Integrated production, sampling quality control and maintenance of deteriorating production systems with AOQL constraint

Abstract – This paper considers the problem of integrated production, preventive maintenance and quality control for a stochastic production system subject to both reliability and quality deteriorations. A make-to-stock production strategy is used to provide protection to the serviceable stock against uncertainties. The quality control is performed by a single acceptance sampling plan by attributes. The preventive maintenance strategy consists of carrying out an imperfect maintenance as a part of the setup activity at the beginning of each lot production, while, a major maintenance (overhaul) is undertaken once the proportion of defectives in a lot rejected reaches or exceeds a given threshold. The main objective of this study is to jointly optimize the production lot size, the inventory threshold, the sampling plan parameters and the overhaul threshold by minimizing the total incurred cost. To meet customer requirements, the optimization problem is subject to a specified constraint on the average outgoing quality limit (AOQL). A stochastic mathematical model is developed and solved using a simulation-based optimization approach. Numerical examples and thorough sensitivity analyses are provided to illustrate the efficiency of the proposed integrated model and the robustness of the resolution approach. Compared with the 100% inspection policy which is widely used in the literature on integrated production, maintenance and quality control, the results obtained show that an economic design of acceptance sampling in such an integrated context can lead to important cost savings of more than 20%.

Keywords – Dynamic process deterioration, production/inventory control, lot sizing, acceptance sampling plan, preventive maintenance, simulation-based optimization.

1. Introduction

Many efforts have been devoted in the past few decades to integrate production planning, quality control and maintenance scheduling and to investigate the hidden interactions between the three aspects. In a recent literature review on this topic, Hadidi et al. (2012) made a difference between the concepts of interrelation and integration between the three fundamental functions: interrelated models are those in which the decisions variables of only one function is considered taking into account the remaining functions as constraints, while, integrated models are those in which two or the three functions are modelled and optimized simultaneously. Based on Hadidi et al.'s definitions, we find that most of the integrated models in the literature consider only two functions at a time. For example, many models integrating only production and preventive maintenance (PM) have been proposed since the second half of the 1990s without considering the quality aspect (see the literature review by Budai et al., 2008). Recent advances in integrated production and PM includes joint determination of the Economic Production Quantity (EPQ) and PM policy (e.g. Sana, 2012; Liao, 2013), joint production and opportunistic PM scheduling (e.g. Xia et al., 2012, 2015) and simultaneous control of production and PM rates (e.g. Berthaut et al., 2011; Assid et al., 2015b). On the other side, research on integrating only the production and quality control policies dates back to the 1970s and 1980s (see the literature review by Goyal et al., 1993). More recently, Inman et al. (2013) surveyed the advances on the interface between quality and production system design in the past two decades.

Research during this period has been concerned with the mutual effects of production and quality settings such as the impact of complexity and technology of production, operations speed, setup planning and tolerance design on the deterioration of process quality (e.g. Sana, 2010a; Pal et *al.*, 2013; Liu et *al.*, 2009; Jeang, 2012) and, conversely, the impact of quality inspection planning on the production flow (e.g. Kim and Gershwin, 2008), etc. In addition, there is a growing interest in the integration of production control design with the Statistical Quality Control (SQC) techniques such as control charts (e.g. Colledani and Tolio, 2011), process capability (e.g. Hajji et *al.*, 2011a), and sampling plans (Bouslah et *al.*, 2013, 2014). Nevertheless, Inman et *al.* (2013) have reported that there are still a large number of sub-areas in quality control (including reliability and maintenance scheduling) that have not been fully integrated with production and they recommended more investigation into the traditional quality control system design in the production context.

Indeed, only a limited number of papers in the literature deal with the simultaneous integration of the three functions. We can classify those papers into two categories based on the quality control policy used. The first category includes studies integrating production and PM design with a 100% inspection policy of all items produced. For example, Liao et *al.* (2009) and Wee and Widyadana (2013) integrated PM programs with the EPQ model for an imperfect production process where all defective items produced must be reworked. Radhoui et *al.* (2009, 2010) suggested an integrated PM and production control policy for an unreliable imperfect process producing a random proportion of non-conforming items. They assumed that each lot produced is subject to an automated quality control of negligible duration and cost. The second category of integrated models corresponds to studies using the SQC tools rather than 100% inspection. For example, Ben-Daya and Makhdoum (1998) and Ben-Daya (1999) presented various integrated models for the joint determination of the EPQ, the economic design of control chart and the optimal PM level. Nevertheless, some other important aspects of the SQC such as the acceptance sampling plans have not yet been integrated simultaneously with production and PM planning. Acceptance sampling plans have been widely used in the industry for a long time to control the outgoing quality especially in situations where 100% inspection of all items produced is technically or economically impractical (Schilling and Neubauer, 2009). In addition, they have significant impacts on production and inventory as shown by Bouslah et *al.* (2013). Unlike 100% inspection and control charts, the interactions between acceptance sampling plans and PM policies have not been investigated yet in the literature.

In the literature on integrated models, many attempts have been made by researchers to adequately pattern the product quality and equipment reliability deteriorations. For example, Rosenblatt and Lee (1986) studied three dynamic patterns of process deterioration (linear, exponential and multi-state) on the EPQ. Moreover, many industrial and academic studies have shown the important impact of production rate on the performance deterioration intensity as in Felix Offodile and Ugwu (1991), Khouja and Mehrez (1994) and Sana (2010b). However, for simplicity, most of the existing integrated models neglect the dynamic aspect of process performance deterioration and the impact of production settings on the deterioration intensity. Generally, the researchers assume that the proportion of defective items produced during the 'out-of-control' periods is constant or follows a prior known distribution.

Furthermore, almost all of the integrated models do not simultaneously consider the quality and reliability deterioration phenomena (Chakraborty et *al.*, 2009). When both phenomena are observed, the PM plays a double role: increasing the reliability of the production equipment and restoring the

product quality to the desired level (Ben-Daya and Duffua, 1995; Rivera-Gomez et *al.*, 2013). Due to the direct impact of deterioration on the production system availability and on the output quality, it is more appropriate to base the PM decision on the actual deterioration state rather than on the equipment age (Grall et *al.*, 2002). An inference on the deterioration state could be based on the equipment condition or on the product quality characteristic (Colledani and Tolio, 2012). Condition-based maintenance has attracted a great deal of attention in the past two decades in conjunction with the technological advances in condition monitoring techniques such as vibration, corrosion, thermography and acoustics analysis (Rao, 1996; Davies, 1998). On the other hand, in situations where quality is directly affected by the degradation of the production system, the quality information feedback, which does not require a costly and high technology for data acquisition and analysis as in the condition monitoring techniques, could represent an alternative solution to recognise the system degradation. Maintenance based on the quality information feedback is becoming more and more attractive field of research especially in the context of maintenance and quality control integration. Tapiero (1986) is among the first who formulated a feedback stochastic control maintenance problem based on the state of the products quality, assuming that quality is a known function of the machine's degradation state. Hsu and Kuo (1995) studied the performance of an inspection and maintenance policy that begins 100% inspection of a production lot after producing a given number of items and then initiates a preventive/corrective maintenance activity when the fraction of defective parts reaches a given threshold. Similarly, Radhoui et *al.* (2009, 2010) also used the 100% inspection policy to determine the proportion of non-conforming items of each lot produced and then compare this proportion to some given thresholds to make decisions on PM and overhaul actions. Recently, Panagiotidou and Tagaras (2010), Pan et *al.* (2012) and Zhang et *al.* (2015) suggested integrating condition-based maintenance and statistical process control strategies where the maintenance decisions are made based on the quality information from the control chart. Nevertheless, the interactions between the acceptance sampling plans and maintenance strategies have never received the same attention in the literature. To the best of our knowledge, there is no published study that investigates the usefulness and relevance of information provided by sampling plans such as the observed percentage of lots accepted/rejected, the current inspection mode (sampling or 100% inspection), etc., for process condition monitoring and maintenance decision-making.

To overcome the limitations of the existing integrated models, in this paper, we intend to develop a new model integrating production lot sizing, production rate control, inventory control, single acceptance sampling plan and PM strategy. Our focus on the acceptance sampling plan techniques in the context of integrated operations management is motivated by three considerations. Firstly, acceptance sampling plans have specific statistical properties (Schilling and Neubauer, 2009) that should be deeply analyzed in order to extract relevant information for process condition monitoring and to make accordingly the appropriate maintenance decisions. Secondly, compared with the 100% inspection policy which is extensively used in the integrated models in the literature, sampling plans are usually more economical and they significantly reduce the unnecessary inspection essentially during periods when the process is in the 'in-control' state (Montgomery, 2008a). Thirdly, it is expected that an economic design of acceptance sampling plans in such a context, instead of using traditional sampling inspection standards such as ANSI/ASQC Z1.4 and ISO 2859, could lead to important economic savings (Nikolaidis and Nenes, 2009). In fact, those standards are purely based on quality considerations and they completely neglect the economic aspect and the interactions with the production, inventory and maintenance in the design of sampling plans.

In this study, we present a stochastic dynamic model considering non-negligible delays and costs of setup, quality control and maintenance operations. The product quality and machine reliability deteriorations depend both on the production equipment usage. We consider that the production setup includes an imperfect PM activity. An overhaul is also required to perfectly restore the performance of the production process. Our objective is to jointly design and optimize the production, quality control and maintenance policies. The optimal integrated solution should minimize the total incurred cost while meeting a predefined restriction on the average outgoing quality limit (AOQL). We used a simulation-based optimization approach to solve such a complex and stochastic problem. Moreover, we present a thorough analysis of the performance and benefits of the proposed integrated model.

The paper is organized as follows. Section 2 presents the notation and the description of the problem under study. The system dynamic modelling and the optimization problem are formulated in Section 3. In Section 4, we present the simulation-based optimization approach. Illustrative numerical examples, and sensitivity and comparative analysis are given in Section 5. Section 6 discusses some managerial implications for the proposed integrated control policy. Finally, Section 7 concludes the paper and provides some directions for future research.

2. Notation and problem description

2.1. Notation

The notations used in this paper are defined as follows.

Decision variables:

Q Production lot size
S Surplus inventory threshold
n Sample size
c Acceptance number
r Overhaul threshold (ratio)

Model parameters:

u_{max} Maximum production rate
d Demand rate
$AOQL_{max}$ Maximum accepted level of the Average Outgoing Quality Limit
τ_{cm} Random variable denoting the corrective maintenance duration
τ_{ovr} Random variable denoting the overhaul duration, ($\overline{\tau}_{ovr} >> \overline{\tau}_{cm}$)
τ_{ins} Unit inspection duration
τ_{set} Setup duration for each production run
C_h Unit inventory holding cost per unit time
C_b Unit backlog cost per unit time ($C_b >> C_h$)
C_{set} Setup cost (including the cost of the imperfect PM)
C_{cm} Corrective maintenance cost
C_{ovr} Overhaul maintenance cost ($C_{ovr} >> C_{cm}$)
C_{ins} Unit inspection cost
C_{rej} Unit rejection cost of a defective item
C_{def} Unit cost of selling a non-inspected defective item

Other notations will be used to model the system deterioration and the dynamic of inventories.

2.2. Problem description and assumptions

The manufacturing system under study consists of a single-product batch-processing production unit supplying a downstream serviceable stock as illustrated in Figure 1. This stock is used to fulfill a continuous and constant market demand d. The production rate $u(.)$ is flexible and can be set at any time at a value between 0 and a maximum level u_{max}. The production unit is subject to a continuous operation-dependent degradation which leads to an increasing failure probability and an increasing proportion of defectives produced. Therefore, maintenance interventions are required to maintain and restore the performances of the production unit. In response to each failure event, a corrective maintenance (minimal repair) is undertaken, which returns the production unit to the 'as-bad-as-old' condition. To preventively cope with the system degradation, an imperfect PM is carried out as a part of the setup activity at the beginning of each production run. We consider that the efficiency of this imperfect PM decreases continuously as the production unit ages. Thus, we assume that the setup reduces the effective age $a(.)$ of the production unit by a certain amount $\phi(.)$ called *improvement factor* which is a decreasing function of the real age $A(.)$ (Wang and Pham, 2006). In addition, a major perfect maintenance (overhaul) is conducted as soon as the rate of defective items produced reaches or exceeds a given threshold r. This feedback overhaul policy is used for two reasons. First, the PM during setups is insufficient to perfectly improve the production unit performance as its perfectness deteriorates with process usage. Second, as the product quality depends intimately on the production unit condition, the rate of defectives that are produced provides a relevant indication of the overall deterioration state and therefore it could be useful as a control parameter for the overhaul scheduling.

In order to ensure that the delivered products meet the outgoing quality requirement, a quality control of lots produced is performed before they arrive to the final serviceable stock. Herein, a single acceptance sampling plan by attributes is used to guarantee an acceptable outgoing quality level. A sample of size n is drawn randomly from each lot produced and inspected item by item. If the number of defectives does not exceed the acceptance number c, then the lot is accepted. Otherwise, the lot is rejected and a 100% inspection is performed in order to sort all the defective items. We assume that the defective items are not rectifiable. Hence, all the defectives, found either in sampling or in 100% inspection, are rejected with no replacement. Depending on the proportion of defective items found in each lot rejected compared to the threshold value r, the decision maker can decide whether or not to immediately initiate the overhaul.

The duration of the setup (including the imperfect PM) is constant. However, the durations of the corrective maintenance (CM) and the overhaul are stochastic, following general probability distributions. Under these assumptions and because the quality control delay of each lot produced is unpredictable and variable depending on the acceptance/rejection decision, shortages may occur. A make-to-stock production strategy is used in order to provide protection to the serviceable stock against uncertainties in production and quality control. Hence, the production rate is controlled over time by a feedback base-stock control policy derived from the well-known hedging point policy (Akella and Kumar, 1986). Our choice of the hedging point method for the production-inventory control is motivated by its optimality, simplicity and ease of implementation features (Hu et al., 1994; Gershwin, 2000; Sarimveis et al., 2008). Finally, we consider that the production of the interrupted lots is always resumed after maintenance interventions.

