7-2020

Study on New Sampling Plans and Optimal Integration with Proactive Maintenance in Production Systems

Sinan Obaidat
University of Arkansas, Fayetteville

Study on New Sampling Plans and Optimal Integration with Proactive Maintenance in Production Systems

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Engineering with a concentration in Industrial Engineering

by

Sinan Obaidat
Jordan University of Science and Technology
Bachelor of Science in Industrial Engineering, 2004
Jordan University of Science and Technology
Master of Science in Industrial Engineering, 2010

July 2020
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Haitao Liao, Ph.D.
Dissertation Director

Chase Rainwater, Ph.D.
Committee Member

Shengfan Zhang, Ph.D.
Committee Member

Yisha Xiang, Ph.D.
Ex-Officio Member

Abstract

Sampling plans are statistical process control (SPC) tools used mainly in production processes. They are employed to control processes by monitoring the quality of produced products and alerting for necessary adjustments or maintenance. Sampling is used when an undesirable change (shift) in a process is unobservable and needs time to discover. Basically, the shift occurs when an assignable cause affects the process. Wrong setups, defective raw materials, degraded components are examples of assignable causes. The assignable cause causes a variable (or attribute) quality characteristic to shift from the desired state to an undesired state.

The main concern of sampling is to observe a process shift quickly by signaling a true alarm, at which, maintenance is performed to restore the process to its normal operating conditions. While responsive maintenance is performed if a shift is detected, proactive maintenance such as age-replacement is integrated with the design of sampling. A sampling plan is designed economically or economically-statistically. An economical design does not assess the system performance, whereas the economic-statistical design includes constraints on system performance such as the average outgoing quality and the effective production rate.

The objective of this dissertation is to study sampling plans by attributes. Two studies are conducted in this dissertation. In the first study, a sampling model is developed for attribute inspection in a multistage system with multiple assignable causes that could propagate downstream. In the second study, an integrated model of sampling and maintenance with maintenance at the time of the false alarm is proposed.

Most of the sampling plans are designed based on the occurrence of one assignable cause. Therefore, a sampling plan that allows two assignable causes to occur is developed in the first study. A multistage serial system of two unreliable machines with one assignable cause that could

occur on each machine is assumed where the joint occurrence of assignable causes propagates the process's shift to a higher value. As a result, the system state at any time is described by one in-control and three out-of-control states where the evolution from a state to another depends on the competencies between shifts. A stochastic methodology to model all competing scenarios is developed. This methodology forms a base that could be used if the number of machines and/or states increase.

In the second study, an integrated model of sampling and scheduled maintenance is proposed. In addition to the two opportunities for maintenance at the true alarm and scheduled maintenance, an additional opportunity for preventive maintenance at the time of a false alarm is suggested. Since a false alarm could occur at any sampling time, preventive maintenance is assumed to increase with time. The effectiveness of the proposed model is compared to the effectiveness of separate models of scheduled maintenance and sampling.

Inspired by the conducted studies, different topics of sampling and maintenance are proposed for future research. Two topics are suggested for integrating sampling with selective maintenance. The third topic is an extension of the first study where more than two shifts can occur simultaneously.

Dedication

This dissertation is dedicated to

The memory of my father, Fayez Obaidat, who passed away before I completed my doctoral studies,

My mother, Intisar Aleksh, for support and encouragement,

My wife, Diana Alshare, for her endless patience and support.

My beloved children, Zain, Rashid, and Haya.

Table of Contents

List of Tables

List of Figures

List of Published Papers

Obaidat, S. and Liao, H. (2019) Optimal Sampling Plan for an Unreliable Multistage Production System Subject to Competing and Propagating Random Shifts. *Submitted to IIE Transactions, under second review.*
Chapter 2.

Obaidat, S. and Liao, H. (2020) Integrated Decision Making for Attributes Sampling and Proactive Maintenance in a Discrete Manufacturing System. *Published online in the International Journal of Production Research*. https://doi.org/10.1080/00207543.2020.1781280.
Chapter 3.

Chapter 1 : Introduction

1. Sampling in production systems

Products are inspected to ensure that they are good enough to proceed to the next manufacturing process or to be forwarded to final stock for shipping to consumers. Lots of products can be accepted, rejected, re-inspected, or rectified based on the result of inspection (Montgomery, 2009).

In quality control, inspection procedures can be classified into three broad categories: no inspection, screening (100% inspection), and sampling. No inspection policy can be applied if incoming products or raw materials are shipped from a highly certified supplier or produced with highly reliable procedures. If a product is expensive, a process is critical, or a new machine or a process is introduced in the production system, screening inspection may be an option, but it may be carried out with high inspection costs. A common procedure of inspection is sampling, a moderate option to compromise between the high cost of screening and the high cost of quality loss of no inspection.

With sampling, inferences about the manufacturing process are drawn. The purpose of sampling is not only to control the quality of products produced, but it is also employed to make decisions regarding production and maintenance. Measures of production such as the economic production quantity and throughput can be determined based on sampling. Furthermore, some maintenance decisions are taken according to the results of sampling. The type of maintenance, the degree of maintenance, and the time to perform maintenance are some examples of maintenance decisions.

Sampling procedures differ according to how products are produced, inventoried, or shipped. Acceptance sampling is designed to inspect from a produced lot (batch). After production ends and products accumulate in lots, a sample is taken from a lot to determine whether a lot should be accepted or rejected. In online sampling, multiple samples are taken from the production line at specified times. The main purpose is to determine if the process goes out-of-control or not, and accordingly, the suitable maintenance action is taken.

Basically, an online sampling scheme is built based on a control chart to monitor the occurrence of an assignable cause that causes a process to shift from the in-control to the out-of-control condition. Other versions of sampling procedures are continuous sampling. The basic model of continuous sampling assumes that the process is always in-control (Dodge, 1943). It alternates between screening and fractional inspection in order to achieve a desired outgoing quality with a minimum inspection.

Sampling could be performed with variable or attribute data. While sampling with variables monitor a variable quality characteristic such as a mean of an ingredient concentration, attributes sampling use counted data such as the number of defective units. Sampling with $\bar{X}$ and R control charts are examples of variable sampling, whereas sampling with np and p control charts are examples of attributes sampling.

Throughout this dissertation, only online sampling with attributes is studied. The "sampling plan" term is used to refer to the setup of sampling parameters used to achieve certain objectives. For instance, the best parameters of a sampling plan with the $\bar{X}$ control chart (sample size, control limit coefficient, and time between samplings) can be determined by minimizing a cost function.

2. Dissertation objectives and organization

This dissertation aims at developing new sampling plans that suit specific production systems designs with the adoption of different maintenance strategies. In this dissertation, two studies are presented in addition to different problems to be addressed for future work.

The first study aims at establishing a methodology of modeling multiple assignable causes in multistage systems. During the inspection cycle, multiple assignable causes could occur. The occurrence of an assignable cause causes a process to shift from a desired “in control state” to an undesired “out-of-control” state. The shift is characterized by an increase of the proportion of nonconforming units to an unacceptable level.

Most sampling plans are built on the assumption that one assignable cause could occur. A few studies are conducted in the case of multiple assignable causes. However, it is assumed that only one of the assignable causes can occur during the inspection cycle. Thus, the joint occurrence of multiple assignable causes in one inspection cycle is investigated. For this purpose, a two-machine serial production system that produces a discrete product is assumed.

Due to the presence of multiple assignable causes, the production system becomes a multistate system with one in-control state and three out-of-control states. Moreover, if one shift is not detected before the occurrence of another shift, a shift magnitude increases “propagates” to a higher value where the magnitude of the propagating shift is a function of the magnitudes of the shifts that occur jointly.

A comprehensive modeling approach is constructed according to the stochastic competition and propagation of shifts. The sampling plan parameters are found by minimizing the long-run cost rate subject to constraints on the system’s availability, the average time to signal, and the effective production rate.

The second study aims to develop an integrated design of sampling and maintenance with multiple maintenance opportunities. This design intent at taking simultaneous decisions regarding the time to perform the scheduled maintenance and the sampling parameters with the aim of minimizing the long-run cost rate of the inspection cycle.

While most of the sampling procedures assume no maintenance is carried out if a false alarm is detected, the proposed design considers the occurrence of a false alarm as a maintenance opportunity. There are a few studies that assume the same concept, but with constant maintenance cost. Also, those studies don't illustrate the benefit of having the additional maintenance opportunity at the false alarm time.

In contrast to those studies, time to accomplish maintenance upon a false alarm is assumed to increase with time. This assumption is made because the likelihood of a shift increases with time. The integrated approach is compared to other alternatives that consider the modeling of sampling and scheduled maintenance separately.

These studies fill some gaps in the literature and provide a basement for further research. The problem of the joint occurrence of multiple assignable causes could be extended to systems composed of more than two machines and/or for more than two states for each machine. Additionally, some subjects of integrating sampling with selective maintenance are suggested. First, a single production unit is assumed. Second, a series-parallel multistate system is examined. For this kind of systems, literature concentrates on developing selective maintenance models to achieve quantitative output and ignore the quality of production. Hence, joint sampling and selective maintenance for such systems is presented.

The remainder of the dissertation is organized as follows. Chapter 2 addresses the problem of the occurrence of multiple assignable causes during a sampling cycle in multistage systems. A detailed methodology of modeling competing and propagating shifts is presented. In Chapter 3, an integrated design of sampling and scheduled maintenance with multiple maintenance opportunities is proposed. Chapter 4 concludes the dissertation and proposes different topics for possible future work.

3. References

Dodge, H. F. (1943) A Sampling inspection plan for continuous production. *The Annals of Mathematical Statistics,* **14**(3), 264-279.

Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*. 6th ed.Wiley: Hoboken, N.J.

Chapter 2 : Optimal Sampling Plan for an Unreliable Multistage Production System Subject to Competing and Propagating Random Shifts

Abstract

Sampling plans play an important role in monitoring production systems and reducing costs related to product quality and maintenance. Most of the existing sampling plans usually focus on one assignable cause. However, multiple assignable causes may occur especially for a multistage production system, and the resulting process shift may propagate downstream. This chapter addresses the problem of finding the optimal sampling plan for a multistage production system subject to competing and propagating random quality shifts. In particular, a serial production system with two unreliable machines that produce a product at a fixed production rate is studied. It is assumed that both machines are subject to random quality shifts due to the presence of assignable causes and can suddenly fail with increasing failure rates. If not failed, each machine operates either in its in-control state or in its out-of-control (i.e., shifted) state with different nonconforming rates. A sampling plan is implemented at the end of the production line to determine whether the system has shifted or not. If a process shift is detected, a necessary maintenance action will be initiated. The optimal sample size, sampling interval, and acceptance threshold are determined by minimizing the long-run cost rate subject to the constraints on average time to signal a true alarm, effective production rate, and system availability. A numerical example is provided to illustrate the application of the proposed sampling plan, and detailed analyses on the effects of key parameters and system constraints are also conducted.

Keywords: Sampling plan, multistage production systems, competing and propagating random shifts.

1. Introduction

Quality improvement is a major concern for the success of a manufacturing enterprise. To be competitive, companies often adopt different procedures to improve their production processes and offer products with high quality. Although the advances in technology and automation enable companies to produce reliable products, the manufacturing environment is always subject to variability and random shift that affect product quality. Monitoring processes and inspecting products aid to take necessary actions at the right times for process adjustments when a product's quality deviates away from the standards.

Product inspection is one of process monitoring methods to determine if a process has shifted or not. The out-of-control state is attributed to the presence of assignable cause(s) such as tool wear, temperature increase, and wrong setups. Specially, an assignable cause makes a process variable, such as the process mean, to deviate from its target, or causes an attribute, such as the proportion of nonconformity, to increase. In addition to the process shift, the production system may fail and stop production. When a process shift or system failure is detected, maintenance actions are initiated. Maintenance could be perfect, imperfect, or minimal. In particular, perfect maintenance restores a production unit to its good-as-new condition, imperfect maintenance restores the unit to a condition between its good-as-new and bad-as-old states, and minimal repair makes the unit operational while keeping the unit in the same health condition as before.

Regarding inspection options, screening (100% inspection), no inspection, sampling plans by control charts (online sampling), acceptance sampling, and continuous sampling are the most widely used. In practice, an inspection policy is adopted according to the type of production and a specific goal. For instance, acceptance sampling is used for batch (lot) production to decide whether a batch should be accepted or not. Such inspection procedures can be employed in both single-stage and multistage systems. Specially, a multistage system is composed of multiple components, machines, processes, or stages required to make the final product (Shi and Zhou, 2009).

A sampling plan is either designed economically or economically-statistically. Economic designs aim at minimizing a cost function without focusing on statistical performance, while economic-statistical designs consider the performance of a process under some practical constraints. The usual performance metrics could be customer-centered such as the average outgoing quality (AOQ). Some measures are more producer-centered such as the average fraction inspected (AFI), process availability, and throughput. Other metrics, such as schedules' delays, are concerning both parties. Studies on these measures can be found in Bouslah et al. (2013), Cao and Subramaniam (2013), and Pandey et al. (2011).

Existing sampling plans are often developed for a single or a multistage system based on the occurrence of one assignable cause. For example, extensive research concentrates on using $\bar{X}$ control chart to monitor single stage systems in the presence of one assignable cause. In these systems, the assignable cause causes a quality characteristic to shift from the in-control state to the out-of-control state, and hence, only two states of the system are considered. Although a few studies consider cases with multiple assignable causes, it is assumed that only one assignable cause can occur during a sampling cycle, and only two states of the production system are considered.

In this chapter, we develop an economic-statistical sampling plan for a serial production system with two unreliable machines by considering the occurrences of more than one assignable cause. The term "stage" can be used in lieu of "machine" to refer to a process or a group of machines (processes). The sampling plan is modeled based on the competency and downstream propagation of process shifts. Sampling parameters are determined by minimizing the long-run cost rate subject to constraints on effective production rate, average time to signal a true alarm and system availability. It is assumed that sampling is performed only after the second stage. For example, in some systems, the synchronized handling of products from one stage to another does not allow any stoppage for inspection after the first stage. In other systems, products are processed sequentially or simultaneously by two different processes on the same machine making quality inspection impractical due to the machine's complex configuration.

Some industrial applications of such a system are as follows. In an automatic blasting and painting line, a fabricated steel unit is first blasted for rust removal and then fed into a painting chamber. Due to degradation, the disc turbines that provide blasting may still leave some rust on the unit's surface that causes poor paint adhesion. On the other hand, the spray nozzles in the painting chamber, if clogged, could cause bad paint coverage. The unit produced is nonconforming if one or both of the quality issues occur. An example of two processes being performed automatically on one machine is the production of purlins for steel structures. Galvanized sheets are fed continuously into a forming machine. Punching holes and bending edges are sequentially or simultaneously processed to produce a purlin. Due to the complex configuration of the machine, any quality imperfection cannot be observed until the whole process is complete. When the punching tips and/or the bending rollers become worn, the purlin is defective because holes, edges, or both are imprecisely made. In some industries, inspection may be performed only after the final stage due to safety or economic reasons. For instance, small steel bars are first heated and then forged to produce small parts such as socket wrenches. Other examples are manufacturing of aluminum cans, automated bakery production, powder coating, automatic riveting for stamping parts, automatic assembling and wire bonding, and multi-material additive manufacturing of electronic devices.

The remainder of this chapter is organized as follows: Section 2 reviews the related literature and illustrates the contributions. Section 3 describes the problem and the assumptions, and provides the notation used throughout this chapter. A comprehensive modeling methodology is developed in Section 4. Section 5 provides the mathematical formulation for the optimal design of the proposed sampling plan. A numerical example and analyses are given in Section 6. Section 7 concludes this chapter and recommends several directions for future research.

2. Literature review and contribution

2.1. Related work

In the context of single-stage production systems, Chiu and Huang (1996) consider an economic model for a preventively maintained process monitored by an $\bar{X}$ control chart. A cost function is minimized without considering statistical performance constraints. Cassady et al. (2000) combine age-replacement preventive maintenance and an $\bar{X}$ control chart in an economic design. Linderman et al. (2005) propose an economic-statistical cost model considering constraints on the average run lengths and three maintenance scenarios. Charongrattanasakul and Pongpullponsak (2011) extend this work by sampling with an exponentially weighted moving average (EWMA) chart with warning limits along with maintenance at the time of a false alarm. Mehrafrooz and Noorossana (2011) consider an additional maintenance scenario due to sudden machine failures. Pandey et al. (2011) use an $\bar{X}$ control chart to determine the sequence of batches produced on a single machine subject to scheduled preventive maintenance. It is worth pointing out that all these studies focus only on one assignable cause. However, this may not be realistic.

Indeed, multiple assignable causes from different sources, such as raw materials, human errors and tool wear, cannot be ignored. Yu and Hou (2006) develop an economic model for an $\bar{X}$ control chart with variable sampling intervals to monitor a process with multiple assignable causes. Yu et al. (2010) construct an economic-statistical model with constraints on type-I and type-II errors. The same constraints are used by Salmasnia et al. (2017). Unlike these studies where only one assignable cause is permitted to occur during an inspection cycle, a case allowing the occurrences of multiple assignable causes during an inspection cycle is examined by Yang et al. (2010). An $\bar{X}$ control chart is designed, but the joint effect of two assignable causes is assumed to be the same. Xiang (2013) study the joint optimization of an $\bar{X}$ control chart and preventive maintenance for a deteriorating production system. The system is assumed to have multiple degraded states that correspond to different assignable causes, and an economic cost model for maintenance, operation, and inspection is provided.

Inspection procedures for multistage systems are diverse. Zantek et al. (2002) assume that the variation of a measurement at a stage depends on both the variation of process parameters (i.e., pressure, temperature, etc.) at the present stage and the variations of measurements taken at preceding stages. Their study aims at identifying which quality and process variables are responsible for the variation at the final stage. Jin and Shi (1999) and Zhou et al. (2003) propose engineering models for a sheet metal assembly line and an automotive engine heads machining line, respectively, to identify sources of variations. Xiang and Tsung (2008) study statistical monitoring with group EWMA control charts based on engineering models. The EWMA control chart is designed for a given in-control average run length to determine the out-of-control condition in a three-stage process where wrong fixturing causes the process to be out-of-control. Without process variables, Lam et al. (2005) develop an engineering model for a four-stage machining process where the last stage has two streams (parallel machines), and each stage or stream is monitored by a separate $\bar{X}$ control chart. It is assumed that only one stage is out-of-control at any time and the probability that a stage is out-of-control is constant. The $\bar{X}$ control charts are only designed to alert out-of-control signals according to a desired average time to signal without addressing whether any adjustment on the process or any rework on defective products is carried out or not.

Inspection allocation is another focus related to multistage systems. Williams and Peters (1989) study inspection allocation in a three-stage serial system monitored by np control charts. Bai and Yun (1996) consider a serial three-stage circuit board manufacturing system with two inspection stations. Inspection locations and inspection level (number of components tested on a circuit board) are determined to minimize the expected total cost of rework, inspection, and defective boards delivered to customers. For the same industry, Chevalier and Wein (1997) study the optimal testing policy that jointly determines the specification limits and inspection locations. Moreover, Rau and Chu (2005) examined a serial multistage system where inspection could be on product variables and attributes. Azadeh et al. (2015) study a batch production system where inspection allocation, inspection tolerances, and full inspection or acceptance sampling are determined. Other studies on inspection allocation are done by Shiau (2002) and Valenzuela et al. (2004).

The quality and quantity are the two main focuses of a multistage production system. Cao and Subramaniam (2013) investigate a serial multistage system where each stage is monitored by a continuous sampling plan (CSP). The CSP alternates between 100% and fractional inspections based on whether or not a consecutive number of conforming units are observed. Additional measures of work in process (WIP) and throughput rate are also considered. Kim and Gershwin (2005) study a two-machine system with one buffer using a Markov process. In their work, a machine is assumed to have three states: operating producing good parts, operating producing bad parts (quality failure state), and complete failure. The effects of quality failure, production rate, and buffer size on the system's yield and effective production rate are analyzed. Kim and Gershwin

(2008) also analyze the performance of flow lines with quality and operational failures. Meerkov and Zhang (2010) investigate different cases for performance analysis of a serial production system with inspection stations and buffers under 100% inspection. Given the number of inspection stations and buffers capacities, the study shows the impact of inspection allocation on bottlenecks, blocked and starving machines, and effective production rate. Colledani and Tolio (2012) develop a Markovian model for a serial system subject to degradation. The critical state that separates the desired degradation states from the undesired states is determined by achieving gains in system's yield and effective production rate. It is worth pointing out that engineering models are analytical tools for identifying sources of variation for quality improvement. Usually, a strategy with 100% inspection of variables is adopted. On the other hand, in most of the inspection allocation models, 100% inspection or acceptance sampling are used with the purposes of locating inspection and determining a testing strategy or inspection level. For both types of models, maintenance is rarely studied.

Liu et al. (2013) study a serial system consisting of two identical units monitored by an $\bar{X}$ control chart. The value of process shift is assumed to be a constant no matter one or both units are in the quality failure state, and an inspection cycle is renewed by one of four maintenance scenarios. The system's performance is evaluated via economic and economic-statistical models with constraints on type-I and type-II errors. Zhu et al. (2016) investigate a serial four-stage process where attributes sampling is carried out at each stage. In their work, only quality failures are considered, and the sampling parameters are found by minimizing the expected total cost of inspection, scrap, and repair with respect to constraints on the average number of produced products between two false alarms. Zhong and Ma (2017) propose a joint control chart for a two-stage dependent serial system where the first and second stages are monitored by an $\bar{X}$ and a residual control chart, respectively. Eight maintenance scenarios are investigated for cost minimization with constraints on the average run lengths. For more studies on part quality inspection in multistage production systems, readers are referred to a recent review by Rezaei-malek et al. (2019).

2.2. Contribution of this work

Clearly, the effects of quality failures, machine failures and maintenance actions on the product quality and the effective production rate of a multistage production system are worthy of investigation. Although a plenty of studies have been conducted on online sampling for single-stage production systems, only a few studies have been done on multistage systems. Specially, there is a lack of research on online sampling of attribute data for multistage systems.

This study aims at developing an attribute sampling plan for a serial multistage system of two unreliable machines for discrete production. Different from the work of Liu et al. (2013), this work considers two nonidentical machines and allows a quality shift to propagate downstream. Indeed, competing process shifts and downstream propagation are two forms of natural interactions in a multistage system. To the best of our knowledge, modeling sampling plans by attributes with competing shifts in a multistage system with unreliable machines have not been studied (Yang et al., 2010; Zhu et al., 2016) in the literature although such a study will have a wide variety of industry applications.

In addition, this work develops a comprehensive economic-statistical model with closed-form formulations and establishes a compromise between quality and quantity performances. Unlike the studies by Yang et al. (2010), Liu et al. (2013) and Xiang (2013) that focus only on quality-related performance, we consider a constraint on system's availability to increase production, and a constraint on effective production rate to increase the fraction of good products. Moreover, a constraint on average time to signal is also included. This model represents a first step that can be extended for a production line with more than two unreliable machines, multiple assignable causes, and different levels of maintenance actions.

3. Problem description

3.1. Problem statement

A serial production system consisting of two unreliable machines that operate continuously to produce discrete units of a product is considered. Each unit of the product is first processed at machine 1 followed by machine 2. Each machine has the proportion of nonconforming (PON) of p_{0m}, $m \in \{1,2\}$ when it is in-control. Due to assignable causes, PON may increase to p_{1m} so that the machine enters its out-of-control state. Each machine is subject to two issues: quality shift when the PON increases from p_{0m} to p_{1m}, and sudden machine breakdown (failure). Failures are observed immediately, whereas quality shifts can be detected only by inspection.

To inspect the finished units, an attribute sampling plan is employed at the end of the production line (i.e., after machine 2) to assess the performance of the production process and to initiate necessary maintenance actions. An inspected unit is classified as either conforming or nonconforming, and if a half-finished unit is nonconforming upstream (after machine 1), it remains nonconforming downstream. The power of detecting a process shift depends on the parameter setting of the sampling plan.

Clearly, sampling may generate two kinds of errors: type I error and type II error. Type I error (false alarm) is generated when a process signals an alarm given that the process has not shifted yet. Type II error is generated when the sampling plan fails to signal a true alarm when the process has already shifted. Determining which machine(s) has/have shifted cannot be done unless the system is shut down for close inspections of the two machines. Therefore, whenever there is a failure or a shift, both machines are stopped for maintenance. However, when machines are shut down because of a false alarm, no maintenance is carried out and production resumes.

It is assumed that the time to shift for machine m follows the exponential distribution with a rate of λ_m, whereas time to failure is assumed to follow the two-parameter Weibull distribution with an increasing failure rate. During operation, if a machine fails, minimal repair is performed, which makes the machine operational but does not reduce its failure rate after repair. If a shift is detected, both machines are restored to their good-as-new conditions with PON of p_{0m} and age 0, and a new inspection cycle begins. Restoration can be either corrective or preventive. Corrective restoration is performed on the machine that has the shift, whereas preventive restoration resets the age of the machine that has not shifted to zero.

Whenever a true alarm is signaled, it is clear that at least one machine has shifted. The sampling plan is designed to detect competing and propagating shifts. A propagating shift occurs if one machine has already shifted but that shift is not detected until the other machine shifts. Hence, the production system is classified as a multistage multistate system. The system at any sampling time can be in one of four states: one in-control state, and three out-of-control states. The system's PON (p_s) can be represented by the set

$$p_s = \{p_0, p_1, p_2, p_3\},$$

where $p_0 = \phi(p_{01}, p_{02})$ is p_s when the system is in the in-control state (i.e., both machines are in control) and $\phi(\cdot,\cdot)$ is a function of machines' PONs. p_1, p_2, and p_3 represent p_s in the out-of-control states where $p_1 = \phi(p_{11}, p_{02})$ is p_s if only machine 1 has shifted, $p_2 = \phi(p_{01}, p_{12})$ is p_s if only machine 2 has shifted, and $p_3 = \phi(p_{11}, p_{12})$ is p_s if the process ends with the propagating shift . Note that p_0 can evolve to either p_1 or p_2, and p_1 or p_2 can evolve to p_3. Basically, p_s can be calculated as:

$$p_s = \phi(p_{f1}, p_{f2}) = 1 - \prod_{m=1}^{2}(1 - p_{fm}), \quad (2.1)$$

where $f = \{0$, machine m is in-control; 1, machine m is out-of-control$\}$.

To study the process with competing and propagating shifts, the sampling plan with one assignable cause proposed by Lorenzen and Vance (1986) is used as the baseline. The sampling plan is illustrated in Figure 2.1. A new inspection cycle starts with both machines being in good-as-new conditions. Inspection continues until a true alarm is signaled. Therefore, the inspection cycle length is defined as the time since the beginning of sampling until the two machines are restored correctively and/or preventively back to their good-as-new conditions after a true alarm. After each "h" time units (called the sampling interval), N units are sampled and inspected. If the number of nonconforming units in this sample exceeds an acceptance threshold r, the two machines are investigated to determine if the out-of-control signal is a false alarm or a true alarm. All the sampled units found to be nonconforming are rejected without replacement.

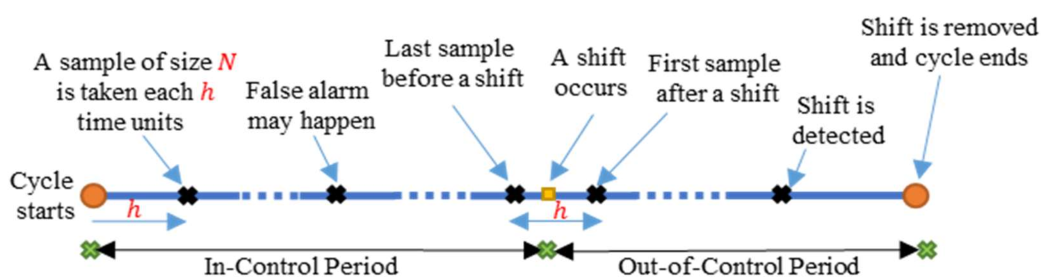

Figure 2.1. Sampling plan proposed by Lorenzen and Vance (1986).

By taking into account competing and propagating shifts, the sampling plan shown in Figure 2.1 is modified in Section 4. The objective is to design an attribute sampling plan considering stochastic competing and propagating shifts. An optimization model is developed to minimize the long-run cost rate and to find the optimal sampling parameters. The assumptions and notation used in this chapter are provided as follows.

3.2. Assumptions

The following assumptions are made throughout this chapter:

- The raw materials are defect free (i.e., incoming quality is perfect). Note that if the incoming quality is not perfect, this effect can be folded into the first stage in-control nonconforming probability.
- Quality shift and machine failure are independent. For example, in an automated painting line, as the ambient temperature decreases, paint becomes more viscous causing undesirable coat quality, but the increased viscosity of paint does not cause a complete machine failure.
- Shifts on the two machines are independent, as the two machines perform different tasks involving different types of shifts.
- The production rates and reliability of the two machines are not significantly different.
- There are enough storage areas for the finished products and WIP so that the production will not be stopped because of lacking storage areas.
- The system is stopped during sampling, which prevents the process with a potential quality shift from running during sampling. This is reasonable if the loss due producing nonconforming units is high. Note that the sampling interval (i.e., h) is an important decision variable in this study.
- The two machines do not deteriorate or shift while being stopped.
- Maintenance requests can only be fulfilled in sequence. In other words, a machine can be maintained only after the current maintenance action is complete. This is reasonable when only one maintenance team is involved.

