{'}\right) \ for\ x_n^{'} > 0, \\ \quad \left(a = -2, -1; k^{'} = 0\right) \vee \\ \quad \left(a = k \neq n > 0, k^{'} = a\right) \vee \\ \quad \left(a \neq k > 0, k^{'} = a\right) \\ P\left(D_n(t) \geq x_n - x_n^{'}\right) \ for\ x_n^{'} = 0, \\ \quad \left(a = -2, -1; k^{'} = 0\right) \vee \\ \quad \left(a = k \neq n > 0, k^{'} = a\right) \vee \\ \quad \left(a \neq k > 0, k^{'} = a\right) \\ P\left(D_n(t) = x_n - x_n^{'} + 1\right) \ for\ x_n^{'} > 1, \\ \quad a = n = k, k^{'} = n \\ P\left(D_n(t) \geq x_n\right) \ for\ x_n^{'} = 1 \\ \quad a = n = k, k^{'} = n \\ \lambda_n / \sum_{j=1}^{N} \lambda_j \ \ for\ x_n^{'} = (x_n - 1)^+ \\ \quad a = 0, k^{'} = 0 \\ 1 - \lambda_n / \sum_{j=1}^{N} \lambda_j \ for\ x_n^{'} = x_n, \\ \quad a = 0, k^{'} = 0 \\ 0 \qquad otherwise, \end{cases} \tag{3}$$

where $t = \tau(s, a)$ and $D_n(t)$ is the random variable with distribution $Poisson(\lambda_n t)$. For $x'_n > 0$, if the action is conducting corrective or preventive maintenance $(a = -2, -1)$, setup for any product $(a > 0 \wedge a \neq k)$ or producing an item other than n $(a > 0 \wedge a = k \neq n)$; then the transition probability to the next state (x'_n, k') is $P(D_n(t) = x_n - x'_n)$; for $x'_n = 0$, then the corresponding probability is $P(D_n(t) \geq x_n - x'_n)$ for the same actions. If production takes place for product n $(a = n = k)$, then the transition probability to the next state is $P(D_n(t) = x_n - x'_n + 1)$ when $x'_n > 1$; for $x'_n = 1$, the transition probability is $P(D_n(t) \geq x_n - x'_n + 1)$. In case the machine is idle $(a = 0)$, then the inventory level at the next decision epoch is either $x'_n = (x_n - 1)^+$ with probability $\lambda_n / \sum_{j=1}^{N} \lambda_j$ (first demand is for item n) or $x'_n = x_n$ with probability $1 - \lambda_n / \sum_{j=1}^{N} \lambda_j$ (the first arrival is demand for another product).

Let $S_n^1, S_n^2, \ldots$ denote demand arrival times for product n. Then, the conditional expectation of S_n^j, given that m demand arrivals occur in period of length t, is given by

$$E\left[S_n^j \middle| D_n(t) = m\right] = \frac{tj}{m+1} \quad for\ 1 \leq j \leq m. \tag{4}$$

By using the conditional expectation of arrival times, the expected inventory holding cost for product n, accumulated in a period of length t, can be expressed as

$$\begin{aligned} &H_n[x_n, t] \\ &= ch_n \sum_{m=0}^{x_n} \sum_{j=1}^{m} \left(\frac{tj}{m+1}\right) P(D_n(t) = m) \\ &+ch_n \sum_{m=0}^{x_n} t(x_n - m) P(D_n(t) = m) \\ &+ch_n \sum_{m=x_n+1}^{\infty} \sum_{j=1}^{x_n} \left(\frac{tj}{m+1}\right) P(D_n(t) = m). \end{aligned} \tag{5}$$

The expected lost sales cost for product n is given by

$$\begin{aligned} &L_n[x_n, t] \\ &= cl_n \sum_{m=x_n+1}^{\infty} (m - x_n) P(D_n(t) = m). \end{aligned} \tag{6}$$