Our objective is to find jointly the optimal production lot size Q, the optimal inventory surplus S, the optimal sampling parameters n and c, and the optimal overhaul threshold r that minimize the total incurred cost. This includes the inventory and backlog costs, quality control costs and maintenance (setup, overhaul and CM) costs. The optimal solution must satisfy a number of constraints related to the system dynamic and the customer-perceived quality.
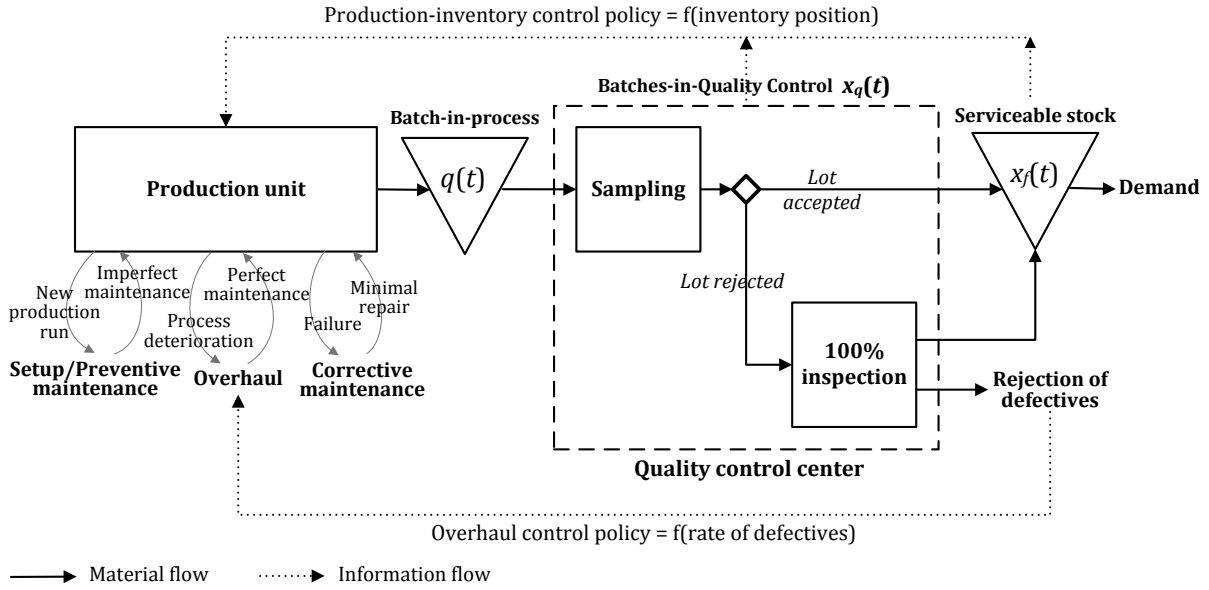

Figure 1. A deteriorating production system with quality control, PM and overhaul.

3. Problem formulation

3.1. Deterioration model

The state of the production unit can be characterized at each instant t by five continuous-time components including:

- A discrete-state stochastic process $\{\alpha(t), t \geq 0\}$ which describes the status of the production unit at each time t, and takes values $\{0,1,2,3\}$ such that : $\alpha(t) = 0$, if the production unit is under CM, $\alpha(t) = 1$, if it is available for production, $\alpha(t) = 2$, if it is under setup, and $\alpha(t) = 3$, if it is under overhaul.
- A piecewise continuous variable, $A(t)$, which represents the cumulative number of items produced since the last overhaul until time t. We call this variable the real age and it is calculated using the following formula:

$$\frac{\partial A(t)}{\partial t} = u\left(t, \alpha(t)\right), \forall\, t \geq T, A(T) = 0 \tag{1}$$

where T is the completion time of the last overhaul.

- A piecewise continuous variable, $a(t)$, which represents the reduced age of the production unit at time t. This variable measures the cumulative effects of the setups on the real age of the production unit and is called the virtual age (also called effective age). It is calculated using these equations:

$$\frac{\partial a(t)}{\partial t} = u\left(t, \alpha(t)\right), \forall\, t \in \left[\theta_k, \theta_{k+1}\right[,\, k = 0,1,..,\infty,\, a(T) = 0 \tag{2}$$

$$a(\theta_k^+) = a(\theta_k^-) - \phi\left(A(\theta_k)\right) \cdot a(\theta_k^-),\, k = 1,..,\infty \tag{3}$$

where θ_k is, at the same time, the end time of the kth setup activity and the start time of the kth production run.

- The improvement factor, $\phi(.)$, which is a decreasing continuous function of the real age $A(.)$ and is described by the following equation:

$$\phi\left(A(t)\right) = \exp\left(-\lambda_{set}\, A(t)^{\gamma_{set}}\right) \tag{4}$$

where λ_{set} and γ_{set} are given positive constants.

- The probability of failure, $F(.)$, which depends instantly on the current virtual age $a(.)$ following a Weibull distribution:

$$F\left(a(t)\right) = 1 - \exp\left(-\lambda_r\, a(t)^{\gamma_r}\right) \tag{5}$$

where λ_r and γ_r are given positive constants.

- The proportion of defective items produced at time t, $p(t)$, which depends also on the current virtual age $a(.)$ as follows:

$$p\left(a(t)\right) = p_0 + \eta\left(1 - \exp\left(-\lambda_q\, a(t)^{\gamma_q}\right)\right) \tag{6}$$

where, p_0 is a very small proportion of defectives produced at the initial condition (i.e., 'as-good-as-new' state), λ_q and γ_q are given positive constants and η is the boundary considered in the quality deterioration.

From Eqs. (4), (5) and (6), the deterioration functions of the improvement factor $\phi(.)$, the probability of failure $F(.)$ and the rate of defectives $p(.)$ belong to the two-parameter exponential family of distributions (Ferguson, 1962). The parameters of those functions can be determined from historical information using estimation methods such as the maximum likelihood and the least squares (Hossain and Zimmer, 2003).

3.2. Integrated control policy of production, quality and maintenance

3.2.1. Quality control policy

The decision on the acceptance/rejection of each lot produced is based on the number of defective items found in the random sample n which itself depends on the number of defective items in the entire lot. Let X_k be the variable denoting the number of defectives in the kth lot produced, $k = 1,2,..,\infty$. X_k can be determined by solving these equations:

$$\frac{\partial X_k(t)}{\partial t} = p\left(a(t)\right) \cdot u\left(t, \alpha(t)\right), \forall t \in \left]\theta_k, \delta_k\right], \; X_k\left(\theta_k\right) = 0 \tag{7}$$

where, δ_k is, the end time of the kth production run (i.e., time when the kth lot is completely processed).

Let Y_k be the variable indicating the number of defective items in the sample n of the kth lot. The probability of finding j defective items, $0 \leq j \leq n$, in the sample n of the kth lot, $k = 1,2,..,\infty$, can be calculated using the binomial distribution as follows:

$$\Pr\left(Y_k = j\right) = \binom{n}{j}\left(\frac{X_k}{Q}\right)^j\left(1 - \frac{X_k}{Q}\right)^{n-j} \tag{8}$$

Then, the probability of acceptance of the kth lot produced $P_a^k(.)$, $k = 1,2,..,\infty$, is calculated as follows

$$P_a^k\left(n,c,Q,X_k(.)\right)=\Pr\left(Y_k\le c\right)=\sum_{j=0}^{c}\Pr\left(Y_k=j\right) \tag{9}$$

As the accepted lots do not undergo a 100% screening inspection, the defective items existing in these lots will be transmitted to the final serviceable stock and therefore will be sold to the customers. The average proportion of defective items transmitted to the serviceable stock through each kth lot produced, also called the Average Outgoing Quality (AOQ_k), $k = 1,2,..,\infty$, is given by

$$AOQ_k\left(n,c,Q,X_k(.)\right)=\frac{\sum_{j=0}^{c}\Pr(Y_k=j)\left(X_k-j\right)}{\sum_{j=0}^{c}\Pr(Y_k=j)\left(Q-Y_k\right)+\sum_{j=c+1}^{n}\Pr(Y_k=j)\left(Q-X_k\right)} \tag{10}$$

The maximum level of AOQ_k, $k = 1,2,..,\infty$, over all possible values of X_k is called the Average Outgoing Quality Limit ($AOQL$) which can be calculated as follows (Schilling and Neubauer, 2009):

$$AOQL\left(n,c,Q\right)=\max_{0\le X_k\le Q}\left\{AOQ_k(.)\right\}=y(c)\left(\frac{1}{n}-\frac{1}{Q}\right) \tag{11}$$

where, $y(c)$ is equal to

$$y(c)=\frac{e^{-np_M}\left(np_M\right)^{c+2}}{c!} \tag{12}$$

and p_M is the value of the ratio X_k/Q at which the $AOQL$ occurs. Tables containing the closed approximated values of $y(c)$ for each given c independently of the sample size n can be found in Schilling and Neubauer (2009).

The manufacturer must select the combinations of production lot size Q and sampling plan parameters n and c in such a way that the $AOQL(.)$ does not exceed a maximum limit imposed by the customers denoted $AOQL_{max}$. Thus, from Eq. (11), we obtain this inequality

$$y(c)\left(\frac{1}{n}-\frac{1}{Q}\right)\le AOQL_{max} \tag{13}$$

3.2.2. Production-inventory control policy

The finished products can be held in three storage locations before being delivered to customers as schematized in Figure 1:
- A production downstream area to cumulate the produced parts of the ongoing batch-in-process until the end of the current production run. This inventory is measured instantly by a piecewise continuous variable denoted $q(.)$, where $0\le q(t)\le Q$, $\forall\ t$. The dynamic of the inventory $q(.)$ can be described by the following equations

$$\frac{\partial q(t)}{\partial t}=u\left(t,\alpha(t)\right),\ q\left(\theta_1\right)=0,\forall t\in\left]\theta_k,\delta_k\right[,\forall k=1,..,\infty \tag{14}$$

$$q\left(\delta_k^+\right)=q\left(\delta_k^-\right)-Q,\forall k=1,..,\infty \tag{15}$$

where, δ_k^- and δ_k^+ denote the left and right boundaries of the kth production run end time δ_k. Recall that each kth production run, $k=1,..,\infty$, can be interrupted many times by the CM and overhaul interventions. As mentioned before, the durations of these interventions are stochastic. In the case

where the production rate $u(.)$ remains unchanged during the production run, δ_k can be estimated as follows

$$\delta_k = \theta_k + Q/u + \sum_{i \in \Lambda(k)} \tilde{\tau}_{cm}^i + \sum_{j \in \Delta(k)} \tilde{\tau}_{ovr}^j \qquad (16)$$

where, $\tilde{\tau}_{cm}^i$ is the duration of the ith CM, $i=1,..,\infty$, and $\tilde{\tau}_{ovr}^j$ is the duration of the jth overhaul, $j=1,..,\infty$. $\Lambda(k)$ and $\Delta(k)$ are respectively the sets of CM and overhauls performed during the kth production run.

- A storage area where each lot produced is temporary held for quality control. This Work-In-Progress (WIP) inventory is measured by the continuous-time variable $x_q(.)$. Its dynamic is given by

$$x_q\left(\delta_k^+\right) = x_q\left(\delta_k^-\right) + Q, \ \forall k = 1,..,\infty \qquad (17)$$

$$x_q\left(\xi_k^+\right) = x_q\left(\xi_k^-\right) - Q, \ \forall k = 1,..,\infty \qquad (18)$$

where ξ_k^- and ξ_k^+ denote the left and right boundaries of the completion time of quality control, ξ_k, of the kth lot produced. ξ_k depends on the lot acceptance/rejection decision as follows:

$$\xi_k = \begin{cases} \delta_k + n\,\tau_{ins} & \text{if } Y_k \le c \text{ (lot accpeted)} \\ \delta_k + Q\,\tau_{ins} & \text{if } Y_k > c \text{ (lot rejected)} \end{cases} \qquad (19)$$

Note that ξ_k indicate also the arrival time of the kth lot to the serviceable stock.

- The final serviceable stock which is used to meet the market demand. This stock (inventory if positive and backlog if negative) is measured by a piecewise continuous variable denoted $x_f(.)$. The dynamic of the serviceable stock is described by these equations

$$\frac{\partial x_f(t)}{\partial t} = -d, x_f(0) = x_f, \forall t \in]\xi_k, \xi_{k+1}[, \forall k = 1,..,\infty \qquad (20)$$

$$x_f\left(\xi_k^+\right) = x_f\left(\xi_k^-\right) + Ind\left\{Y_k \le c\right\}\left(Q - Y_k\right) + Ind\left\{Y_k > c\right\}\left(Q - X_k\right), \ \forall k = 1,..,\infty \qquad (21)$$

where, $Ind\{.\}$ is an indicator function defined as follows: $Ind\{\Theta(.)\}=1$ if $\Theta(.)$ is true, and $Ind\{\Theta(.)\}=0$ if $\Theta(.)$ is false. Thus, this function is used in Eq. (21) to indicate if the kth lot has been accepted (i.e., $Y_k \le c$) or not (i.e., $Y_k > c$).