3.3. Notation

The notation used throughout this chapter is listed in Table 2.1.

Table 2.1: Notation list.

Decision variables	
h	Sampling interval measured in hours.
N	Sample size
r	Acceptance threshold
Objective function	
$LRCR$	Long-run cost rate measured in \$/hour
Other variables, constants and indices	
j	Index referring to the sample number at which an inspection cycle ends
i, k, q, w	Indices
m	Index for a machine, $m \in \{1,2\}$
G	Inspection cycle operational time excluding false alarms, minimal repairs, true alarm, and restoration times
S_m	Shift of machine m, $m \in \{1,2\}$
S_{12}	Propagating shift
λ_m	Shift rate of machine m, $m \in \{1,2\}$
T_m	Time to shift of machine m, exponentially distributed $T_m \sim \text{Exp}(\lambda_m)$, $m \in \{1,2\}$
τ_{S_m}	Time of occurrence of S_m since the last sampling
PON	Proportion of nonconforming
p_{fm}	PON of machine m, $m \in \{1,2\}$, $f = \{0$, machine m is in-control; 1, machine m is out-of-control$\}$
p_s	PON of the production system
$\phi(\cdot,\cdot)$	A function that represents p_s in terms of machines' PONs
d	Number of nonconforming units found in a sample of size N
α	Type I error due to a false signal
T_{in}	Time process stays in the in-control state
T_{s_1}	Time the process is running with $p_s = p_1 = \emptyset(p_{11}, p_{02})$
T_{s_2}	Time the process is running with $p_s = p_2 = \emptyset(p_{01}, p_{12})$
$T_{s_{12}}$	Time the process is running with $p_s = p_2 = \emptyset(p_{11}, p_{12})$
β_{p_s}	Type II error when $p_s \in \{p_1, p_2, p_3\}$
ARL_0	Average run length while the process is in-control
$ARL_{s_{12}}$	Average run length while the process is out-of-control with propagating shift
Q_{in}	Number of samples taken while the process is in-control
$Q_{p_1}(Q_{p_2})$	Number of samples taken while the process is operating with $p_s = p_1(p_2)$
$V_{in}(V_{out})$	Number of rejected units found during sampling in the in-control (out-of-control)
RJU	Total number of rejected units during sampling
t_s	Average time of inspecting one unit of the product

Table 2.1(Cont.)

Other variables, constants and indices	
$T_{FA}(T_{TA})$	Average time to search for a false (true) alarm on each machine
T_{MRm}	Average time to perform a minimal repair on machine m, $m \in \{1,2\}$
$CRT_m(PRT_m)$	Average corrective (preventive) restoration time on machine m, $m \in \{1,2\}$
S_t	Total time of sampling in an inspection cycle
TT_{FA}	Total time of searching for false alarms in one inspection cycle
TT_{TA}	Average total time of searching for a true alarm in an inspection cycle
MRT	Total time of minimal repairs in an inspection cycle
RT	Total restoration time in an inspection cycle
C_s	Average inspection cost per unit time
$C_{FA}(C_{TA})$	Average cost per unit time of searching for a false (true) alarm
C_{MR}	Average cost per unit time of performing a minimal repair
$C_{Cm}(C_{Pm})$	Average corrective (preventive) restoration cost per unit time for machine m, $m \in$
C_{LP}	Average lost production cost per one unit of the product
C_{RJ}	Average cost of a rejected unit found during sampling
C_{NC}	Average cost of a nonconforming unit received by a consumer
S_c	Total cost of sampling in an inspection cycle
FA_c	Total cost of searching for false alarms in an inspection cycle
TA_c	Average total cost of searching for a true alarm in an inspection cycle
MR_c	Total cost of minimal repairs in an inspection cycle
$RC_{S_1}(RC_{S_2})$	Average restoration cost if an inspection cycle ends with $S_1(S_2)$
$RC_{S_{12}}$	Average restoration cost if an inspection cycle ends with S_{12}
RC	Total restoration cost in an inspection cycle
LP_c	Lost production cost in an inspection cycle
CRJ	Total cost of rejected units during sampling
CNC	Total cost of nonconforming units received by customers
$\theta_m(\gamma_m)$	Shape (scale) factor of Weibull distribution of machine m, $m \in \{1,2\}$, $\theta_m > 1$
g_m	Production rate of stage m
g_s	Production rate of the system, $\min_{m\in\{1,2\}}\{g_m\}$
$h_m(t)$	Failure rate of machine m, $m \in \{1,2\}$
$M_m(t)$	Expected number of failures of machine m, $m \in \{1,2\}$ in time interval $[0,t]$
MN_m	Number of minimal repairs on machine m, $m \in \{1,2\}$ in an inspection cycle
AV	System's availability
PR_{eff}	Effective production rate
ATS	Average time to signal
$CP(NCP)$	Number of conforming (nonconforming) products produced in one inspection cycle
TP	Total number of products produced in one inspection cycle
CC	Inspection cycle total cost
CT	Inspection cycle total time

4. Model development

4.1. Stochastic cases

Let G be the time at which the inspection cycle terminates due to detecting a shift. The random variable $G \in \{h, 2h, \cdots\cdots, \infty\}$ is the operational time that does not include the stoppage times of inspection, false alarms, minimal repairs, true alarms, and restorations, where the sampling interval h is the time between two successive inspections. Clearly, the shortest length of G is h. Since the production process has competing and propagating shifts, G can be derived based on the following three cases:

- Case I: Machine 2 shift (S_2) and machine 1 shift (S_1) occur in the same sampling interval, i.e., between $(i-1)h^{th}$ and ih^{th} sampling points as shown in Figure 2.2.
- Case II: S_2 is not detected before the occurrence of S_1 given that S_2 occurs between $(i-1)h^{th}$ and ih^{th} sampling points, and S_1 occurs after the ih^{th} sampling point as shown in Figure 2.3.
- Case III: S_2 is detected before the occurrence of S_1 as shown in Figure 2.4.

It is worth pointing out that the above cases also apply when S_1 occurs before S_2.

4.1.1. Case I

Let T_1 and T_2 be the times to shift of machines 1 and 2, respectively, and T_1 and T_2 follow the exponential distributions with rates λ_1 and λ_2, respectively. Moreover, let τ_{S_1} and τ_{S_2} be the times of occurrence of S_1 and S_2, respectively, since the most recent sampling. As shown in Figure 2.2, when $T_1 > T_2$, S_2 is missed because it is followed by S_1 before taking the next sampling. Then, the production process starts to produce units with propagating shift at the time of occurrence of S_1.

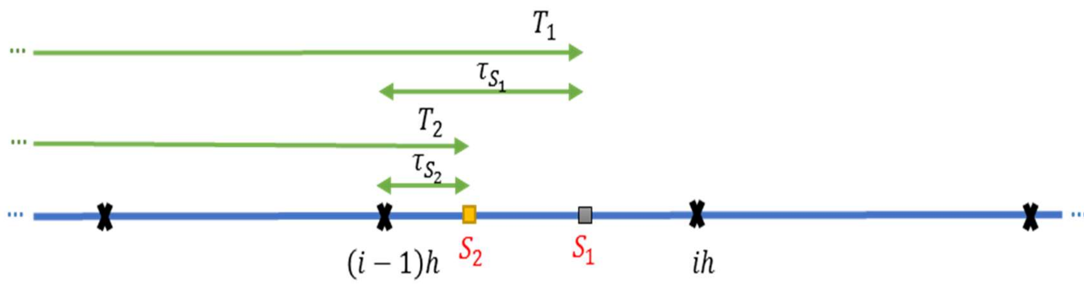

Figure 2.2. Case I, $T_1 > T_2$.

The probability that S_2 and S_1 happen in the same sampling interval given that $T_1 > T_2$ is

$$P\big((i-1)h \le T_2 \le T_1 < ih\big) = \int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_1} \lambda_2 e^{-\lambda_2 t_2}\,\lambda_1 e^{-\lambda_1 t_1} dt_2\, dt_1$$

$$= e^{-\lambda_2 (i-1)h}\big(e^{-\lambda_1 (i-1)h} - e^{-\lambda_1 ih}\big) + \frac{\lambda_1}{\lambda_1+\lambda_2}\big(e^{-(\lambda_1+\lambda_2)ih} - e^{-(\lambda_1+\lambda_2)(i-1)h}\big).$$

Thus, the probability that $G = jh$ given that S_1 and S_2 happen between the $(i-1)^{th}$ and i^{th} sampling points and $T_1 > T_2$ is

$$P\big(G = jh, \text{Case I}_{T_1>T_2}\big) = \sum_{i=1}^{j} P\big((i-1)h \le T_2 \le T_1 < ih\big)\,\beta_{p_3}^{j-i}\big(1-\beta_{p_3}\big), j = 1,\cdots,\infty \quad (2.2)$$

where β_{p_3} is the type II error resulting from that the system is producing units with $p_s = p_3 = p_{11} + p_{12} - p_{11}p_{12}$ according to equation 1. Let d be the number of nonconforming units in the sample, then the type II error $\beta_{p_s\in\{p_1,p_2,p_3\}}$ for $p_s \in \{p_1, p_2, p_3\}$ is given as

$$\beta_{p_s\in\{p_1,p_2,p_3\}} = \sum_{d=0}^{r}\binom{N}{d} p_s^d\,(1-p_s)^{N-d}. \quad (2.3)$$

For instance, in Case I and $T_1 > T_2$, $G = 2h$ if $0 \le T_2 \le T_1 < h$ and a shift is not detected until $j = 2$, or $h \le T_2 \le T_1 < 2h$ and a shift is detected at $j = 2$. Then, the probability that $G = 2h$ is

$$\left\{\left(1-e^{-\lambda_1 h}\right)+\frac{\lambda_1}{\lambda_1+\lambda_2}\left(e^{-(\lambda_1+\lambda_2)h}-1\right)\right\}\beta_{p_3}\left(1-\beta_{p_3}\right)$$

$$+\left\{e^{-\lambda_2 h}\left(e^{-\lambda_1 h}-e^{-\lambda_1 2h}\right)+\frac{\lambda_1}{\lambda_1+\lambda_2}\left(e^{-(\lambda_1+\lambda_2)2h}-e^{-(\lambda_1+\lambda_2)h}\right)\right\}\left(1-\beta_{p_3}\right).$$

The same procedure is followed for $T_2 > T_1$. Hence, $P\left((i-1)h \le T_1 \le T_2 < ih\right)$ is given as

$$P\left((i-1)h \le T_1 \le T_2 < ih\right) = \int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_2} \lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2} dt_1\, dt_2$$

$$= e^{-\lambda_1 (i-1)h}\left(e^{-\lambda_2 (i-1)h}-e^{-\lambda_2 ih}\right)+\frac{\lambda_2}{\lambda_1+\lambda_2}\left(e^{-(\lambda_1+\lambda_2)ih}-e^{-(\lambda_1+\lambda_2)(i-1)h}\right),$$

Thus, the probability that $G = jh$ given that S_1 and S_2 happen between the $(i-1)^{th}$ and i^{th} sampling points and $T_2 > T_1$ is

$$P\left(G = jh, \text{Case I}_{T_2>T_1}\right) = \sum_{i=1}^{j} P\left((i-1)h \le T_1 \le T_2 < ih\right)\beta_{p_3}^{j-i}\left(1-\beta_{p_3}\right), j = 1, \cdots, \infty \quad (2.4)$$

4.1.2. Case II

For this case, as shown in Figure 2.3, S_1 occurs at least one sample after the occurrence of S_2. Due to the type II error, S_2 is always undetected until after the occurrence of S_1. The minimum value of G is $2h$ as a result that S_2 happens before taking the first sample (i.e., before time h) but is not detected, S_1 occurs afterwards, and the total shift is detected at time $2h$.

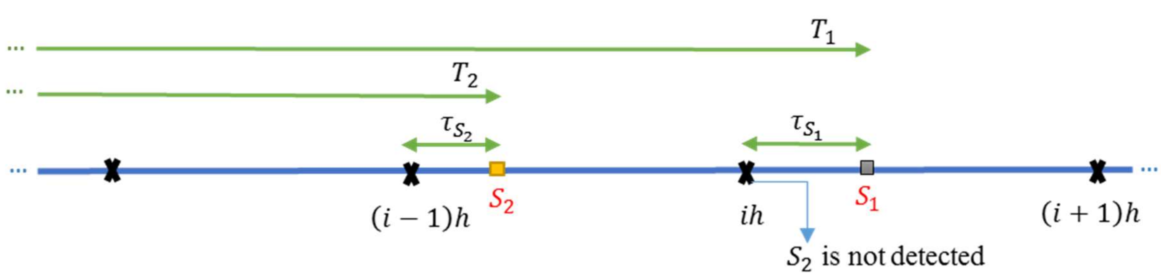

Figure 2.3. Case II, $T_1 > T_2$.

Note that Figure 2.3 only shows that S_1 occurs after one sample after the occurrence of S_2.

However, in this case, S_1 can occur more than one sample after S_2.

The probability that $G = jh$ in Case II and $T_1 > T_2$ is

$$P(G = jh, \text{Case II}_{T_1>T_2}) =$$

$$\sum_{i=1}^{j-1} \sum_{k=0}^{j-i-1} \left(e^{-\lambda_2 (i-1)h} - e^{-\lambda_2 ih}\right)\left(e^{-\lambda_1 (k+i)h} - e^{-\lambda_1 (k+1+i)h}\right)\beta_{p_2}^{k+1}\beta_{p_3}^{j-i-k-1}\left(1 - \beta_{p_3}\right), \tag{2.5}$$

$$j = 2, \dots, \infty,$$

where β_{p_2} is the type II (obtained by equation 2.3) that could result if the system is producing units with $p_s = p_2 = p_{01} + p_{12} - p_{01}p_{12}$. For instance, $P\left(G = h, \text{Case II}_{T_1>T_2}\right) = 0$, and $P\big(G = 2h$, Case $\text{II}_{T_1>T_2}\big) = \left(1 - e^{-\lambda_2 h}\right)\left(e^{-\lambda_1 h} - e^{-\lambda_1 2h}\right)\beta_{p_2}\left(1 - \beta_{p_3}\right)$, and so on.

The same procedure can be followed for $T_2 > T_1$, and $P(G = jh, \text{Case II}_{T_2>T_1})$ is obtained as

$$P(G = jh, \text{Case II}_{T_2>T_1}) =$$

$$\sum_{i=1}^{j-1} \sum_{k=0}^{j-i-1} \left(e^{-\lambda_1 (i-1)h} - e^{-\lambda_1 ih}\right)\left(e^{-\lambda_2 (k+i)h} - e^{-\lambda_2 (k+1+i)h}\right)\beta_{p_1}^{k+1}\beta_{p_3}^{j-i-k-1}\left(1 - \beta_{p_3}\right), \tag{2.6}$$

$$j = 2, \dots, \infty,$$

where β_{p_1} is the type II error (obtained by equation 2.3) that could result if the system is producing units with $p_s = p_1 = p_{11} + p_{02} - p_{11}p_{02}$.

4.1.3. Case III

In this case, as shown in Figure 2.4, S_2 is always detected before the occurrence of S_1, and the probability that $G = jh$ given Case III and $T_1 > T_2$ can be expressed as

$$P\left(G = jh, \text{Case III}_{T_1>T_2}\right) = e^{-\lambda_1 jh} \sum_{i=1}^{j} \left(e^{-\lambda_2 (i-1)h} - e^{-\lambda_2 ih}\right) \beta_{p_2}^{j-i}\left(1 - \beta_{p_2}\right), \tag{2.7}$$

$$j = 1, \cdots, \infty.$$

For example, $P(G = h, \text{Case III}_{T_1>T_2}) = e^{-\lambda_1 h}(1-e^{-\lambda_2 h})(1-\beta_{p_2})$, and $P(G = 2h, \text{Case III}_{T_1>T_2}) = e^{-\lambda_1 2h}\{(1-e^{-\lambda_2 h})\beta_{p_2}(1-\beta_{p_2}) + (e^{-\lambda_2 h} - e^{-\lambda_2 2h})(1-\beta_{p_2})\}$, and so on.

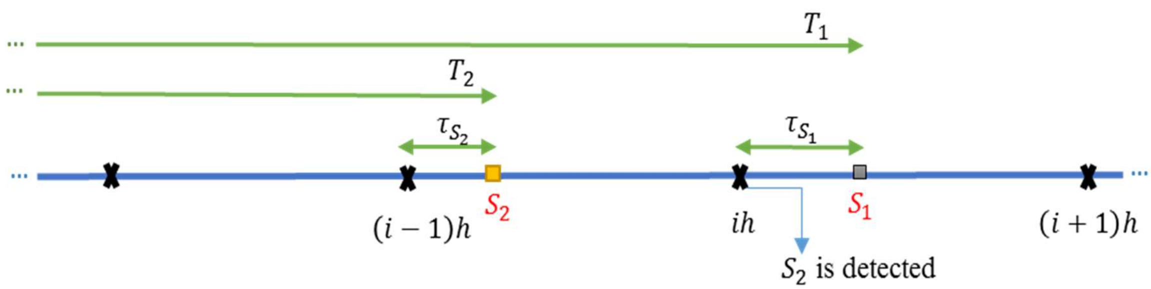

Figure 2.4. Case III, $\mathrm{T}_1 > \mathrm{T}_2$.

Similarly, when $T_2 > T_1$, $P(G = jh, \text{Case III}_{T_2>T_1})$ can be obtained as

$$P(G = jh, \text{Case III}_{T_2>T_1}) = e^{-\lambda_2 jh} \sum_{i=1}^{j} (e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih})\, \beta_{p_1}^{j-i}(1-\beta_{p_1}), \quad j = 1, \cdots, \infty. \tag{2.8}$$

4.1.4. The operational time

According to the cases illustrated above, the expected value of the operational time $E[G]$ can be given as

$$E[G] = A_1 + A_2 + A_3 + A_4 + A_5 + A_6, \tag{2.9}$$

where A_1 to A_6 are the weighted expected values of the cycle length given all cases. A_1 to A_6 are obtained as follows, respectively:

$A_1 = \sum_{j=1}^{\infty} jh \cdot P(G = jh, \text{Case I}_{T_1>T_2}) = \frac{h(e^{(\lambda_1+\lambda_2)h} - \beta_{p_3})(\lambda_2 e^{\lambda_2 h}(e^{\lambda_1 h}-1) - \lambda_1(e^{\lambda_2 h}-1))}{(\lambda_1+\lambda_2)(1-\beta_{p_3})(e^{(\lambda_1+\lambda_2)h}-1)^2}$,

$A_2 = \sum_{j=1}^{\infty} jh \cdot P(G = jh, \text{Case I}_{T_2>T_1}) = \frac{h(e^{(\lambda_1+\lambda_2)h} - \beta_{p_3})(\lambda_1 e^{\lambda_1 h}(e^{\lambda_2 h}-1) - \lambda_2(e^{\lambda_1 h}-1))}{(\lambda_1+\lambda_2)(1-\beta_{p_3})(e^{(\lambda_1+\lambda_2)h}-1)^2}$,

$A_3 = \sum_{j=2}^{\infty} jh \cdot P\big(G = jh, \text{Case II}_{T_1>T_2}\big) =$

$\frac{h\beta_{p_2}(e^{\lambda_1 h}-1)(e^{\lambda_2 h}-1)(e^{\lambda_1 h}+(\beta_{p_3}-2)e^{(2\lambda_1+\lambda_2)h}+\beta_{p_2}(e^{(\lambda_1+\lambda_2)h}-\beta_{p_3}))}{(\beta_{p_3}-1)(e^{(\lambda_1+\lambda_2)h}-1)^2(e^{\lambda_1 h}-\beta_{p_2})^2}$,

$A_4 = \sum_{j=2}^{\infty} jh \cdot P\big(G = jh, \text{Case II}_{T_2>T_1}\big) =$

$\frac{h\beta_{p_1}(e^{\lambda_1 h}-1)(e^{\lambda_2 h}-1)(e^{\lambda_2 h}+(\beta_{p_3}-2)e^{(\lambda_1+2\lambda_2)h}+\beta_{p_1}(e^{(\lambda_1+\lambda_2)h}-\beta_{p_3}))}{(\beta_{p_3}-1)(e^{(\lambda_1+\lambda_2)h}-1)^2(e^{\lambda_2 h}-\beta_{p_1})^2}$,

$A_5 = \sum_{j=1}^{\infty} jh \cdot P(G = jh, \text{Case III}_{T_1>T_2}) = \frac{h(\beta_{p_2}-1)e^{\lambda_1 h}(e^{\lambda_2 h}-1)(\beta_{p_2}-e^{(2\lambda_1+\lambda_2)h})}{(e^{(\lambda_1+\lambda_2)h}-1)^2(e^{\lambda_1 h}-\beta_{p_2})^2}$,

$A_6 = \sum_{j=1}^{\infty} jh \cdot P\big(G = jh, \text{Case III}_{T_2>T_1}\big) = \frac{h(\beta_{p_1}-1)e^{\lambda_2 h}(e^{\lambda_1 h}-1)(\beta_{p_1}-e^{(\lambda_1+2\lambda_2)h})}{(e^{(\lambda_1+\lambda_2)h}-1)^2(e^{\lambda_2 h}-\beta_{p_1})^2}$.

4.2. Time and cost of sampling

The average number of samples taken during the inspection cycle equals to $E[G]/h$. Then, the expected time of sampling $E[S_t]$ can be expressed as

$$E[S_t] = \frac{t_s \cdot N \cdot E[G]}{h}, \tag{2.10}$$

where t_s is the average time of inspecting one unit of the product. Let C_s be the average cost per unit time of sampling, then the expected cost of sampling $E[S_c]$ is

$$E[S_c] = C_s\, E[S_t]. \tag{2.11}$$

4.3. Time and cost of false alarms

The process is out-of-control once any of the two shifts occurs. Consequently, the time period that the process is in-control T_{in} follows the exponential distribution with $T_{in} = \text{Min}(T_1, T_2) \sim \text{Exp}(\lambda_1 + \lambda_2)$. Therefore, the expected time that the process is in-control $E[T_{in}]$ is

$$E[T_{in}] = \frac{1}{\lambda_1 + \lambda_2}.$$

Let Q_{in} be the number of samples taken while the system is in-control. Then the expected value $E[Q_{in}]$ can be calculated as

$$E[Q_{in}] = \sum_{i=0}^{\infty} i \cdot \left(e^{-(\lambda_1+\lambda_2)ih} - e^{-(\lambda_1+\lambda_2)(i+1)h}\right) = \frac{1}{e^{(\lambda_1+\lambda_2)h} - 1}.$$

As a result, the expected total time of false alarms $E[TT_{FA}]$ is given as

$$E[TT_{FA}] = 2\, T_{FA} \frac{E[Q_{in}]}{ARL_0}, \tag{2.12}$$

where T_{FA} is the average time to search for a false alarm on each machine , ARL_0 is the average run length when the process is in-control (i.e., the average number of samples taken until a false alarm is alerted), and $E[Q_{in}]/ARL_0$ is the average number of false alarms in one cycle. ARL_0 is given as (Montgomery, 2009)

$$ARL_0 = \frac{1}{\alpha},$$

where the type I error α is reported when $p_s = p_0 = p_{01} + p_{02} - p_{01}p_{02}$, and it is given by

$$\alpha = 1 - \sum_{d=0}^{r} \binom{N}{d} p_0^d\, (1 - p_0)^{N-d}.$$

The direct cost of false alarms is due to the effort taken for identifying false alarms and inspecting machines. Let C_{FA} be the average cost per unit time of searching for a false alarm. Then, the expected total cost of searching for false alarms $E[FA_c]$ can be simply expressed as

$$E[FA_c] = C_{FA}\, E[TT_{FA}]. \tag{2.13}$$

4.4. Time and cost of searching for a true alarm

Let C_{TA} be the average cost per unit time of searching for a true alarm, then the average total time TT_{TA} and cost TA_c of searching for a true alarm are given as follows, respectively:

$$TT_{TA} = 2\, T_{TA}, \tag{2.14}$$

$$TA_c = C_{TA}\ TT_{TA}. \tag{2.15}$$

4.5. Restoration time and cost

Restoration time is the time required for machine maintenance and shift removal(s). Since inspection ends with a shift, at least one of the two machines needs corrective restoration. Three possible scenarios are described next.

- ***Inspection cycle ends only with S_1***

For this scenario, machine 1 is correctively restored, and machine 2 is preventively restored. The probability that the inspection cycle ends with this scenario equals the probability that S_1 is detected before the occurrence of S_2. Let CRT_1 and PRT_2 be the average corrective restoration time of machine 1 and the average preventive restoration time of machine 2, respectively, and C_{C1} and C_{P2} be the average costs per unit time of corrective and preventive restorations on machines 1 and 2, respectively. Then, the average restoration cost of this scenario RC_{S_1} is

$$RC_{S_1} = C_{C1}\, CRT_1 + C_{P2}\, PRT_2.$$

- ***Inspection cycle ends only with S_2***

In this scenario, machine 2 is correctively restored, and machine 1 is preventively restored. The probability that the inspection cycle ends in this scenario is the probability that S_2 is detected before the occurrence of S_1. Let PRT_1 and CRT_2 be the average preventive restoration time of machine 1 and the average corrective restoration time of machine 2, respectively, and C_{C2} and C_{P1} be the average costs per unit time of corrective and preventive restorations on machines 2 and 1, respectively. Then, the average restoration cost of this scenario RC_{S_2} is

$$RC_{S_2} = C_{P1}\, PRT_1 + C_{C2}\, CRT_2.$$

- ***Inspection cycle ends with propagating shift S_{12}***

In this scenario, both machines have shifted, and corrective restorations are carried out on both machines. The average cost of restoration of this scenario $RC_{S_{12}}$ is given as

$$RC_{S_{12}} = C_{C1}\, CRT_1 + C_{C2}\, CRT_2.$$

Hence, the expected total restoration cost $E[RC]$ and time $E[RT]$ are given as follows, respectively:

$$E[RC] = RC_{S_1}\, B_6 + RC_{S_2}\, B_5 + RC_{S_{12}}\, B, \tag{2.16}$$

$$E[RT] = (CRT_1 + PRT_2)\, B_6 + (PRT_1 + CRT_2)\, B_5 + (CRT_1 + CRT_2)\, B, \tag{2.17}$$

where $B_1(B_2)$ is the probability of Case I given $T_1 > T_2(T_2 > T_1)$, $B_3(B_4)$ is the probability of Case II given $T_1 > T_2(T_2 > T_1)$, and $B_5(B_6)$ is the probability of Case III given $T_1 > T_2(T_2 > T_1)$. B, and B_1 to B_6 are given as follows, respectively:

$B = B_1 + B_2 + B_3 + B_4$,

$B_1 = \sum_{j=1}^{\infty} P\big(G = jh, \text{Case I}_{T_1>T_2}\big) = \frac{\lambda_1(1-e^{\lambda_2 h})+\lambda_2(e^{(\lambda_1+\lambda_2)h}-e^{\lambda_2 h})}{(\lambda_1+\lambda_2)(e^{(\lambda_1+\lambda_2)h}-1)}$,

$B_2 = \sum_{j=1}^{\infty} P\big(G = jh, \text{Case I}_{T_2>T_1}\big) = \frac{\lambda_2(1-e^{\lambda_1 h})+\lambda_1(e^{(\lambda_1+\lambda_2)h}-e^{\lambda_1 h})}{(\lambda_1+\lambda_2)(e^{(\lambda_1+\lambda_2)h}-1)}$,

$B_3 = \sum_{j=2}^{\infty} P\big(G = jh, \text{Case II}_{T_1>T_2}\big) = \frac{\beta_{p_2}(e^{\lambda_1 h}-1)(e^{\lambda_2 h}-1)}{(e^{(\lambda_1+\lambda_2)h}-1)(e^{\lambda_1 h}-\beta_{p_2})}$,

$B_4 = \sum_{j=2}^{\infty} P\big(G = jh, \text{Case II}_{T_2>T_1}\big) = \frac{\beta_{p_1}(e^{\lambda_2 h}-1)(e^{\lambda_1 h}-1)}{(e^{(\lambda_1+\lambda_2)h}-1)(e^{\lambda_2 h}-\beta_{p_1})}$,

$B_5 = \sum_{j=1}^{\infty} P\big(G = jh, \text{Case III}_{T_1>T_2}\big) = \frac{e^{\lambda_1 h}(e^{\lambda_2 h}-1)(1-\beta_{p_2})}{(e^{(\lambda_1+\lambda_2)h}-1)(e^{\lambda_1 h}-\beta_{p_2})}$,

$B_6 = \sum_{j=2}^{\infty} P\big(G = jh, \text{Case III}_{T_2>T_1}\big) = \frac{e^{\lambda_2 h}(e^{\lambda_1 h}-1)(1-\beta_{p_1})}{(e^{(\lambda_1+\lambda_2)h}-1)(e^{\lambda_2 h}-\beta_{p_1})}$.