$C(s, a)$ denotes the total expected cost incurred until the next decision epoch if action a is taken in state s. It can be expressed as

$$C(s,a) = \begin{cases} \dfrac{\sum_{n=1}^{N} x_n ch_n}{\sum_{n=1}^{N} \lambda_n} + \dfrac{\sum_{n=1}^{N} cl_n (1 - x_n)^+ \ \lambda_n}{\sum_{n=1}^{N} \lambda_n} & for\ a = 0 \\ cc + \sum_{n=1}^{N} H_n[x_n, CT] + \sum_{n=1}^{N} L_n[x_n, CT] & for\ a = -2 \\ cp + \sum_{n=1}^{N} H_n[x_n, PT] + \sum_{n=1}^{N} L_n[x_n, PT] & for\ a = -1 \\ \sum_{n=1}^{N} H_n[x_n, \rho_m] \ + \sum_{n=1}^{N} L_n[x_n, \rho_m] & for\ a = m, m = k \\ cs_m + \sum_{n=1}^{N} H_n[x_n, ST_m] + \sum_{n=1}^{N} L_n[x_n, ST_m] & for\ a = m, m \neq k. \end{cases} \tag{7}$$

The average cost optimality equation is as follows

$$V^*(s) = \min_{a \in A(s)} \left\{ C(s,a) - g^* \tau(s,a) + \sum_{s' \in S} P(s'|s,a) V^*(s') \right\} \quad \forall s \in S, \tag{8}$$

where $s = (i, k, x_1, \ldots, x_N)$ and $s' = (i', k', x'_1, \ldots, x'_N)$ and the transition probability is

$$P(s'|s,a) = P_{ii'}^{k,a} \prod_{n=1}^{N} T_n(k', x'_n | k, x_n, a). \tag{9}$$

A solution to these equations gives the minimum expected average cost per time unit g^* and optimal value functions $V^*(s)$. Any policy that minimizes the right-hand side of (8) for all $s \in S$ is optimal.

4 SOLUTION METHOD

Dynamic programming methods provide exact optimal policies by iteratively solving the Bellman equations. However, they are not computationally feasible for large and even for moderate-size problems. Hence, we apply the Q-learning algorithm to our problem, which is a model-free reinforcement learning algorithm proposed by Watkins (1989). Q-learning estimates the optimal state-action values (Q-values) by the sampled values instead of using complete transition probabilities to make expected updates.

```
1. Initialize a starting state $s$, and $Q(s,a)$ for all $s \in S$ and action $a \in A(s)$
2. For $t = 1, 2, \dots, T$
  2.1. Choose action $a \in A(s)$ for state $s$ ($\epsilon - greedy$)
  2.2. Sample the cost $c(s,a)$, the sojourn time $\tau$ and the next state $s'$ based on the current state $s$ and action $a$
  2.3. Update the Q-value by the equation:
       $Q(s,a) \leftarrow Q(s,a) + \alpha \left( c(s,a) + e^{-\gamma\tau} \min_{a' \in A(s')} Q(s',a') - Q(s,a) \right)$
  2.4. Update $s \leftarrow s'$
3. Return $\mu(s) = arg \min_{a \in A(s)} Q(s,a) \quad for\ all\ s \in S$
```

Figure 1: Steps of Q-learning for SMDP.

We use the Q-learning algorithm adapted for the SMDP (Bradtke and Suff, 1994), which is proven to converge to an optimal policy if certain conditions are satisfied (Parr, 1998). Figure 1 shows the steps of the algorithm. First, the Q-values are initialized to 0. Then, in Step 2, to balance exploration and exploitation, an action is chosen for current state s according to the $\epsilon - greedy$ policy in which the greedy action is taken with probability $(1-\epsilon)$ and a random action is taken with probability ϵ. In Step 2.2, the realization of the degradation path on the current state-action pair and the product demands are sampled. By using these values, the one period cost $c(s,a)$ and the next state s' are determined by the system equations given in Appendix. Once these values are obtained, the state-action value of the visited state-action pair, $Q(s,a)$, is updated with the temporal difference $\left(c(s,a) + e^{-\gamma\tau} \min_{a' \in A(s')} Q(s',a') - Q(s,a)\right)$, where τ is the sampled sojourn time to the next iteration, and γ is the discount factor which is chosen to be sufficiently small in order to minimize the average cost. Note that the learning rate, α, satisfies the necessary conditions for optimality (Parr, 1998).

5 NUMERICAL EXAMPLE

In this section, we consider a manufacturing system that produces three products $(N = 3)$. For each product n, the storage capacity is $\bar{x}_n = 20$. The setup costs are set to zero as they mainly correspond to the time lost when the system in not operating, which is represented by the setup times. For each product, the randomly generated parameters within the specific ranges are shown in Table 1. The machine degradation follows a gamma Process with shape $\alpha = 0.5$ and scale parameter $\beta = 1$, and the failure threshold is $L = 10$. Based on the procedure proposed by De Jonge (2019), the gamma process is approximated by a discrete-time Markov Chain with 21 states where 1 is the as-new state and 21 is the failure state F.