From Eqs. (16), (19) and (21), one can see that the final serviceable stock is affected by two sources of disruption: the uncertainty in the duration of production runs due to the stochastic maintenance interventions, and the variability in the duration of quality control activities due to the uncertain decision on acceptance/rejection of lots produced. A surplus inventory S is used to protect the serviceable stock against stochastic variability and to mitigate the risk of shortage. According to the classical hedging point policy (HPP), the production rate $u(.)$ should be set at its maximum level during the build-up of the buffer stock S, which shall be maintained by setting $u(.)$ at the same level of the demand rate. In our context, some considerations should be included in the production-inventory control policy. First, the control of the surplus inventory should take into account the total amount of the on-hand lots including those under sampling and 100% inspection. Second, as the setup is a regular activity with a non-negligible delay, the loss in production during each setup should be systemically recovered in the subsequent production run. Otherwise, the surplus inventory S will be prematurely depleted. Third, for practical purposes, we consider some restrictions on the variation of

the production rate during each production run. In fact, the classical HPP assumes that there is an infinite surplus of raw material that allows the instantaneous augmentation of the production rate. In the case of unreliable supply systems, the production controllers are more concerned with the availability of the raw material (Hajji *et al.*, 2011b). In order to reduce the effects of the operations speed variation on the upstream supply chain, we assume that the production rate setting is determined only at the beginning of each production run $[\theta_k, \delta_k[, k= 1,2,..,\infty$.

Thus, by dividing the production-planning horizon into consecutive periods $\{[\theta_k, \theta_{k+1}[, k= 1,2,..,\infty\}$, we suggest the following production-inventory control policy:

$$u\left(t \in [\theta_k, \theta_{k+1}[, x, \alpha\right) = \begin{cases} u_{\max} & \text{if } \{x(\theta_k) < S \text{ or } \Gamma^k(t) = 1\} \text{ and } \{\alpha(t) = 1\} \text{ and } \{t \in [\theta_k, \delta_k]\} \\ \dfrac{d}{1 - \tau_{set} \cdot d/Q} & \text{if } \{x(\theta_k) = S\} \text{ and } \{\alpha(t) = 1\} \text{ and } \{t \in [\theta_k, \delta_k]\} \\ 0 & \text{if } \{\alpha(t) \in \{0,2,3\}\} \text{ or } \{t \in]\delta_k, \theta_{k+1}[\} \end{cases} \quad (22)$$

where $x(.)$ is called the inventory position and it is calculated instantly as follows

$$x(t) = x_q(t) + x_f(t), \forall t \geq 0 \quad (23)$$

and, $\Gamma^k(t)$ is binary function with 1 if a maintenance (CM or overhaul) occurs in $[\theta_k, t[$, and 0 if not. In fact, if the inventory position $x(.)$ at the beginning of each new kth production run is strictly below the threshold S, then the corresponding kth lot is manufactured at the maximum production rate u_{max}. Otherwise, if the inventory position $x(.)$ is exactly equal to the threshold S, then the production rate of the kth lot is set to an adjusted-demand rate $d/(1- \tau_{set} \cdot d/Q)$. This adjustment is required to compensate the loss in production during setups and to therefore maintain the surplus S. The production rate setting, either u_{max} or $d/(1- \tau_{set} \cdot d/Q)$, is preserved until the end of the production run. However, if a maintenance occurs during the production run, then the production is immediately stopped (i.e., $u(.)=0$) and, once it is completed ($\Gamma^k(t) = 1$), the production is resumed at the maximum production rate u_{max}. Finally, the production rate is reset to 0 once the production run is completed and during setups (i.e., $\delta_k < t < \theta_{k+1}$).

3.2.3. Maintenance policy

Once the surplus inventory S is built by setting the production rate at the maximum level u_{max}, it must be maintained through two mechanisms. First, the production rate has to be set at the adjusted-demand rate $d/(1- \tau_{set} \cdot d/Q)$ as previously explained. Second, the setup activities should be controlled in such a way that the inventory position is equal to the threshold S before starting a new production run. This means that the setup should be started before the inventory position is depleted to $S+\tau_{set} \cdot d$, where $\tau_{set} \cdot d$ is the amount of inventory consumed during the setup. Let $\Pi_k(t,x,\alpha)$ denote a binary function with 1 if the kth setup (including the PM) has to be carried out at time t, and 0 if not. Thus, the setup control policy is given by:

$$\Pi_k\left(t, x(t), \alpha(t)\right) = \begin{cases} 1 & \text{if } \{x(t) \leq S + \tau_{set} \cdot d\} \text{ and } \{t \geq \delta_{k-1}\} \text{ and } \{\alpha(t) = 1\} \\ 0 & \text{Otherwise} \end{cases}, \quad k = 1,..,\infty \quad (24)$$

From Eq. (24), a new kth setup is executed only when the following three conditions are satisfied: the inventory position is less or equal to $S+\tau_{set} \cdot d$, the previous production run is finished (i.e., $t \geq \delta_{k-1}$) and the production unit is available (i.e., $\alpha(t)=1$).

On the other hand, the overhaul is carried out when the 100% inspection of a kth lot rejected (i.e., $Y_k>c$) results to a rejection rate greater or equal to the threshold r. Let $\Omega_k(.)$ denote a binary function with 1 if an overhaul has to be performed based on the proportion of defectives in the kth lot produced, and 0 if not. Then, the overhaul control policy is represented by the following equation:

$$\Omega_k\left(n,c,Q,X_k(.)\right)=\begin{cases}1 & \text{if }\{Y_k>c\}\text{ and }\left\{\dfrac{X_k}{Q}\geq r\right\} \\ 0 & \text{Otherwise}\end{cases}, \quad k=1,..,\infty \tag{25}$$

Although the overhaul decision is made only based on the quality of rejected lots, the relevance of this policy to recognize the real state of the production process quality and whether to react accordingly by undertaking or not the overhaul lies in the fact that the frequency of rejection of lots produced reflects in itself the degree of quality deterioration. Indeed, one of the intrinsic characteristic of sampling plans is that the probability of rejection $1-P_a^k$, $k=1,2,..,\infty$, increases systematically as the quality deteriorates. Figure 2 depicts the deterioration of process quality with respect to the effective age $a(.)$, and its impacts on the quality of lots produced and the probability of rejection, as described by Eqs. (6), (7) and (9). In practice, regardless of the values of the sampling parameters n and c, the quality deterioration generates an increasing number of lots rejected which improves the availability of information about the rate of defectives and therefore increases the visibility on the process condition.
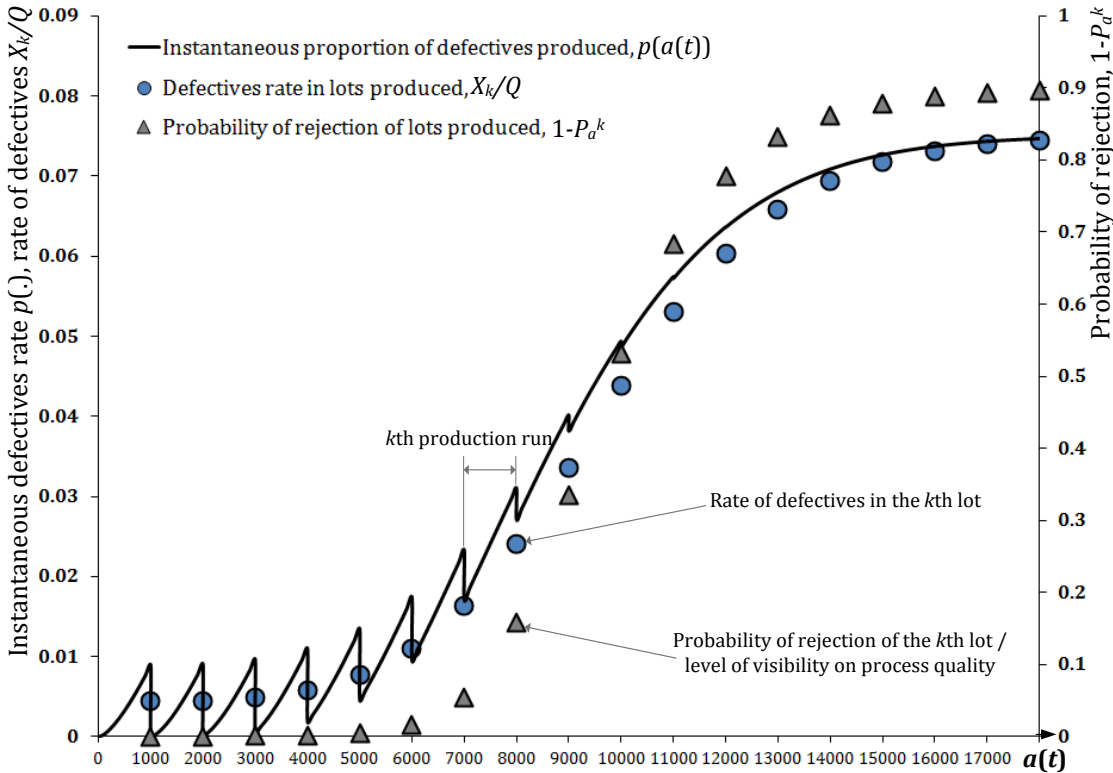

Figure 2. Impacts of process usage-deterioration on the quality of lots produced and on the probability of rejection.

3.3. Optimization problem

From the above mathematical formulation, many complex interactions between production, inventory, quality and maintenance have been highlighted. For example, the quality deterioration is influenced by the production run length and the production rate which depend respectively on the production lot size Q and the surplus threshold S, as shown in Eqs. (2), (6), (16), (7) and (22). The production lot size Q impacts also the frequency of the PM which improves partially the process performance as in Eq. (3). The overhaul decisions are impacted by the design of the sampling plan (n, c) as explained in section 3.4. The outgoing quality depends on the production lot size Q, on the quality deterioration state and on the sampling plan (n, c) as shown in Eq. (10). Hence, a trade-off solution resulting from a joint optimization approach can take into consideration all these interactions and accordingly increases the overall performance of the manufacturing system. Indeed, our aim is to determine jointly the optimal values of the production lot size Q, the inventory threshold S, the sampling plan parameters c and n and the overhaul threshold r which minimize the expected total cost per unit time, $ETC(.)$ and at the same time satisfy the $AOQL$ constraint (i.e., Ineq. (13)). The ETC consists of the sum of the inventory holding and backlog costs, the costs related to quality, the setup cost and the maintenance costs.

Let $G(t)$ denote the total cost of inventory holding and backlog in the period [0,t]. $G(t)$ is given at any time t by

$$G(t) = \int_0^t \left(C_h \left(q(z) + x_q(z) + x_f^+(z) \right) + C_b x_f^-(z) \right) \cdot dz \tag{26}$$

where $x_f^+(t) = \max\left(x(t), 0 \right)$ and $x_f^-(t) = \max\left(-x(t), 0 \right)$.

The expected quality cost $Q(t)$ in the period [0,t] includes the sampling cost, the 100% inspection cost of rejected lots, the cost of rejected items and the cost of accepting/selling the non-inspected defective items as follows:

$$
\begin{aligned}
Q(t) = &\, C_{ins} n N(t) && \text{(sampling cost)} \\
&+ C_{ins} \sum_{k=1}^{N(t)} \left(1 - P_a^k \right) (Q - n) && \text{(100\% inspection cost)} \\
&+ C_{rej} \sum_{k=1}^{N(t)} \left(1 - P_a^k \right) X_k && \text{(cost of rejected items)} \\
&+ C_{def} \sum_{k=1}^{N(t)} \sum_{j=0}^{c} \Pr(Y_k = j) \left(X_k - j \right) && \text{(cost of accepting/selling defective items)}
\end{aligned}
\tag{27}
$$

where, $N(t)$ is the total number of lots produced in the period [0,t]. $N(t)$ can be calculated as follows

$$N(t) = \frac{d \cdot t + x(t)}{Q} \tag{28}$$

The expected maintenance cost $M(t)$ during the same period [0,t] includes the costs of setup, CM and overhaul activities. It is given by

$$M(t) = C_{set} N(t) + C_{cm} f(t) + C_{ovr} \, m(t) \tag{29}$$

where $f(t)$ and $m(t)$ are respectively the expected numbers of CM and overhaul interventions in the period [0, t].

Thus, the $ETC(.)$ is obtained

$$ETC\left(Q, S, c, n, r\right) = \lim_{t \to +\infty} \frac{1}{t} \left(G(t) + Q(t) + M(t) \right) \tag{30}$$

The optimization problem is to solve the following mixed-integer, non-linear and stochastic model:

$$
\begin{cases}
\text{Minimize} \quad ETC\left(Q, S, c, n, r\right) \\[2ex]
\text{Subject to} \quad y(c)\left(\dfrac{1}{n} - \dfrac{1}{Q}\right) \leq AOQL_{\max}, \qquad (AOQL \text{ constraint}) \\[3ex]
\qquad\qquad\quad \text{Eqs. (1)-(9),} \qquad\qquad\quad \text{(dynamics of quality and reliability deteriorations)} \\[2ex]
\qquad\qquad\quad \text{Eqs. (14)-(21),} \qquad\qquad\; \text{(dynamics of inventories)} \\[2ex]
\qquad\qquad\quad \text{Eqs. (22)-(25),} \qquad\qquad\; \text{(production, setup and overhaul control policies)} \\[2ex]
\qquad\qquad\quad c < n < Q, \\[2ex]
\qquad\qquad\quad 0 < r < 1, \\[2ex]
\qquad\qquad\quad Q, S, c, n, r \geq 0; \quad Q, S, c, n : \text{integers}
\end{cases}
$$

Since it is extremely hard to solve this problem either analytically or numerically, because of the complexity of the dynamic constraints (i.e., continuous-state equations with impulsive 'jumps' as in Eqs. (1)-(3), (7), (15) and (17)-(21)) and the difficulty of computing the inventory/backlog cost as in Eq. (26), and some probabilistic elements such as the quality and maintenance costs, as in Eqs. (27) and (29) respectively, the simulation-based optimization approaches are more suitable herein to find the optimal solution (Fu, 1994). Hence, using simulation as a powerful tool to imitate the dynamic and stochastic aspects of complex systems, it is possible to run instantly the continuous-time variables, to calculate accurately the expected levels of inventories and backlog, the amount of defectives produced, the number of lots rejected and the number of overhaul and CM interventions, and to accordingly compute the $ETC(.)$.