4.6. Time and cost of minimal repairs

Minimal repair is performed each time a machine fails unless a shift is detected. By nature, minimal repair does not change the failure rate of a failed machine. The failure rate $h_m(t)$ of machine m is given as

$$h_m(t) = \frac{\theta_m}{\gamma_m}\left(\frac{t}{\gamma_m}\right)^{\theta_m - 1},$$

where $\theta_m > 1$ and γ_m are the corresponding shape and scale parameters of the Weibull distribution, respectively. Then, the expected number of failures (i.e., minimal repairs) $M_m(t)$ of machine m during the interval $[0, t]$ can be obtained as

$$M_m(t) = \int_0^t h_m(u)du = \left(\frac{t}{\gamma_m}\right)^{\theta_m}.$$

Since machines do not age during downtime, the expected number of minimal repairs on machine m in each inspection cycle $E[MN_m]$ can be expressed as

$$E[MN_m] = \sum_{j=1}^{\infty} \left(\frac{jh}{\gamma_m}\right)^{\theta_m} P(G = jh), \tag{2.18}$$

where

$$P(G = jh) = P\left(G = jh, \text{Case I}_{T_1>T_2}\right) + P\left(G = jh, \text{Case I}_{T_2>T_1}\right) + P\left(G = jh, \text{Case II}_{T_1>T_2}\right) + P\left(G = jh, \text{Case II}_{T_2>T_1}\right) + P\left(G = jh, \text{Case III}_{T_1>T_2}\right) + P\left(G = jh, \text{Case III}_{T_2>T_1}\right).$$

Since the purpose of minimal repair is to make a failed machine operational again with minimal resources, the PON of the system will be the same as that right before the failure. Let T_{MRm} and $C_{MRm}, m \in \{1,2\}$ be the average time and cost per unit time to perform a minimal repair on machine m, respectively. Then, the expected total time $E[MRT]$ and the expected total cost of performing minimal repairs $E[MR_c]$ are given as follows, respectively:

$$E[MRT] = T_{MR1}E[MN_1] + T_{MR2}E[MN_2], \tag{2.19}$$

$$E[MR_c] = C_{MR1}T_{MR1}E[MN_1] + C_{MR2}T_{MR2}E[MN_2]. \tag{2.20}$$

4.7. Cost of lost production

The time due to stoppages for searching for false alarms and true alarms, sampling, minimal repairs, and restoration causes loss in production. Let C_{LP} be the average cost of lost production per one unit of the product, then the expected cost of lost production $E[LP_c]$ can be expressed as

$$E[LP_c] = C_{LP}g_s\{E[TT_{FA}] + TT_{TA} + E[S_t] + E[MRT] + E[RT]\}, \tag{2.21}$$

where g_s is the system's production rate, and it is given by

$$g_s = \min_{m\in\{1,2\}}\{g_m\},$$

where g_m is the production rate of machine m.

4.8. Cost of units rejected in all samples

Any nonconforming unit found in a sample is rejected without replacement, and the production process at each sampling time should be in one of the following states: in-control state and three out-of-control states. To find the cost of rejected units in all samples, we first define the following quantities:

$$a_{p_s} = \sum_{d=r+1}^{N} d\binom{N}{d}p_s^d\,(1-p_s)^{N-d}, p_s \in \{p_0, p_1, p_2, p_3\},$$

$$b_{p_s} = \sum_{d=0}^{r} d\binom{N}{d}p_s^d\,(1-p_s)^{N-d}, p_s \in \{p_0, p_1, p_2, p_3\},$$

where a_{p_s} represents the expected number of nonconforming units found in a sample if a false or a true alarm is alerted, whereas b_{p_s} refers to the expected number of nonconforming units found in a sample taken if no alarm is alerted. For instance, $a_{p_1}$is the expected number of nonconforming units found in the last sample that alerts the true alarm when the process is operating with S_1 ,whereas b_{p_0} is the expected number of nonconforming units found in a sample taken while the process is in control and no false alarm is alerted.

Any sample taken in the in-control period may indicate no alarm or false alarm, and the expected number of samples with false alarms equals to the expected number of false alarms. Then, the expected number of rejected units found during inspection when the process is in-control $E[V_{in}]$ is

$$E[V_{in}] = \alpha E[Q_{in}]a_{p_0} + (1-\alpha)E[Q_{in}]b_{p_0}.$$

The expected number of rejected units found during inspection when the process is out-of-control $E[V_{out}]$ is derived as follows.

In Case I, units are produced with $p_s = p_3$ .The expected number of samples taken until a true alarm is alerted is $ARL_{s_{12}}$ where $ARL_{s_{12}}$ is the average run length when the process is operating with S_{12}, and it is given as (Montgomery, 2009):

$$ARL_{s_{12}} = \frac{1}{1-\beta_{p_3}}.$$

The average length in the out-of-control state is defined as the average number of samples taken since the occurrence of a shift until a true alarm is alerted.

The last sample which alerts the true signal has $r < d \leq N$. Hence, the expected number of rejected units found during sampling when the process is out-of-control given Case I $E[V_{out}|\text{Case I}]$ is expressed as

$$E[V_{out}|\text{Case I}] = \{a_{p_3} + (ARL_{s_{12}} - 1)b_{p_3}\},$$

where $ARL_{s_{12}} - 1$ are the samples that don't alert a true alarm if S_{12} occurs

In Cases II & III, at least one sample is taken with $p_s = p_2$ if $T_1 > T_2$, or with $p_s = p_1$ if $T_2 > T_1$. Let Q_{p_2} and Q_{p_1} be the number of samples taken with $p_s = p_2$, and $p_s = p_1$, respectively. Then $E[Q_{p_2}|\text{Case II}_{T_1>T_2}]$ and $E[Q_{p_1}|\text{Case II}_{T_2>T_1}]$ are given as follows, respectively:

$$E[Q_{p_2}|\text{Case II}_{T_1>T_2}] = \frac{\sum_{q=1}^{\infty}\sum_{i=1}^{\infty} q\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)\left(e^{-\lambda_1(i+q-1)h} - e^{-\lambda_1(i+q)h}\right)\beta_{p_2}^{q}}{\sum_{q=1}^{\infty}\sum_{i=1}^{\infty} \left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)\left(e^{-\lambda_1(i+q-1)h} - e^{-\lambda_1(i+q)h}\right)\beta_{p_2}^{q}} = \frac{e^{\lambda_1 h}}{(e^{\lambda_1 h} - \beta_{p_2})},$$

$$E[Q_{p_1}|\text{Case II}_{T_2>T_1}] = \frac{\sum_{q=1}^{\infty}\sum_{i=1}^{\infty} q\left(e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih}\right)\left(e^{-\lambda_2(i+q-1)h} - e^{-\lambda_2(i+q)h}\right)\beta_{p_1}^{q}}{\sum_{q=1}^{\infty}\sum_{i=1}^{\infty} \left(e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih}\right)\left(e^{-\lambda_2(i+q-1)h} - e^{-\lambda_2(i+q)h}\right)\beta_{p_1}^{q}} = \frac{e^{\lambda_2 h}}{(e^{\lambda_2 h} - \beta_{p_1})},$$

where q denotes the number of samples taken between the occurrence times of S_1 and S_2. In Case

III, $S_2(S_1)$is always detected before the occurrence of $S_1(S_2)$, and hence, $E\left[Q_{p_2}\middle|\text{Case III}_{T_1>T_2}\right]$ and $E\left[Q_{p_1}\middle|\text{Case III}_{T_2>T_1}\right]$ are given as follows, respectively:

$$E\left[Q_{p_2}\middle|\text{Case III}_{T_1>T_2}\right] = \frac{\sum_{w=1}^{\infty}\sum_{i=1}^{\infty} w\left(e^{-\lambda_2(i-1)h}-e^{-\lambda_2 ih}\right)e^{-\lambda_1(i+w-1)h}\beta_{p_2}^{w-1}(1-\beta_{p_2})}{\sum_{w=1}^{\infty}\sum_{i=1}^{\infty}\left(e^{-\lambda_2(i-1)h}-e^{-\lambda_2 ih}\right)e^{-\lambda_1(i+w-1)h}\beta_{p_2}^{w-1}(1-\beta_{p_2})} = \frac{e^{\lambda_1 h}}{(e^{\lambda_1 h}-\beta_{p_2})},$$

$$E\left[Q_{p_1}\middle|\text{Case III}_{T_2>T_1}\right] = \frac{\sum_{w=1}^{\infty}\sum_{i=1}^{\infty} w\left(e^{-\lambda_1(i-1)h}-e^{-\lambda_1 ih}\right)e^{-\lambda_2(i+w-1)h}\beta_{p_1}^{w-1}(1-\beta_{p_1})}{\sum_{w=1}^{\infty}\sum_{i=1}^{\infty}\left(e^{-\lambda_1(i-1)h}-e^{-\lambda_1 ih}\right)e^{-\lambda_2(i+w-1)h}\beta_{p_1}^{w-1}(1-\beta_{p_1})} = \frac{e^{\lambda_2 h}}{(e^{\lambda_2 h}-\beta_{p_1})},$$

where w represents the number of samples that process undergoes with S_2 until a successful detection. The term $e^{-\lambda_1(i+w-1)h}$ indicates that S_2 is detected at the sampling time $(i+w-1)h$, at which, S_1 still has not occurred yet.

Consequently, the expected number of rejected units during the inspection when the process is out-of-control $E[V_{out}]$ can be obtained as

$$\begin{aligned}E[V_{out}] = {} & E[V_{out}|\text{Case I}]\{B_1+B_2\} + E\left[Q_{p_2}\middle|\text{Case II}_{T_1>T_2}\right]b_{p_2}B_3 + E\left[Q_{p_1}\middle|\text{Case II}_{T_2>T_1}\right]b_{p_1}B_4 \\ & + \left\{\left(ARL_{s_{12}}-1\right)b_{p_3}+a_{p_3}\right\}\{B_3+B_4\} + \left(E\left[Q_{p_2}\middle|\text{Case III}_{T_1>T_2}\right]-1\right)b_{p_2}B_5 + \\ & \left(E\left[Q_{p_1}\middle|\text{Case III}_{T_2>T_1}\right]-1\right)b_{p_1}B_6 + a_{p_2}B_5 + a_{p_1}B_6.\end{aligned}$$

In the above equation, $E\left[Q_{p_2}\middle|\text{Case II}_{T_1>T_2}\right]$ $\left(E\left[Q_{p_1}\middle|\text{Case II}_{T_2>T_1}\right]\right)$ is the expected number of samples that don't alert a true alarm in Case II when a process operates with $S_2(S_1)$, $\left(ARL_{s_{12}}-1\right)$ is the average number of samples that don't alert a true alarm when the process operates with S_{12} in Case II, and a_{p_3}represents the expected number of rejected units in the last sample that alert a true alarm given Case II. In Case III, $E\left[Q_{p_2}\middle|\text{Case III}_{T_1>T_2}\right]-1\left(E\left[Q_{p_1}\middle|\text{Case III}_{T_2>T_1}\right]-1\right)$ is the expected number of samples that don't alert a true alarm when the process operates with $S_2(S_1)$, and $a_{p_2}\left(a_{p_1}\right)$ is the average number of rejected units found in the last sample that detects $S_2(S_1)$.

Accordingly, the expected total number and cost of rejected units during inspection $E[V]$ and $E[CRJ]$, are given as follows, respectively:

$$E[V] = E[V_{in}] + E[V_{out}], \tag{2.22}$$

$$E[CRJ] = C_{RJ}E[V], \tag{2.23}$$

where C_{RJ} denotes the average cost of a rejected unit.

4.9. Cost of nonconforming units delivered to customers

A nonconforming unit found by a customer may cost more than a nonconforming unit found during the inspection. Let C_{NC} be the average cost of a nonconforming unit received by a customer, then the expected cost of nonconforming units received by customers $E[CNC]$ is given by

$$E[CNC] = C_{NC}\{g_s(p_0E[T_{in}] + p_1E[T_{s_1}] + p_2E[T_{s_2}] + p_3E[T_{s_{12}}]) - E[V]\}, \tag{2.24}$$

where $E[T_{s_1}]$, $E[T_{s_2}]$, and $E[T_{s_{12}}]$ are the expected values of times that the process could operate with S_1, S_2, and S_{12}, respectively. The details of these terms are given in Section 5.

4.10. Expected total cycle cost and time

Based on the above calculations, the expected total cycle cost $E[CC]$ and the expected total cycle time $E[CT]$ can be obtained as follows, respectively:

$$E[CC] = E[S_c] + E[FA_c] + TA_c + E[RC] + E[MR_c] + E[LP_c] + E[CRJ] + E[CNC], \tag{2.25}$$

$$E[CT] = E[G] + E[S_t] + E[TT_{FA}] + TT_{TA} + E[RT] + E[MRT]. \tag{2.26}$$

5. Optimal design of the sampling plan

5.1. Mathematical formulation

The optimal sampling parameters are determined by minimizing the long-run cost rate $LRCR = E[CC]/E[CT]$, which is the ratio between the expected total cycle cost and the expected total cycle time. The mathematical formulation of the problem is given as follows.

$$\min_{N,r,h} \qquad LRCR = \frac{E[CC]}{E[CT]} \tag{2.27}$$

Subject to $AV \geq A$ (2.27.1)

$PR_{eff} \geq W$ (2.27.2)

$ATS \leq L$ (2.27.3)

$N \leq (h - u_l)g_s, \quad l \in \{1,4,5,6\}$ (2.27.4)

$N > r$ (2.27.5)

$N, r \in$ integers, $r \geq 0, h > 0.$ (2.27.6)

The formulation belongs to a Mixed Integer Nonlinear Programming (MINLP) problem. Equation (2.27) states that $LRCR$ is minimized with respect to the three decision variables $N, r,$ and h. Equations (2.27.1) - (2.27.3) specify three performance constraints. In equation (2.27.1), the system availability AV must be greater than or equal to a predefined threshold A in order to increase the expected total number of units produced in one inspection cycle. However, with increased availability, both the expected numbers of conforming and nonconforming units increase. Since the latter is undesirable, equation (2.27.2) imposes another constraint on the effective production rate PR_{eff} to ensure that the fraction of expected number of conforming units produced is above a certain level W. Moreover, equation (2.27.3) is used to ensure the speed of detecting process shifts in terms of the average time to signal ATS. ATS is defined as the average time taken to alert a true alarm since the occurrence of a shift. In practice, ATS could be short to avoid excess losses when producing products in the out-of-control state (i.e., ATS should be less than or equal to a threshold L).

Inspection at each sampling time is carried out from the last unit produced, and a group of constraints given by equation (2.27.4) is provided to ensure that units are sampled from only one population (i.e., with the same p_s). These constraints also guarantee that N is always less than the number of units produced between two inspections. Note that because $u_1 > u_2$ when $T_1 > T_2$, we have $h - u_1 < h - u_2$. Moreover, because $u_4 > u_3$ when $T_2 > T_1$, we have $h - u_4 < h - u_3$ ($u_1$to u_6 are defined below). Therefore, the constraints corresponding to $l \in \{2,3\}$ are redundant. Lastly, the decision variables r and $N(> r)$ are integers where $r \geq 0$, and h is a positive continuous variable as specified in equations (2.27.5) and (2.27.6), respectively.

Since the three performance measures are essential to the operation of this system, they will be elaborated next.

5.2. System's availability

The system's availability AV is defined as:

$$AV = \frac{E[G]}{E[CT]}, \tag{2.28}$$

which is the ratio between the expected operational time in a cycle and the expected total cycle length.

5.3. Effective production rate

The effective production rate PR_{eff} is the fraction of the expected numbers of conforming units produced $E[CP]$ in the inspection cycle. PR_{eff} can be obtained as

$$PR_{eff} = \frac{E[CP]}{E[TP]} = 1 - \frac{E[NCP]}{E[TP]},$$

where $E[TP]$ and $E[NCP]$ are the expected total number and the expected number of nonconforming units produced in one cycle, respectively. $E[NCP]$ is the sum of the number of nonconforming units produced in the in-control and the other three out-of-control states. Since each state has a different p_s, $E[NCP]$ and $E[TP]$ are given as follows, respectively:

$$E[NCP] = g_s\{p_0 E[T_{in}] + p_1 E[T_{s_1}] + p_2 E[T_{s_2}] + p_3 E[T_{s_{12}}]\},$$

$$E[TP] = g_s E[G].$$

Therefore, PR_{eff} is

$$PR_{eff} = 1 - \frac{\{p_0 E[T_{in}] + p_1 E[T_{s_1}] + p_2 E[T_{s_2}] + p_3 E[T_{s_{12}}]\}}{E[G]}, \quad (2.29)$$

where $E[T_{in}] = \frac{1}{\lambda_1+\lambda_2}$, $E[T_{s_1}]$, $E[T_{s_2}]$, and $E[T_{s_{12}}]$ are obtained as follows.

5.3.1. $E[T_{s_1}]$, $E[T_{s_2}]$, and $E[T_{s_{12}}]$

Let us first define the followings:

$u_1(u_3)$ represent the conditional expectation of τ_{S_1} given Case I, $T_1 > T_2(T_2 > T_1)$.

$u_2(u_4)$ represent the conditional expectation of τ_{S_2} given Case I, $T_1 > T_2(T_2 > T_1)$.

$u_5(u_6)$ represent the conditional expectation of $\tau_{S_1}(\tau_{S_2})$ given Case II/III.

$C_1(C_2)$ represent the corresponding probability of Case I, $T_1 > T_2(T_2 > T_1)$,

Case I. Given that S_2 and S_1 occur in the same sampling interval as shown in Figure 2.2, u_1, u_2 are given as follows:

$$u_1 = E\left[\tau_{S_1}\middle|(i-1)h \le T_2 \le T_1 < ih\right] = \frac{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_1}(t_1-(i-1)h)\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_2 dt_1}{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_1}\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_2 dt_1}$$

$$= \frac{\lambda_2^2 e^{\lambda_2 h}(e^{\lambda_1 h}-1) - \lambda_1^3 h(e^{\lambda_2 h}-1) - \lambda_1\lambda_2 e^{\lambda_2 h}(2-2e^{\lambda_1 h}+\lambda_2 h) + \lambda_1^2(1+\lambda_2 h - e^{\lambda_2 h}(1+2\lambda_2 h))}{\lambda_1(\lambda_1+\lambda_2)(\lambda_1 - e^{\lambda_2 h}(\lambda_1+\lambda_2-\lambda_2 e^{\lambda_1 h}))},$$

$$u_2 = E\left[\tau_{S_2}\middle|(i-1)h \le T_2 \le T_1 < ih\right] = \frac{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_1}(t_2-(i-1)h)\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_2 dt_1}{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_1}\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_2 dt_1}$$

$$= \frac{e^{\lambda_2 h}(\lambda_2^2 e^{\lambda_1 h} - (\lambda_1+\lambda_2)^2) + \lambda_1(\lambda_1+2\lambda_2+\lambda_2(\lambda_1+\lambda_2)h)}{\lambda_2(\lambda_1+\lambda_2)(\lambda_1 - e^{\lambda_2 h}(\lambda_1+\lambda_2-\lambda_2 e^{\lambda_1 h}))}.$$

Since S_2 occurs before S_1 in the same sampling interval, S_2 propagates to S_{12} at the time of S_1 occurrence and prior to the next sampling time. Therefore,

$$E\left[T_{s_2}\middle|\text{Case I}_{T_1>T_2}\right] = u_1 - u_2,$$

$$E\left[T_{s_1}\middle|\text{Case I}_{T_1>T_2}\right] = 0,$$

$$E\left[T_{s_{12}}\middle|\text{Case I}_{T_1>T_2}\right] = hARL_{s_{12}} - u_1,$$

If $T_2 > T_1$, then u_3 and u_4 are given as follows, respectively:

$$u_3 = E\left[\tau_{S_1} \middle| (i-1)h \leq T_1 \leq T_2 < ih\right] = \frac{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_2}(t_1-(i-1)h)\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_1 dt_2}{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_2}\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_1 dt_2}$$

$$= \frac{e^{\lambda_1 h}({\lambda_1}^2 e^{\lambda_2 h}-(\lambda_1+\lambda_2)^2)+\lambda_2(\lambda_2+2\lambda_1+\lambda_1(\lambda_1+\lambda_2)h)}{\lambda_1(\lambda_1+\lambda_2)(\lambda_2-e^{\lambda_1 h}(\lambda_1+\lambda_2-\lambda_1 e^{\lambda_2 h}))},$$

$$u_4 = E\left[\tau_{S_2} \middle| (i-1)h \leq T_1 \leq T_2 < ih\right] = \frac{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_2}(t_2-(i-1)h)\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_1 dt_2}{\int_{(i-1)h}^{ih}\int_{(i-1)h}^{t_2}\lambda_2 e^{-\lambda_2 t_2}\lambda_1 e^{-\lambda_1 t_1}dt_1 dt_2}$$

$$= \frac{{\lambda_1}^2 e^{\lambda_1 h}(e^{\lambda_2 h}-1)-{\lambda_2}^3 h(e^{\lambda_1 h}-1)-\lambda_1\lambda_2 e^{\lambda_1 h}(2-2e^{\lambda_2 h}+\lambda_1 h)+{\lambda_2}^2(1+\lambda_1 h-e^{\lambda_1 h}(1+2\lambda_1 h))}{\lambda_2(\lambda_1+\lambda_2)(\lambda_2-e^{\lambda_1 h}(\lambda_1+\lambda_2-\lambda_1 e^{\lambda_2 h}))}.$$

Since S_1 occurs before S_2 in the same sampling interval, S_1 propagates to S_{12} at the time of S_2 occurrence and prior to the next sampling time. Therefore,

$$E\left[T_{s_2} \middle| \text{Case I}_{T_2>T_1}\right] = 0,$$

$$E\left[T_{s_1} \middle| \text{Case I}_{T_2>T_1}\right] = u_4 - u_3,$$

$$E\left[T_{s_{12}} \middle| \text{Case I}_{T_2>T_1}\right] = hARL_{s_{12}} - u_4.$$

C_1 and C_2 are given as follows, respectively:

$$C_1 = \sum_{i=1}^{\infty} P\big((i-1)h \leq T_2 \leq T_1 < ih\big) = \frac{\lambda_1\left(1-e^{\lambda_2 h}\right)+\lambda_2(e^{(\lambda_1+\lambda_2)h}-e^{\lambda_2 h})}{(\lambda_1+\lambda_2)(e^{(\lambda_1+\lambda_2)h}-1)},$$

$$C_2 = \sum_{i=1}^{\infty} P\big((i-1)h \leq T_1 \leq T_2 < ih\big) = \frac{\lambda_2\left(1-e^{\lambda_1 h}\right)+\lambda_1(e^{(\lambda_1+\lambda_2)h}-e^{\lambda_1 h})}{(\lambda_1+\lambda_2)(e^{(\lambda_1+\lambda_2)h}-1)}.$$

Cases II & III. In Case II and Case III, S_2 and S_1 occur in different sampling intervals as shown in Figures 2.3 and 2.4 where $0 \leq \tau_{S_1} \leq h$, and $0 \leq \tau_{S_2} \leq h$. Therefore, u_5 and u_6 are given as follows, respectively:

$$u_5 = E\left[\tau_{S_1} \middle| (i-1)h \leq T_1 < ih\right] = \frac{\int_{(i-1)h}^{ih}(t_1-(i-1)h)\lambda_1 e^{-\lambda_1 t_1}dt_1}{\int_{(i-1)h}^{ih}\lambda_1 e^{-\lambda_1 t_1}dt_1} = \frac{1-(1+\lambda_1 h)e^{-\lambda_1 h}}{\lambda_1(1-e^{-\lambda_1 h})},$$

$$u_6 = E\left[\tau_{S_2}\middle|(i-1)h \le T_2 < ih\right] = \frac{\int_{(i-1)h}^{ih}(t_2-(i-1)h)\lambda_2 e^{-\lambda_2 t_2}dt_2}{\int_{(i-1)h}^{ih}\lambda_2 e^{-\lambda_2 t_2}dt_2} = \frac{1-(1+\lambda_2 h)e^{-\lambda_2 h}}{\lambda_2(1-e^{-\lambda_2 h})}.$$

Cases II. $E[T_{s_1}]$ and $E[T_{s_2}]$ depend on how many samples $q, q = \{1, \cdots, \infty\}$ are between T_1 and T_2. For instance, if S_1 occurs three samples after the occurrence of S_2, then $E[T_{s_2}] = 3h - u_6 + u_5$ given that S_2 is not detected until the occurrence of S_1. Therefore, if $T_1 > T_2$, then

$$C_3 = E\left[T_{s_2}, \text{Case II}_{T_1>T_2}\right] =$$

$$\sum_{q=1}^{\infty}\sum_{i=1}^{\infty}(qh - u_6 + u_5)\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)\left(e^{-\lambda_1(i+q-1)h} - e^{-\lambda_1(i+q)h}\right)\beta_{p_2}^q =$$

$$= \frac{\beta_{p_2}(e^{\lambda_1 h}-1)(e^{\lambda_2 h}-1)(e^{\lambda_1 h}(h+u_5-u_6)+\beta_{p_2}(u_6-u_5))}{(e^{(\lambda_1+\lambda_2)h}-1)\left(e^{\lambda_1 h}-\beta_{p_2}\right)^2},$$

$$E\left[T_{s_1}\middle|\text{Case II}_{T_1>T_2}\right] = 0,$$

$$E\left[T_{s_{12}}\middle|\text{Case II}_{T_1>T_2}\right] = hARL_{s_{12}} - u_5,$$

In C_3, S_2 occurs in the sampling interval $[(i-1)h, ih\,]$ and S_1 occurs in the sampling interval $[(i+q-1)h, (i+q)h\,]$ afterwards. For instance, if S_2 occurs in $[0, h\,]$, then S_1 could occur one sample afterwards, i.e., $[h, 2h\,]$, or two samples afterwards, i.e., $[2h, 3h\,]$, and so on. For any q, the sampling plan always fails to detect S_2 until the occurrence of S_1 resulting in $\beta_{p_2}^q$ type II error. If $T_2 > T_1$, then

$$C_4 = E\left[T_{s_1}, \text{Case II}_{T_2>T_1}\right] =$$

$$\sum_{q=1}^{\infty}\sum_{i=1}^{\infty}(qh - u_5 + u_6)\left(e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih}\right)\left(e^{-\lambda_2(i+q-1)h} - e^{-\lambda_2(i+q)h}\right)\beta_{p_1}^q =$$

$$= \frac{\beta_{p_1}(e^{\lambda_2 h}-1)(e^{\lambda_1 h}-1)(e^{\lambda_2 h}(h+u_6-u_5)+\beta_{p_1}(u_5-u_6))}{(e^{(\lambda_1+\lambda_2)h}-1)\left(e^{\lambda_2 h}-\beta_{p_1}\right)^2},$$

$$E\left[T_{s_2}\middle|\text{Case II}_{T_2>T_1}\right] = 0,$$

$$E\left[T_{s_{12}}\middle|\text{Case II}_{T_2>T_1}\right] = hARL_{s_{12}} - u_6,$$

Cases III.