Table 1: Problem parameters.

Parameter	Value
Demand rates	U(0.1, 0.25)
Production rates	U(1,2)
Preventive maintenance time	U(4,8)
Corrective maintenance time	U(10,16)
Setup time	U(0.5,2)
Preventive maintenance cost	U(300,500)
Corrective maintenance cost	U(800,1000)
Holding costs	U(0.5,2)
Lost sales costs	U(100,200)

In the Q-learning algorithm, the actions are selected according to $\epsilon - greedy$ policy. Initially, we set ϵ to 0.1, then every 10^5 iterations, it is reduced by $\epsilon \leftarrow 0.9 \times \epsilon$. In updating the Q-values, the harmonic learning rate, $\alpha = b/(b + N_t(s,a) - 1)$ with $b = 5$, is used, where $N_t(s,a)$ is the number of times that a state action pair (s,a) has been visited up to the time t; the parameter b is tuned based on the procedure in Powell (2011). Figure 2 shows the learning curve for the Q-learning. The convergence of the average cost rate towards the end of the simulation indicates that the algorithm has reached a stable policy. The solution algorithm is implemented in Wolfram Mathematica 12. The computational experiment took 223.35 minutes on a quad-core Intel i7 processor running at 1.80 GHz with 16 GB RAM.

6 CONCLUSIONS

In this paper, we have addressed the joint optimization of production, inventory and CBM for a stochastically degrading multi-product production system with considerable setup and maintenance times under stochastic product demands. The goal is to find a dynamic production and maintenance policy so as to minimize the long run average cost per time

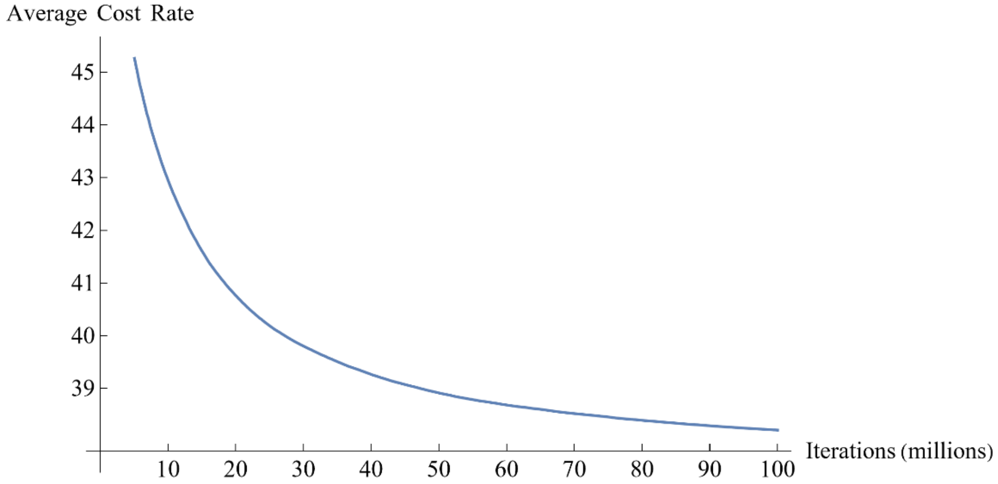

Figure 2: Learning curve of the Q-learning.

unit, which contains setup, holding, lost sales, preventive maintenance and corrective maintenance costs. The problem has been formulated as a SMDP, in which decisions are made based on the current machine status and stock levels of the products. To solve this problem, we have proposed a Q-learning algorithm and illustrate its performance by a numerical example.

REFERENCES

Aghezzaf E-H., Jamali M., and Ait-Kadi D. (2007). An integrated production and preventive maintenance planning model. European Journal of Operational Research, 181 (2), 679–685.

Ben-Daya M., and Makhdoum M. (1998). Integrated production and quality model under various preventive maintenance policies. Journal of the Operational Research Society, 49(8), 840-853.

Ben-Daya M. (2002). The economic production lot-sizing problem with imperfect production processes and imperfect maintenance. International Journal of Production Economics, 76 (3), 257–264.

Bradtke J.A, and Moore, A.W. (1995). Reinforcement learning methods for continuous-time Markov decision problems. In *Advances in Neural Information Processing Systems 7: Proceedings of the 1994 Conference* Denver, Colorado. MIT Press.

Cheng G.Q., Zhou B.H., and Li L. (2017). Joint optimization of lot sizing and condition-based maintenance for multi-component production systems. *Computers & Industrial Engineering, 110*, 538-549.