4. Resolution approach

4.1. Simulation-based optimization approach

Simulation-based optimization approaches consist of combining computer simulation with optimization techniques such as evolutionary algorithms, the Response Surface Methodology and stochastic approximation algorithms to heuristically solve problems which are analytically and numerically intractable (Gosavi, 2003; Tekin and Sabuncuoglu, 2004). Computer simulation has been successfully and widely applied to various real-world manufacturing problems in order to provide practical and implementable solutions (Jahangirian et *al.*, 2010). However, most of the existing simulation models in the literature are limited to the discrete-event simulation (Jahangirian et *al.*, 2010). In our study, we use a combined discrete-continuous simulation to model both discrete events and continuous variables and to therefore solve the optimization problem formulated above (Berthaut et *al.*, 2011; Assid et *al.*, 2015a). We suggest the following optimization approach (Figure 3):

a. *Mathematical model*: Formulate analytically the problem as shown in Section 3. This provides an accurate modeling of the system dynamic as a function of its state, and the formulation of the optimization problem.

b. *Simulation model*: Transform the mathematical model into a discrete-continuous simulation model according to the following logic: the continuous-time equations (i.e., Eqs. (1), (2), (4)-(7), (14), (20) and (23)) are modelled and calculated instantly with C++ subroutines, and the difference

equations which can be called also discrete-time equations (i.e., Eqs. (3), (8), (9), (15), (17)-(19), (21), (22), (24) and (25)) are transformed into discrete events using the SIMAN simulation language. Hence, for given values of the decision variables Q, S, c, n and r, the system performance and the costs incurred are obtained from simulation.

c. *Optimization*: Use an optimization algorithm to conduct experiments and to find the optimal values of the decision variables Q, S, c, n and r which minimize the *ETC* under the *AOQL* constraint (see Section 4.3).

(a) Mathematical model

(b) Combined discrete-continuous simulation model

Differential equations

Difference equations

Continuous simulation with C++

Discrete-event simulation with SIMAN

Experimental plan (values of Q, S, c, n, r)

(c) Optimization algorithm

No

Optimal solution found ?

Yes

Optimal solution (Q^*, S^*, c^*, n^*, r^*)
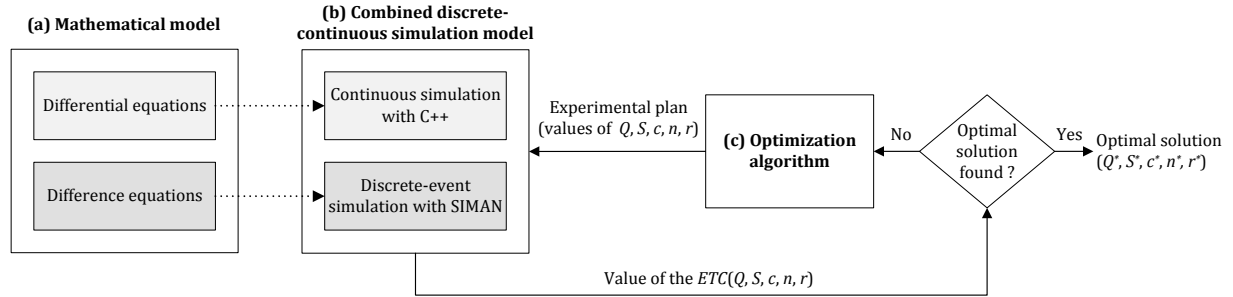
Value of the $ETC(Q, S, c, n, r)$

Figure 3. Simulation-based optimization procedure.

4.2. Simulation model

A combined discrete-continuous simulation model has been developed and executed through the *Arena Simulation* software. The discrete model imitates both the material flow and the logic of the integrated production, quality control and maintenance policy as described in Section 3. The differential equations, i.e. Eqs. (1), (2), (7), (14) and (20), are integrated continuously using the Runge–Kutta–Fehlberg method (Pegden *et al.*, 1995; Cheney & Kincaid, 2013), while the remaining continuous-time equations are calculated instantly using the C++ mathematical functions and operators. Both discrete and continuous parts of the simulation model work synchronously to calculate the variations in the real age $A(.)$, the virtual age $a(.)$, the number of defectives X_k in each kth lot produced and the inventories $q(.)$, $x_q(.)$ and $x_f(.)$. Accordingly, the improvement factor $\phi(.)$, the probability of failure $F(.)$ and the proportion of defectives $p(.)$ are instantly updated in the C++ subroutines using Eqs. (4), (5) and (6), respectively. The surplus $x_f^+(.)$ and the backlog $x_f^-(.)$ are also instantly derived from the instantaneous level of the final inventory $x_f(.)$. The duration of simulation runs, t_∞, is set in such a way to ensure that the steady-state is reached. At the end of each simulation run, the total inventory/backlog cost $G(t_\infty)$, the quality cost $Q(t_\infty)$ and the maintenance cost $M(t_\infty)$ are calculated respectively using Eqs. (26), (27) and (29). The stochastic durations of the CM and overhaul are randomly generated following predefined probability distributions.

To check the accuracy of the simulation model, we used a set of verification and validation techniques such as tracing the model's operation, testing for reasonableness, model structure and data, and using the animation and debug features of *Arena Simulation* (Pegden at al., 1995; Law, 2008). For example, Figure 4 shows that the integrated production, quality control and maintenance policy operates properly as intended: Figures 4.(a), 4.(b) and 4.(c) confirm that the setup and production decisions are adequately controlled with respect to the inventory position $x(.)$ and the production unit state $\alpha(.)$ as in Eqs. (22) and (24). Figures 4.(d), 4.(e) and 4.(f) depict, respectively, the impact of the production equipment usage on the deterioration of reliability, quality of lots produced and PM efficiency, as described in Eqs. (4)-(7). Those Figures show also the effects of the overhaul interventions on the

restoration of the production process to the initial conditions. Figure 4.(e) shows that the rate of defectives in lots produced, X_k/Q, $k=1,..,\infty$, increases as the production progresses and it can exceed the threshold r. However, according to the overhaul control policy (i.e., Eq. (25)), if it is found that the defective rate in a rejected lot exceeds the threshold r, then an overhaul is immediately undertaken as shown in Figure 4.(a). The quality deterioration implies an increase of the probability of rejection of lots produced, $1 - P_a^k$, $k=1,2,..,\infty$, as shown in Figure 4.(f). The impact of the 100% inspection of rejected lots on the inventory is clearly shown on the time lag between the inventory position $x(.)$ and the serviceable inventory level $x_f(.)$ as in Figure 4.(c).
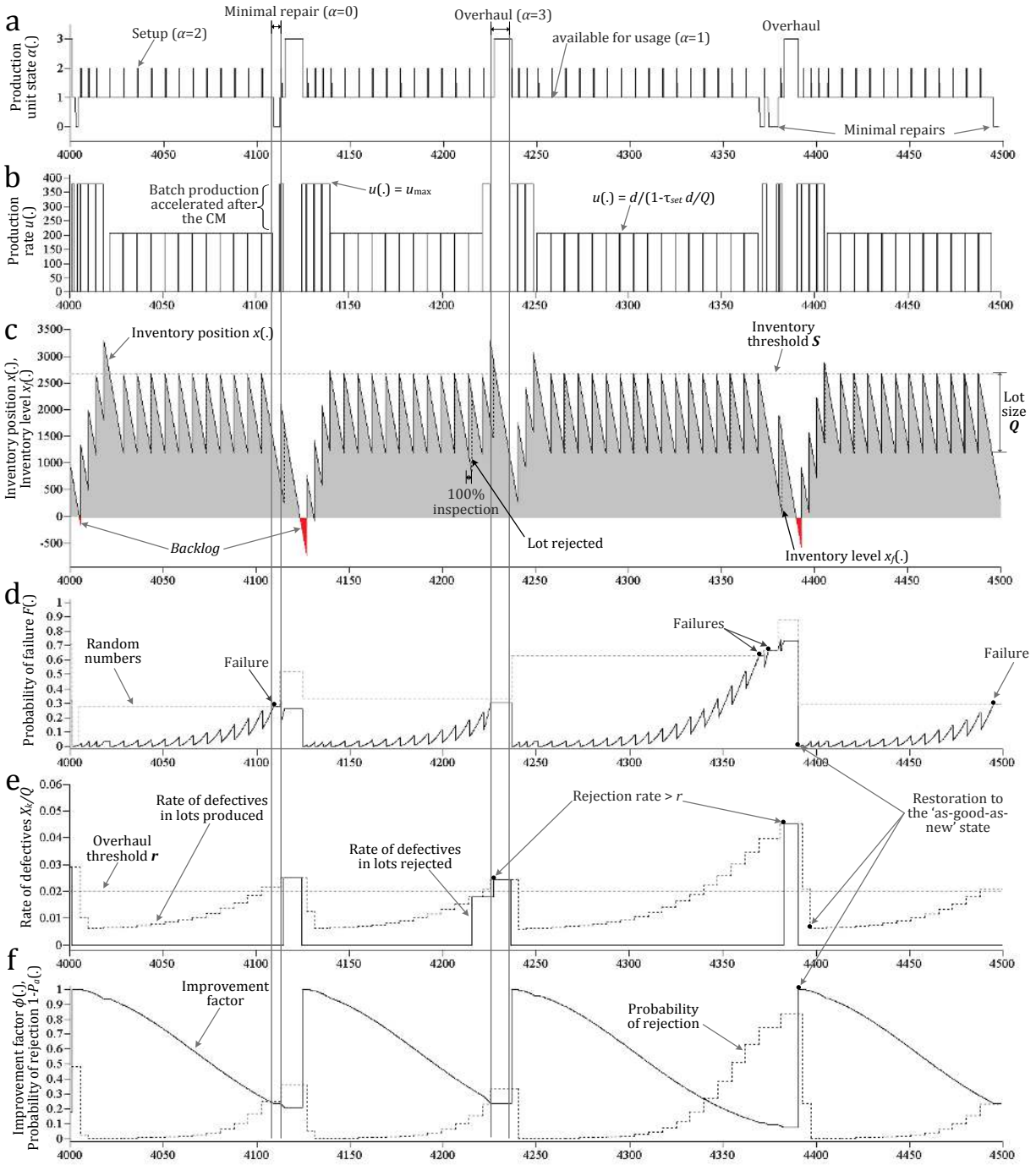
Figure 4. Evolution of production, inventory and operations performance during the simulation run.

4.3. Optimization algorithm

The optimization algorithm consists of a combination of an enumeration procedure with respect to the acceptance number c, a design of experiments (DOE), a statistical analysis and the Response Surface Methodology (RSM) to find a solution close to the global optimum. This algorithm can be summarized in the following steps:

Step 0. Set $c = 0$.

Step 1. For a fixed acceptance number c, determine $\psi_c(.)$ a quadratic approximation function of the expected total cost $ETC_c(Q, S, n, r)$ using a combination of DOE, regression analysis and RSM (Myers et al., 2009). $\psi_c(.)$ is called the response surface and should take the following equation:

$$\psi_c(Q, S, n, r) = \beta_0 + \beta_1 Q + \beta_2 S + \beta_3 n + \beta_4 r + \beta_{12} QS + \beta_{13} Qn + \beta_{14} Qr + \beta_{23} Sn + \beta_{24} Sr + \beta_{34} nr$$
$$+ \beta_{11} Q^2 + \beta_{22} S^2 + \beta_{33} n^2 + \beta_{44} r^2 + \varepsilon \tag{31}$$

where, β_0, β_i, β_{ii} and β_{ij}, $(i, j) \in \{1,2,3,4\}$, are unknown parameters to be estimated from the collected simulation data, and ε is a random error.

Step 2. Find Q_c^*, S_c^*, n_c^* and r_c^* the optimal solution of the following non-linear constrained problem

$$
\begin{cases}
\text{Minimize} & \psi_c(Q, S, n, r) \\[2mm]
\text{Subject to} & y(c)\left(\dfrac{1}{n} - \dfrac{1}{Q}\right) \leq AOQL_{\max}, \\[4mm]
& c < n < Q, \\[2mm]
& 0 < r < 1, \ S \geq 0
\end{cases}
$$

and calculate $\psi_c\left(Q_c^*, S_c^*, n_c^*, r_c^*\right)$. If $c = 0$, then set $c = 1$ and go to Step 1.

Step 3. If $\psi_c\left(Q_c^*, S_c^*, n_c^*, r_c^*\right) \leq \psi_{c-1}\left(Q_{c-1}^*, S_{c-1}^*, n_{c-1}^*, r_{c-1}^*\right)$, then set $c = c + 1$ and go to Step 1. Otherwise, find the optimal acceptance number c^* such that

$$\psi_{c-1}\left(Q_{c-1}^*, S_{c-1}^*, n_{c-1}^*, r_{c-1}^*\right) \geq \psi_c\left(Q_c^*, S_c^*, n_c^*, r_c^*\right) \leq \psi_{c+1}\left(Q_{c+1}^*, S_{c+1}^*, n_{c+1}^*, r_{c+1}^*\right).$$

Thus, the optimal values of the production lot size, the surplus inventory, the sample size and the overhaul threshold are respectively $Q_{c^*}^*$, $S_{c^*}^*$, $n_{c^*}^*$ and $r_{c^*}^*$.

The enumeration procedure is used since the acceptance number c is usually a very small discrete which cannot be approximated by a continuous variable. In step 1, we check the fitness of the second-order regression model $\psi_c(.)$ in the local region of the optimal solution using three ways as in Myers et al. (2009). First, the model's overall performance is evaluated. This is referred to as the coefficient of multiple determination R-squared and the adjusted R-squared which represent the proportion of total variation explained by the regression model. The values of these two coefficients should be close to 1. Second, a complete residual analysis should be done to check the normality assumption and the homogeneity of residuals. Third, once the optimization is performed, the optimal solution is cross-checked to ensure the validity. In step 2, the minimization problem can be solved using the non-linear constrained optimization techniques such as the penalty and barrier methods (Luenberger and Ye, 2008). It can be also solved using the MS-Excel Solver. In step 3, for practical implementation, the optimal values $Q_{c^*}^*$, $S_{c^*}^*$ and $n_{c^*}^*$ should be rounded to the nearest integers.