If $T_1 > T_2$, then sampling plan is always able to detect S_2 before the occurrence of S_1 as shown in Figure 2.4. Therefore, the system is only operating with S_2. For instance, $E[T_{s_2}] = h - u_6$, if S_2 is immediately detected at the next sampling time and before the occurrence of S_1. $E[T_{s_2}] = 2h - u_6$, if S_2 is detected two sampling times since its occurrence and before the occurrence of S_1. Sampling fails to detect S_2 at the first sampling time, but it can detect it at the second sampling time. The following formula generalizes this situation.

$$\begin{aligned}
C_5 &= E\left[T_{s_2}, \text{Case III}_{T_1>T_2}\right] \\
&= \sum_{w=1}^{\infty}\sum_{i=1}^{\infty} (wh - u_6)\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)e^{-\lambda_1(i+w-1)h}\beta_{p_2}^{w-1}\left(1-\beta_{p_2}\right) \\
&= \frac{\left(1-\beta_{p_2}\right)e^{\lambda_1 h}\left(e^{\lambda_2 h}-1\right)\left(e^{\lambda_1 h}(h-u_6)+\beta_{p_2}u_6\right)}{\left(e^{(\lambda_1+\lambda_2)h}-1\right)\left(e^{\lambda_1 h}-\beta_{p_2}\right)^2},
\end{aligned}$$

where w represents the number of samples that process undergoes with S_2 until a success detection. The term $e^{-\lambda_1(i+w-1)h}$ indicates that S_2 is detected at the sampling time $(i+w-1)h$, at which, S_1 still has not occurred yet. For example, if S_2 occurs in the time interval $[h, 2h\,]$, then $E\left[T_{s_2}\,\right] = h - u_6$ if S_2 is detected at time $2h$, and hence, $i = 2, w = 1$, and

$$\begin{aligned}
&(wh-u_6)\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)e^{-\lambda_1(i+w-1)h}\beta_{p_2}^{w-1}\left(1-\beta_{p_2}\right) \\
&= (h-u_6)\left(e^{-\lambda_2 h} - e^{-\lambda_2 2h}\right)e^{-\lambda_1 2h}\left(1-\beta_{p_2}\right).
\end{aligned}$$

$E\left[T_{s_2}\,\right] = 2h - u_6$ if S_2 is detected at time $3h$, and hence, $i = 2, w = 2$, and

$$\begin{aligned}
&(wh-u_6)\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)e^{-\lambda_1(i+w-1)h}\beta_{p_2}^{w-1}\left(1-\beta_{p_2}\right) \\
&= (2h-u_6)\left(e^{-\lambda_2 h} - e^{-\lambda_2 2h}\right)e^{-\lambda_1 3h}\beta_{p_2}\left(1-\beta_{p_2}\right).
\end{aligned}$$

Note that

$$E\left[T_{s_1}, \text{Case III}_{T_1>T_2}\right] = E\left[T_{s_{12}}, \text{Case III}_{T_1>T_2}\right] = 0.$$

If $T_2 > T_1$, then sampling plan is always able to detect S_1 before the occurrence of S_2. Therefore, the system is only operating with S_1. The same derivation approach as in $T_1 > T_2$ is followed, and hence,

$$\begin{aligned}
C_6 &= E\left[T_{s_1}, \text{Case III}_{T_2>T_1}\right] \\
&= \sum_{w=1}^{\infty}\sum_{i=1}^{\infty}(wh - u_5)\left(e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih}\right)e^{-\lambda_2(i+w-1)h}\,\beta_{p_1}^{w-1}\left(1-\beta_{p_1}\right) \\
&= \frac{(1-\beta_{p_1})e^{\lambda_2 h}(e^{\lambda_1 h}-1)(e^{\lambda_2 h}(h-u_5)+\beta_{p_1}u_5)}{(e^{(\lambda_1+\lambda_2)h}-1)(e^{\lambda_2 h}-\beta_{p_1})^2}.
\end{aligned}$$

Note that

$$E\left[T_{s_2}, \text{Case III}_{T_2>T_1}\right] = E\left[T_{s_{12}}, \text{Case III}_{T_2>T_1}\right] = 0.$$

Based on the above calculations, $E[T_{s_1}]$, $E[T_{s_2}]$, and $E[T_{s_{12}}]$, are given as follows, respectively:

$$E[T_{s_1}] = \{u_4 - u_3\}C_2 + C_4 + C_6,$$

$$E[T_{s_2}] = \{u_1 - u_2\}C_1 + C_3 + C_5,$$

$$\begin{aligned}
E[T_{s_{12}}] = &\left(hARL_{s_{12}} - u_1\right)C_1 + \left(hARL_{s_{12}} - u_4\right)C_2 + \left(hARL_{s_{12}} - u_5\right)C_7 \\
&+ \left(hARL_{s_{12}} - u_6\right)C_8,
\end{aligned}$$

where C_7 is the probability that the time needed is $hARL_{s_{12}} - u_5$ to alert a true alarm since the occurrence of a shift given Case II, $T_1 > T_2$, whereas C_8 is the probability that the time needed is $hARL_{s_{12}} - u_6$ to alert a true alarm since the occurrence of a shift given Case II, $T_2 > T_1$. C_7 and C_8 are given as follows, respectively:

$$C_7 = \sum_{q=1}^{\infty}\sum_{i=1}^{\infty}\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)\left(e^{-\lambda_1(i+q-1)h} - e^{-\lambda_1(i+q)h}\right)\beta_{p_2}^{q}$$

$$= \frac{\beta_{p_2}e^{-(4\lambda_1+\lambda_2)h}\left(e^{\lambda_1 h} - 1\right)\left(e^{\lambda_2 h} - 1\right)\left(\beta_{p_2}e^{(4\lambda_1+2\lambda_2)h} - e^{(4\lambda_1+\lambda_2)h}\right)}{\left(e^{(\lambda_1+\lambda_2)h} - 1\right)\left(e^{\lambda_1 h} - \beta_{p_2}\right)\left(\beta_{p_2}e^{\lambda_2 h} - 1\right)},$$

$$C_8 = \sum_{q=1}^{\infty}\sum_{i=1}^{\infty}\left(e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih}\right)\left(e^{-\lambda_2(i+q-1)h} - e^{-\lambda_2(i+q)h}\right)\beta_1^{q}$$

$$= \frac{\beta_{p_1}e^{-(\lambda_1+4\lambda_2)h}(e^{\lambda_1 h} - 1)(e^{\lambda_2 h} - 1)(\beta_{p_1}e^{(2\lambda_1+4\lambda_2)h} - e^{(\lambda_1+4\lambda_2)h})}{(e^{(\lambda_1+\lambda_2)h} - 1)(e^{\lambda_2 h} - \beta_{p_1})(\beta_{p_1}e^{\lambda_1 h} - 1)}.$$

5.4. Average time to signal

As defined earlier, ATS is the average time taken until the sampling plan is successful to alert a true alarm since the occurrence of a shift. However, the process could run with two shifts (propagating shift), and hence, the exact definition of ATS will be the average time taken to alert a true alarm since the occurrence of the earlier shift.

In Case I, as shown in Figure 2.2, S_1 or S_2 occurs first, and then, it propagates and becomes S_{12} until it is detected. The average number of samples taken to alert a true alarm is $ARL_{s_{12}}$, and hence, $ATS|\text{Case I}$ is

$$ATS|\text{Case I} = \begin{cases} hARL_{s_{12}} - u_2, & T_1 > T_2 \\ hARL_{s_{12}} - u_3, & T_2 > T_1. \end{cases}$$

As shown in Figure 2.3 ($T_1 > T_2$), S_2 occurs τ_{S_2} time units since time $(i-1)h$. Therefore, $qh - u_6 + u_5$ is the elapsed time between the occurrences of S_2 and S_1. At the time of the occurrence of S_1 , the process starts operating with S_{12} until true detection, i.e., $hARL_{s_{12}} - u_5$ is the time needed to alert a true alarm. Summing up these times, $h(q + ARL_{s_{12}}) - u_6$ is the ATS since the occurrence of S_2. The same applies when $T_2 > T_1$, but with $h\left(q + ARL_{s_{12}}\right) - u_5$, and therefore, $ATS|$Case II is given as

$$ATS|\text{Case II} = \begin{cases} h\left(q + ARL_{s_{12}}\right) - u_6, & T_1 > T_2\,, q = \{1, \cdots, \infty\} \\ h\left(q + ARL_{s_{12}}\right) - u_5, & T_2 > T_1\,, q = \{1, \cdots, \infty\}, \end{cases}$$

where q refers to the number of samples taken between the occurrence times of the two shifts.

For Case III, as shown in Figure 2.4, there is no S_{12}. Therefore, $ATS|$Case III is

$$ATS|\text{Case III} = \begin{cases} wh - u_6, & T_1 > T_2\;, w = \{1, \cdots, \infty\} \\ wh - u_5, & T_2 > T_1\;, w = \{1, \cdots, \infty\}, \end{cases}$$

where w represents the number of samples that process undergoes with $S_2(S_1)$ until a successful detection.

Note that $ATS\big|\text{Case III}_{T_1>T_2}$ and $ATS\big|\text{Case III}_{T_2>T_1}$ equal to the conditional expectations of T_{s_2} and T_{s_1}, respectively, given Case III. Therefore C_5 and C_6 are used in the equation below.

Considering all cases, ATS is given by

$$ATS = (ATS|\text{Case I})\; C_1 + (ATS|\text{Case I})\; C_2 + D_1 + D_2 + C_5 + C_6, \tag{2.30}$$

where

$$D_1 = \sum_{q=1}^{\infty}\sum_{i=1}^{\infty} \left(ATS\big|\text{Case II}_{T_1>T_2}\right)\left(e^{-\lambda_2(i-1)h} - e^{-\lambda_2 ih}\right)\left(e^{-\lambda_1(i+q-1)h} - e^{-\lambda_1(i+q)h}\right)\beta_2^q =$$

$$\frac{\beta_2\left(h\left(ARL_{s_{12}}+1\right)-u_6\right)\left(e^{\lambda_1 h}-e^{2\lambda_1 h}-e^{(\lambda_1+\lambda_2)h}+e^{(2\lambda_1+\lambda_2)h}\right)+\beta_2^2\left(hARL_{s_{12}}-u_6\right)\left(e^{\lambda_1 h}+e^{\lambda_2 h}-e^{(\lambda_1+\lambda_2)h}-1\right)}{\left(e^{(\lambda_1+\lambda_2)h}-1\right)\left(e^{\lambda_1 h}-\beta_2\right)^2},$$

$$D_2 = \sum_{q=1}^{\infty}\sum_{i=1}^{\infty} \left(ATS\big|\text{Case II}_{T_2>T_1}\right)\left(e^{-\lambda_1(i-1)h} - e^{-\lambda_1 ih}\right)\left(e^{-\lambda_2(i+q-1)h} - e^{-\lambda_2(i+q)h}\right)\beta_1^q =$$

$$\frac{\beta_1\left(h\left(ARL_{s_{12}}+1\right)-u_5\right)\left(e^{\lambda_2 h}-e^{2\lambda_2 h}-e^{(\lambda_1+\lambda_2)h}+e^{(\lambda_1+2\lambda_2)h}\right)+\beta_1^2\left(hARL_{s_{12}}-u_5\right)\left(e^{\lambda_1 h}+e^{\lambda_2 h}-e^{(\lambda_1+\lambda_2)h}-1\right)}{\left(e^{(\lambda_1+\lambda_2)h}-1\right)\left(e^{\lambda_2 h}-\beta_1\right)^2}.$$

6. Numerical example

We consider an automatic shot blasting and painting system as shown in Figure 2.5. Small fabricated steel parts such as cleats or rails are first loaded into the conveyor (or hanged on a monorail) and fed into the shot blasting chamber to remove rust from the surface of each part and texturizes it for better paint adhesion. Afterwards, parts are moved to the painting chamber for coating. Both blasting and painting are performed in closed environments. In the blasting machine, turbine disks that blow shot blasting balls on part surface are subject to degradation. Degradation of those disks reduces the amount of balls that hit the surface, so that possible rust could be left on the part's surface. On the other hand, the nozzles of spray guns in the painting chamber may be clogged so that they cannot uniformly spray paint and may dip some frozen paint particles on the part's surface. Indeed, painting on a rusty surface and dipping frozen paint particles cause a rough paint appearance.

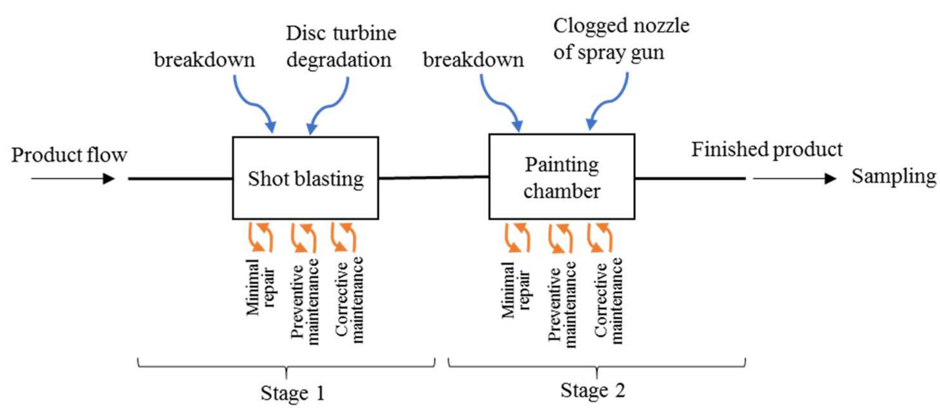

Figure 2.5. Automatic production line of shot blasting and painting.

At the end of the line, a sampling plan by attributes explained previously is employed for inspecting the painted products. The deteriorated turbine disks and spray guns are considered as the sources of assignable causes, but they do not cause machines to breakdown. Instead, machine failures can be caused by other reasons such as overheating and power outage.

Tables 2.2-2.4 show the parameters of shifts, failures, production rate, costs, time elements, and bounds of different constraints. T_{FA} is chosen to be greater than T_{TA}, as it is often easier to detect a shift when a process actually has shifted, whereas more time may be spent on a false alarm to make sure that there is no shift. C_{FA} and C_{TA} are assumed to be equal since the same tooling and practices are used for searching. The time and cost of maintenance increase as the degree of a maintenance action increases. Specially, corrective restoration may include replacing some components (e.g., turbine disk, spray gun, filter, nozzle) and thus require more tooling than other types of maintenance. However, a minimal repair needs the minimum resources to make the failed machine operational again. Therefore, we have $C_{cm} > C_{pm} > C_{MR}$ and $CRT_m > PRT_m > T_{MR}$. Moreover, since C_{NC} may include indirect costs such as claims and a company's goodwill, it is assumed that C_{NC} is greater than C_{LP} and C_{RJ}.

Table 2.2: Shift and failures parameters, and production rate.

p_{01}	p_{11}	p_{02}	p_{12}	λ_1	λ_2	θ_1	θ_2	γ_1	γ_2	g_1, g_2
0.03	0.10	0.05	0.10	0.01 hr^{-1}	0.03 hr^{-1}	1.5	2.0	10 hr	10 hr	100,100 unit/hr

Table 2.3: Cost parameters.

C_s	C_{c1}	C_{p1}	C_{c2}	C_{p2}	C_{MR1}	C_{MR2}	C_{FA}	C_{TA}	C_{LP}	C_{RJ}	C_{NC}
100 $/hr	1200 $/hr	600 $/hr	1200 $/hr	600 $/hr	150 $/hr	150 $/hr	200 $/hr	200 $/hr	3.00 $/unit	3.00 $/unit	4.50 $/unit

Table 2.4: Parameters of key time elements and bounds of constraints.

t_s	CRT_1	PRT_1	CRT_2	PRT_2	T_{MR1}	T_{MR2}	T_{FA}	T_{TA}	L	A	W
0.5 min/unit	50 min	25 min	50 min	25 min	15 min	15 min	15 min	7.5 min	3.00 hr	0.800	0.900

The objective function of the MINLP formulated in Section 5 is mathematically complex. Thus, Genetic Algorithms (GA) solver in MATLAB R2019b is used to obtain the solution. GA is a stochastic method that doesn't require derivatives and is able to search for different solutions within one operation, and hence, the chance of finding a global optimum and avoiding of being trapped in the local optimum increases (Charongrattanasakul and Pongpullponsak 2011). The population size is set to 20 since only three decision variables are to be determined. The integer GA solver in MATLAB overrides settings supplied for creation, crossover, and mutation functions. Instead, GA uses special creation, crossover, and mutation functions (MATLAB & Simulink, 2019). To make the search process for an optimal solution efficient, strict constraint and function tolerances are used. Both are set to default values, i.e., 1×10^{-3} and 1×10^{-6}, respectively. Moreover, UseParallel option is used to compute the fitness value and the feasibility of the nonlinear constraints in parallel to speed up computation. GA is designed to stop if any of the following criteria is met:

- The maximum number of generations (iterations) is reached. Here, the default number is used (i.e., 100×number of decision variables).
- The average change in the penalty fitness value is less than the function tolerance over stall generations where the maximum stall generations is 50.
- Time limit is reached. Here, the default setting is used (i.e., infinity).
- There is no improvement in the objective function during an interval of time called stall time limit. Here, the default setting of the stall time limit is used (i.e., infinity).

The optimal solution is $LRCR^*$ =$ 141.61/hr, r^* =1, $N^* =$ 5, and h^* =0.428 hrs. The optimization problem is solved several times, and on average, the computational time is 133 seconds. To investigate the effects of some input parameters and the bounds of performance constraints on the optimal solution, the following analyses are conducted.

6.1. Effect of C_{FA}

Table 2.5 illustrates how the change in C_{FA} affects the optimal solution. Naturally, the expected cost of false alarms increases as C_{FA} increases while keeping the same sampling parameters. $E[FA_c]$ increases from \$219 when $C_{FA} = 100$ to \$329 when $C_{FA} = 150$. However, $E[FA_c]$ decreases to \$304 when $C_{FA} = 200$ (optimal solution), and then increases again. The increase in C_{FA} from 150 to 200 allows r to increase in order to avoid frequent false alarms by accepting nonconforming units during inspection. With the increase in r, N increases to reduce type I error α and to achieve the desired PR_{eff}. Since with $r = 0$ and $N = 2$, α becomes high, the only way to reduce the average number of false alarms is to reduce the average number of samples taken by having longer h. This justifies why h is higher for $C_{FA} = 100$ and 150, and why it is lower when $C_{FA} = 200$, 250, and 300.

As seen in Table 2.5, there are two setups can be used for inspection: for $C_{FA} < 200$, the setup $(r, N, h) =$ (0, 2, 0.847) is appropriate, and for $C_{FA} \geq 200$, the setup (1, 5, 0.428) is more economical. Practitioners can choose between the two setups for a given a value of C_{FA} without needing to solve the problem again, i.e., the two setups are usable for wide range of C_{FA}. Moreover, more solutions can be created from those setups by changing the decision variables slightly to get further reductions in $LRCR$ especially if the constraints are not violated significantly. This strategy allows more flexibility in selecting the most appropriate solution to cope with possible uncertainties and specific conditions. For instance, if a product is produced for a new customer, management may decide to reduce h slightly to 0.800 instead of 0.847 ($C_{FA} < 200$) to increase customer satisfaction by increasing inspection frequency regardless the increase in $LRCR$.

Table 2.5: Effect of C_{FA} on the optimal sampling plan compared to the current setting.

C_{FA}	r	N	h	$LRCR$
100	0	2	0.847	135.16
150	0	2	0.847	138.78
200	1	5	0.428	141.61
250	1	5	0.428	144.19
300	1	5	0.428	146.86

6.2. Effect of C_{LP}

The effect of C_{LP} is depicted in Table 2.6. Since the expected total cost increases with the increase in wasted time due to non-productive times such as sampling and false alarms, high C_{LP} values ($C_{LP} = 4$,$C_{LP} = 5$) decrease N and increase h in order to increase AV. For instance, the achieved AV when C_{LP} =5 is 83.2% whereas AV =82.0% when C_{LP} =1 A lower value of N means less time will be spent at each sampling, and a higher value of h means a smaller number of samples will be taken in each cycle, and hence, resulting in higher AV. Therefore, for $C_{LP} = 4$ and $C_{LP} = 5$, $LRCR$ has a less sampling cost but a higher cost of rejected units received by customers. On the contrary, a low C_{LP}, such as $C_{LP} = 1$ and $C_{LP} = 2$, permits inspecting more units each time of sampling but with a lower h.

The higher values of N in the first two scenarios reduce the number of false alarms by accepting nonconforming units during the inspection ($r = 1$), and the lower h reduces the cost of rejected units received by customers. Again, as shown in Table 2.6, practitioners can choose the setup (0, 2, 0.847) for any $C_{LP} \geq 4$ and (1, 5, 0.428) for any $C_{LP} < 4$. Hence, given the value of C_{LP}, the corresponding setup can be immediately identified.

Table 2.6: Effect of C_{LP} on the optimal sampling plan compared to the current setting.

C_{LP}	r	N	h	$LRCR$
1	1	5	0.428	105.71
2	1	5	0.428	123.67
3	1	5	0.428	141.61
4	0	2	0.847	159.22
5	0	2	0.847	176.04

6.3. Effect of quality shift parameters

The effect of quality shift parameters is illustrated in Table 2.7. With low shift rates as in the first two scenarios, the inspection process tends to be intensive. This results in a higher number of false alarms compared to scenarios with high shift rates, and hence, the costs, such as lost production and nonconforming units received by customers, also increase. However, the increases in those costs are absorbed by a longer operational time and a longer in-control period, and thus cause a reduction in $LRCR$.

Compared to the current optimal solution, the increases in the total operational and the in-control times for the first scenario are 47.11% and 60.00%, respectively. In addition, 93.02% of the operational time is in the in-control period for the first scenario whereas it is 73.25% for the optimal solution and 68.24% for the last scenario. Lower shift rates increase both AV and PR_{eff}. For instance, in the first scenario, AV and PR_{eff} equal 90.00% and 92.10%, respectively. On the contrary, in the last scenario, AV and PR_{eff} are 80.30% and 91.30%, respectively. As seen in the first scenario, the value of $LRCR$ is 39% and 47.2% less than $LRCR$ of the optimal solution and the last scenario, respectively.

Since the shift rate is one of the features of a machine, the decision maker can focus on how to reduce the shift rate. Redesigning or replacing machines to achieve a cost reduction could be a valuable option. For example, an automated painting chamber can be reinsulated with a better insulation material to avoid spraying products with high viscous paint in a cold environment and hence, avoiding undesirable coating.

Table 2.7: Effect of shift parameters on the optimal sampling plan compared to the current setting.

λ_1	λ_2	r	N	h	$LRCR$
0.0025	0.0225	0	2	0.860	86.26
0.005	0.025	0	2	0.832	128.20
0.01	0.03	1	5	0.428	141.61
0.015	0.035	1	5	0.435	152.78
0.02	0.04	1	5	0.440	163.49

6.4. Influence of ATS constraint L.

Table 2.8 illustrates the influence of ATS constraint L on $LRCR$ and the parameters of the sampling plan while keeping $AV \geq 0.80$ and $PR_{eff} \geq 0.90$. Clearly, $LRCR$ significantly decreases, and AV has a noticeable increase while PR_{eff} slightly decreases with the increase in L. Consequently, L can be further increased to get more reduction in $LRCR$. Actually, with $L = 13.95$ hours, $LRCR = 111.24$, $AV = 0.900$, $PR_{eff} = 0.900$, r=1, N=3, and h=0.730.

Any increment beyond 13.95 hours violates the constraint on PR_{eff}. The constraint on AV is violated for any value of L that's less than 1.75 hours, at which $AV = 0.800$, $PR_{eff} = 0.918$, $LRCR$=156.53, r=0, N=2, and h=0.491. To conclude, further reductions in $LRCR$ can be gained if ATS is increased from 3 to 13.95 while keeping other constraints unviolated. If more interest is in signaling an early true alarm, ATS can be further reduced down to 1.75 without affecting other constraints but increasing $LRCR$.

Table 2.8: Influence of ATS on the optimal sampling plan compared to the current setting.

L	r	N	h	$LRCR$	AV	PR_{eff}
1.75	0	2	0.491	156.53	0.800	0.918
2.00	0	2	0.560	151.56	0.812	0.918
2.50	0	2	0.700	145.42	0.826	0.917
3	1	5	0.428	141.61	0.821	0.916
3.50	1	5	0.500	137.09	0.833	0.915
4.00	1	4	0.367	133.71	0.840	0.914
13.95	1	3	0.730	111.24	0.900	0.900

6.5. The marginal effect of h.

Figure 2.6 shows how the change in the sampling interval h affects $LRCR$ and the performance measures when keeping other parameters and constraints unchanged. As shown in Figure 2.6.a, AV increases as h increases from 0.214 hours to 0.856 hours, and then decreases as h goes beyond 0.856 hours. Since ATS is a function of h and $ARL_{s_{12}}$, ATS is an increasing linear function of h for given values of r and N (constant $ARL_{s_{12}}$) as shown in Figure 2.6.b. Moreover, with lower values of h, more inspection is carried out, and therefore, PR_{eff} tends to be higher as illustrated in Figure 2.6.c.

As h increases, the number of nonconforming units produced between two inspection increases, and hence, PR_{eff} gets lower. Lastly, $LRCR$ significantly decreases when h increases from 0.214 hours to 0.856 hours, achieves the lowest value of 130.21 at h= 0.856 hours, and starts to increase slowly beyond h = 0.856 hours as illustrated in Figure 2.6.d. With a low value of h, $LRCR$ tends to be high since costs of false alarms, inspection, and lost production significantly increase. Overall, if more interest is in reducing $LRCR$, h can be increased beyond $h^* = 0.428$ by violating some constraints. This may be satisfying if the violations are not significant. For instance, with h= 0.856, $LRCR$ is reduced to 130.21, but ATS increases to 5.

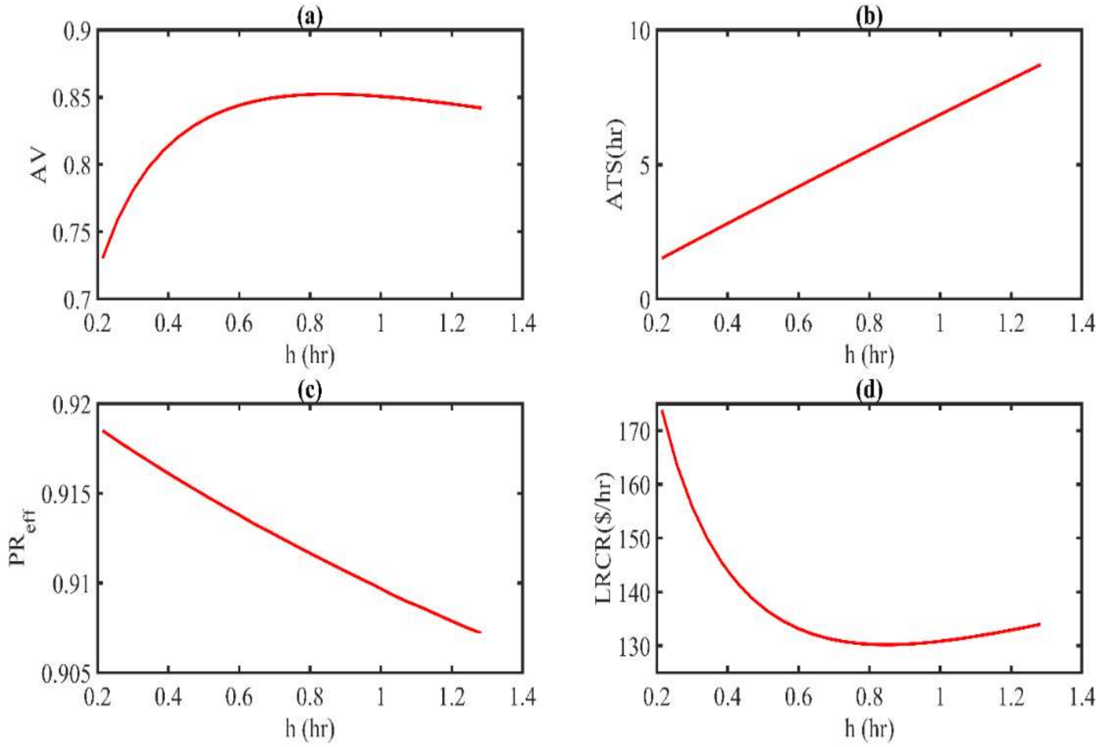

Figure 2.6. The marginal effect of h when r =1, N = 5.

6.6. The marginal effect of r

Figure 2.7 illustrates the marginal effect of r on $LRCR$ and the performance measures. Compared to the optimal setting $r^* = 1$, as seen in Figure 2.7.a, AV drops to 0.650 when r = 0, and ATS decreases to 0.63 as shown in Figure 2.7.b. As r increases, ATS increases quite fast. Moreover, a higher value of r tends to accept more nonconforming units during the inspection, and hence, PR_{eff} is lower for higher r as illustrated in Figure 2.7.c. Having r = 0, the corresponding number of false alarms is about 8 and 70 times of the numbers of false alarms for r = 1 and r = 2, respectively. This drastically increases $LRCR$ to 219 due to poor AV as depicted in Figure 2.7.d. Basically, r is not flexible to change compared to h, as changing r causes significant violations on the constraints. Therefore, attention should be paid when changing the value of r.

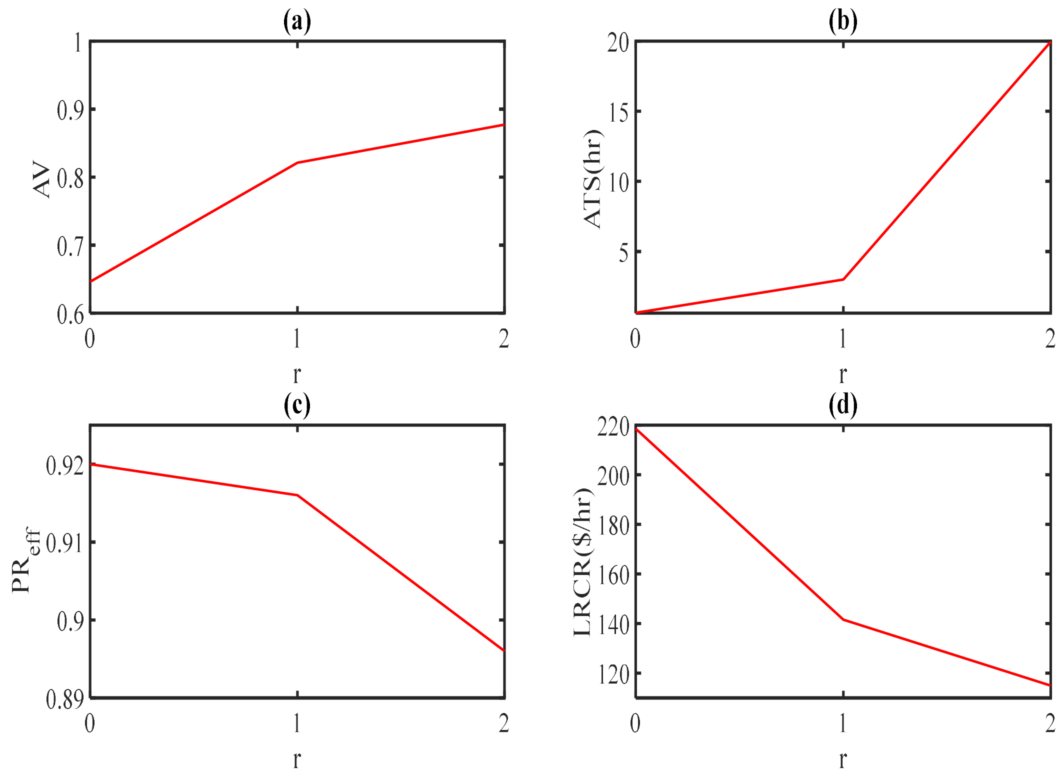

Figure 2.7. The marginal effect of r when h = 0.428, N = 5.

6.7. The marginal effect of N.

In Figure 2.8, the marginal effect of N on $LRCR$ and the corresponding performance measures are provided. In Figure 2.8.b, ATS has a noticeable increase when N decreases to 4 and 3, then it slowly decreases as N goes to 6 and 7. Since ATS increases with the increase in h and $ARL_{s_{12}}$, a low value of N increases type II error, and hence, increases $ARL_{s_{12}}$. The increase in ATS is observable as N decreases such as when N=3.