Cheng G.Q., Zhou B.H., and Li L. (2018). Integrated production, quality control and condition-based maintenance for imperfect production systems. *Reliability Engineering & System Safety, 175*, 251–264.

Darendeliler A., Claeys D., and E.H. Aghezzaf (2022). Integrated condition-based maintenance and multi-item lot-sizing with stochastic demand. *Journal of Industrial and Management Optimization*. (In press).

De Jonge B. (2019). Discretizing continuous-time continuous-state deterioration processes, with an application to condition-based maintenance. *Reliability Engineering & System Safety, 188*, 1-5.

El-Ferik S. (2008). Economic production lot-sizing for an unreliable machine under imperfect age-based maintenance policy. *European Journal of Operational Research, 186(1)*, 150–163.

Iravani S., and Duenyas I. (2002). Integrated maintenance and production control of a deteriorating production system. *IIE Transactions, 34(5)*, 423–435.

Jafari L., and Makis V. (2015). Joint optimal lot sizing and preventive maintenance policy for a production facility subject to condition monitoring. *International Journal of Production Economics, 169*, 156-168.

Jafari L., and Makis V. (2016). Joint optimal lot sizing and maintenance policy for a partially observable two-unit system. *The International Journal of Advanced Manufacturing Technology, 87*, 1621-1639.

Jafari L., and Makis V. (2019). Optimal production and maintenance policy for a partially observable production system with stochastic demand. *International Journal of Industrial and Systems Engineering, Vol:13,* No:7.

Jardine A., Lin D., and Banjevic D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*, 1483–1510.

Khatab A., Diallo C., Aghezzaf E-H., and Venkatadri U. (2019). Integrated production quality and condition-based maintenance for a stochastically deteriorating manufacturing system. *International Journal Production Research, 57(8)*, 2480-2497.

Liao G.L., and Sheu S.H. (2011). Economic production quantity model for randomly failing production process with minimal repair and imperfect maintenance.

International Journal of Production Economics, 130, 118-124.

Löhndorf N., and Minner S. (2013). Simulation optimization for the stochastic economic lot scheduling problem. *IIE Transactions, 45*, 796-810.

Parr R.E. (1998). Hierarchical control and learning for Markov decision processes. PhD diss., University of California at Berkeley.

Peng H., and van Houtum G.-J. (2016). Joint Optimization of Condition-based Maintenance and Production Lot-sizing. *European Journal of Operational Research, 253*, 94–107.

Powell W.B. (2011). *Approximate Dynamic Programming*. Second edition. John Wiley & Sons.

Qiu J., and Loulou R. (1995). Multiproduct production/inventory control under random demands. *IEEE Transactions on Automatic Control, 40(2)*, 350–356.

Shamsaei F., and Van Vyve M. (2017). Solving integrated production and condition-based maintenance planning problems by MIP modeling. *Flexible Services and Manufacturing Journal, 29*, 184-202.

Sloan T.W. (2004). A Periodic Review Production and Maintenance Model with Random Demand, Deteriorating Equipment, and Binomial Yield. *Journal of the Operational Research Society, 55(6)*, 647–656.

Sox C.R., Jackson P.L., Bowman A., and Muckstadt J.A. (1999). A review of the stochastic lot scheduling problem. *International Journal of Production Economics, 62(3),* 181-200.

Suliman S.M., and Jawad S.H. (2012). Optimization of preventive maintenance schedule and production lot size. *International Journal of Production Economics, 137*, 19-28.

Wang. J., Xueping L., and Xiotan Z. (2012). Intelligent dynamic control of stochastic economic lot scheduling by agent-based reinforcement learning. *International Journal of Production Research, 50(16)*, 4381-4395.

Watkins C.J. (1989). Learning from delayed rewards. PhD diss, Cambridge: Kings College.

Winands E., Adan I., and van Houtum G.J. (2011). The stochastic economic lot scheduling problem: *A survey. European Journal of Operational Research, 201(1)*, 1–9.

Xiang Y., Cassady C.R., Jin T., and Zhang C.W. (2014). Joint production and maintenance planning with machine deterioration and random yield. *International Journal Production Research, 52(6)*,1644–1657.