5. Experimentation and analysis of the results

5.1. Numerical example

A hypothetical example of the proposed model is provided for illustration. Let us consider the following parameters in the appropriate units : u_{max}=380, d=200, C_h=0.1, C_b=5, C_{set}=2500, C_{ovr}=30000, C_{cm}=7500, C_{ins}=2.5, C_{rej}=20, C_{def}=35, τ_{ins}=5×10^{-4}, τ_{set}=0.15, τ_{cm}~Log-Normal(3,1), τ_{ovr}~Gamma(1,9), λ_{set}= 5×10^{-10}, γ_{set}=2.55 , λ_r=8×10^{-10}, γ_r=2.4, λ_q=4×10^{-6}, γ_q=1.4, p_0=0.3%, η= 0.075 and $AOQL_{max}$= 2.0%.

For each fixed acceptance number c, simulation runs are conducted according to a four factors Box-Behnken experimental plan (27 runs) for each combination of factors Q, S, n and r. This type of design is suitable because of its rotatable feature and its efficiency in terms of number of required runs (Montgomery, 2008b). To adequately select the levels of the experimental design plan factors, we repeat the DOE, simulation and RSM, narrowing the domain of (Q, S, n, r) until it is centered about the optimum design point. Through this sequential procedure, the admissible experimentation region is fully explored, and therefore the obtained solution will be a global optimum. In order to ensure that the steady-state is reached, the duration of each simulation run t_∞ is set to 500,000 units of time (it takes on average 2.5 seconds on a computer with a 2.80 GHz CPU).

Table 1 presents the results obtained from the step-by-step application of the resolution approach procedure to the present numerical example. The adjusted R-squared for all acceptance number is always greater than 97%. This states that about 97% of the observed variability in the $ETC_c(.)$ is explained by the second-order models $\psi_c(.)$. It should be mentioned here that the ANOVA of fitting models for all acceptance number showed that the linear and quadratic effects of the factors Q, S, n and r and their interactions are significant for the response variable at a 5% level of significance. For example, Table 2 shows the ANOVA of standardized effects for the Box-Behnken design when the acceptance number c is equal to 4.

Table 1. Optimum solutions with respect to the acceptance number c.

c	R^2-adj	n_c^*	Q_c^*	S_c^*	r_c^*	$\psi_c^*(.)$	$y(c)$	$AOQL$	$AOQ(\infty)$	$P_a(\infty)$	$\bar{f}(\infty)$	$\bar{m}(\infty)$	$Av(\infty)$
0	0.9701	57	1324	2933	2.990%	1 890.6 \$	0.3679	0.62%	0.24%	0.375	0.0359	0.0217	0.807
1	0.9710	84	1341	2632	2.838%	1 839.2 \$	0.84	0.94%	0.44%	0.645	0.0330	0.0213	0.814
2	0.9740	129	1411	2549	2.526%	1 791.4 \$	1.371	0.97%	0.50%	0.712	0.0251	0.0229	0.818
3	0.9713	150	1474	2636	2.199%	1 723.2 \$	1.942	1.16%	0.60%	0.758	0.0312	0.0224	0.811
4	**0.9740**	**157**	**1484**	**2631**	**1.996%**	**1 681.5 \$**	**2.544**	**1.45%**	**0.74%**	**0.829**	**0.0333**	**0.0227**	**0.808**
5	0.9749	168	1451	2762	1.864%	1 696.1 \$	3.168	1.67%	0.89%	0.842	0.0388	0.0219	0.797

For each combination of c, Q_c^*, S_c^*, n_c^* and r_c^* in Table 1, we used the simulation to calculate some performance measures such as the long-term proportion of acceptance of lots produced denoted by $P_a(\infty)$, the long-term average outgoing quality denoted by $AOQ(\infty)$, the long-term frequency of overhauls denoted by $\bar{m}(\infty)$, the long-term frequency of CM denoted by $\bar{f}(\infty)$ and the long-term system availability denoted by $Av(\infty)$, using the formula in Appendix A. Thus, we see that the variation of the acceptance number c, which systematically affects the severity of the sampling plan (i.e., measured by $P_a(\infty)$), has very significant impacts on the remaining decision variables (i.e., optimal values of Q, S, n and r), on the optimal expected cost $\psi_c^*(.)$, on the frequency of failures $\bar{f}(\infty)$ and obviously on the outgoing quality $AOQ(\infty)$. This emphasizes the relevance of the acceptance number optimization in an integrated production, quality control and maintenance context. The optimal

acceptance number c^* is 4 because it corresponds to the minimum expected total cost which is $1681.5. Using the statistical software STATISTICA the related second order cost function $\psi_4(.)$ is given by:

$$\psi_4\left(Q,S,n,r\right) = 4908.37 - 277.62\times10^{-3}\,Q - 784.92\times10^{-3}\,S - 24.55\ n - 6.27\times10^{-3}\,r - 252.35\times10^{-6}Q\,S$$
$$- 6.59\times10^{-3}\,Q\,n + 3.99\times10^{-6}\,Q\,r + 3.92\times10^{-3}\,S\,n - 1.14\times10^{-6}\,S\,r - 8.83\times10^{-6}\,n\,r \qquad (32)$$
$$+ 638.67\times10^{-6}Q^2 + 107.76\times10^{-6}\,S^2 + 77.11\times10^{-3}\,n^2 + 11.9\times10^{-8}\,r^2$$

Table 2. The ANOVA table for the total expected cost (c = 4).

Factor	SS	d.f.	MS	F-Ratio	P-value	Significant
Q (Linear + quadratic)	1094586	2	547292.9	12273.34	0.000000	Yes
S (Linear + quadratic)	1033646	2	516822.9	11590.04	0.000000	Yes
n (Linear + quadratic)	84681	2	42340.4	949.51	0.000000	Yes
r (Linear + quadratic)	31655	2	15827.4	354.94	0.000000	Yes
Q . S	297388	1	297387.6	6669.08	0.000000	Yes
Q . n	105329	1	105328.5	2362.05	0.000000	Yes
Q . r	379	1	379.3	8.51	0.014027	Yes
S . n	80598	1	80597.9	1807.45	0.000000	Yes
S . r	14366	1	14366.1	322.17	0.000000	Yes
n . r	1187	1	1186.9	26.62	0.000314	Yes
Error	73869	21	7337.8			
Total SS	2835513	35			R^2-Adjusted=0.97395	

Figure 5 presents the projection of the cost response surface $\psi_4(.)$ on two-dimensional spaces. The region with gray-shaded contours in the (n, Q) two-dimensional space represents the set of the infeasible solutions where the *AOQL* constraint is not satisfied. The minimum expected total cost $1681.5 is located at $Q^* = 1484$, $S^* = 2631$, $n^* = 157$ and $r^* = 1.996\%$. These values represent the best approximation of the optimal solution of the integrated (Q, S, c, n, r) policy which should be applied to jointly control the setup operations, the production rate, the outgoing quality and the overhaul interventions. From 20 replications of the simulation, we validated the solution by verifying that the estimated optimal cost $\psi_4^*(.) = \$1681.5$ is within the 95% Confidence Interval [$1675.7, $1682.4].

Also, from Table 1, we observe a high correlation (about -95%) between the optimal threshold r_c^* and the severity of the corresponding sampling plan measured by $P_a(\infty)$: the optimal overhaul threshold r_c^* decreases as the optimal sampling plan (c, n_c^*) becomes more and more reduced (i.e., $P_a(\infty)$ increases), and vice-versa. In fact, because a reduced inspection narrows increasingly the visibility on the process condition and implies an increasing outgoing quality, r_c^* decreases gradually in order to maintain the frequency of the overhauls at an optimal level (about an average frequency of 0.0221). Such an observation illustrates how the overhaul control policy and sampling plan technique interact together to monitor the process deterioration (as discussed in Section 3.4).
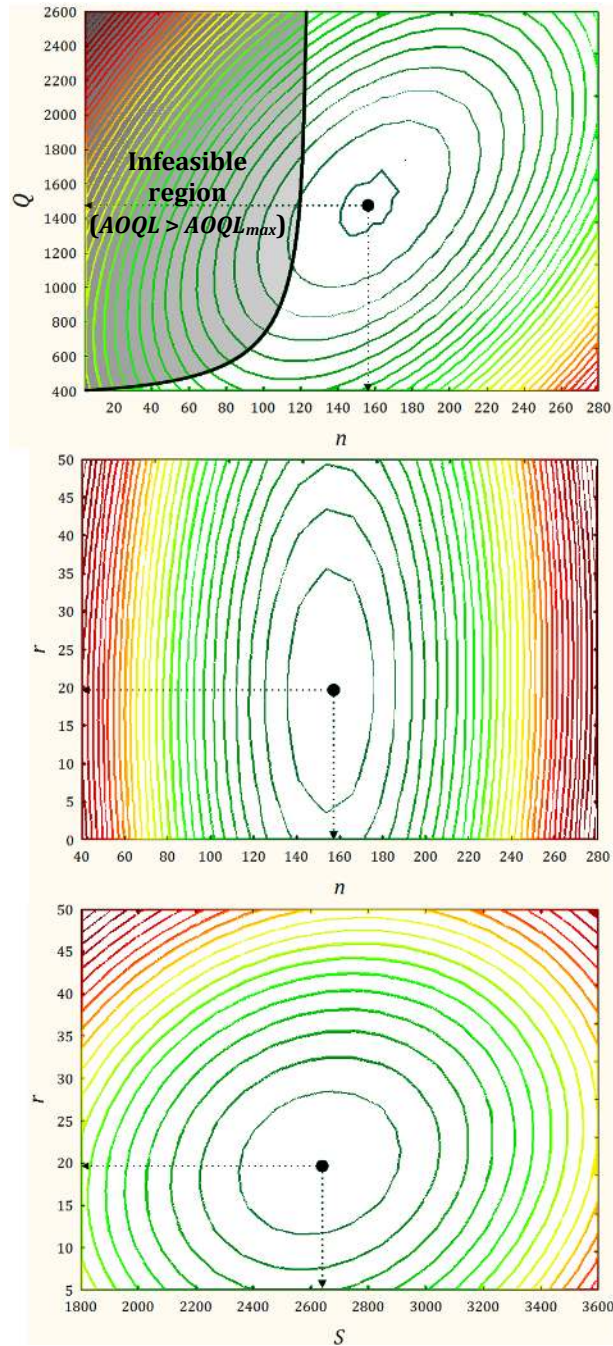
Figure 5. Contour plots of the estimated expected total cost $\psi_4(.)$.

5.2. Sensitivity analysis

Another set of experiments has been conducted to measure and analyse the sensitivity of the proposed integrated policy with respect to ranges of system parameters. The purpose of this analysis is to validate the simulation-based resolution approach and to study the reaction of model outputs and the variation of the optimal solution in response to changes of model parameters (inputs). Table 3 presents twenty four configurations of system parameters derived from the basic case by varying

significantly their values above and below one at a time. The results obtained make sense as expected and can be explained as follows:

• *Variation of the holding cost*: When the holding cost C_h increases (case 1), the optimal hedging threshold S^* decreases in order to reduce the global inventory cost. The optimal lot size Q^* decreases in order to reduce the WIP inventory in production and quality control centers. Due to the fact that the 100% inspection delay decreases proportionally to the decrease of the lot size Q^*, the optimal sampling plan becomes tighter (as n^* increases and $P_a(\infty)$ decreases). However, a tightened plan involves an increasing cost of rejected items. For that reason, the optimal threshold r^* decreases in order to perform more frequently the overhaul interventions (i.e., $\bar{m}(\infty)$ increases). Note that the decrease in the inventory cost produces the opposite effects on the control variables (case 2).

• *Variation of the backlog cost*: When the backlog cost C_b increases (case 3), the optimal inventory surplus S^* increases in order to enhance protection to the serviceable stock against shortages. Moreover, the optimal lot size Q^* decreases in order to reduce the production delay and therefore ensure a better supply to the final stock. In addition, the optimal sampling plan becomes slightly reduced (as n^* decreases) in order to reduce the quality control delay. The optimal threshold r^* decreases to maintain the outgoing quality at an acceptable level. Note that the decrease in the backlog cost has the opposite effects (case 4).

• *Variation of the setup cost*: When the setup cost C_{set} increases (case 5), the optimal lot size Q^* increases in order to reduce the total number of setup operations. As the production cycle and the 100% inspection become both longer proportionally to the increase in the lot size Q^*, the optimal inventory surplus S^* increases in order to provide better protection to the serviceable stock. Note that a bigger lot size reduces the effects of PM during setups, increases the quality deterioration rate and therefore contains more defective items. Thus, the optimal sampling plan becomes tighter (i.e., $P_a(\infty)$ decreases). In addition, the optimal threshold r^* decreases (so $\bar{m}(\infty)$ increases) to improve both the reliability and quality of the production process. Note that the decrease of the setup cost has the opposite effects (case 6).

• *Variation of the CM cost*: When the corrective maintenance cost C_{cm} increases (case 7), the optimal lot size Q^* and the optimal inventory threshold S^* decrease both in order to reduce the reliability deterioration rate. In fact, setting the inventory surplus S^* at a lower level restrains the usage of the production unit during the period of the surplus build up and therefore slows down the reliability detrioration. In addition, the optimal overhaul threshold r^* decreases in order to restore more frequently the production unit to the 'as-good-as-new' state and accordingly reduces the probability of failure. Note that a decrease in the CM cost has the opposite effects (case 8).