As seen in Figure 2.8.c, PR_{eff} increases with the increase in N. As N increases, type II error decreases and a smaller number of nonconforming units are produced. The linear trends in Figures 8.a and 8.d are expected since as N increases, the times and costs of inspection and false alarms increase causing $LRCR$ to increase and AV to decrease. Like h, N is flexible to change for benefit to some extent. For instance, $LRCR$ can be reduced to 130 if ATS is violated and increased to 4.7 when N is reduced to 4. In addition, N can be increased to 6 in order to reduce ATS to less than 2.5 hours resulting in a slight decrease in AV but an increase in $LRCR \approx 150$.

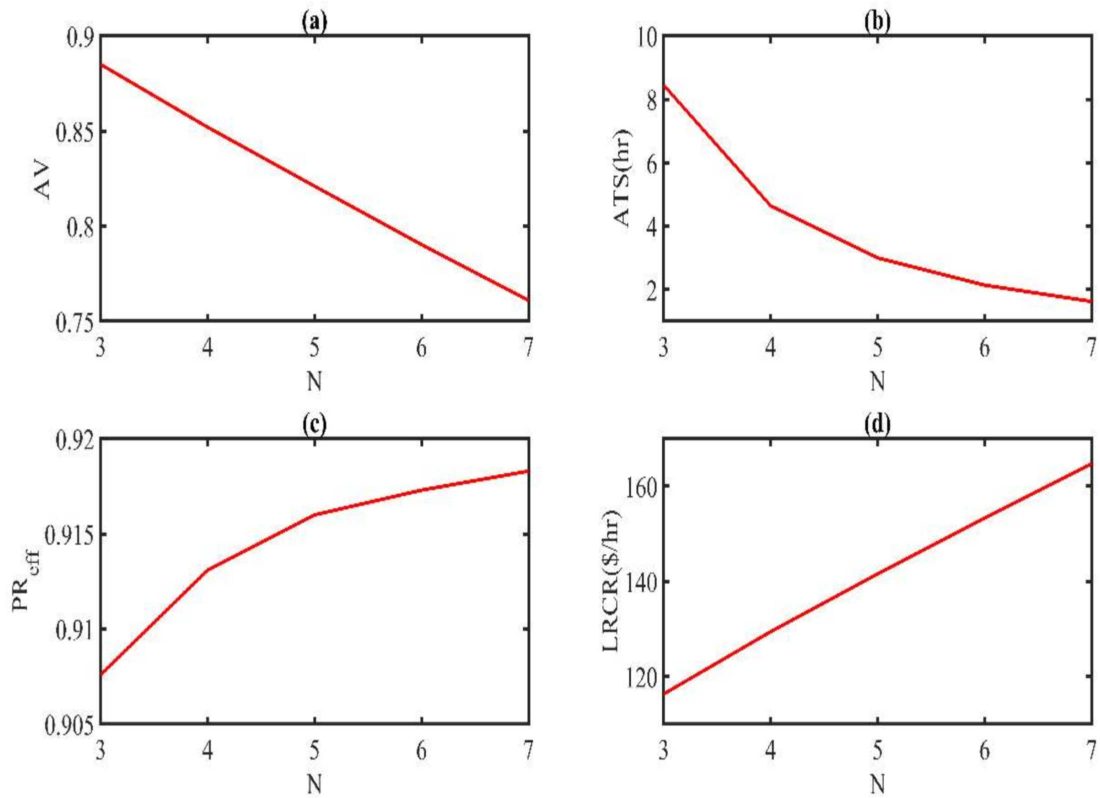

Figure 2.8. The marginal effect of N when h = 0.428, r = 1.

7. Conclusion and future work

Most of online sampling studies investigating multiple assignable causes are conducted on single-stage system. A few studies consider the multiplicity of assignable causes in multistage systems. However, those studies assume identical stages, $\bar{X}$ control chart, same shift level, economic model, no failures, or no quality related costs. This chapter presents a sampling plan for attributes for a serial production system consisting of two unreliable machines where each machine is subject to sudden failure and shift in quality.

A comprehensive economic-statistical model is developed to investigate the joint effect of different shifts by considering the stochastic competency and propagation of the shifts during manufacturing. The developed model generalizes all previous works and compromises between the quality and the quantity performances. The proposed sampling plan minimizes the long-run cost rate subject to constraints on system availability, effective production rate, and average time to signal. A thorough analysis is conducted on some input parameters, the constraint on average time to signal, and the marginal effects of decision variables.

Different managerial insights are provided. Specially, investigating the effects of process parameters, such as shift rates, helps management take long-term decisions (e.g., system overhaul and replacement). Moreover, the analysis shows that when some decision variables are flexible to change, some adjustments can be made to emphasize specific needs.

There are some situations where the assumptions given in Section 3 do not hold. First, if the production rates and reliability of the two machines are significantly different and there are limited areas for storing WIP, the faster and the more reliable machine may have to be stopped to reduce WIP and reducing the related inventory costs. Then, issues like starving and blocking arise. As a result, the developed model in this work is unsuitable, and a new model must be developed to include additional decisions about buffer size and inventory control. Second, if the two machines are dependent (i.e., a failure or a shift of one machine affects the other), a more complex model and different maintenance strategies are needed. Third, to avoid producing more nonconforming units, we assume the system will be preventively stopped during sampling. This is worthwhile if the sampling interval is long (the chance for the system to have a shift is high) and measuring the sampled units takes a while. If the production is allowed to continue during sampling, a delay time due to searching for a true alarm must be added to the average time to signal, and an additional cost due to potentially producing more nonconforming units must be considered.

Beyond these, this work can be extended in other directions. In particular, a multistage system with more than two machines can be considered. Moreover, more than two states of product quality and multiple deterioration states of each machine can be considered. Clearly, the number of system states exponentially increases as the number of machines and/or the number of states of each machine get bigger. For such a complex situation, a simulation-based optimization approach may be utilized. Finally, other system configurations, such as a series-parallel system and parallel-series system, can be studied to deal with cases involving multiple identical machines that perform the same actions during production.

8. References

Azadeh, A., Sangari, M.S., Sangari, E. and Fatehi, S.(2015) A particle swarm algorithm for optimising inspection policies in serial multistage production processes with uncertain inspection costs. *International Journal of Computer Integrated Manufacturing,* **28**(7), 766-780.

Bai, D. S., and Yun, H. J. (1996) Optimal allocation of inspection effort in a serial multistage production system. *Computers & Industrial Engineering,* **30** (3), 387-396.

Bouslah, B., Gharbi, A. and 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.

Cao,Y. and Subramaniam, V. (2013) Improving the performance of manufacturing systems with continuous sampling plans. *IIE Transactions,* **45**(6), 575-590.

Cassady, C.R., Bowden, R.O., Liew, L. and Pohl, E.A. (2000) Combining preventive maintenance and statistical process control: A preliminary investigation. *IIE Transactions,* **32**(6), 471-478.

Charongrattanasakul, P. and Pongpullponsak, A. (2011) Minimizing the cost of integrated systems approach to process control and maintenance model by EWMA control chart using genetic algorithm. *Expert Systems with Applications,* **38**(5), 5178-5186.

Chevalier, B. C. and Wein, L. M. (1997) Inspection for circuit board assembly. *Management Science,* **43**(9), 1198-1213.

Chiu, H-N. and Huang, B-H. (1996) The economic design of $\overline{\mathrm{X}}$ control charts under preventive maintenance policy. *International Journal of Quality & Reliability Management,* **13**(1), 61-71.

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

Jin, J. and Shi, J. (1999) State space modeling of sheet metal assembly for dimensional control. *ASME Transactions, Journal of Manufacturing Science and Engineering,* **121**, 756-762.

Kim, J. and Gershwin S. B. (2005) Integrated quality and quantity modeling of a production Line. *OR Spectrum,* **27**, 287-314.

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

Lam, Y. C., M., S., Zhang, S. and Wu, Z. (2005) Integrated control chart system - optimization of sample Sizes, sampling intervals and control limits. *International Journal of Production Research,* **43** (3), 563-582.

Linderman, K., McKone-Sweet, K.E. and Anderson, J.C. (2005) An integrated system approach to process control and maintenance. *European Journal of Operational Research,* **164**(2), 324-340.

Liu, L., Yu, M., Ma, Y. and Tu, Y. (2013) Economic and economic-statistical designs of an $\bar{X}$ control chart for two-unit series systems with condition-based maintenance. *European Journal of Operational Research,* **226**(3), 491-499.

Lorenzen, T.J. and Vance, L.C. (1986) The economic design of control charts: A unified approach. *Technometrics,* **28**(1), 3-10.

MATLAB & Simulink. (2019) Mixed Integer Optimization. https://www.mathworks.com/help/gads/mixed-integer-optimization.html#bs1cifg. Accessed October 22, 2019.

Meerkov, S. M. and Zhang, L. (2010) Product quality inspection in Bernoulli lines: analysis, bottlenecks, and design. *International Journal of Production Research,* **48** (16), 4745-4766.

Mehrafrooz, Z. and Noorossan, R. (2011) An integrated model based on statistical process control and maintenance. *Computers & Industrial Engineering,* **61**(4),1245-1255.

Montgomery, D.C. (2009). *Introduction to Statistical Quality Control*. 6th ed. Wiley: Hoboken, N.J.

Pandey, D., Kulkarni, M.S. and Vrat, P. (2011) A methodology for joint optimization for maintenance planning, process quality, and production scheduling. *Computers and Industrial Engineering*, **61**(4), 1098-1106.

Rau, H. and Chu, Y.-H. (2019) Inspection allocation planning with two types of workstation: WVD and WAD. *International Journal of Advanced Manufacturing Technology,* **25** (9-10), 947-953.

Rezaei-Malek, M., Mohammadi, M., Jean-Yves, Siadat, A. and Tavakkoli-Moghaddam, R. (2019) A review on optimisation of part quality inspection planning in a multi-stage manufacturing system. *International Journal of Production Research,* **57** (15-16), 4880-7897.

Salmasnia, A., Abdzadeh, B. and Namdar, M. (2017) A joint design of production run length, maintenance policy and control chart with multiple assignable causes. *Journal of Manufacturing Systems,* **42**, 44-56.

Shi, J., and Zhou, S. (2009) Quality control and improvement for multistage systems : A survey. *IIE Transactions*, **41**, 744-753.

Shiau, Y-R. (2002) Inspection resource assignment in a multistage manufacturing system with an inspection error model. *International Journal of Production Research,* **40**(8), 1787-1806.

Valenzuela, J. F., Smith, J. S. and Evans, J. L. (2004) Allocating solder-paste printing inspection in high- volume electronics manufacturing. *IIE Transactions,* **36**(12),1171-1181.

Williams, W.W. and Peters, M.H. (1989) Economic design of an attribute control system for a multistage serial production process. *International Journal of Production Research*, **27**(8), 1269-1286.

Xiang, L. and Tsung, F. (2008) Statistical monitoring of multi-stage processes based on engineering models. *IIE Transactions*, **40**(10), 957-970.

Xiang, Y. (2013) Joint optimization of $\bar{X}$ control chart and preventive maintenance policies: A Discrete-Time Markov Chain approach. *European Journal of Operational Research,* **229**(2), 382-390.

Yang,Y.M., Su, C.Y. and Pearn,W.L. (2010) Economic design of $\bar{X}$ control charts for continuous flow process with multiple assignable causes. *International Journal of Production Economics,* **128**(1),110-117.

Yu, F.J. and Hou, J. L. (2006) Optimization of design parameters for $\bar{X}$ control charts with multiple assignable causes. *Journal of Applied Statistics,* **33**(3), 279-290.

Yu, F.J., Tsou, C.S., Huang, K.I. and Wu, Z. (2010) An economic-statistical design of $\bar{X}$ control charts with multiple assignable causes. *Journal of Quality,* **17**(4), 327-338.

Zantek, P. F.,Wright, G. P. and Plante, R. D. (2002) Systems with correlated stages process and product improvement in manufacturing systems with correlated stages. *Management Science,***48**(5), 591-606.

Zhong, J. and Ma, Y. (2017) An integrated model based on statistical process control and maintenance for two-stage dependent processes.*Communications in Statistics: Simulation and Computation,* **46**(1), 106-126.

Zhou, S., Huang, Q. and Shi, J. (2003) State space modeling of dimensional variation propagation in multistage machining process using differential motion vectors. *IEEE Transactions on Robotics and Automation,* **19**(2), 296-309.

Zhu, H., Zhang, C. and Deng, Y. (2016) Optimization design of attribute control charts for multi-station manufacturing system subjected to quality shifts. *International Journal of Production Research,* **54**(6), 1804-1821.

Chapter 3 : Integrated Decision Making for Attributes Sampling and Proactive Maintenance in a Discrete Manufacturing System

Abstract

An integrated optimal design of attributes sampling and scheduled maintenance for a discrete manufacturing system is proposed in this chapter. It is assumed that the failure of a critical component causes the process to shift, and the time to failure follows the Weibull distribution with an increasing failure rate. The developed model for integrated decision making is focused on online sampling with the binomial and truncated negative binomial distributions. Since the likelihood of the process to have a shift increases with time, multiple maintenance opportunities are offered to assist a manager in making a timely and economical maintenance decision based on product inspection results. In addition to performing scheduled maintenance and unscheduled corrective maintenance at the time of a true alarm, an additional maintenance opportunity when a false alarm occurs is also considered. The optimal scheduled maintenance time and sampling parameters are determined by minimizing the long-run cost rate. A numerical example is provided to illustrate the proposed integrated maintenance and attributes sampling plan. The results show that the integrated approach outperforms the alternatives that consider different models separately. More importantly, showing the benefit of doing maintenance upon a false alarm provides a manager with a new idea in managing a deteriorating manufacturing system.

Keywords: Integrated design of sampling and maintenance, attributes sampling, truncated negative binomial distribution, multiple maintenance opportunities.

1. Introduction

The performance of a production process can be assessed by its ability to deliver products with acceptable quality. To improve quality, companies continuously invest in labor and new technologies. However, such investments increase the costs of operations. Hence, the proper understanding of the interactions between operations is crucial to avoid additional costs.

Maintenance and quality are two key operational areas that have been extensively studied. Quality of products basically depends on the condition of the manufacturing system on which products are produced. It is practical to control quality through planning for maintenance where maintenance can be scheduled periodically as a proactive procedure. Also, a manufacturing system can be monitored by Statistical Process Control (SPC) tools. These tools such as sampling are widely used to control manufacturing processes by providing proper information for immediate maintenance if processes are adversely affected.

Sampling is an inspection procedure that compromises between the high cost of 100% inspection and the high cost of quality loss if no inspection is carried out. It aims at ensuring the conformity of products and monitoring processes for unusual failures. A sampling plan refers to the setup of sampling parameters chosen to achieve a certain objective. For instance, sample size, control limit coefficient, and time between two samplings for inspection with $\bar{X}$ control chart are elected by minimizing a cost function.

A Sampling plan is mainly used to monitor a production system when a failure in a process is unobservable. This type of failure is called a shift, and generally, it doesn't cause the system to stop suddenly. Instead, production continues with a shift until it is detected by sampling. Process shifts have different causes and forms. For example, degradation of a machine could increase the percentage of nonconforming units produced, whereas a wrong setup could cause the average concentration of an ingredient of a drink to increase above a desired level. Any sampling plan is designed in such a way an alarm is alerted when a predefined inspection criterion is not met. If a shift is confirmed, the cause of the shift is eliminated by maintenance interventions. Maintenance actions are versatile form minor adjustments to overhauls where major repairs that may include replacing failed or degraded components are carried out with considerable time and cost.

Sampling and maintenance are designed separately in many studies. In sampling models, maintenance is performed only if a shift in a process is detected, and no scheduled (planned) maintenance is carried out. Some examples are the models developed by Li et al. (2016) and Yeong et al. (2013). On the other hand, studies such as Samrout et al. (2009) and Moghaddam and Usher (2010) consider only scheduled maintenance without sampling.

Considering only scheduled maintenance, the process may operate with a shift for a considerable duration before the time of maintenance is reached whereas, in a sampling model, the type II error could delay shift detection. Furthermore, preventive maintenance can be carried out at the scheduled time with less time and cost before a process undergoes significant degradation. Therefore, integrating sampling and scheduled maintenance in one model could be economically beneficial since two maintenance opportunities are provided. Integrated (joint) modeling of sampling and maintenance can be found in studies conducted by Yeung et al. (2008) and Zhong and Ma (2017).

In integrated models, maintenance is carried out at a scheduled time called maintenance interval or at the time the sampling plan detects a shift, whichever occurs first. The decisions about how long the maintenance interval is and what are the best sampling parameters are taken simultaneously. It has been shown in some studies that the integrated models ordinarily outperform the standalone models of sampling and maintenance.

Basically, in an integrated model, there are two maintenance opportunities: the scheduled maintenance, and maintenance at the time of shift detection (true alarm). A few studies consider the time at which a false alarm is alerted as another maintenance opportunity. All those studies assume that, at this opportunity, the maintenance time is the same regardless of at what time the false alarm is signaled. Furthermore, there are no explanations about how it is beneficial to perform maintenance at the time of the false alarm.

The objective of this chapter is to develop an integrated model of maintenance and sampling by attributes in which multiple maintenance opportunities are offered and the time of maintenance at the false alarm opportunity increases with time. The performance of the integrated model is compared to the performance of the separate models of maintenance and sampling. Three economic models are developed, and the long-run cost rate of each model is minimized.

The remainder of this chapter is organized as follows. Section 2 reviews the literature and illustrates the contributions of this study. Section 3 describes the problem and provides the assumptions and notation used throughout this chapter. In Section 4, the integrated and separate models of sampling and maintenance are developed. Section 5 shows the experimental work and sensitivity analyses. Last, Section 6 concludes this study and recommends future work.

2. Literature review and contribution

2.1. Related works

Sampling models for production systems are versatile. A sampling procedure is designed to suit the type of production and serve a certain purpose. Acceptance sampling is used in lot (batch) production in order to decide on accepting or rejecting a produced lot. Some studies on acceptance sampling are Kaya (2009) and Duarte and Saraiva (2008). Continuous sampling (CS) is another sampling procedure that is first developed by Dodge (1943). Basically, a CS model assumes that the process is always in control, and it alternates between 100% and fractional inspections in order to achieve a desired outgoing quality with a minimum inspection. The most common sampling procedure is the online sampling, i.e., sampling that uses control charts to monitor production processes and alert for maintenance if unusual failures in a process are detected.

Duncan (1956) designs a sampling plan of $\bar{X}$ control chart that maximizes an income function. Lorenzen and Vance (1986) generalize Duncan's model for various control charts. Sampling by the exponentially weighted moving average (EWMA) chart with a quality loss function is studied by Serel (2009). All these studies propose economic models, i.e., no constraints are considered. Saniga (1989) introduces an economic-statistical model with constraints on type I error, type II error, and the average time to signal for sampling design with $\bar{X}$ and R control charts. Yeong et al. (2013) propose economic and economic-statistical designs with constraints on the average run lengths (ARLs) for the synthetic $\bar{X}$ control chart under different quality loss functions. Safaei et al. (2015) study the uncertainty of process parameters in the design of sampling with $\bar{X}$ control chart. More studies on sampling designs are Rahim (1993), Lee and Rahim (2001), Ben-Daya and Duffuaa (2003), Christopher et al. (2010), and Seif et al. (2015).

Models for scheduled maintenance have also been extensively developed. Barlow and Hunter (1960) develop a periodic replacement model for a system that is minimally repaired if a failure occurs during the maintenance period. Das and Sarkar (1999) propose a preventive maintenance model for a production system with (S,s) inventory policy. A deteriorating system that is preventively repaired if reliability reaches a threshold and replaced after a successive number of preventive maintenance times is investigated by Liao et al. (2010). Lin and Huang (2010) suggest non-periodic scheduled preventive maintenance for a deteriorated system. Huynh et al. (2012) construct an age-based maintenance model for a system subject to continuous degradation and shock failures. Zong et al. (2013) determine the optimal replacement policy for a deteriorating system subject to shocks and increasing repair times. Further studies on scheduled maintenance are El-Ferik and Ben-Daya (2006), Lee et al. (2006), and Mujahid and Rahim (2010).

In the context of integrated models, Cassady et al. (2000) design an integrative model of $\bar{X}$ control chart and age-replacement preventive maintenance. The process shift is assumed to be attributed to a component's failure where time to failure follows the Weibull distribution with an increasing failure rate. Linderman et al. (2005) propose three maintenance scenarios by which the process is renewed. The process shift follows the Weibull distribution, and a cost function is constructed with constraints on ARL(s). Another model for the multivariate exponentially weighted moving average (MEWMA) control chart is developed by Ardakan et al. (2016). Rasay et al. (2018) study Chi-square sampling in a two-stage dependent process. Eight maintenance scenarios are developed based on the states of the two processes.

Rahim and Ben-Daya (1998) design an integrated plan with variable sampling intervals to determine the economic production quantity (EPQ), maintenance interval, and the $\bar{X}$ control chart parameters. The same procedure is assumed by Ben-Daya and Rahim (2000). Imperfect preventive maintenance is carried out at each sampling time to reduce the shift rate by an amount proportional to the level of maintenance performed. Pandey et al. (2012) consider a system monitored by $\bar{X}$ control chart and subject to complete failures and quality shifts. Imperfect preventive maintenance is performed at a scheduled time, whereas minimal repair is carried out to restore the system when a failure occurs, or a shift is detected. The same design procedure is followed by Shrivastava et al. (2016) for the joint modeling of the cumulative sum (CUSUM) control chart and preventive maintenance. Liu et al. (2017) assume that imperfect preventive maintenance is performed each time the $\bar{X}$ control chart alerts for maintenance, and the system is replaced after a specified number of imperfect preventive maintenance or after a complete failure.

Zhou and Zhu (2008) add a new maintenance scenario to the scenarios developed by Linderman et al. (2005). In their model, maintenance is carried out when a false alarm is alerted. This work is extended by Charongrattanasakul and Pongpullponsak (2011) for the integrated design of the EWMA control chart. Six maintenance scenarios are built due to considering warning zones and maintenance at the time of the false alarm. False alarm maintenance is also assumed by Mehrafrooz and Noorossana (2011). Six maintenance scenarios that consider complete failure are developed. The Shewhart individual-residual control chart is used to monitor a system of two stages (Zhong and Ma, 2017). Considering maintenance at the time of false alarm and all states of the two stages, eight maintenance scenarios are developed.

A deteriorating production system with multiple out-of-control states and monitored by $\bar{X}$ control chart is studied by Xiang (2013). For this system, imperfect preventive maintenance is carried out when a true alarm is alerted or at a scheduled time, whereas corrective maintenance is performed when a complete failure occurs. Multiple out-of-control states are also assumed by Tagaras (1988). Yin et al. (2015) present the concept of delayed monitoring. It assumed that inspection at the beginning of the operation can be delayed for avoiding unnecessary inspection costs since the process is highly likely to be in control and less prone to breakdowns. Both quality shift and equipment failure are assumed to be independent stochastic processes that follow Weibull distributions with increasing failure rates. More literature on integrated sampling and maintenance plans can be found in Panagiotidou and Tagaras (2007) and Radhoui et al. (2009).

As reviewed above, different versions and subjects of the integrated models are presented. While most of the models are constructed based on $\bar{X}$ control chart, a few studies investigate other charts such as Chi-square and EWMA. Further, some models consider issues such as multiple out-of-control states, complete failures, delayed monitoring, imperfect maintenance, false alarm maintenance, and variable sampling intervals. Such issues increase the number of scenarios of renewing an inspection (or production) cycle.

2.2. Contribution of this work

A common assumption in most of existing models is that no maintenance is performed upon a false alarm. This assumption is reasonable if the time to shift follows the exponential distribution (i.e., with a constant failure rate). However, when the time to shift has an increasing failure rate, only a few studies allow performing maintenance upon a false alarm, and these studies always assume a constant maintenance time and cost regardless of at what time a false alarm is alerted.

This chapter focuses on three main issues that have not been well addressed in the intergrated models. First, assuming a constant maintenance time when a false alarm is alerted may not be practical. Since a false alarm may happen earlier in an inspection cycle, the likelihood of a shift is small, and hence, maintenance may be unnecessary. Instead, relating the maintenance time to the likelihood of the occurrence of a shift is more practical and economical. In other words, more maintenance time may be spent if the likelihood of shift is higher, and less time may be spent if the likelihood of shift is lower. Second, the benefit of taking the maintenance opportunity on a false alarm is not clarified in previous studies. In this chapter, a detailed analysis is conducted to show the value of this opportunity and the merit of having multiple maintenance opportunities in the integrated model. This is illustrated by comparing the integrated model with the individual models for sampling and scheduled maintenance in terms of their performance. Third, integrated modeling of maintenance and sampling by attributes such as sampling with *np* control charts has not or rarely been investigated.

Another important technical contribution is on sampling. Since an inspection cycle may end with a true alarm, false alarm, or scheduled maintenance, sampling is carried out according to a combination of binomial and truncated negative binomial distributions. It is worth pointing out that inspections based on the negative binomial distribution is rarely studied. Although Huang, Lo, and Ho (2008) and Huang, Lin, and Ho (2013) develop inspection procedures based on the negative binomial distribution, their procedures involve approximations and are used for acceptance sampling, not online sampling, as investigated here.

3. Problem description

3.1. Problem statement

Consider a production process that operates continuously and produces discrete units of one product. The production process begins in the in-control state and produces the product with a proportion of nonconforming (PON) equals to p_0. Due to usage and aging, the process may start producing products with an undesirable PON equals to p_1 ($p_1 > p_0$), at which, the process shifts to the out-of-control state. Unlike a sudden breakdown which is noticed immediately, the shift in the process is unobservable. Therefore, a sampling plan by attributes is used for inspecting products and detecting the shift in the process.

It is assumed that the process's shift is related to the failure of a critical component in the production unit. Such an assumption can be found in Cassady et al. (2000). It is supposed that the failure of the critical component doesn't cause the production unit to break down. Instead, it only causes PON to increase from p_0 to p_1. Since the chance of the shift's occurrence increases with the increase of the production unit usage, it is assumed that the time to shift follows the two-parameter Weibull distribution with an increasing failure rate.

The proposed integrated sampling and maintenance plan is shown in Figure 3.1. At each time of sampling, one unit is inspected to see if it is conforming or nonconforming. The process is assumed to keep operating at the time of inspection, and if an inspected unit is nonconforming, it is rejected without replacement. The time between two inspections h is called the sampling interval. Sampling continues until the number of nonconforming units found during the inspection X exceeds a predefined acceptance number r by one unit or until n units are inspected, whichever occurs first. Therefore, the shortest inspection cycle length is $(r+1)h$ due to $r+1$ consecutive nonconforming units are found since the beginning of the inspection, whereas the longest length is $t_n = nh$, at which, the scheduled maintenance is performed.

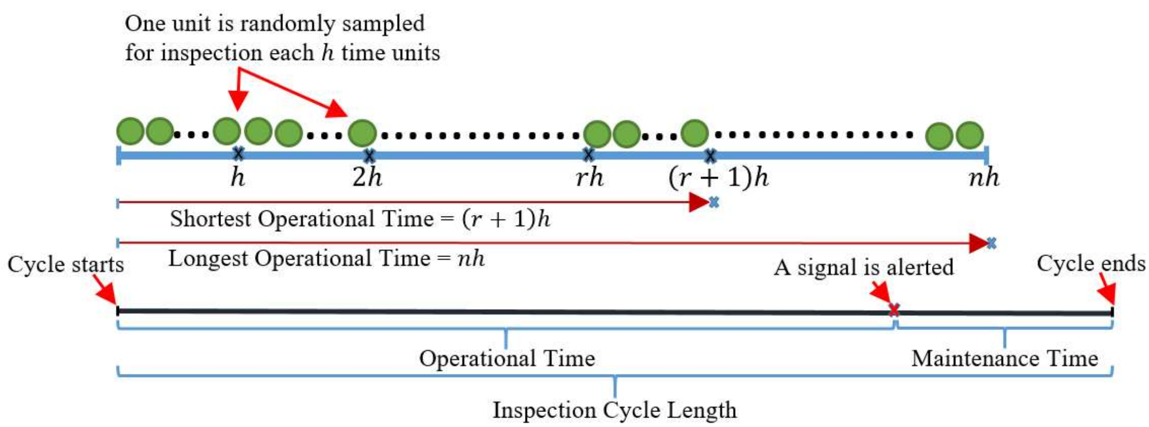

Figure 3.1. The proposed integrated sampling and maintenance plan.

Whenever $X = r+1$, an alarm is signaled, and the process is stopped for investigation. The outcome of the investigation is either a false alarm or a true alarm. With a false alarm, the process is still in control, whereas, with true alarm, the process is out-of-control. For both alarms, the production process is halted for maintenance. If the production process continues without any signal (i.e., $X < r+1$) for n inspections, the production process is shut down for the scheduled maintenance. Since maintenance is scheduled at time t_n if no prior signal is alerted, it may not be necessary to sampling at time t_n. However, in this study, we assume that inspection is carried out at time t_n to eliminate the inspected unit if it is found nonconforming. This is reasonable especially if the inspection cost and time are minimal and the aftersales costs such as warranty and claims are significant.