APPENDIX

The system evolves as follows: At a decision epoch, upon observing the state $s = (i, k, x_1, \dots, x_N)$, one of the following actions can be chosen: (1) perform corrective maintenance if and only if failure occurs; (2) perform preventive maintenance; (3) keep the machine idle; (4) produce item m. Then, based on state $s = (i, k, x_1, \dots, x_N)$, the selected action $a \in A(s)$, the sampled degradation path and the demand values (also the arrival times), the one-period cost $c(s, a)$ and the next state $s' = (i', k', x_1', \dots, x_N')$ are determined by the following equations:

Let,

τ: time until the next decision epoch,
d_n: sampled demand for product n from $Poisson(\lambda_n \tau)$ for $n = 1, \dots, N$,
S_n^r: sampled r^{th} demand arrival time for $r = 1, \dots, d_n$ and for product $n = 1, \dots, N$,
$w_n = min\{x_n, d_n\}$ for product $n = 1, \dots, N$,

1. If $i = F$ and corrective maintenance is being performed ($a = -2$), then

$$\tau = CT,$$
$$c(s,a) = c_c + \sum_{n=1}^{N} \sum_{r=1}^{w_n} \frac{\left(1 - e^{-\gamma S_n^r}\right)}{\gamma} ch_n + \frac{\left(1 - e^{-\gamma\tau}\right)}{\gamma} \sum_{n=1}^{N} ch_n (x_n - w_n) + \sum_{n=1}^{N} \sum_{r=w_{n+1}}^{d_n} e^{-\gamma S_n^r} cl_n,$$
$$i' = 1, k' = 0, x_1' = (x_1 - d_1)^+, \dots, x_N' = (x_N - d_N)^+.$$

2. If $i < F$ and preventive maintenance is being performed ($a = -1$), then

$$\tau = PT,$$
$$c(s,a) = c_p + \sum_{n=1}^{N} \sum_{r=1}^{w_n} \frac{\left(1 - e^{-\gamma S_n^r}\right)}{\gamma} ch_n + \frac{\left(1 - e^{-\gamma\tau}\right)}{\gamma} \sum_{n=1}^{N} ch_n (x_n - w_n) + \sum_{n=1}^{N} \sum_{r=w_{n+1}}^{d_n} e^{-\gamma S_n^r} cl_n,$$
$$i' = 1, k' = 0, x_1' = (x_1 - d_1)^+, \dots, x_N' = (x_N - d_N)^+.$$

3. If $i < F$ and the machine is being kept idle ($a = 0$), then, let $S^* = min\{S_1^1, \dots, S_N^1\}$, and $S_r^1 = S^*$, for some $r \in \{1, \dots, N\}$, then

$$\tau = S^*,$$
$$c(s,a) = \frac{\left(1 - e^{-\gamma\tau}\right)}{\gamma} \sum_{n=1}^{N} ch_n x_n + e^{-\gamma\tau} cl_r (1 - x_r)^+,$$
$$i' = i, k' = 0, x_r' = (x_r - 1)^+.$$
$$x_m' = x_m \, for \; m \in \{1, \dots, N\} \setminus r$$

4. If $i < F$ and product m is going to be produced $(a = m)$, then

4.1. If $a = m = k$,

$$\tau = \rho_m,$$

$$c(s,a) = \sum_{n=1}^{N}\sum_{r=1}^{w_n}\frac{\left(1 - e^{-\gamma S_n^r}\right)}{\gamma}ch_n$$

$$+\frac{(1 - e^{-\gamma\tau})}{\gamma}\sum_{n=1}^{N}ch_n(x_n - w_n)$$

$$+\sum_{n=1}^{N}\sum_{r=w_n+1}^{d_n}e^{-\gamma S_n^r}cl_n,$$

$$i' \geq i,\ k' = m, x_1{}' = (x_1 - d_1)^+,$$
$$\dots, (x_m - d_m)^+ + 1, \dots, x_N{}' = (x_N - d_N)^+.$$

4.2. If $a = m \neq k$,

$$\tau = ST_m,$$

$$c(s,a) = cs_m + \sum_{n=1}^{N}\sum_{r=1}^{w_n}\frac{\left(1 - e^{-\gamma S_n^r}\right)}{\gamma}ch_n$$

$$+\frac{(1 - e^{-\gamma\tau})}{\gamma}\sum_{n=1}^{N}ch_n(x_n - w_n)$$

$$+\sum_{n=1}^{N}\sum_{r=w_{n+1}}^{d_n}e^{-\gamma S_n^r}cl_n,$$

$$i' = i, k' = m, x_1{}' = (x_1 - d_1)^+,$$
$$\dots, (x_m - d_m)^+, \dots, x_N{}' = (x_N - d_N)^+.$$