• *Variation of the overhaul cost*: When the overhaul cost C_{ovr} increases (case 9), the optimal threshold r^* increases in order to reduce the number and therefore the total cost of the overhauls. The optimal sampling plan becomes reduced (as $P_a(\infty)$ increases) in order to lower the number of overhaul interventions based on the quality of rejected lots. As the reduction of the overhaul maintenance makes system breakdowns more frequently, the optimal inventory surplus S^* increases in order to improve the protection of the serviceable stock against shortages during the CM operations. Note that the decrease of the overhaul cost produces the opposite effects (case 10).

• *Variation of the inspection cost*: When the inspection cost C_{insp} increases (case 11), the optimal sampling becomes reduced (as $P_a(\infty)$ increases) in order to minimize the cost of 100% inspection of lots rejected. In order to mitigate with the risk of accepting a higher number of poor quality lots, more

perfect maintenance operations (overhaul) are required to improve the process quality, which explains the decrease of the optimal threshold r^*. The optimal inventory surplus S^* increases in order to meet the increasing shortage risk as the overhaul maintenance becomes more frequent. Note that a lower inspection cost produces the opposite effects (case 12).

• *Variation of the rejection cost*: When the rejection cost C_{rej} increases (case 13), the optimal overhaul threshold r^* decreases in order to improve the process quality and to accordingly reduce the number and the cost of rejected items. As the total rejection cost represents only a small portion of the total operating cost, the optimal lot size Q^* and the optimal inventory surplus S^* decreases slightly in order to reduce the quality deterioration rate. Also, the optimal sampling plan becomes more reduced (i.e., $P_a(\infty)$ increases) in order to diminish the number of rejected lots. Because a reduced inspection narrows the visibility on the process quality, the optimal threshold r^* decreases to maintain the frequency of overhaul maintenances at the same level. Note that the decrease of the rejection cost produces the opposite effects (case 14).

• *Variation of the cost of selling a defective item*: When the cost of selling a defective item C_{def} increases (case 15), the severity of the optimal sampling plan increases slightly (as $P_a(\infty)$ decreases) in order to improve the quality of lots produced. Accordingly, the optimal overhaul threshold r^* increases as the visibility on the process condition increases. The optimal inventory surplus S^* decreases slightly in order to reduce the quality deterioration rate. Note that a lower cost of a defective item sold has the opposite effects (case 16).

• *Variation of the inspection delay*: When the unit inspection delay τ_{insp} increases (case 17), the optimal inventory surplus S^* increases in order to provide additional protection to the serviceable stock against the increasing quality control delay. The optimal lot size Q^* decreases in order to reduce the 100% inspection delay. Although the optimal sampling plan becomes tightened, the long-term proportion of acceptance $P_a(\infty)$ increases due to the important decrease of the optimal overhaul threshold r^*. In fact, the mechanism (c^*, n^*, r^*) reacts by improving the process quality and reduces the full inspection operations of rejected lots accordingly. Note that a lower unit inspection delay leads to the opposite effects (case 18).

• *Variation of the quality deterioration rate*: When the quality deterioration rate increases (case 19), the optimal lot size Q^* decreases in order to perform more frequently the setup operations (i.e., imperfect PM). Moreover, the optimal sampling plan (c^*, n^*) and the optimal overhaul threshold r^* vary in such a way that more lots produced are effectively rejected (i.e., $P_a(\infty)$ decreases significantly) in order to increase the full inspection activities and to improve the outgoing quality. The optimal inventory surplus S^* decreases in order reduce the usage-deterioration of the production unit during the periods of buffer stock build up. Note that the decrease of the quality deterioration rate produces the opposite effects (case 20).

• *Variation of the reliability deterioration rate*: When the reliability deterioration increases (case 21), the optimal threshold r^* decreases in order to more frequently carry out the overhauls. In addition, the optimal lot size Q^* decreases in order to increase the setup activities and to reduce the degradation rate between setups. The increase of the maintenance activities improves the process quality which explains the increase of the long-term proportion of acceptance $P_a(\infty)$, whereas, these activities reduce the production unit availability $Av(\infty)$. As a result, the optimal inventory surplus S^* increases in order to mitigate the high risk of shortage during periods of system unavailability. Note that a lower reliability deterioration rate has the opposite effects (case 22).

• *Variation of the PM efficiency*: The decrease of the improvement factor $\phi(.)$ intensifies both the process quality and reliability deteriorations (case 23). The optimal lot size Q^* decreases in order to reduce the production unit aging between setups which also implies an increase in the number of the setups to therefore slow down deterioration of the efficiency of the PM. The optimal settings of the sampling plan and the overhaul threshold lead to increasing the 100% inspection operations (i.e., long-term proportion of acceptance $P_a(\infty)$ decreases), in order to cope with the increasing number of defectives produced. As the increasing number of failures reduces the production unit availability $Av(\infty)$, the optimal inventory surplus increases in order to ensure better protection to the serviceable stock against shortages. Note that the increase of the PM efficiency produces the opposite effects (case 24).

[Insert Table 3]

5.3. Influence of the *AOQL* constraint

Table 4 presents the optimal solutions of the integrated policy for different levels of the $AOQL_{max}$. For values of $AOQL_{max}$ from 0.1% to 1.45%, the *AOQL* constraint is active (i.e., $AOQL = AOQL_{max}$). However, for all values of $AOQL_{max} > 1.45\%$, the *AOQL* constraint is inactive as the optimal solution obtained at $AOQL_{max} = 1.45\%$ realizes the minimum possible cost (i.e., 1681.5 $) among all solutions obtained for all given acceptance numbers and $AOQL_{max}$ values. Thus, we see that the total expected cost increases as the $AOQL_{max}$ decreases (with $AOQL_{max} < 1.45\%$), while, it remains the same for $AOQL_{max} > 1.45\%$. Moreover, faced with a decrease of the $AOQL_{max}$ level, the optimal solutions (Q^*,S^*,c^*,n^*,r^*) lead to increase the severity of the optimal sampling plan (i.e., as shown by the decreasing values of $P_a(\infty)$) and to increase the frequency of the overhauls (i.e., as shown by the increasing values of $\bar{m}(\infty)$) in order to improve the quality of lots produced (i.e., as shown also by the decreasing values of $AOQ(\infty)$) and to accordingly satisfy the *AOQL* constraint. When the optimal acceptance number c^* remains unchanged (e.g. cases when $AOQL_{max}$ takes values from 0.1% to 0.5%, and values greater than 1.25%), we remark that, as the $AOQL_{max}$ decreases, the optimal sample size n^* increases to tighten the quality control, the optimal threshold r^* decreases in order to more frequently perform the overhauls, the optimal inventory threshold S^* decreases in order to reduce the usage of the production unit during the periods of the inventory surplus build up and to consequently slow down the process quality deterioration, and finally the optimal lot size Q^* increases in order to reduce the setup activities as the overhauls, which are more efficient than the setups, become more frequent.

Table 4. Sensitivity analysis for the *AOQL* constraint.

$AOQL_{max}$	c^*	n^*	Q^*	S^*	r^*	$\psi_{c^*}^*(.)$	$AOQL$	$AOQ(\infty)$	$P_a(\infty)$	$\bar{f}(\infty)$	$\bar{m}(\infty)$	$Av(\infty)$	Δ-Cost
0.10%	0	294	1485	2024	1.945%	2 103.8 $	0.10%	0.01%	0.089	0.0108	0.0273	0.826	-0.6%
0.25%	0	134	1458	2481	2.303%	1 946.5 $	0.25%	0.06%	0.246	0.0212	0.0243	0.821	-8.0%
0.50%	0	70	1394	2841	2.610%	1 898.9 $	0.50%	0.19%	0.450	0.0215	0.0238	0.820	-10.3%
0.75%	2	162	1431	2117	2.425%	1 801.4 $	0.75%	0.38%	0.635	0.0227	0.0232	0.819	-14.9%
1.00%	3	172	1485	2246	2.182%	1 734.4 $	1.00%	0.52%	0.719	0.0276	0.0230	0.813	-18.1%
1.25%	4	179	1508	2262	1.852%	1 701.3 $	1.25%	0.66%	0.806	0.0280	0.0228	0.809	-19.6%
≥1.45%	4	157	1484	2631	1.996%	1 681.5 $	1.45%	0.74%	0.829	0.0333	0.0226	0.808	-20.6%

In addition, Table 4 shows that, for a highly restricted *AOQL* constraint (i.e., $AOQL_{max} \leq 0.5\%$), the zero-acceptance number sampling plans are more economical than the non-zero acceptance sampling plans, as they provide a higher discriminatory power (Schilling and Neubauer, 2009). However, for a higher acceptable level of *AOQL* (e.g. $AOQL_{max} \geq 0.75\%$), non-zero acceptance number plans are more economical as they reduce the extra 100% inspection cost. For example, for an $AOQL_{max} = 2.0\%$ (basic case), we observe in Table 1 that the optimal zero-acceptance number sampling plan ($c=0$, $n_c^*=57$) provides the best quality protection for the customer (i.e., with the lowest values of *AOQL* and $AOQ(\infty)$) but it is too far to be economically optimal for the manufacturer (i.e., 11.38% more costly than the optimal sampling plan ($c^*=4$, $n_c^*=157$)). These results are in line with previous findings in the literature, showing that the total inspection cost of zero acceptance sampling plans are generally significantly higher than non-zero acceptance sampling plans (Baker, 1988).

5.4. Comparative study

In this section, we compare the performance of the integrated (Q, S, c, n, r) policy, called Policy-I, with a similar integrated model proposed in the literature by Radhoui et *al*. (2009, 2010) and others where the quality control consists of a 100% inspection of all lots produced instead of using the acceptance sampling plan techniques. In reality, the 100% inspection policy can be viewed as a special case of sampling inspection where each lot produced is fully inspected during the sampling step (i.e., $n=Q$) and the acceptance number is set to zero. Thus, a simplified version of the optimization approach described in section 4 can be used to find the optimal values of the lot size Q, the inventory surplus S and the overhaul threshold r for fixed parameters $c=0$ and $n=Q$ (called Policy-II). Therefore, we used a combination of a three factors Box-Behnken experimental plan for each combination of Q, S and r, regression analysis and RSM to fit the total expected cost by a quadratic model denoted by $\psi_{100\%}(.)$.

Using the same basic case data, the ANOVA of this model leads to an R-squared adjusted equal to 0.9766. In addition, the three design factors (including the linear and the quadratic effects) and their interactions are significant (P-value < 0.05). Thus, we get the following quadratic function:

$$\psi_{100\%}(Q,S,r) = 3346.87 - 109.57 \times 10^{-2} Q - 16.25 \times 10^{-2} S - 93.06 \times 10^{-4} r - 11.45 \times 10^{-5} QS$$
$$+ 589.65 \times 10^{-8} Q r - 452.01 \times 10^{-8} S r + 433.23 \times 10^{-6} Q^2 + 64.32 \times 10^{-6} S^2 + 37.795 \times 10^{-8} r^2 \quad (33)$$

The minimization of the $\psi_{100\%}(.)$ function provides the following optimal solution: $Q^* = 1573$, $S^* = 3374$ and $r^* = 2.022\%$ and the corresponding expected total cost is \$2116.7. Herein, Policy-I (with acceptance sampling plan) is 20.6% more economical than Policy-II (with 100% inspection). Table 5 presents complementary performance measures obtained by simulation with the optimal solutions of both policies I and II for the basic case. According to Policy-I, only 17.1% (i.e., 1-$P_a(\infty)$) of lots produced are fully inspected to economically meet the output quality requirement and to provide sufficient visibility on the process quality for the overhaul control system. Under this policy, the long-term average outgoing quality of all lots produced is 0.74%, while, one lot could contain, on average, a maximum outgoing quality of 1.45%. In the other side, Policy-II ensures always the delivery of defect-free products to customers through the 100% inspection of all lots produced (i.e., $AOQ(\infty)=0$). The significant gap between the costs of both policies is due to two factors. First, Policy-II generates extra quality control costs especially during periods when the process quality is perfect. Second, the 100% inspection operations (Policy-II) lead to incur extra holding costs due to the increase of the WIP in the Quality Control center, and the increase of the optimal inventory surplus S^* in the serviceable stock as the total quality control delay increases. From Table 5, the WIP inventory $E[x_q]$ under Policy-II is 4

times greater than the $E[x_q]$ under Policy-I. Similarly, the average positive inventory $E[x_f^+]$ in the serviceable stock under Policy-II is 1.5 greater than the $E[x_f^+]$ under Policy-I. However, the average negative inventory $E[x_f^-]$ under Policy-I is much higher than the $E[x_f^-]$ under Policy-II. This is due to the increasing degree of uncertainty in Policy-I caused by the effects of the probabilistic acceptance/rejection decision of lots produced. Also from Table 5, we remark that the optimal lot size Q^* under Policy-I is smaller than that under Policy-II. In fact, smaller lots reduce the number of defectives transmitted to the serviceable stock and improve the visibility on the process quality as discussed in Section 5.2.

Table 5. Comparison of the optimal control policies.

Policy	Optimal solution						Quality control			Inventory		
	c^*	n^*	Q^*	S^*	r^*	ETC^*	$AOQL$	$AOQ(\infty)$	$P_a(\infty)$	$E[x_q]$	$E[x_f^+]$	$E[x_f^-]$
Policy-I	4	157	1485	2631	1.996%	1 681.5 \$	1.45%	0.74%	82.9%	39.0	1642.9	27.5
Policy-II	-	-	1573	3374	2.022%	2 116.7 \$	0.0%	0.0%	0.0%	159.3	2482.9	0.6

Additional comparisons are also provided in Tables 3 and 4. For each configuration of the system parameters in those tables, we calculated the cost savings resulting from using the acceptance sampling plan rather than 100% inspection as follows:

$$\Delta\text{-}Cost\left(\%\right)=\frac{\psi_{c^*}^{*}(.)-\psi_{100\%}^{*}(.)}{\psi_{100\%}^{*}(.)}\times100 \tag{34}$$

For all configurations of model parameters presented in both Tables 3 and 4, Policy-I is always more economical than Policy-II. However, the level of savings *Δ-Cost* achieved under Policy-I depends on the settings of the system parameters. For example, the savings are much more important for high inspection cost, as well as the process quality is improved by increasing the efficiency and frequency of maintenance activities (setups and overhauls) or by decelerating the quality deterioration. The savings are negligible in situations where the maximum allowed *AOQL* is very low (e.g. $AOQL_{max} \leq$ 0.1%). In addition, all the results obtained in both Tables 3 and 4 confirm the observations mentioned above: the optimal inventory threshold S^* and the optimal lot size Q^* under Policy-I are always smaller than those under Policy-II.