Maintenance could be preventive or corrective. Preventive maintenance is performed if a false alarm is alerted or if the scheduled maintenance time t_n is reached with no shift. Since the failure rate is increasing, the time spent for preventive maintenance at the time of the false alarm is assumed to increase with the increase of the likelihood of shift occurrence. On the other hand, the scheduled preventive maintenance is performed with constant time at time t_n. Regardless of where the false alarm occurs, the false alarm maintenance time is always assumed to be less than the scheduled preventive maintenance time. If a shift is observed, either with a true alarm or at time t_n, corrective maintenance is carried out to replace the critical component. Any maintenance action retrieves the process to the in-control state, and by the completion of maintenance, the inspection cycle ends, and a new cycle begins with PON = p_0. Figure 3.2 illustrates how the integrated sampling and maintenance plan works.

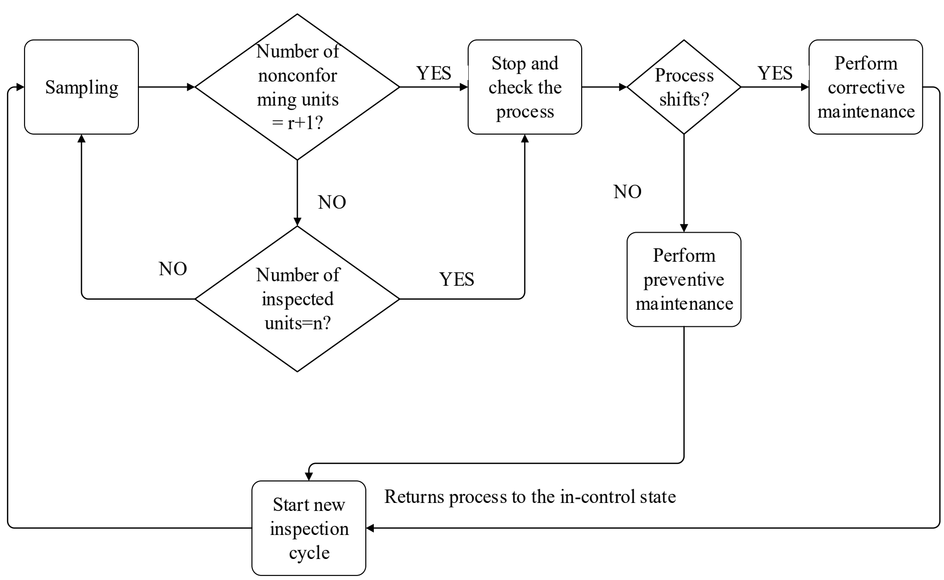

Figure 3.2. Decision flowchart of the integrated plan

As illustrated in Figures 3.1 and 3.2, the integrated design of attribute sampling and scheduled maintenance offers different maintenance opportunities. The decision variables r, n, and h are determined by minimizing the long-run cost rate ($LRCR$). The assumptions about system operation and the notation used in this paper are provided next.

3.2. Assumptions

The following assumptions are made throughout this chapter:

- A failure of a critical component in the production unit causes quality to shift. A "production unit" refers to a machine or a production line. It is common that some components fail (or degrade) more frequently than others. For instance, in an automatic painting process, spraying nozzles clog by time causing an unacceptable coat applied to products.
- Time to shift (failure) follows the two-parameter Weibull distribution with an increasing failure rate.
- Production continues at the time of sampling. This assumption can be made to get gains in production especially if the production rate is high, the shift rate is low, or PON is small.
- A nonconforming inspected product is rejected without replacement.
- When an alarm is alerted, the production unit is stopped since close investigation is needed. Also, maintenance cannot be performed while a production unit is operating.
- The raw materials are defect free (i.e., incoming quality is perfect). Note that if the incoming quality is not perfect, this effect can be folded into the in-control nonconforming probability.
- The production unit doesn't deteriorate while being stopped.
- There is enough storage area for the finished products so that the production will not be stopped because of lacking storage areas.

3.3. Notation

The notation that is used throughout this chapter is shown in Table 3.1.

Table 3.1: Notation list

Decision variables	
h	Sampling interval measured in hours, a decision variable, $h > 0$.
n	Number of samples until the scheduled maintenance, a decision variable, $n > r$.
r	Acceptance number, a decision variable, $r \geq 0$.
m	Time until scheduled maintenance in the maintenance policy, a decision variable, $m > 0$.
Objective functions	
$LRCR$, $LRCR_s$, $LRCR_m$	Long-run cost rate of the integrated, sampling, and maintenance policies, respectively. $LRCR$ is measured by \$/hr.
Other variables, constants and indices	
PON	Proportion of nonconforming: $PON = p_0$ in the in-control state, $PON = p_1$ in the in the out-of-control state.
n_1, n_2	Number of samples taken with p_0 and p_1, respectively, given no signal is alerted, $n_1 + n_2 = n$.
X	Number of nonconforming units found during inspection, $X \in \{0, \cdots, r\}$.
X_1, X_2	Number of nonconforming units found in n_1 and n_2, respectively. $X_1 + X_2 = X$.
Y	Total number of inspected units in an inspection cycle.
j	Index refers to the sample's number.
i	Number of the conforming inspected units in the inspection cycle.
t_{r+1+i}	Time at which a signal is alerted, $t_{r+1+i} = (r+1+i)h$.
k_1, k_2	Number of samples taken with p_0 and p_1, respectively, given a signal is alerted and. inspection ends before or right upon t_{r+1+i}, $k_1 = j - 1$, $k_2 = r + 2 + i - j$, $k_1 + k_2 = k = r + 1 + i$.
$\acute{k_2}$	$\acute{k_2} = k_2 - 1$.
Q_1	A set that represents the possible values of δ: $Q_1 = \{1, x, E[V \mid \Omega^j_{t_n}]\}, \delta \in Q_1$.
Q_2	A set that represents the possible values of π: $Q_2 = \{1, t_{r+1+i}, E[V \mid \Omega^i_1], E[V \mid \Omega^{i,j}_2], E[V \mid \Omega^i_2]\}, \pi \in Q_2$.
μ	Production rate measured in units per hour.
β, η	Shape and scale factors of the Weibull distribution, η measured in hours.
$c_{ins}(c_{rej})$	Average cost of inspecting (rejecting) one unit of a product, measured in \$/unit.
c_{cun}	Average cost of a nonconforming uninspected unit, measured in \$/unit.
c_{lp}	Average cost per each unit unproduced, measured in \$/unit.
$c_{PM}(c_{CM})$	Average cost per hour (\$/hr) of performing preventive (corrective) maintenance
c_0	Average cost per hour (\$/hr) for of searching a true or false alarm

Table 3.1 (Cont.)

Other variables, constants and indices	
C_{CM}	Average cost of corrective maintenance
c_{COM}	Average cost of a new component
$T_S(T_{NS})$	Average time in hours to conclude that there is a shift (no shift)
T_0	Average time to perform scheduled preventive maintenance, measured in hours
t_0	Fixed maintenance time in hours
$t_{PM,t_{r+1+i}}, t_{PM,t_j}$	Time of preventive maintenance at the false alarm opportunity in the integrated and sampling policies, respectively.
t_{CM}	Average time in hours of performing corrective maintenance
MT, MT_s, MT_m	Total maintenance time in the inspection cycle of the integrated, sampling, and maintenance policies, respectively.
MC, MC_s, MC_m	Total maintenance cost in the inspection cycle of the integrated, sampling, and maintenance policies, respectively.
V, V_s, V_m	Total number of uninspected units that found nonconforming in the inspection cycle of the integrated, sampling, and maintenance policies, respectively
CUN, CUN_s, CUN_m	Total number of uninspected nonconforming units produced in the inspection cycle of the integrated, sampling, and maintenance policies, respectively.
SC, SC_s	Sampling cost of the integrated and sampling policies, respectively
RC, RC_s	Rejection cost of the integrated and sampling policies, respectively
LPC, LPC_s, LPC_m	Lost production cost in the integrated, sampling, and maintenance policies, respectively
CL, CL_s, CL_m	Inspection cycle length in hours of the integrated, sampling, and maintenance policies, respectively.
CC, CC_s, CC_m	Inspection cycle total cost of the integrated, sampling, and maintenance policies, respectively.
T_{in}	Time to shift, i.e., time that process stays in control
$T_{j-1,j}^{in}$	Time to shift given a shift occurs in time interval $[(j-1)h, jh]$
$T_{r+i,r+1+i}^{in}$	Time to shift given a shift occurs in time interval $[(r+i)h, (r+1+i)h]$
$\tau_{j-1,j}$	The elapsed time for shift occurrence since the last sampling time given that the shift occurs in the time interval $[(j-1)h, jh]$
$T_{in,m}$	Time to shift in the maintenance policy
FA_j	False alarm occurs at sample j
N_{j,x_1}	Number of samples taken until a true alarm is alerted given that x_1 nonconforming units found in the $j-1$ samples taken before the occurrence of the shift.
TA_{j,x_1}	A true alarm is alerted at sample j given that x_1 nonconforming units found in the $j-1$ samples taken before the occurrence of the shift.
C	The operational time of the inspection cycle under the sampling policy.

4. Model development

In this section, the integrated model, as well as the separate models of sampling and maintenance, are constructed. Moreover, the optimization problem of each model is formulated.

4.1. The integrated model of sampling and scheduled maintenance

Let X be the total number of nonconforming units found during inspection. Then, $X \in \{0, \cdots, r, r + 1\}$. If $X = r + 1$ at any time of sampling, a signal is alerted, and the production process is stopped for maintenance. If $X < r + 1$, the production process continues to the next sampling time. Sampling continues if there is no signal until n units are inspected. At that time, the production process is shut down for scheduled maintenance whether a signal is alerted or not. Hence, the shortest cycle length is $t_{r+1} = (r + 1)h$ at which $r + 1$ inspected units are consecutively found nonconforming since the beginning of inspection, and the longest cycle length is $t_n = nh$. Accordingly, the following cases are defined.

- Case 1: Inspection cycle ends at time t_n and $X < r + 1$. In this case, the inspection cycle length (CL) equals $t_n = nh$ plus the maintenance time.

  Case 1.1: After investigation, the process is still in control.

  Case 1.2: After investigation, the process has shifted (out-of-control).

- Case 2: Inspection cycle ends before or right upon n inspections and $X = r + 1$. For this case, $CL \in \{t_{r+1}, \cdots, t_n\}$ plus the maintenance time.

  Case 2.1: After investigation, the process is still in control, and a false alarm is reported.

  Case 2.2: After investigation, the process has shifted, and a true alarm is confirmed.

Figure 3.3 shows the maintenance actions corresponding to the above two cases. Cases 1.1, 1.2, 2.1, and 2.2 are explained below.

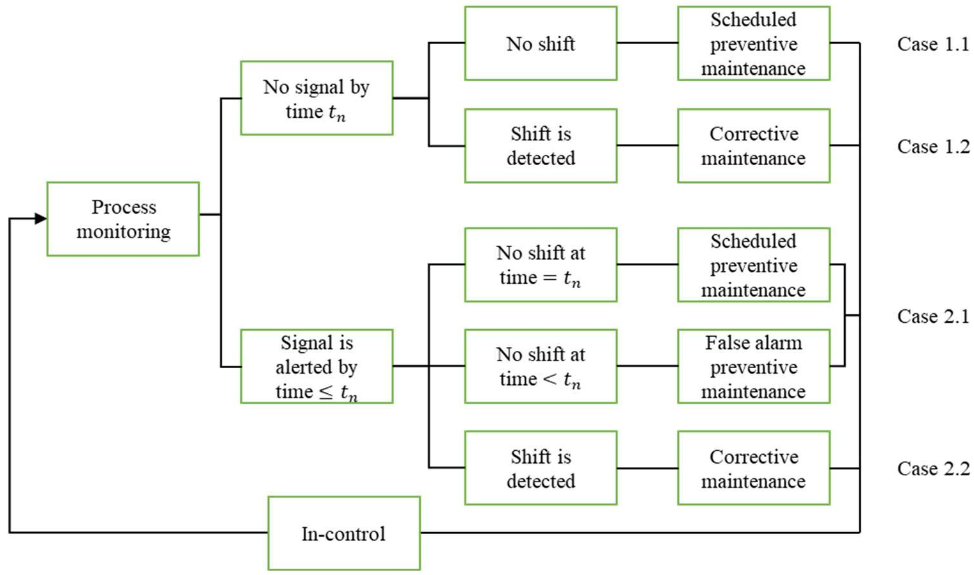

Figure 3.3. Maintenance actions of the integrated model.

4.1.1. Case 1.1: Process has not shifted by time t_n

As shown in Figure 3.4, the process continues with no signal is alerted until time t_n, at which, the scheduled maintenance time is reached. At that time, no shift is detected, and hence, preventive maintenance is carried out. Here, $CL = t_n + T_0$, where T_0 is the preventive maintenance time of the scheduled maintenance.

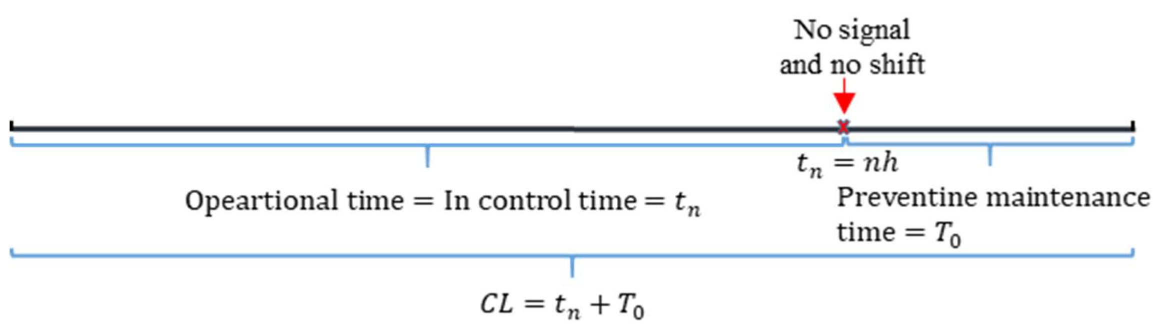

Figure 3.4. Inspection cycle with no signal alerted nor shift observed by time t_n.

Since no shift is found, all units are produced with $PON = p_0$. Let us define $A_{0,\delta}$ as follows:

$$A_{0,\delta} = e^{-\left(\frac{nh}{\eta}\right)^{\beta}} \sum_{x=0}^{r} \delta \binom{n}{x} {p_0}^{x} (1-p_0)^{n-x}, \tag{3.1}$$

where $\delta \in Q_1, Q_1 = \{1, x, E[V|\Omega_{t_n}^{j}]\}$. Then, $A_{0,\delta=1}$ represents the probability that inspection ends at time t_n with no signal and no shift are recorded.

4.1.2. Case 1.2: Process has shifted before or right upon t_n

In Figure 3.5, the process has shifted prior to or right upon time t_n, but the sampling plan fails to alert a true signal at any sampling time. The operational time consists of the in-control time in which $PON = p_0$ and the out-of-control time in which $PON = p_1$, and hence, $CL = t_n + t_{CM}$, where t_{CM} is corrective maintenance time.

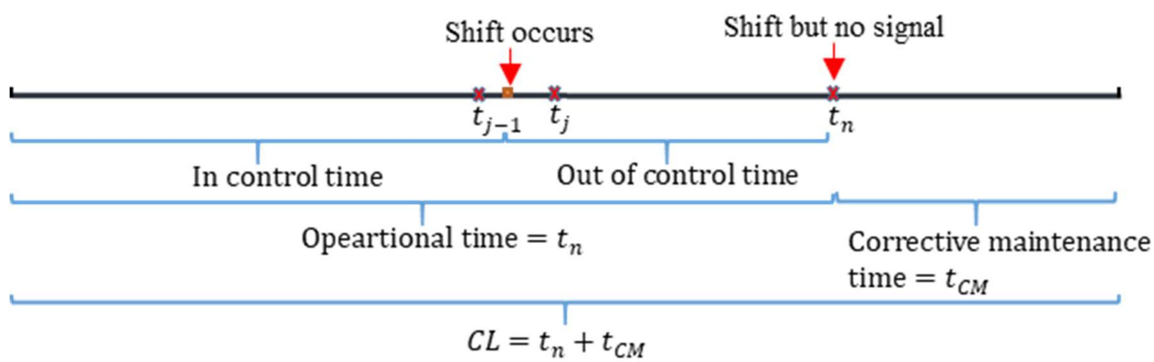

Figure 3.5. Inspection cycle with shift observed at time t_n and no signal alerted.

Assume that the shift has occurred between t_{j-1} and t_j sampling times where $j \in \{1, \cdots, r+1, \cdots, n\}$, $t_{j-1} = (j-1)h$, and $t_j = jh$. Let n_1 and n_2 denote the number of samples "units" taken with p_0 and the number of samples taken with p_1, respectively. Then, $n_1 = j-1$ and $n_2 = n - j + 1$ where $n_1 + n_2 = n$. Also, define X_1 and X_2 as the number of nonconforming units found in n_1 and the number of nonconforming units found in n_2, respectively, where $X_1 + X_2 = X$, $X \in \{0, \cdots, r\}$. Apparently, the sizes of n_1 and n_2 vary with the shift's occurrence location, and therefore, the following scenarios are defined.

***Scenario 1.1.* $\boldsymbol{n \geq 2r}$**

***1. Shift occurs between* $\mathbf{t_{j-1}}$ *and* $\mathbf{t_j}$ *such that* $\boldsymbol{n_1 \geq r}$ *and* $\boldsymbol{n_2 \geq r}$**

Here, the shift occurs at least r samples since the beginning of inspection and at least r samples far away from t_n. Since $n_1 \geq r$ and $n_2 \geq r$ imply that $j \geq r+1$ and $j \leq n-r+1$, the range in which the shift could occur is $r+1 \leq j \leq \min\{n, n-r+1\}$. The upper bound should be $n-r+1$, but because r could equal 0, the upper bound is modified to be $\min\{n, n-r+1\}$. Let us define $A_{11,\delta}$ as follows:

$$A_{11,\delta} = \sum_{j=r+1}^{\min\{n,n-r+1\}} \sum_{x=0}^{r} \sum_{x_1=0}^{x} \delta \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right) \binom{j-1}{x_1} {p_0}^{x_1} (1 - p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} {p_1}^{x-x_1} (1-p_1)^{n-j+1-x+x_1}, \tag{3.2}$$

where $A_{11,\delta=1}$ is the probability of this subscenario. As illustrated above, this subscenario always applies if $n \geq 2r, r \geq 0$.

***2. Shift occurs between* $\mathbf{t_{j-1}}$ *and* $\mathbf{t_j}$ *such that* $\boldsymbol{n_1 < r}$ *and* $\boldsymbol{n_2 \geq r}$**

$j-1 < r$ and $n-j+1 \geq r$ imply that $j < r+1$ and $j \leq n-r+1$, and hence, $1 \leq j \leq \min\{r, n-r+1\}$. Since $n \geq 2r$, substituting $\min\{n\} = 2r$ reveals that $n-r+1 = r+1$, and therefore, the shift could occur in the range $1 \leq j \leq r$. Because $n_1 < r$ and r is a nonnegative integer, r must be greater than 0, which means that this subscenario exists if $n \geq 2r, r \geq 1$. As seen in the quantity $A_{21,\delta}$ defined below, all nonconforming units can be fully found in n_1 or in n_2 given that $0 \leq x \leq j-1$ as shown in the first term. The second term represents the situation when the number of nonconforming units exceeds n_1 given that $j \leq x \leq r$.

$$A_{21,\delta} = \sum_{j=1}^{r} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right)$$

$$\left\{ \sum_{x=0}^{j-1} \sum_{x_1=0}^{x} \delta \binom{j-1}{x_1} {p_0}^{x_1} (1-p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} {p_1}^{x-x_1} (1-p_1)^{n-j+1-x+x_1} \right.$$

$$\left. + \sum_{x=j}^{r} \sum_{x_1=0}^{j-1} \delta \binom{j-1}{x_1} {p_0}^{x_1} (1-p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} {p_1}^{x-x_1} (1-p_1)^{n-j+1-x+x_1} \right\}, \tag{3.3}$$

where $A_{21,\delta=1}$ refers to the probability of this subscenario.

3. Shift occurs between $\mathbf{t_{j-1}}$ and $\mathbf{t_j}$ such that $\mathbf{n_1 \geq r}$ and $\mathbf{n_2 < r}$

$n_1 \geq r$ and $n_2 < r$ means that $j \geq r + 1$ and $j > n - r + 1$, and hence, the range in which the shift could occur is $\max\{r + 1, n - r + 2\} \leq j \leq n$. Substituting the minimum $n = 2r$ makes $n - r + 2$ always greater than $r + 1$. Therefore, the range of the shift's occurrence is $n - r + 2 \leq j \leq n$. In the quantity $A_{31,\delta}$ given below, X such that $0 \leq x \leq n_2$ can be found in $n_1$and/or n_2 without restrictions, whereas for $n_2 + 1 \leq x \leq r$, X cannot be fully found in n_2.

$$A_{31,\delta} = \sum_{j=n-r+2}^{n} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right)$$

$$\left\{ \sum_{x=0}^{n-j+1} \sum_{x_2=0}^{x} \delta \binom{j-1}{x-x_2} {p_0}^{x-x_2} (1-p_0)^{j-1-x+x_2} \binom{n-j+1}{x_2} {p_1}^{x_2} (1-p_1)^{n-j+1-x_2} \right.$$

$$+ \sum_{x=n-j+2}^{r} \sum_{x_2=0}^{n-j+1} \delta \binom{j-1}{x-x_2} {p_0}^{x-x_2} (1-p_0)^{j-1-x+x_2} \binom{n-j+1}{x_2} {p_1}^{x_2} (1$$

$$\left. - p_1)^{n-j+1-x_2} \right\}, \tag{3.4}$$

where $A_{31,\delta=1}$ stands for the probability of this subscenario. Note that $r \geq 2$ to satisfy the condition of $n - r + 2 \leq j \leq n$, and therefore, this subscenario applies if $n \geq 2r$ and $r \geq 2$.

Scenario 2.1. $\boldsymbol{n < 2r}$

1. Shift occurs between $\mathbf{t_{j-1}}$ ***and*** $\mathbf{t_j}$ ***such that*** $\boldsymbol{n_1 < r}$ ***and*** $\boldsymbol{n_2 < r}$

Since $j-1 < r$ and $n-j+1 < r$, a shift could occur in those intervals such that $n-r+2 \leq j \leq r$ as shown in the quantity $A_{12,\delta}$ defined below. Finding all x in $n_1$or in n_2 implies that $0 \leq x \leq \min\{j-1, n-j+1\}$ as seen in the first summation. In the second term, there is a minimum number of nonconforming units must be found in n_1 or in n_2. For instance, if $n = 10$ and $r = 7$, then $5 \leq j \leq 7$, and if the shift happens between t_6 and t_7, then to find $x = 7$, the range of X_1 is $3 \leq x_1 \leq 6$

$$A_{12,\delta} = \sum_{j=n-r+2}^{r} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right)$$

$$\left\{ \sum_{x=0}^{\min\{j-1,n-j+1\}} \sum_{x_1=0}^{x} \delta \binom{j-1}{x_1} {p_0}^{x_1} (1-p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} {p_1}^{x-x_1} (1-p_1)^{n-j+1-x+x_1} \right.$$

$$+ \sum_{x=\min\{j-1,n-j+1\}+1}^{r} \sum_{x_1=\max\{0,x-(n-j+1)\}}^{\min\{j-1,x\}} \delta \binom{j-1}{x_1} {p_0}^{x_1} (1-p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} {p_1}^{x-x_1} (1$$

$$\left. - p_1)^{n-j+1-x+x_1} \right\}, \tag{3.5}$$

where $A_{12,\delta=1}$ represents the probability of this subscenario. Since $n < 2r$ implies that $n_1 + n_2 \leq 2r-1$, but because $\max\{n_1\} = \max\{n_2\} = r-1$, this subscenario doesn't exist for $n = 2r-1$, and it only exists for $n < 2r-1$. Also, $r \geq 3$ to meet that $n-r+2 \leq j \leq r$, $n > r$, and $n < 2r$.

2. Shift occurs between $\mathbf{t_{j-1}}$ ***and*** $\mathbf{t_j}$ ***such that*** $\boldsymbol{n_1 < r}$ ***and*** $\boldsymbol{n_2 \geq r}$

Here, $1 \leq j \leq \min\{r, n-r+1\}$. Since $n < 2r$, $\max\{n\} = 2r-1$, and hence, $\max\{n-r+1\} = r$. Therefore, $n-r+1 \leq r$ and the range of j become $1 \leq j \leq n-r+1$. Let us define $A_{22,\delta}$ as follows:

$$A_{22,\delta} = \sum_{j=1}^{n-r+1} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right)$$

$$\left\{ \sum_{x=0}^{j-1} \sum_{x_1=0}^{x} \delta \binom{j-1}{x_1} p_0^{x_1} (1-p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} p_1^{x-x_1} (1-p_1)^{n-j+1-x+x_1} \right.$$

$$\left. + \sum_{x=j}^{r} \sum_{x_1=0}^{j-1} \delta \binom{j-1}{x_1} p_0^{x_1} (1-p_0)^{j-1-x_1} \binom{n-j+1}{x-x_1} p_1^{x-x_1} (1-p_1)^{n-j+1-x+x_1} \right\}, \qquad (3.6)$$

where $A_{22,\delta=1}$ is the probability of this subscenario. The conditions $1 \le j \le n-r+1$, $n > r, n_1 < r, n_2 \ge r$, and $n < 2r$ imply that $r \ge 3$ if $n < 2r-1$, and $r \ge 2$ if $n = 2r-1$.

3. Shift occurs between $\mathbf{t_{j-1}}$ and $\mathbf{t_j}$ such that $\mathbf{n_1 \ge r}$ and $\mathbf{n_2 < r}$

For this subscenario, $j-1 \ge r$ and $n-j+1 < r$, and therefore, $j \ge r+1$ and $j > n-r+1$. For $\max\{n\} = 2r-1$, $j \ge r+1$, and hence, the range of j is $r+1 \le j \le n$. Define $A_{32,\delta}$ as

$$A_{32,\delta} = \sum_{j=r+1}^{n} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right)$$

$$\left\{ \sum_{x=0}^{n-j+1} \sum_{x_2=0}^{x} \delta \binom{j-1}{x-x_2} p_0^{x-x_2} (1-p_0)^{j-1-x+x_2} \binom{n-j+1}{x_2} p_1^{x_2} (1-p_1)^{n-j+1-x_2} + \right.$$

$$\left. \sum_{x=n-j+2}^{r} \sum_{x_2=0}^{n-j+1} \delta \binom{j-1}{x-x_2} p_0^{x-x_2} (1-p_0)^{j-1-x+x_2} \binom{n-j+1}{x_2} p_1^{x_2} (1-p_1)^{n-j+1-x_2} \right\}, \qquad (3.7)$$

then $A_{32,\delta=1}$ is the probability of this subscenario. Again $r \ge 3$ if $n < 2r-1$ to satisfy that $r + 1 \le j \le n$, $n > r, n_1 \ge r$, $n_2 < r$, and $n < 2r$. Also, $r \ge 2$ if $n = 2r-1$ to meet the same conditions.

For scenarios 1.1 and 2.1, let us define the followings:

(1) a_0 returns 1 if $n \ge 2r$ & $r \ge 0$, and = 0 otherwise.

(2) a_1 returns 1 if $n \ge 2r$ & $r \ge 1$, and = 0 otherwise.

(1) a_2 returns 1 if $n \geq 2r$ & $r \geq 2$, and = 0 otherwise.

(2) a_3 returns 1 if $n < 2r - 1$ & $r \geq 3$, and = 0 otherwise.

(3) a_4 returns 1 if $n = 2r - 1$ & $r \geq 2$, and = 0 otherwise.

Consequently, the quantity A_δ is defined as follows:

$$A_\delta = a_0 A_{11,\delta} + a_1 A_{21,\delta} + a_2 A_{31,\delta} + a_3\left(A_{12,\delta} + A_{22,\delta} + A_{32,\delta}\right) + a_4\left(A_{22,\delta} + A_{32,\delta}\right), \quad (3.8)$$

where $A_{\delta=1}$ refers to the total probability of Case 1.2.

The above equation states that only one of Scenarios 1.1 and 2.1 is applied. Moreover, if $n \geq 2r$ and because of the equality condition, there must be at least one point such that $n_1 \geq r$ and $n_2 \geq r$ and it is not always true to have $n_1 < r$, $n_2 \geq r$ and/or $n_1 \geq r$, $n_2 < r$. For instance, if $r = 0$, a shift always occurs such that $n_1 \geq r$ and $n_2 \geq r$. If $n < 2r$, it is not always necessary to have $n_1 < r$ and $n_2 < r$. But it must have $n_1 < r$, $n_2 \geq r$ and $n_1 \geq r$, $n_2 < r$. For example, if $n = 5$, $r = 3$, there are no points such that $n_1 < r$ and $n_2 < r$. Instead, there are always some points such that $n_1 < r$, $n_2 \geq r$ and other points such that $n_1 \geq r$, $n_2 < r$. Accordingly, the first three terms in the above equation cover the Scenario 1.1, the fourth term covers the Scenario 2.1 given $n_1 < r$ and $n_2 < r$ and the last term covers the Scenario 2.1 given there are no points such that $n_1 < r$ and $n_2 < r$.