6. Managerial implications

The integrated production, quality control and maintenance policy proposed in this paper can be implemented in batch processing manufacturing systems where both acceptance sampling techniques and dynamic production-inventory control are effective. For example, in the pharmaceutical industry, single acceptance sampling plan is used to reduce the delivery lead time of drugs which are safe and unique in the market and can improve the survival of some vulnerable patients. The cost of 100% inspection of a lot of such a drug is far greater than accepting a bad lot. This is because delayed treatment could result in a patient's death while, in accepting the bad lot, the drug could still have some beneficial placebo effect on the patient (Yang and Carlin, 2001). Likewise, dynamic production-inventory control based on the concept of the Hedging Point Policy has various applications in the pharmaceutical industry are reported in Assid et *al.* (2015). Other potential applications of the

integrated control policy include also military, electronic, semiconductor, paper pulp and food industries (see MIL-STD-105 series; Pearn and Wu, 2007; Anthony, 2004; GonzÁLez and Palomo, 2003; Gershwin, 2000).

In order to facilitate the implementation of the proposed integrated control policy in practice, an implementation logic chart is presented in Figure 6 for the basic case (Section 5.1), which shows how the integrated decisions on setup, production rate setting, lots acceptance/rejection and overhaul should be made. Managerial implications into business practice for the integrated policy require full information about the state of the production unit and the inventory position. The manager should monitor the production quality by observing the defective rate of batches rejected in sampling inspection and predictably make the decision on the overhaul intervention based on the quality of the latest batch rejected. In addition, the maintenance team should utilize the setup times as opportunities to perform PM actions. This should help industrials to eliminate production stoppages for PM and overcome complexity of system scheduling (Xia et al., 2015).
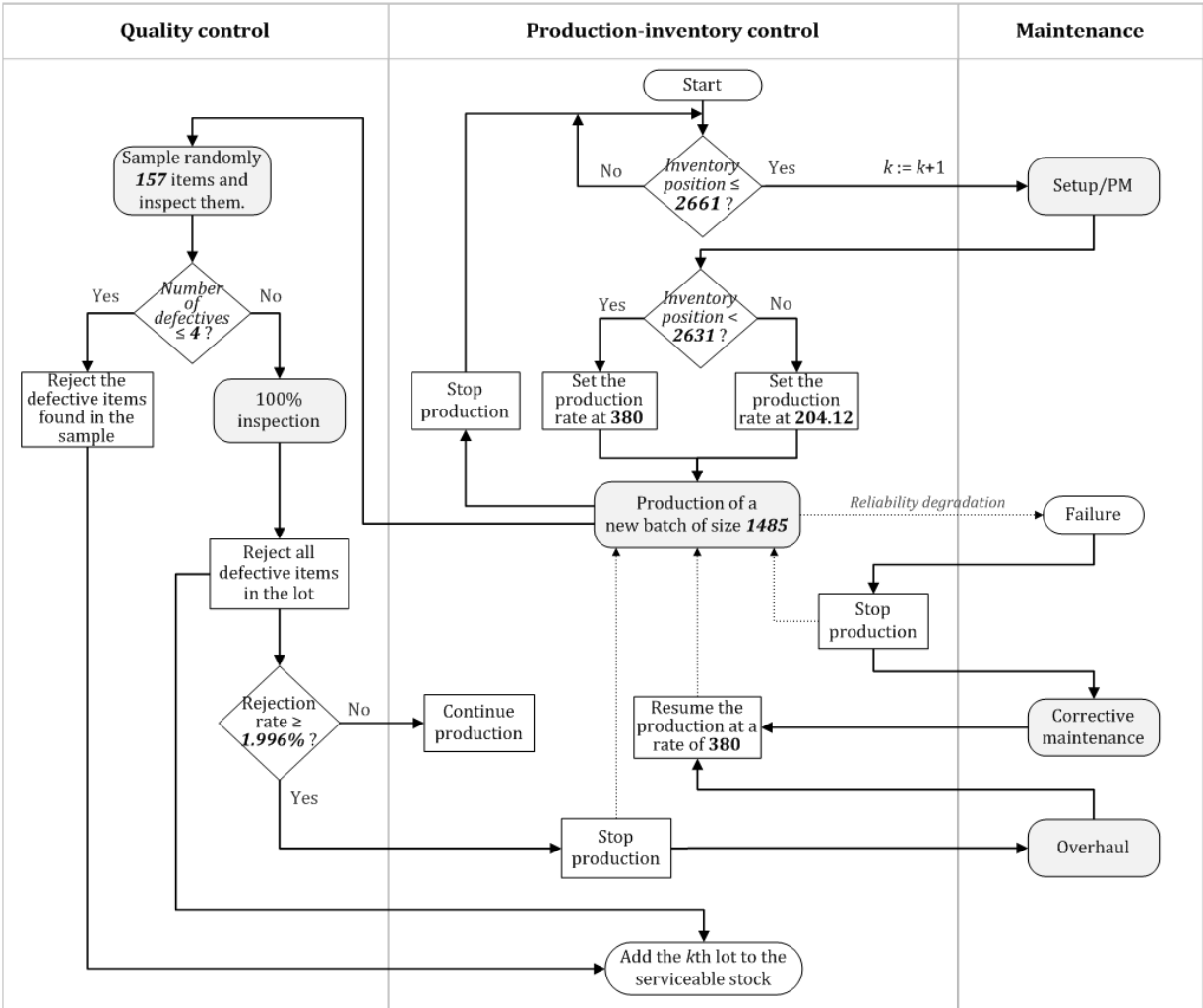

Figure 6. Implementation logic chart of the integrated control policy of production, quality and maintenance.

7. Conclusion

The joint design and optimization of production, PM and quality control using the acceptance sampling plans have never been studied before in the literature. In this paper, we proposed a new holistic approach to the joint optimization of the production lot size, the safety stock, the acceptance sampling plan and the overhaul scheduling considering an outgoing quality constraint for degrading production systems. The suggested approach contributes to research on integrated production, maintenance and quality control in three ways. First, in the context of correlated quality and reliability deteriorations, we investigated the intrinsic statistical characteristics of the single acceptance sampling plan in order to show the relevance of quality information resulting from such a quality control technique to support maintenance decision-making. Second, we provided a new modeling framework combining stochastic mathematical formulation and discrete-continuous simulation in order to model complex interactions between degradation phenomena, operations planning and settings, product quality and process reliability. Thus, this modelling framework can be employed to relax many unrealistic assumptions used in the literature, to overcome the limitations of classical resolution approaches and to solve such hard optimization problems in manufacturing systems. Third, due to the fact that acceptance sampling plans adapt systemically the level of quality inspection with the degree of process deterioration, we showed experimentally that important cost savings could be realized by using these plans rather than 100% inspection. At the practical level, operations managers should figure out from this study the strong and deep links between production, maintenance and quality control (acceptance sampling plans). Practitioners should recognize that the capacity of satisfying the demand without back orders and the level of quality perceived by the final customers, which impact both the marketability of the final products, are the results of the whole operations settings and not only the results of one separated function among production, PM or quality control.

One limitation of our model is that we assume a single quality attribute deteriorating with age. Modern products are complex with numerous quality attributes that can deteriorate with different rates. In such a situation, various quality-control tests could be required. Moreover, separate AOQL could be defined for each test. Future research can be conducted to investigate the optimal sampling inspection and preventive maintenance settings for multi-attribute products. In addition, our study should stimulate further research on the interpretation and usefulness of quality information feedback from acceptance sampling plans in integrated operations management and control. In fact, other sampling techniques such as multiple sampling plans and sampling plans by variables have specific inspection procedures and more particular statistical characteristics that should be extensively explored in order to integrate additional quality measures in managerial decision-making. Further research could be carried out to integrate the production, inventory and process reliability aspects in the design of acceptance sampling schemes. Sampling schemes are widely used in industry to adapt the inspection severity to the variation of quality of lots produced. The switching rules procedures as described in MIL-STD-105E and ISO 2859-1 can be improved by including quality history, process reliability and inventory state in order to enhance responsiveness and adaptive control face to production process deterioration.

Appendix A – Formula used to calculate the quality and availability performances

The long-term proportion of acceptance of lots produced, $P_a(\infty)$, is given by

$$P_a(\infty) = \frac{\sum_{k=1}^{N(t_\infty)} Ind\{Y_k \leq c\}}{N(t_\infty)} \tag{A.1}$$

The calculation of the long-term average outgoing quality, $AOQ(\infty)$, is derived from Eq. (10) as follows

$$AOQ(\infty) = \frac{\sum_{k=1}^{N(t_\infty)} Ind\{Y_k \leq c\} \cdot (X_k - Y_k)}{\sum_{k=1}^{N(t_\infty)} Ind\{Y_k \leq c\} \cdot (Q - Y_k) + \sum_{k=1}^{N(t_\infty)} Ind\{Y_k > c\} \cdot (Q - X_k)} \tag{A.2}$$

The long-term frequencies of overhauls and CM, denoted respectively by $\bar{m}(\infty)$ and $\bar{f}(\infty)$, are determined as follows

$$\bar{m}(\infty) = \frac{1}{t_\infty} \cdot m(t_\infty) = \frac{1}{t_\infty} \cdot \sum_{k=1}^{N(t)} Ind\{Y_k > c\} \cdot Ind\left\{\frac{X_k}{Q} \geq r\right\} \tag{A.3}$$

$$\bar{f}(\infty) = \frac{1}{t_\infty} \cdot f(t_\infty) = \frac{1}{t_\infty} \cdot \sum Ind\{\alpha(t + \delta t) = 0 | \alpha(t) = 1\} \tag{A.4}$$

The long-term system availability $Av(\infty)$ is given by

$$Av(\infty) = 1 - \frac{N(t_\infty)\tau_{set} + \sum_{i=1}^{f(t_\infty)} \tilde{\tau}_{cm}^i + \sum_{j=1}^{m(t_\infty)} \tilde{\tau}_{ovr}^j}{t_\infty} \tag{A.5}$$

References

Akella, R., & Kumar, P. (1986). Optimal control of production rate in a failure prone manufacturing system. *Automatic Control, IEEE Transactions on, 31*(2), 116-126.

Anthony, R. M. (2004). *Analyzing sampling methodologies in semiconductor manufacturing.* M.Sc. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Assid, M., Gharbi, A., & Dhouib, K. (2015a). Joint production and subcontracting planning of unreliable multi-facility multi-product production systems. *Omega, 50*, 54-69.

Assid, M., Gharbi, A., & Hajji, A. (2015b). Joint production, setup and preventive maintenance policies of unreliable two-product manufacturing systems. *International Journal of Production Research*. doi: 10.1080/00207543.2015.1030468

Baker, R. (1988). Zero Acceptance Sampling Plans: Expected Cost Increases. *Quality Progress, 21*(1), 43-46.

Ben-Daya, M. (1999). Integrated production maintenance and quality model for imperfect processes. *IIE transactions, 31*(6), 491-501.

Ben-Daya, M., & Duffuaa, S. (1995). Maintenance and quality: the missing link. *Journal of Quality in Maintenance Engineering, 1*(1), 20-26.

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

Berthaut, F., Gharbi, A., & Dhouib, K. (2011). Joint modified block replacement and production/inventory control policy for a failure-prone manufacturing cell. *Omega, 39*(6), 642-654.

Bouslah, B., Gharbi, A., & Pellerin, R. (2013). Joint optimal lot sizing and production control policy in an unreliable and imperfect manufacturing system. *International Journal of Production Economics, 144*(1), 143-156.

Bouslah, B., Gharbi, A., & Pellerin, R. (2014). Joint production and quality control of unreliable batch manufacturing systems with rectifying inspection. *International Journal of Production Research, 52*(14), 4103-4117.

Budai, G., Dekker, R., & Nicolai, R. P. (2008). Maintenance and production: a review of planning models *Complex System Maintenance Handbook* (pp. 321-344). London: Springer-Verlag.

Chakraborty, T., Giri, B., & Chaudhuri, K. (2009). Production lot sizing with process deterioration and machine breakdown under inspection schedule. *Omega, 37*(2), 257-271.

Cheney, E., & Kincaid, D. (2013). *Numerical mathematics and computing*. Boston, MA: Brooks/Cole.

Colledani, M., & Tolio, T. (2011). Joint design of quality and production control in manufacturing systems. *CIRP Journal of Manufacturing Science and Technology, 4*(3), 281-289.

Colledani, M., & Tolio, T. (2012). Integrated quality, production logistics and maintenance analysis of multi-stage asynchronous manufacturing systems with degrading machines. *CIRP Annals-Manufacturing Technology, 61*(1), 455-458.

Davies, A. (1998). *Handbook of condition monitoring: techniques and methodology*. New York: Chapman & Hall.

Felix Offodile, O., & Ugwu, K. (1991). Evaluating the effect of speed and payload on robot repeatability. *Robotics and computer-integrated manufacturing, 8*(1), 27-33.

Ferguson, T. S. (1962). Location and scale parameters in exponential families of distributions. *The Annals of Mathematical Statistics, 33*(3), 986-1001.

Fu, M. C. (1994). Optimization via simulation: A review. *Annals of Operations Research, 53*(1), 199-247.

Gershwin, S. B. (2000). Design and operation of manufacturing systems: the control-point policy. *IIE transactions, 32*(10), 891-906.