4.1.3. Case 2.1: Process has not shifted by time t_{r+1+i}

As depicted in Figure 3.6, a signal is alerted at time t_{r+1+i} since $X = r + 1$, but the process has found not shifted, i.e., a false alarm is alerted. The stoppage could be before or right upon t_n, and CL is the time until the first false alarm has occurred plus the preventive maintenance time, and therefore, sampling until the first false alarm occurs follows the truncated negative binomial distribution with $PON = p_0$ as shown by equation (3.9).

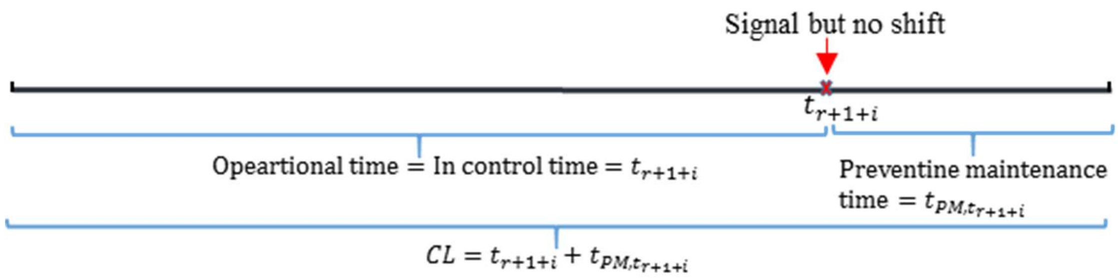

Figure 3.6. Inspection cycle ends with a false alarm.

Let $\varphi_{t_{r+1+i}}$ denotes the probability that a process ends at t_{r+1+i} with a signal alerted and no shift is observed. Then, $\varphi_{t_{r+1+i}}$ is given as

$$\varphi_{t_{r+1+i}} = e^{-\left(\frac{t_{r+1+i}}{\eta}\right)^{\beta}} \binom{r+i}{r} {p_0}^{r+1}(1-p_0)^{i}, i \in \{0, \cdots, n-r-1\}, \tag{3.9}$$

where i represents the number of conforming units found during inspection until the time of alerting a signal. For instance, $i = 0$ means that sampling is stopped when finding $r + 1$ nonconforming units consecutively since the beginning of sampling. In other words, $r + 1$ units are found nonconforming out of $r + 1$ units inspected. Also, $i = n - r - 1$ refers to that $r + 1$ nonconforming units are found by time t_n. Define $B_{1,\pi}$ as

$$B_{1,\pi} = \sum_{i=0}^{n-r-1} \pi\, \varphi_{t_{r+1+i}}, \tag{3.10}$$

where $\pi \in Q_2$, $Q_2 = \{1, t_{r+1+i}, E[V|\Omega_1^i], E[V|\Omega_2^{i,j}], E[V|\Omega_2^i]\}$, and $B_{1,\pi=1}$ refers to the probability that inspection ends with a false alarm.

4.1.4. Case 2.2: Process has shifted before or right upon t_{r+1+i}

For this subcase, the sampling plan alerts a signal at time t_{r+1+i} and a shift is confirmed, i.e., a true alarm is alerted. Figure 3.7 shows that CL is the sum of the time until the true alarm is alerted and the corrective maintenance time.

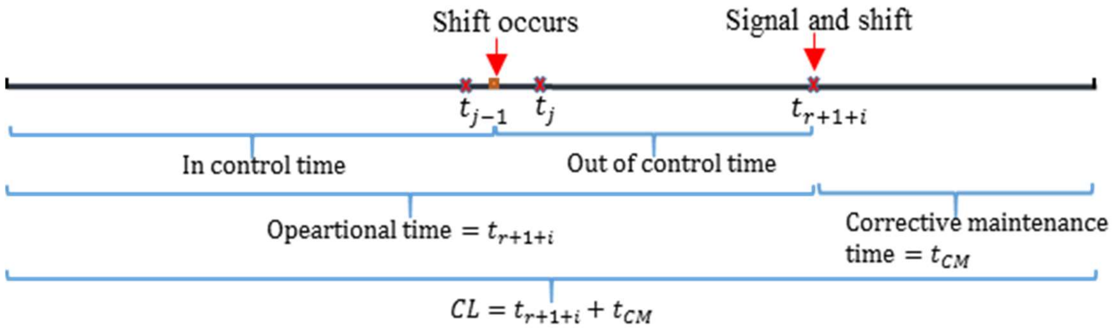

Figure 3.7. Inspection cycle ends with a true alarm.

Under this subcase, the total number of the inspected units is $k = r + 1 + i$, and the last nonconforming unit is found at t_{r+1+i}. If the shift occurs between t_{j-1} and t_j, $j \in \{1, \cdots, r + 1 + i\}$, then the number of samples "units" taken with p_0 is $k_1 = j - 1$, and the number of samples taken with p_1 is $k_2 = r + 1 + i - k_1 = r + 2 + i - j$. For all subscenarios illustrated below either in scenario 1.2 or scenario 2.2, the occurrence of a shift between t_{r+i} and t_{r+1+i} is not considered. Therefore, k_2 is reduced to $\acute{k_2} = r + 1 + i - j$. Since, for this case, inspection continues until a true alarm is signaled, sampling from $\acute{k_2}$ follows the truncated negative binomial distribution with $PON = p_1$, whereas sampling form k_1 follows the binomial distribution with $PON = p_0$.

Scenario 1.2. $r + i \geq 2r$

1. Shift occurs between t_{j-1} and t_j, $j \in \{1, \cdots, r + i\}$ such that $k_1 \geq r$ and $\acute{k_2} \geq r$.

In this subscenario, $j - 1 \geq r$ and $r + 1 + i - j \geq r$, and hence, $r + 1 \leq j \leq i + 1, i \geq r$. Because r could be 0, the upper bound is modified to $\min\{r + i, i + 1\}$. Again, the upper bound is modified to $\max\{\min\{r + i, i + 1\}, 1\}$ to consider the case when $r = 0$ and sampling ends at t_1 due to shift detection. Since $X = r + 1$ and the last nonconforming unit must be found in $\acute{k_2}$, the range of X_1 is $0 \leq x_1 \leq r$. The probability that inspection ends at t_{r+1+i} considering this subscenario $\phi_{11,t_{r+1+i}}$ is

$$\phi_{11,t_{r+1+i}} = \sum_{j=r+1}^{\max\{\min\{r+i,i+1\},1\}} \sum_{x_1=0}^{r} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}}\right) \binom{j-1}{x_1} {p_0}^{x_1} (1 - p_0)^{j-1-x_1} \binom{r+1+i-j}{r-x_1} {p_1}^{r+1-x_1} (1-p_1)^{i-j+1+x_1}.$$

Since $\max\{i\} = n-r-1$, this subscenario exists if $r + \max\{i\} \geq 2r$. In other words, it exists if $n - 2r - 1 \geq 0$ and $r \geq 0$.

Let us define $B_{11,\pi}$ as

$$B_{11,\pi} = \sum_{i=r}^{n-r-1} \pi \cdot \phi_{11,t_{r+1+i}}, \tag{3.11}$$

where $B_{11,\pi=1}$ represents the probability of this subscenario.

2. Shift occurs between $\mathbf{t_{j-1}}$ and $\mathbf{t_j}$, $\boldsymbol{j \in \{1, \cdots, r+i\}}$, such that $\boldsymbol{k_1 < r}$ and $\boldsymbol{\acute{k_2} \geq r}$

$j - 1 < r$ and $r + 1 + i - j \geq$ reveals that $1 \leq j \leq \min\{r, i+1\}$. Since $\min\{i\} = r$, then $\min\{i+1\} = r+1$, and hence, $1 \leq j \leq r$. Since $1 \leq j \leq r$, this subscenario applies if $n - 2r - 1 \geq 0$ and $r \geq 1$. The probability that inspection ends at t_{r+1+i} considering this subscenario $\phi_{21,t_{r+1+i}}$ is

$$\phi_{21,t_{r+1+i}} = \sum_{j=1}^{r} \sum_{x_1=0}^{j-1} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}}\right) \binom{j-1}{x_1} {p_0}^{x_1} (1 - p_0)^{j-1-x_1} \binom{r+1+i-j}{r-x_1} {p_1}^{r+1-x_1} (1-p_1)^{i-j+1+x_1}.$$

Define $B_{21,\pi}$ as

$$B_{21,\pi} = \sum_{i=r}^{n-r-1} \pi \cdot \phi_{21,t_{r+1+i}}, \tag{3.12}$$

where $B_{21,\pi=1}$ is the probability of the above subscenario.

3. Shift occurs between t_{j-1} and t_j, $j \in \{1, \cdots, r+i\}$, such that $k_1 \geq r$ and $\acute{k_2} < r$

$j-1 \geq r$ and $r+1+i-j < r$ imply that $\max\{r+1, i+2\} \leq j \leq r+i$. Since $\min\{i\} = r$ entails that $\min\{i+2\} = r+2$, $i+2$ is always greater than $r+1$, and then, the range of j is $i+2 \leq j \leq r+i$. The probability that inspection ends at t_{r+1+i} considering this subscenario $\phi_{31,t_{r+1+i}}$ is

$$\phi_{31,t_{r+1+i}} = \sum_{j=i+2}^{r+i} \sum_{x_2=0}^{r+1+i-j} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}}\right) \binom{j-1}{r-x_2} p_0{}^{r-x_2} (1 - p_0)^{j-1-r+x_2} \binom{r+1+i-j}{x_2} p_1{}^{x_2+1} (1-p_1)^{r+i-j+1-x_2}.$$

Define $B_{31,\pi}$ as

$$B_{31,\pi} = \sum_{i=r}^{n-r-1} \pi \cdot \phi_{31,t_{r+1+i}}, \tag{3.13}$$

where $B_{31,\pi=1}$ refers to the probability of the above subcenario. Because of $i+2 \leq j \leq r+i$, this subscenario applies if $n-2r-1 \geq 0$ and $r \geq 2$.

Scenario 2.2. $r+i < 2r$

The above subscenarios apply only if $r+i \geq 2r$, i.e., they only apply if $n-2r-1 \geq 0$ and $i \geq r$. However, an inspection cycle may end at t_{r+1+i} where $0 \leq i < r$ and $n-2r-1 \geq 0$. Also, if $n-2r \leq 0$, none of the above subscenarios exists because there is no i such that $i \geq r$. Thus, the subscenarios explained below may exist when $n-2r-1 \geq 0$ and $0 \leq i < r$ or when $n-2r \leq 0$, $0 \leq i \leq r$.

1. Shift occurs between t_{j-1} and t_j , $j \in \{1, \cdots, r+i\}$, such that $k_1 < r$ and $\acute{k_2} < r$

Since $j-1 < r$ and $r+1+i-j < r$, the range of j is $i+2 \leq j \leq r$. Because the last nonconforming unit must be found at t_{r+1+i}, the other r nonconforming units are found right upon or before t_{r+i}. It is impossible to find r nonconforming units in $\acute{k_2}$ since $\acute{k_2} < r$, and hence, there is a minimum number of nonconforming units that must be found in k_1. This equal $r - (r+1+i-j) = j-1-i$, and therefore, the range of X_1 is $j-1-i \leq x_1 \leq j-1$. The probability that sampling ends at t_{r+1+i} considering this subscenario $\phi_{12,t_{r+1+i}}$ is

$$\phi_{12,t_{r+1+i}} = \sum_{j=i+2}^{r} \sum_{x_1=j-1-i}^{j-1} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right) \binom{j-1}{x_1} {p_0}^{x_1} (1 - p_0)^{j-1-x_1} \binom{r+1+i-j}{r-x_1} {p_1}^{r+1-x_1} (1-p_1)^{i-j+1+x_1}.$$

Solving $r + i < 2r$ and $i + 2 \leq j \leq r$ entails that $0 \leq i \leq r-2$, and hence, $r \geq 2$ since $\min\{i\} = 0$. If $n - 2r - 1 \geq 0$ or $n = 2r$, then $0 \leq i \leq r-2$. If $n < 2r$, then $0 \leq i \leq n-r-1$. Thus, the range of i is $0 \leq i \leq \min\{r-2, n-r-1\}$. Define $B_{12,\pi}$ as

$$B_{12,\pi} = \sum_{i=0}^{\min\{r-2,n-r-1\}} \pi \cdot \phi_{12,t_{r+1+i}}, \tag{3.14}$$

where $B_{12,\pi=1}$ represents the probability of this subscenario.

2. Shift occurs between $\mathbf{t_{j-1}}$ and $\mathbf{t_j}$, $\boldsymbol{j = \{1, \cdots, r+i\}}$, such that $\mathbf{k_1 < r}$ and $\boldsymbol{\acute{k_2} \geq r}$

Because $j < r+1$ and $j \leq i+1$, the range of j is $1 \leq j \leq \min\{r, i+1\}$. Since $\max\{i\} = r-1$ as $r + i < 2r$, then $\max\{i+1\} = r$. Therefore, the range of j becomes $1 \leq j \leq i+1$, and the probability that inspection ends at t_{r+1+i} of this subscenario $\phi_{22,t_{r+1+i}}$ is

$$\phi_{22,t_{r+1+i}} = \sum_{j=1}^{i+1} \sum_{x_1=0}^{j-1} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right) \binom{j-1}{x_1} {p_0}^{x_1} (1 - p_0)^{j-1-x_1} \binom{r+1+i-j}{r-x_1} {p_1}^{r+1-x_1} (1-p_1)^{i-j+1+x_1}.$$

Solving $1 \leq j \leq i+1, r+i < 2r$, and $k_1 < r$ reveals that $0 \leq i \leq r-1$ and $r \geq 1$ if $n-2r-1 \geq 0$ or $1 \leq i \leq \min\{r-1, n-r-1\}$ and $r \geq 1$ if $n-2r \leq 0$. Thus, $B_{22,\pi}$ is defined as

$$B_{22,\pi} = \sum_{i=0}^{\min\{r-1,n-r-1\}} \pi\, \phi_{22,t_{r+1+i}}, \tag{3.15}$$

where $B_{22,\pi=1}$ denotes the probability of the above subscenario.

3. Shift occurs between $\mathbf{t}_{j-1}$ and $\mathbf{t}_j$, $j = \{1, \cdots, r+i\}$, such that $k_1 \geq r$ and $\acute{k}_2 < r$

Here, $j \geq r+1$ and $j > i+1$, and therefore, the range of j is $\max\{r+1, i+2\} \leq j \leq r+i$. Because $\max\{i\} = r-1$ as $r+i < 2r$, then $\max\{i+2\} = r+1$. Hence, $r+1 \leq j \leq r+i$ is the range of j. The probability that sampling ends at t_{r+1+i} in this subscenario $\phi_{32,t_{r+1+i}}$ is

$$\phi_{32,t_{r+1+i}} = \sum_{j=r+1}^{r+i} \sum_{x_2=0}^{r+1+i-j} \left(e^{-\left(\frac{t_{j-1}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_j}{\eta}\right)^{\beta}} \right) \binom{j-1}{r-x_2} {p_0}^{r-x_2} (1 - p_0)^{j-1-r+x_2} \binom{r+1+i-j}{x_2} {p_1}^{x_2+1} (1-p_1)^{r+i-j+1-x_2}.$$

Having $r+i < 2r$ and $r+1 \leq j \leq r+i$ needs $1 \leq i \leq r-1$ and $r \geq 2$ if $n-2r-1 \geq 0$ or $1 \leq i \leq \min\{r-1, n-r-1\}$ and $r \geq 2$ if $n-2r \leq 0$. Therefore, $B_{32,\pi}$ is defined as

$$B_{32,\pi} = \sum_{i=1}^{\min\{r-1,n-r-1\}} \pi\, \phi_{32,t_{r+1+i}}\,, \tag{3.16}$$

where $B_{32,\pi=1}$ expresses the probability of this subscenario.

Let us define the quantity $B_{2,\pi}$ as

$$B_{2,\pi} = b_0 B_{11,\pi} + b_1 B_{21,\pi} + b_2 B_{31,\pi} + b_3 \left(B_{12,\pi} + B_{32,\pi}\right) + \ b_4 B_{22,\pi}, \tag{3.17}$$

where $B_{2,\pi=1}$ represents the probability of Case 2 given that a shift has occurred within (0, t_{r+i}] and a true alarm is alerted at t_{r+1+i}, and

(1) b_0 returns 1 if $n - 2r - 1 \geq 0$ & $r \geq 0$, and = 0 otherwise.

(2) b_1 returns 1 if $n - 2r - 1 \geq 0$ & $r \geq 1$, and = 0 otherwise.

(3) b_2 returns 1 if $n - 2r - 1 \geq 0$ & $r \geq 2$, and = 0 otherwise.

(4) b_3 returns 1 if $r \geq 2$, and 0 otherwise.

(5) b_4 returns 1 if $r \geq 1$, and 0 otherwise.

Shift occurs between $\mathbf{t}_{r+i}$ and $\mathbf{t}_{r+1+i}$

Scenarios 1.2 and 2.2 exclude the occurrence of a shift in the last sampling interval, i.e., between t_{r+i} and t_{r+1+i}. Let $\sigma_{t_{r+1+i}}$ represent the probability that sampling ends at time t_{r+1+i} given that the shift occurs between t_{r+i} and t_{r+1+i}. Then, $\sigma_{t_{r+1+i}}$ is given by

$$\sigma_{t_{r+1+i}} = \left(e^{-\left(\frac{t_{r+i}}{\eta}\right)^{\beta}} - e^{-\left(\frac{t_{r+1+i}}{\eta}\right)^{\beta}}\right)\binom{r+i}{r} {p_0}^r (1 - p_0)^i p_1.$$

Define $B_{3,\pi}$ as follows:

$$B_{3,\pi} = \begin{cases} \sum_{i=0}^{n-r-1} \pi\, \sigma_{t_{r+1+i}}, & r > 0 \\ \sum_{i=1}^{n-r-1} \pi\, \sigma_{t_{r+1+i}}, & r = 0, \end{cases} \tag{3.18}$$

where $B_{3,\pi=1}$ is the probability of this scenario. Note that when $r = 0$, i starts from 1 because the case $r = 0$, $i = 0$ is already considered in the previous scenarios.

4.1.5. Cost of sampling

Let Y be the number of inspected units in each inspection cycle. In Case 1, $Y = n$, and in Case 2, $Y \in \{r + 1, \cdots, n\}$. Let c_{ins} denote the average cost of inspecting one unit, then the expectation $E[Y]$ and the expected cost of sampling $E[SC]$ are given as follows, respectively:

$$E[Y] = n\{A_{0,\delta=1} + A_{\delta=1}\} + \frac{B_{1,\pi=t_{r+1+i}} + B_{2,\pi=t_{r+1+i}} + B_{3,\pi=t_{r+1+i}}}{h}, \tag{3.19}$$

$$E[SC] = c_{ins}E[Y]. \tag{3.20}$$

4.1.6. Cost of rejected units found during inspection

Any nonconforming unit found during inspection is rejected without replacement with an average cost of c_{rej} per unit. In Case 1, $X \in \{0, \cdots, r\}$, and in Case 2, $X = r + 1$. Therefore, the expectation $E[X]$ and the expected cost of the rejected units $E[RC]$ are given as follows, respectively:

$$E[X] = A_{0,\delta=x} + A_{\delta=x} + (r+1)\{B_{1,\pi=1} + B_{2,\pi=1} + B_{3,\pi=1}\}, \tag{3.21}$$

$$E[RC] = c_{rej}E[X]. \tag{3.22}$$

4.1.7. Cost of uninspected nonconforming units

One risk associated with inspection is delivering a high percentage of nonconforming units to customers. Usually, the cost of a nonconforming unit received by a customer could cost more than the cost of producing the unit itself. Costs related to the producer's reputation, return, and decline in sales may be considered.

Let $T^{in}_{j-1,j}$ represent the time to shift from the in-control state to the out-of-control state given $(j-1)h < T^{in} < jh$, where T^{in} denotes the time to shift, i.e., time process stays in control. Then, the expectation $E\left[T^{in}_{j-1,j}\right]$ is given by

$$E\left[T^{in}_{j-1,j}\right] = \frac{\int_{(j-1)h}^{jh} t\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} dt}{e^{-\left(\frac{(j-1)h}{\eta}\right)^{\beta}} - e^{-\left(\frac{jh}{\eta}\right)^{\beta}}}.$$

Define the following events:

(1) Ω_{t_n}: inspection ends at time $t_n$with no shift and no alarm.

(2) $\Omega^{j}_{t_n}$: a shift occurs such that $(j-1)h \le T^{in} < jh,$ but no alarm is signaled by time t_n.

(3) Ω_1^i: inspection ends at time t_{r+1+i} with an alarm alerted and no shift is observed.

(4) $\Omega_2^{i,j}$: inspection ends at time t_{r+1+i} with an alarm alerted and a shift has occurred such that $(j-1)h \le T^{in} < jh$.

(5) Ω_3^i: inspection ends at time t_{r+1+i} with an alarm alerted and a shift has occurred such that $(r+i)h \le T^{in} < (r+1+i)h$.

Let μ be the average production rate, then the expected total number of products produced by time t_n is $nh\mu$. Given Ω_{t_n}, n units are inspected, and the expected total number of uninspected units that could be nonconforming $E\left[V \middle| \Omega_{t_n}\right]$ is obtained as

$$E\left[V \middle| \Omega_{t_n}\right] = (nh\mu - n)p_0 = np_0(h\mu - 1).$$

If the process has shifted before or right upon t_n, then the cycle length consists of the in-control period in which $PON = p_0$ and the out-of-control period in which $PON = p_1$. Thus, given $\Omega_{t_n}^j$, the expected total number of uninspected units that could be nonconforming $E\left[V \middle| \Omega_{t_n}^j\right]$ is

$$E\left[V \middle| \Omega_{t_n}^j\right] = \left(E\left[T_{j-1,j}^{in}\right]\mu - \left\lfloor \frac{E\left[T_{j-1,j}^{in}\right]}{h} \right\rfloor\right)p_0 + \left\{\left(nh - E\left[T_{j-1,j}^{in}\right]\right)\mu - \left\lceil \frac{nh - E\left[T_{j-1,j}^{in}\right]}{h} \right\rceil\right\}p_1,$$

$$\forall\, j \in J_{A_{11}}, j \in J_{A_{21}}, j \in J_{A_{31}}, j \in J_{A_{12}}, j \in J_{A_{22}}, j \in J_{A_{32}},$$

where,

$J_{A_{11}} = \{j: r+1 \le j \le \min\{n, n-r+1\}\}$,

$J_{A_{21}} = \{j: 1 \le j \le r\}, J_{A_{31}} = \{j: n-r+2 \le j \le n\}$,

$J_{A_{12}} = \{j: n-r+2 \le j \le r\}$,

$J_{A_{22}} = \{j: 1 \le j \le n-r+1\}$,

$J_{A_{32}} = \{j: r+1 \le j \le n\}$,

$\left\lfloor \frac{E\left[T_{j-1,j}^{in}\right]}{h} \right\rfloor$ represents the expected number of the inspected units in the in-control period, and

$\left\lceil \frac{E\left[T_{j-1,j}^{in}\right]}{h} \right\rceil$ represents the expected number of the inspected units in the out-of-control period.

Given Ω_1^i, the expected total number of uninspected units that could be nonconforming $E\left[V \middle| \Omega_1^i\right]$ is

$$E\left[V \middle| \Omega_1^i\right] = (r+1+i)(h\mu - 1)p_0,$$

$$\forall i \in I_{B_1}, I_{B_1} = \{i: 0 \le i \le n-r-1\}.$$

The expected total number of the uninspected nonconforming units $E\left[V \middle| \Omega_2^{i,j}\right]$ given $\Omega_2^{i,j}$ and $E\left[V \middle| \Omega_3^i\right]$ given Ω_3^i are obtained as follows, respectively:

$$E\left[V \middle| \Omega_2^{i,j}\right] =$$

$$\left(E\left[T_{j-1,j}^{in}\right]\mu - \left\lfloor \frac{E\left[T_{j-1,j}^{in}\right]}{h} \right\rfloor\right)p_0 + \left\{\left((r+1+i)h - E\left[T_{j-1,j}^{in}\right]\right)\mu - \left\lceil \frac{(r+1+i)h - E\left[T_{j-1,j}^{in}\right]}{h} \right\rceil\right\}p_1,$$

$\forall i \in I_{B_{11}} \& j \in J_{B_{11}}, I_{B_{21}} \& j \in J_{B_{21}}, \forall i \in I_{B_{31}} \& j \in J_{B_{31}}, \forall i \in I_{B_{12}} \& j \in J_{B_{12}}, \forall i \in I_{B_{22}} \& j \in J_{B_{22}}, \forall i \in I_{B_{32}} \& j \in J_{B_{32}}$,

$$E\left[V \middle| \Omega_3^i\right] = \left(E\left[T_{r+i,r+1+i}^{in}\right]\mu - \left\lfloor \frac{E\left[T_{r+i,r+1+i}^{in}\right]}{h} \right\rfloor\right)p_0 +$$

$$\left\{\left((r+1+i)h - E\left[T_{r+i,r+1+i}^{in}\right]\right)\mu - \left\lceil \frac{(r+1+i)h - E\left[T_{r+i,r+1+i}^{in}\right]}{h} \right\rceil\right\}p_1, \quad \forall i \in I_{B_3},$$

where

$I_{B_{11}} = \{i: r \le i \le n-r-1\}, J_{B_{11}} = \{j: r+1 \le j \le \max\{\min(r+i, i+1), 1\}, i \in I_{B_{11}}\}$,

$I_{B_{21}} = \{i: r \leq i \leq n-r-1\}, J_{B_{21}} = \{j: 1 \leq j \leq r\}, i \in I_{B_{21}}\}$,

$I_{B_{31}} = \{i: r \leq i \leq n-r-1\}, J_{B_{31}} = \left\{j: i+2 \leq j \leq r+i, i \in I_{B_{31}}\right\}$,

$I_{B_{12}} = \{i: 0 \leq i \leq \min{(r-2, n-r-1)}\}, J_{B_{12}} = \left\{j: i+2 \leq j \leq r, i \in I_{B_{12}}\right\}$,

$I_{B_{22}} = \{i: 0 \leq i \leq \min{(r-1, n-r-1)}\}, J_{B_{22}} = \left\{j: 1 \leq j \leq i+1, i \in I_{B_{22}}\right\}$,

$I_{B_{32}} = \{i: 1 \leq i \leq \min{(r-1, n-r-1)}\}, J_{B_{32}} = \left\{j: r+1 \leq j \leq r+i, i \in I_{B_{32}}\right\}$,

$I_{B_3} = \{i: 0 \leq i \leq n-r-1 \text{ if } r > 0\,,\;\; 1 \leq i \leq n-r-1 \text{ if } r = 0\,\}$, and

$$E\left[T^{in}_{r+i,r+1+i}\right] = \frac{\int_{(r+i)h}^{(r+1+i)h} t\,\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} dt}{e^{-\left(\frac{(r+i)h}{\eta}\right)^{\beta}} - e^{-\left(\frac{(r+1+i)h}{\eta}\right)^{\beta}}}, \qquad i \in I_{B_3}.$$

Based on the above events and calculations, the expected total number of uninspected nonconforming units produced in one inspection cycle $E[V]$ is obtained as

$$E[V] = E\left[V\middle|\Omega_{t_n}\right]A_{0,\delta=1} + A_{\delta=E\left[V\middle|\Omega^{j}_{t_n}\right]} + B_{1,\pi=E\left[V\middle|\Omega^{i}_{1}\right]} + B_{2,\pi=E\left[V\middle|\Omega^{i,j}_{2}\right]} + B_{3,\pi=E\left[V\middle|\Omega^{i}_{2}\right]}. \quad (3.23)$$

Let c_{cun} denote the average cost per each uninspected unit that found nonconforming, then the expected cost of the uninspected nonconforming units $E[CUN]$ in one inspection cycle is

$$E[CUN] = c_{cun}E[V]. \quad (3.24)$$

4.1.8. Time and cost of maintenance

The scheduled preventive maintenance is performed if no shift is observed at time t_n. Since it is scheduled (planned), it is performed with known and pre-allocated resources. Therefore, a constant average time T_0 is assumed for the scheduled maintenance. As explained previously, preventive maintenance at the time of false alarm depends on how much the process has shifted. Hence, the time of maintenance can be expressed as a function of the cumulative shift at time t_{r+1+i}. This equivalent to the cumulative density function (CDF) of the Weibull distribution by time t_{r+1+i}, and therefore, the average time of performing preventive maintenance $t_{PM,t_{r+1+i}}$ at sampling time t_{r+1+i} if a false alarm is alerted is obtained as

$$t_{PM,t_{r+1+i}} = t_0 + \left(1 - e^{-\left(\frac{t_{r+1+i}}{\eta}\right)^{\beta}}\right) T_0, \ i \in \{0, \cdots, n-r-2\}, \tag{3.25}$$

where $t_0 (< T_0)$ is fixed maintenance time.

For all $i \in \{0, \cdots, n-r-2\}$, $t_{PM,t_{r+1+i}} < T_0$. To satisfy this condition, $t_{PM,t_{n-1}} < T_0$, where $t_{PM,t_{n-1}}$ is the time of preventive maintenance performed if a false alarm is alerted in the previous sample before the scheduled maintenance time.