GonzÁLez, C., & Palomo, G. (2003). Bayesian acceptance sampling plans following economic criteria: an application to paper pulp manufacturing. *Journal of Applied Statistics, 30*(3), 319-333.

Gosavi, A. (2003). *Simulation-based optimization: parametric optimization techniques and reinforcement learning*. Boston: Kluwer Academic Publishers.

Goyal, S., Gunasekaran, A., Martikainen, T., & Yli-Olli, P. (1993). Integrating production and quality control policies: A survey. *European journal of operational research, 69*(1), 1-13.

Grall, A., Berenguer, C., & Dieulle, L. (2002). A condition-based maintenance policy for stochastically deteriorating systems. *Reliability Engineering & System Safety, 76*(2), 167-180.

Hadidi, L. A., Al-Turki, U. M., & Rahim, A. (2012). Integrated models in production planning and scheduling, maintenance and quality: a review. *International Journal of Industrial and Systems Engineering, 10*(1), 21-50.

Hajji, A., Gharbi, A., Kenne, J.-P., & Pellerin, R. (2011b). Production control and replenishment strategy with multiple suppliers. *European Journal of Operational Research, 208*(1), 67-74.

Hajji, A., Mhada, F., Gharbi, A., Pellerin, R., & Malhamé, R. (2011a). Integrated product specifications and productivity decision making in unreliable manufacturing systems. *International Journal of Production Economics, 129*(1), 32-42.

Hossain, A., & Zimmer, W. (2003). Comparison of estimation methods for Weibull parameters: complete and censored samples. *Journal of Statistical Computation and Simulation, 73*(2), 145-153.

Hsu, L.-F., & Kuo, S. (1995). Design of optimal maintenance policies based on on-line sampling plans. *European journal of operational research, 86*(2), 345-357.

Hu, J.-Q., Vakili, P., & Yu, G.-X. (1994). Optimality of hedging point policies in the production control of failure prone manufacturing systems. *IEEE Transactions on Automatic Control, 39*(9), 1875-1880.

Inman, R. R., Blumenfeld, D. E., Huang, N., Li, J., & Li, J. (2013). Survey of recent advances on the interface between production system design and quality. *IIE transactions, 45*(6), 557-574.

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., & Young, T. (2010). Simulation in manufacturing and business: A review. *European Journal of Operational Research, 203*(1), 1-13.

Jeang, A. (2012). Simultaneous determination of production lot size and process parameters under process deterioration and process breakdown. *Omega, 40*(6), 774-781.

Khouja, M., & Mehrez, A. (1994). Economic production lot size model with variable production rate and imperfect quality. *Journal of the Operational Research Society, 45*(12), 1405-1417.

Kim, J., & Gershwin, S. B. (2008). Analysis of long flow lines with quality and operational failures. *IIE transactions, 40*(3), 284-296.

Law, A. M. (2008). *Simulation Modeling and Analysis*. New York: McGraw Hill.

Liao, G.-L. (2013). Optimal economic production quantity policy for randomly failing process with minimal repair, backorder and preventive maintenance. *International Journal of Systems Science, 44*(9), 1602-1612.

Liao, G. L., Chen, Y. H., & Sheu, S. H. (2009). Optimal economic production quantity policy for imperfect process with imperfect repair and maintenance. *European journal of operational research, 195*(2), 348-357.

Liu, J., Shi, J., & Hu, S. J. (2009). Quality-assured setup planning based on the stream-of-variation model for multi-stage machining processes. *IIE transactions, 41*(4), 323-334.

Luenberger, D. G., & Ye, Y. (2008). *Linear and nonlinear programming*. New York: Springer.

Montgomery, D. C. (2008a). *Introduction to statistical quality control*. New York: John Wiley & Sons.

Montgomery, D. C. (2008b). *Design and analysis of experiments.* New York: John Wiley & Sons Inc.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology: process and product optimization using designed experiments*. Hoboken, NJ: John Wiley & Sons.

Nikolaidis, Y., & Nenes, G. (2009). Economic evaluation of ISO 2859 acceptance sampling plans used with rectifying inspection of rejected lots. *Quality Engineering, 21*(1), 10-23.

Pal, B., Sana, S. S., & Chaudhuri, K. (2013). A mathematical model on EPQ for stochastic demand in an imperfect production system. *Journal of Manufacturing Systems, 32*(1), 260-270.

Pan, E., Jin, Y., Wang, S., & Cang, T. (2012). An integrated EPQ model based on a control chart for an imperfect production process. *International Journal of Production Research, 50*(23), 6999-7011.

Panagiotidou, S., & Tagaras, G. (2010). Statistical Process Control and Condition-Based Maintenance: A Meaningful Relationship through Data Sharing. *Production and Operations Management, 19*(2), 156-171.

Pearn, W. L., & Wu, C.-W. (2007). An effective decision making method for product acceptance. *Omega, 35*(1), 12-21.

Pegden, C. D., Sadowski, R. P., & Shannon, R. E. (1995). *Introduction to simulation using SIMAN*. New York: McGraw-Hill.

Radhoui, M., Rezg, N., & Chelbi, A. (2009). Integrated model of preventive maintenance, quality control and buffer sizing for unreliable and imperfect production systems. *International Journal of Production Research, 47*(2), 389-402.

Radhoui, M., Rezg, N., & Chelbi, A. (2010). Integrated maintenance and control policy based on quality control. *Computers & Industrial Engineering, 58*(3), 443-451.

Rao, B. K. N. (1996). *Handbook of condition monitoring.* Oxford: Elsevier.

Rivera-Gomez, H., Gharbi, A., & Kenné, J. P. (2013). Joint control of production, overhaul, and preventive maintenance for a production system subject to quality and reliability deteriorations. *The International Journal of Advanced Manufacturing Technology, 69*(9-12), 2111-2130.

Rosenblatt, M. J., & Lee, H. L. (1986). Economic production cycles with imperfect production processes. *IIE transactions, 18*(1), 48-55.

Sana, S. S. (2010a). A production–inventory model in an imperfect production process. *European Journal of Operational Research, 200*(2), 451-464.

Sana, S. S. (2010b). An economic production lot size model in an imperfect production system. *European Journal of Operational Research, 201*(1), 158-170.

Sana, S. S. (2012). Preventive maintenance and optimal buffer inventory for products sold with warranty in an imperfect production system. *International Journal of Production Research, 50*(23), 6763-6774.

Sarimveis, H., Patrinos, P., Tarantilis, C. D., & Kiranoudis, C. T. (2008). Dynamic modeling and control of supply chain systems: A review. *Computers & Operations Research, 35*(11), 3530-3561.

Schilling, E. G., & Neubauer, D. V. (2009). *Acceptance sampling in quality control.* Boca Raton: Chapman & Hall/CRC.

Tapiero, C. S. (1986). Continuous quality production and machine maintenance. *Naval Research Logistics Quarterly, 33*(3), 489-499.

Tekin, E., & Sabuncuoglu, I. (2004). Simulation optimization: A comprehensive review on theory and applications. *IIE transactions, 36*(11), 1067-1081.

Wang, H., & Pham, H. (2006). Imperfect Maintenance and Dependence *Reliability and Optimal Maintenance* (pp. 13-30). New York: Springer.

Wee, H. M., & Widyadana, G. A. (2013). A production model for deteriorating items with stochastic preventive maintenance time and rework process with FIFO rule. *Omega, 41*(6), 941-954.

Xia, T., Jin, X., Xi, L., & Ni, J. (2015). Production-driven opportunistic maintenance for batch production based on MAM–APB scheduling. *European Journal of Operational Research, 240*(3), 781-790.

Xia, T., Xi, L., Zhou, X., & Lee, J. (2012). Dynamic maintenance decision-making for series–parallel manufacturing system based on MAM–MTW methodology. *European Journal of Operational Research, 221*(1), 231-240.

Yang, H., & Carlin, D. (2001). Acceptance Sampling Plans by Attributes *Applied Statistics in the Pharmaceutical Industry* (pp. 475-502): Springer.

Zhang, G., Deng, Y., Zhu, H., & Yin, H. (2015). Delayed maintenance policy optimisation based on control chart. *International Journal of Production Research*, *53*(2), 341-353.

Table 3. Sensitivity analysis for model parameters.

Case Number	Parameter	Variation	Policy I (with acceptance sampling plan)												Policy II (with 100% inspection)				Δ-Cost
			c^*	n^*	Q^*	S^*	r^*	$\psi^*_{c^*}(.)$	AOQL	$AOQ(\infty)$	$P_a(\infty)$	$\bar{f}(\infty)$	$\bar{m}(\infty)$	$Av(\infty)$	Q^*	S^*	r^*	$\psi^*_{100\%}(.)$	
basic	-	-	4	157	1485	2631	1.996%	1 681.5 $	1.45%	0.74%	0.829	0.0333	0.0227	0.808	1573	3374	2.022%	2 116.7 $	-20.6%
1	C_h	+50%	4	168	1451	2161	1.828%	1 798.9 $	1.34%	0.69%	0.821	0.0291	0.0234	0.809	1523	2870	1.766%	2 262.5 $	-20.5%
2		-50%	4	147	1518	3085	2.143%	1 551.3 $	1.56%	0.79%	0.834	0.0377	0.0223	0.803	1625	3897	2.268%	1 946.2 $	-20.3%
3	C_b	+50%	4	153	1418	2812	1.967%	1 768.3 $	1.48%	0.75%	0.831	0.0346	0.0219	0.809	1563	3595	1.937%	2 189.3 $	-19.2%
4		-50%	4	162	1561	2155	2.085%	1 608.0 $	1.41%	0.73%	0.813	0.0333	0.0227	0.805	1589	2776	2.140%	2 051.5 $	-21.6%
5	C_{set}	+50%	4	160	1605	2706	1.833%	1 825.9 $	1.43%	0.74%	0.802	0.0362	0.0228	0.802	1677	3453	1.980%	2 252.4 $	-18.9%
6		-50%	5	156	1004	2286	2.752%	1 498.9 $	1.72%	0.80%	0.880	0.0329	0.0204	0.818	1231	3114	2.144%	1 953.9 $	-23.3%
7	C_{cm}	+50%	4	159	1449	2565	1.859%	1 777.9 $	1.42%	0.73%	0.822	0.0318	0.0229	0.809	1548	3155	1.390%	2 188.3 $	-18.8%
8		-50%	5	151	1522	2963	2.043%	1 629.9 $	1.89%	0.93%	0.859	0.0464	0.0212	0.794	1622	3770	2.970%	2 027.4 $	-19.6%
9	C_{ovr}	+50%	5	165	1472	2983	2.631%	2 072.8 $	1.71%	0.84%	0.847	0.0433	0.0210	0.801	1530	3670	2.957%	2 483.4 $	-16.5%
10		-50%	3	164	1546	2380	0.805%	1 277.2 $	1.06%	0.58%	0.784	0.0227	0.0263	0.802	1652	2716	0.505%	1 631.8 $	-21.7%
11	C_{insp}	+50%	5	154	1451	2822	1.618%	1 777.4 $	1.84%	0.90%	0.868	0.0441	0.0253	0.798	1572	3368	2.006%	2 370.0 $	-25.0%
12		-50%	4	162	1503	2554	2.255%	1 589.1 $	1.40%	0.71%	0.792	0.0323	0.0218	0.808	1574	3381	2.037%	1 863.4 $	-14.7%
13	C_{rej}	+50%	4	155	1471	2613	1.897%	1 699.0 $	1.46%	0.75%	0.834	0.0333	0.0227	0.806	1565	3325	1.899%	2 142.7 $	-20.7%
14		-50%	4	159	1498	2648	2.106%	1 663.5 $	1.44%	0.73%	0.820	0.0334	0.0226	0.807	1581	3424	2.144%	2 089.8 $	-20.4%
15	C_{def}	+50%	4	158	1482	2614	2.031%	1 706.9 $	1.44%	0.74%	0.821	0.0325	0.0227	0.807	1573	3374	2.022%	2 116.7 $	-19.4%
16		-50%	4	155	1486	2651	1.963%	1 656.0 $	1.47%	0.75%	0.834	0.0332	0.0226	0.806	1573	3374	2.022%	2 116.7 $	-21.8%
17	τ_{insp}	+50%	3	166	1444	2859	1.497%	1 730.3 $	1.04%	0.70%	0.841	0.0235	0.0246	0.805	1475	3488	2.182%	2 193.7 $	-21.1%
18		-50%	4	151	1508	2488	2.062%	1 664.3 $	1.52%	0.67%	0.725	0.0351	0.0224	0.800	1632	3319	1.930%	2 023.3 $	-17.7%
19	γ_q	+15%	4	138	1325	2144	4.329%	1 813.0 $	1.65%	0.76%	0.703	0.0083	0.0294	0.809	1549	2790	1.201%	2 275.7 $	-20.3%
20		-15%	3	126	1640	3946	0.411%	1 572.8 $	1.42%	0.61%	0.969	0.0747	0.0176	0.747	1659	4779	3.319%	2 053.0 $	-23.4%
21	γ_r	+15%	2	134	651	3453	1.004%	2 659.2 $	0.81%	1.55%	0.924	0.0915	0.0318	0.625	900	4052	1.516%	3 065.4 $	-13.3%
22		-15%	7	198	1718	2258	2.628%	1 228.2 $	2.00%	0.59%	0.800	0.0015	0.0192	0.882	1801	2842	2.222%	1 799.6 $	-31.7%
23	γ_{set}	+15%	3	118	1277	3418	2.242%	2 243.1 $	1.49%	0.79%	0.773	0.0486	0.0327	0.719	1419	5670	2.931%	2 438.9 $	-8.0%
24		-15%	2	127	1629	2537	1.678%	1 162.1 $	1.00%	0.62%	0.841	0.0040	0.0093	0.937	1687	3293	1.723%	1 547.0 $	-24.9%