By corrective maintenance, the failed component is replaced. Therefore, the average cost of corrective maintenance C_{CM} is given by

$$C_{CM} = c_0 T_S + c_{CM} t_{CM} + c_{COM}, \tag{3.26}$$

where c_0 is the average cost for searching a false or a true alarm, T_S is the average time to conclude there is a shift, c_{CM} is the average cost per unit time of replacing the component, t_{CM} is the average time of replacement, and c_{COM} is the average cost of a new component.

Considering all kinds of maintenance, the expected total time and cost of performing maintenance $E[MT]$ and $E[MC]$ are given as follows, respectively:

$$\begin{aligned} E[MT] = {} & (T_{NS} + T_0)\left(A_{0,\delta=1} + \varphi_{t_n}\right) + \sum_{i=0}^{n-r-2} \left(T_{NS} + t_{PM,t_{r+1+i}}\right) \varphi_{t_{r+1+i}} \\ & + (T_S + t_{CM})\left\{A_{\delta=1} + B_{2,\pi=1} + B_{3,\pi=1}\right\}, \end{aligned} \tag{3.27}$$

$$E[MC] = (c_0 T_{NS} + c_{PM} T_0)\left(A_{0,\delta=1} + \varphi_{t_n}\right)$$

$$+\sum_{i=0}^{n-r-2}\left(c_0 T_{NS}+c_{PM} t_{PM,t_{r+1+i}}\right)\varphi_{t_{r+1+i}}+C_{CM}\{A_{\delta=1}+B_{2,\pi=1}+B_{3,\pi=1}\}, \tag{3.28}$$

where c_{PM} is the average cost per unit time of performing preventive maintenance, T_{NS} is the average time to conclude there is no shift if a false signal is alerted, and φ_{t_n} is the probability of signaling a false alarm at time t_n.

The first term in $E[MT]$ equation refers to the total time of scheduled preventive maintenance. This includes the time of detecting no shift and time of maintenance. Note that a false alarm could be alerted at the end of the cycle since sampling is carried out at time t_n, but because the scheduled maintenance time is reached, no false alarm maintenance is performed. The second term implies that false alarm maintenance could be carried at all sampling times except for the last sampling time, i.e., $0 \leq i \leq n-r-2$. The last term entails that total corrective maintenance time includes the time of detecting a shift and time of replacement.

4.1.9. Cost of lost production

Let c_{lp} be the average loss cost per each unit unproduced, then the expected cost of lost production due to the lost time in maintenance $E[LPC]$ is

$$E[LPC]=c_{lp}\mu E[MT]. \tag{3.29}$$

4.1.10. Expected inspection cycle length and cost

The length of the inspection cycle CL consists of the operational time plus the time spent in maintenance. The time of inspection is not included since the process is not stopped for inspection. The inspection cycle total cost CC consists of costs of sampling, rejected units, uninspected nonconforming units, maintenance, and lost production. According to all scenarios explained in

Cases 1 and 2, the expected cycle length $E[CL]$ and the expected total cycle cost $E[CC]$ are given as follows, respectively:

$$E[CL] = t_n\{A_{0,\delta=1} + A_{\delta=1}\} + B_{1,\pi=t_{r+1+i}} + B_{2,\pi=t_{r+1+i}} + B_{3,\pi=t_{r+1+i}} + E[MT], \qquad (3.30)$$

$$E[CC] = E[SC] + E[RC] + E[CUN] + E[MC] + E[LPC]. \qquad (3.31)$$

The first term in $E[CL]$ equation represents the cycle length if no signal is alerted, i.e., $CL = t_n = nh$. The second term refers to the weighted cycle length when a false or a true alarm is alerted at any sampling time, i.e., $CL = \{t_{r+1}, \cdots, t_n\}$. The last term is the expected downtime due to maintenance intervention.

4.1.11. Mathematical formulation of the integrated model

The sampling parameters and the scheduled maintenance time are determined by minimizing the long-run cost rate $LRCR$. By the renewal reward theory (Ross, 2003), $LRCR$ is the ratio of the expected total cost $E[CC]$ to the expected total time $E[CL]$ as

$$LRCR = \frac{E[CC]}{E[CL]}. \qquad (3.32)$$

The mathematical formulation of the optimization problem is

$$\min_{n,r,h} \quad LRCR \qquad (3.33)$$

$$\text{subject to} \quad (n-1)h < Z, \qquad (3.33.1)$$

$$n > r, \qquad (3.33.2)$$

$$n, r \in \text{integers}, r \geq 0, \;\; h > 0. \qquad (3.33.3)$$

In this mixed integer nonlinear programming (MINLP) problem, $LRCR$ is minimized with respect to $n, r,$ and h. The constraint shown by equation (3.33.1) is explained below. As illustrated in equations (3.33.2) - (3.33.3), n is a positive integer whereas r is a nonnegative integer, for which n is greater than r, and h is a continuous positive variable.

Constraint $(n-1)h < Z$

To meet that $t_{PM,t_{r+1+i}} < T_0$, the condition $t_{PM,t_{n-1}} < T_0$ must hold since $t_{PM,t_{n-1}}$is the longest time of preventive maintenance that can be performed at the time of false alarm. According to equation (3.25),

$$t_{PM,t_{n-1}} = t_0 + \left(1 - e^{-\left(\frac{t_{n-1}}{\eta}\right)^{\beta}}\right) T_0 = t_0 + \left(1 - e^{-\left(\frac{(n-1)h}{\eta}\right)^{\beta}}\right) T_0.$$

Solving

$$\left\{t_0 + \left(1 - e^{-\left(\frac{nh}{\eta}\right)^{\beta}}\right) T_0\right\} < T_0$$

reveals that

$$(n-1)h < \eta\left(-\ln\left(\frac{t_0}{T_0}\right)\right)^{1/\beta},$$

where

$$Z = \eta\left(-\ln\left(\frac{t_0}{T_0}\right)\right)^{1/\beta}.$$

4.2. Sampling model

In this model, there is no scheduled maintenance, and inspection continues until the first false alarm or a true alarm is alerted. The sampling model can be obtained by considering only Case 2 in the integrated model and by setting the upper bounds in all equations of Case 2 to ∞. However, a simpler formulation is provided below.

Let j denote the sample number at which the first false alarm occurs. Then, the probability that the first false alarm $P(FA_j)$ occurs at time t_j is

$$P(FA_j) = e^{-\left(\frac{jh}{\eta}\right)^{\beta}} \binom{j-1}{r} {p_0}^{r+1}(1-p_0)^{j-r-1}, j \in \{r+1, \cdots, \infty\}.$$

If the inspection cycle ends with a true alarm, the average number of samples until a true alarm is alerted equals the average number of trials γ until $r+1$ successes given a probability of success p. Generally, γ is the expected value of the negative binomial distribution, and it is given as (Montgomery, 2009)

$$\gamma = \frac{r+1}{p}.$$

Let N be the number of samples taken until a true alarm is alerted. Given that the shift occurs in the time interval $[(j-1)h, jh]$, $j \in \{1, \cdots, \infty\}$, and there are x_1 nonconforming units found in the $(j-1)$ samples, the conditional expectation $E[N|J=j, X_1=x_1]$ and $P(J=j, X_1=x_1)$ are given by:

$$E[N|J=j, X_1=x_1] = (j-1) + \frac{r+1-x_1}{p_1},$$

$$P(J=j, X_1=x_1) = \left(e^{-\left(\frac{(j-1)h}{\eta}\right)^{\beta}} - e^{-\left(\frac{jh}{\eta}\right)^{\beta}}\right)\binom{j-1}{x_1} {p_0}^{x_1}(1-p_0)^{j-1-x_1},$$

where $x_1 \in \{0, \cdots, j-1\}$ if $j \in \{1, \cdots, r\}$, and $x_1 \in \{0, \cdots, r\}$ if $j \in \{r+1, \cdots, \infty\}$.

4.2.1. Cost of sampling

Let C be the operational time in one inspection cycle, then its expected value $E[C]$ is

$$E[C] =$$

$$\sum_{j=r+1}^{\infty} jhP(FA_j) + \sum_{j=r+1}^{\infty}\sum_{x_1=0}^{r} E[N|J=j, X_1=x_1]h\,P(J=j, X_1=x_1)$$

$$+\,\omega \sum_{j=1}^{r}\sum_{x_1=0}^{j-1} E[N|J=j, X_1=x_1]h\,P(J=j, X_1=x_1), \tag{3.34}$$

where

$$\omega = \begin{cases} 0 & , \quad r = 0 \\ 1 & , \quad r \geq 1. \end{cases}$$

Then, the expected cost of inspection $E[SC_s]$ is given by

$$E[SC_s] = c_{ins}\, E[C]/h. \tag{3.35}$$

4.2.2. Cost of rejected units found during inspection

In the sampling model, always $r+1$ units are rejected, and hence, the expected rejection cost $E[RC_s]$ is

$$E[RC_s] = c_{rej}(r+1). \tag{3.36}$$

4.2.3. Cost of uninspected nonconforming units

If the inspection cycle ends at time jh because of a false alarm, the expected number of uninspected nonconforming units produced given a false alarm $E\left[V_s \middle| FA_j\right]$ is obtained as

$$E\left[V_s \middle| FA_j\right] = p_0 j(h\mu - 1), j \in \{r+1, \cdots, \infty\}.$$

Let $\tau_{j-1,j}$ denote the time elapsed since time $(j-1)h$ until shift's occurrence given that the shift occurs in $[(j-1)h, jh]$. Then, the expectation $E\left[\tau_{j-1,j}\right]$ is given by

$$E\left[\tau_{j-1,j}\right] = \frac{\int_{(j-1)h}^{jh} (t-(j-1)h)\, \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} dt}{e^{-\left(\frac{(j-1)h}{\eta}\right)^{\beta}} - e^{-\left(\frac{jh}{\eta}\right)^{\beta}}}.$$

Define TA_{j,x_1} as the event of alerting a true alarm given that the shift occurs in time interval $[(j-1)h, jh]$ and there are x_1 nonconforming units found in the $(j-1)$ samples. Then, the expected number of uninspected nonconforming units given a true alarm is

$$E\left[V_s \middle| TA_{j,x_1}\right] = \left(\left((j-1)h + E\left[\tau_{j-1,j}\right]\right)\mu - (j-1)\right)p_0 + \{(E[N|J=j, X_1=x_1]h -$$

$$E\left[\tau_{j-1,j}\right])\mu - E[N|J=j, X_1=x_1]\}p_1.$$

Therefore, the expected number $E[V_s]$ and the expected cost $E[CUN_s]$ of uninspected nonconforming units produced in one inspection cycle can be expressed as

$$E[V_s] = \sum_{j=r+1}^{\infty} E\left[V_s \middle| FA_j\right] P\left(FA_j\right) + \sum_{j=r+1}^{\infty} \sum_{x_1=0}^{r} E\left[V_s \middle| TA_{j,x_1}\right] P(J=j, X_1=x_1)$$
$$+\omega \sum_{j=1}^{r} \sum_{x_1=0}^{j-1} E\left[V_s \middle| TA_{j,x_1}\right] P(J=j, X_1=x_1), \tag{3.37}$$

$$E[CUN_s] = c_{cun} E[V_s]. \tag{3.38}$$

4.2.4. Time and cost of maintenance

As explained previously in the integrated model, the preventive maintenance at the time of the false alarm is a function of the Weibull distribution CDF and T_0. Although scheduled maintenance is not performed here, the variable preventive maintenance time is expressed as a fraction of T_0 as follows:

$$t_{PM,t_j} = t_0 + \left(1 - e^{-\left(\frac{jh}{\eta}\right)^{\beta}}\right) T_0,$$

where t_{PM,t_j} represents the time of preventive maintenance performed if a false alarm is alerted at time t_j.

The expected maintenance time and cost, $E[MT_s]$ and $E[MC_s]$, are given as follows, respectively:

$$E[MT_s] = \sum_{j=r+1}^{\infty} \left(T_{NS} + t_{PM,t_j}\right)P\left(FA_j\right) + \ (T_S + t_{CM})P(TA), \tag{3.39}$$

$$E[MC_s] = \sum_{j=r+1}^{\infty} \left(c_0 T_{NS} + c_{PM} t_{PM_j}\right)P\left(FA_j\right) + C_{CM}P(TA), \tag{3.40}$$

where $P(TA)$ denotes the total probability of alerting a true alarm and is given by

$$P(TA) = \sum_{j=r+1}^{\infty} \sum_{x_1=0}^{r} P(J = j, X_1 = x_1) + \omega \sum_{j=1}^{r} \sum_{x_1=0}^{j-1} P(J = j, X_1 = x_1).$$

4.2.5. Cost of lost production

The expected cost of lost production is

$$E[LPC_s] = c_{lp} \mu E[MT_s]. \tag{3.41}$$

4.2.6. Expected inspection cycle length and cost

The expected cycle length and cost, $E[CL_s]$ and $E[CC_s]$, are given as follows, respectively:

$$E[CL_s] = E[C] + E[MT_s], \tag{3.42}$$

$$E[CC_s] = E[SC_s] + E[RC_s] + E[CUN_s] + E[MC_s] + E[LPC_s] \tag{3.43}$$

4.2.7. Mathematical formulation of the sampling model

The mathematical formulation of the optimization problem under the sampling policy is given by

$$\min_{r,h} \qquad LRCR_s = \frac{E[CC_s]}{E[CL_s]} \tag{3.44}$$

Subject to $\quad r \in$ integers, $r \geq 0, \ \ h > 0,$ (3.44.1)

where $LRCR_s$ is the long-run cost rate of the sampling policy. In this formulation, the long-run cost rate of the sampling model $LRCR_s$ is minimized with respect to the two decision variables r and h as illustrated in equation (3.44). Like the integrated r is nonnegative and h is continuous positive as seen in equation (3.44.1).

4.3. Scheduled maintenance model

In this model, no sampling is carried out and maintenance is only performed at the scheduled time. The objective of this model is to find the optimal maintenance interval m at which maintenance is carried out and the long-run cost rate is minimized.

4.3.1. Cost of uninspected nonconforming units

Let $T_{in,m}$ represent the time until a shift has occurred, then $E[T_{in,m}]$ is obtained as

$$E[T_{in,m}] = \frac{\int_0^m t\left(\frac{t}{\eta}\right)^{\beta-1}\frac{\beta}{\eta}\, e^{-\left(\frac{t}{\eta}\right)^{\beta}}\,dt}{1-e^{-\left(\frac{m}{\eta}\right)^{\beta}}}.$$

At time m, the process is either found shifted or not and therefore, the expected number and cost of nonconforming units produced, $E[V_m]$ and $E[CUN_m]$, are given as follows, respectively:

$$E[V_m] = m\mu p_0 e^{-\left(\frac{m}{\eta}\right)^{\beta}} + \left(E[T_{in,m}]\mu p_0 + \left(m - E[T_{in,m}]\right)\mu p_1\right)\left(1-e^{-\left(\frac{m}{\eta}\right)^{\beta}}\right), \tag{3.45}$$

$$E[CUN_m] = c_{cun}E[V_m], \tag{3.46}$$

where $m\mu p_0$ represents the number of uninspected nonconforming units produced if the process has not shifted by time m, whereas $E[T_{in,m}]\mu p_0$ and $\left(m - E[T_{in,m}]\right)\mu p_1$ represent the expected number of nonconforming units produced in the in-control and out-of-control periods, respectively, if the process has shifted by time m.

4.3.2. Time and cost of maintenance

At time m, an investigation is performed to determine whether the process has shifted or not, and hence, preventive or corrective maintenance is carried out, accordingly. The expected time and cost of maintenance, $E[MT_m]$ and $E[MC_m]$, are given by

$$E[MT_m] = (T_{NS} + T_0)e^{-\left(\frac{m}{\eta}\right)^{\beta}} + (T_S + t_{CM})\left(1 - e^{-\left(\frac{m}{\eta}\right)^{\beta}}\right), \tag{3.47}$$

$$E[MC_m] = (c_0 T_{NS} + c_{PM} T_0)e^{-\left(\frac{m}{\eta}\right)^{\beta}} + C_{CM}\left(1 - e^{-\left(\frac{m}{\eta}\right)^{\beta}}\right). \tag{3.48}$$

As shown above, scheduled preventive maintenance with time T_0 is performed if the investigation reveals that there is no shift. Otherwise, scheduled corrective maintenance is carried out with time t_{CM}. For each kind of maintenance, the time of investigation (concluding no shift or shift) is added.

4.3.3. Cost of lost production

The expected cost of lost production $E[LPC_m]$ is obtained as

$$E[LPC_m] = c_{lp}\mu E[MT_m] \tag{3.49}$$

4.3.4. Expected inspection cycle length and cost

In the maintenance model, the length of the inspection cycle consists of the time of operation (time until scheduled maintenance) m and the time of maintenance. Since no sampling is carried out, the cycle cost consists only of costs of uninspected nonconforming units, maintenance, and lost production. The expected cycle length and cost of the maintenance model, $E[CL_m]$ and $E[CC_m]$, are given as follows, respectively:

$$E[CL_m] = m + E[MT_m], \tag{3.50}$$

$$E[CC_m] = E[CUN_m] + E[MC_m] + E[LPC_m] \tag{3.51}$$

4.3.5. Mathematical formulation of the maintenance model

The mathematical formulation of the optimization problem under the maintenance policy is

$$\min_{m} \qquad LRCR_m = \frac{E[CC_m]}{E[CL_m]}, \tag{3.52}$$

subject to $\quad m > 0.$ (3.52.1)

In this formulation, the long-run cost rate of the maintenance model $LRCR_m$ is minimized with respect to m as shown in equation (3.52). The decision variable m is continuous and positive as represented by equation (3.52.1).

5. Numerical example

In this section, a numerical example is provided to illustrate the proposed integrated maintenance and sampling plan. Sensitivity analysis is carried out to analyze the influence of model parameters on the obtained solutions. To depict the performance of the integrated plan, the optimal solutions are compared to the optimal solutions of the other two individual models of sampling and maintenance.

In an automatic powder coating line as shown in Figure 3.8, small fabricated steel products such as cleats and base plates are powder coated in a closed chamber where sampling by attributes is employed. The nozzles of the guns that spray powder on products could wear out due to usage and accumulation of dry powder and dirt. As a result, the contaminated powder will nonuniformly sprayed coatings.

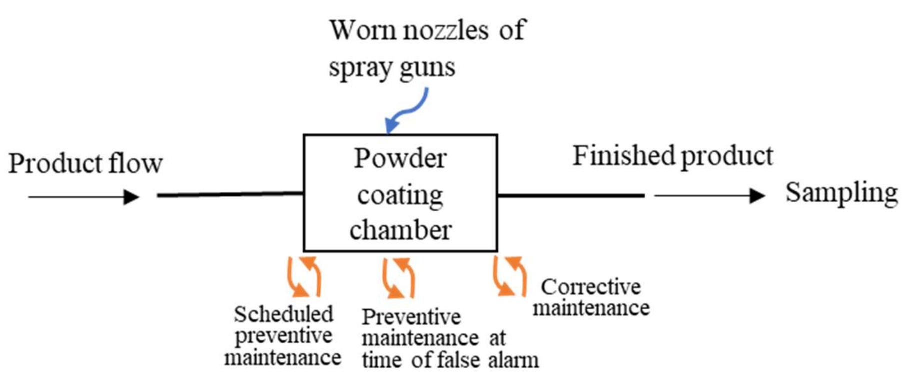

Figure 3.8. Maintenance and sampling in a powder coating system.

For this system, corrective maintenance is performed to replace worn nozzles, clean unworn ones, and perform other necessary adjustments. On the contrary, upon scheduled preventive maintenance, all nozzles are cleaned, and necessary adjustments are made. For both kinds of maintenance, the guns are dismantled and the whole spraying system is flushed. Preventive maintenance at the false alarm time aims at performing minimum cleaning of nozzles, and it depends on the accumulated dirt and power. The parameters related to this system are shown in Table 3.2. Specially, we have $c_{CM} > c_{PM}$, $t_{CM} > T_0$, and it is assumed that $T_{NS} > T_S$ as more time is often needed to assure exactly that there is no shift in the system.

Table 3.2: Parameters used in the numerical example

Cost parameters		Time parameters		Process parameters	
c_{lp} =5	c_0 =500	t_0 =0.5	T_S =0.25	p_0 =0.05	β =1.5
c_{rej} =5	c_{COM} =500	T_0 =1.5	T_{NS} =0.5	p_1 =0.10	η =25
c_{ins} =20	c_{PM} =1000	t_{CM} =3		μ =1000	
c_{cun} =35	c_{CM} =2000				

The objective function of the integrated model formulated in Section 4 is mathematically complex. Thus, in this study, the genetic algorithm (GA) is used to solve this problem. GA is a stochastic method that doesn't require derivatives and is able to search for different solutions within one operation, and hence, the chance of finding a global optimum and avoiding being trapped in the local optimum increases (Charongrattanasakul and Pongpullponsak, 2011).

5.1. Solution procedure

The GA solver in MATLAB R2019b is used to solve the mixed-integer nonlinear problem (MINLP) of the integrated model and the other two separate models of sampling and maintenance. The population size is chosen to be 10 since only three decision variables are to be determined.

Any supplied settings for mutation, crossover, and creation functions are ignored by the GA solver for integer problems. Instead, special mutation, crossover, and creation functions are used (MATLAB & Simulink, 2019). In order to make the search process for an optimal solution efficient, the constraint tolerance and the function tolerance are set to their default values, i.e., 1×10^{-3} and 1×10^{-6}, respectively. Moreover, UseParallel option is used to compute the fitness value and the nonlinear constraint feasibility in parallel in order to speed up the computation. GA is designed to stop if any of the following criteria is met:

- The maximum number of generations (iterations) is reached. Here, the default number is used (i.e., 100×number of decision variables).
- The average change in the penalty fitness value is less than the function tolerance over stall generations where the maximum stall generations is 50.
- Time limit is reached, i.e., GA runs for a specified time. The default setting is used, i.e., infinity.
- There is no improvement in the objective function during an interval of time called stall time limit. Here, the default setting of the stall time limit is used (i.e., infinity).

The relationship between n and h is given by the constraint $(n-1)h < Z$. As n increases, h decreases, and vice versa. If $n = 1$, h falls in the range $0 < h < \infty$. If $n = 2$, h falls in the range $0 < h < Z$, and if $n = 3$, h falls in the range $0 < h < Z/2$, and so on. Since h is a continuous variable, the ranges of n and h are very large, and hence, the solution space is huge. To facilitate computation, the problem is divided into subproblems with respect to r. At each level of r, $r = 0, 1, 2, \cdots$, GA is used to solve each subproblem according to the solution procedure illustrated below in Table 3.3. The same procedure is used to solve the individual models of sampling and scheduled maintenance.

When $r = 0$, $n \geq 1$ (since $n > r$) and $h \to \infty$. For $r = 1$, $n \geq 2$ and $h < Z$, and for $n \geq 3$, $0 < h < Z/2$. The upper bound of h when $r = 0$ is set as multiples of the upper bound of h when $r = 1$. Let h_r^u and n_r^u denote the upper bounds of h and n, respectively, at a given r. Also, let e be a positive integer multiplier. Then, $h_0^u = eZ$. As shown in the solution procedure in Table 3.3, $h_r^u = \frac{Z}{n_r^l - 1} = \frac{Z}{r}$, where n_r^l is the lowest value of n given r, i.e., $n_r^l = r + 1$. n_r^u and e are chosen by the user.

Table 3.3: Solution procedure to solve the MINLP of the integrated model.

Step 0. Begin with $r \leftarrow 0$. *Set* $h_r^u = \begin{cases} eZ, & r = 0 \\ \frac{Z}{r}, & r \geq 1 \end{cases}$, *and* $n_r^u = \begin{cases} > 1, & r = 0 \\ > n_{r-1}^u, & r \geq 1. \end{cases}$

Step 1. Solve the problem using GA.

Step 2. If $n^* = n_r^u$, *increase* n_r^u *and return to Step 1.*

Otherwise, if $1 \leq n^* < n_r^u$, *record the best* $LRCR(r, n^*, h^*)$, *and go to Step 3.*

Step 3. Increment r *by 1 and return to Step 0. Record the best* $LRCR(r + 1, n^*, h^*)$

Step 4. If $LRCR(r + 1, n^*, h^*) < LRCR(r, n^*, h^*)$, *then return to Step 3. ; otherwise, output the best solution* $LRCR(r, n^*, h^*)$.

The problem is solved with processor Intel(R) Core(TM) i7-7500U CPU @2.7GHz 2.90 GHz. 20 instances are solved at each r level. Table 3.4 shows the solution for each instance and the corresponding computational time (CPUT). Searching for an optimal solution stops at $r = 2$, at which, the optimal solution exceeds the optimal solution at $r = 1$. The problem is also solved for $r = 3$, and five instances are run, but $LRCR$ keeps increasing. The optimal solution at $r = 3$ is $n = 123$, $h = 0.203$, and $LRCR = 2652.64$, and the average CPUT=5653 seconds.

Therefore, the optimal solution obtained is r^* =1, n^* =53, and h^* =0.402. The optimal long-run cost rate is $LRCR^*$ =2637.26. The optimal solution of the corresponding sampling model is r^* =1, h^* =0.338, and $LRCR_s^*$ =2644.17, whereas the optimal solution of the corresponding maintenance model is m^* =15.56 and $LRCR_m^*$ =2655.20.

Table 3.4: Optimal solutions at different levels of r.

Instance number	$r=0$				$r=1$				$r=2$			
	n	h	$LRCR$	CPUT (sec)	n	h	$LRCR$	CPUT (sec)	n	h	$LRCR$	CPUT (sec)
1	20	0.946	2648.23	76.86	18	0.931	2651.56	112.85	88	0.282	2643.00	1416.23
2	50	0.481	2681.15	340.69	25	0.717	2646.79	228.49	55	0.361	2649.11	370.25
3	6	2.748	2647.67	26.22	49	0.424	2637.42	546.77	76	0.294	2643.25	1042.85
4	10	1.724	2645.99	31.55	20	0.851	2650.07	263.17	90	0.262	2642.04	2637.08
5	12	1.466	2645.80	36.09	36	0.506	2640.38	181.97	91	0.261	2642.04	2662.65
6	15	1.200	2646.20	54.05	92	0.276	2650.20	1696.2	66	0.321	2645.41	1313.13
7	16	1.159	2646.37	54.87	51	0.411	2637.29	443.90	92	0.258	2642.06	2868.17
8	13	1.366	2645.83	47.96	55	0.395	2637.29	639.74	89	0.265	2642.06	2367.75
9	13	1.375	2645.83	40.07	52	0.406	2637.70	532.77	91	0.260	2642.04	2160.75
10	5	3.278	2648.43	19.57	51	0.411	2637.29	635.13	72	0.278	2643.62	1604.80
11	11	1.585	2645.85	36.13	53	0.402	2637.26	622.84	84	0.270	2642.16	2257.04
12	57	0.471	2683.86	317.93	54	0.398	2637.27	716.70	90	0.261	2642.04	2609.31
13	4	4.037	2649.38	162.10	60	0.379	2637.59	770.12	89	0.263	2642.04	2515.38
14	14	1.287	2645.94	74.82	52	0.406	2637.70	659.54	32	0.540	2660.86	536.30
15	9	1.894	2646.24	33.79	58	0.379	2637.49	729.35	27	0.646	2664.00	275.22
16	12	1.472	2645.79	40.75	51	0.413	2637.30	216.85	93	0.257	2642.08	2798.39
17	6	2.748	2647.67	26.13	53	0.402	2637.26	606.17	89	0.263	2642.04	2372.55
18	14	1.291	2645.94	52.11	53	0.402	2637.27	719.82	93	0.257	2642.08	2822.81
19	21	0.906	2648.90	84.72	92	0.276	2650.20	1794.5	77	0.289	2642.95	1769.28
20	14	1.275	2645.98	49.26	51	0.411	2637.29	632.96	87	0.267	2642.09	2084.17
		Average (CPUT)		**80.28**				**637.50**				**1924.21**

5.2. Sensitivity analysis

To investigate the effect of the model parameters on the optimal solutions, the model parameters given in Table 3.2 are changed by ±25%. The sensitivity analysis is performed for the three models, and the results are illustrated in Table 3.5. As shown in Table 3.5, the integrated model generally outperforms the separate models of sampling and maintenance except for two cases. For the first case, decrementing p_1 by 25% makes p_1 not large relative to p_0, and hence, losses due to producing nonconforming units in the out-of-control state reduce. It may be unnecessary to perform a frequent inspection, and it is more economical to follow the maintenance or the integrated policies since the process becomes more stable. For the other case, as β is reduced by 25%, the sampling policy becomes more economical. Since the shift is a result of a component failure, low $\beta(> 1)$ means that the failure rate increases slower than a failure rate of higher β, and therefore, the maintenance policy becomes less preferable. In both cases, the integrated policy can be used since $LRCR$ is very close to $LRCR_m$ or $LRCR_s$.

Table 3.5: Sensitivity analysis on model parameters.

Incremented /decremented by	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
$+25\%T_0$	16.766	2762.88	1	0.328	2730.71	1	66	0.362	2727.24*
$-25\%T_0$	14.269	2536.91	2	0.251	2551.94	0	3	4.920	2535.37*
$+25\%\mu$	15.275	3268.41	2	0.233	3238.44	1	60	0.368	3235.66*
$-25\%\mu$	16.045	2041.63	1	0.360	2045.27	0	9	1.955	2037.03*
$+25\%\eta$	17.695	2544.52	2	0.271	2521.02	1	61	0.425	2517.36*
$-25\%\eta$	13.268	2817.17	1	0.302	2824.19	0	6	2.344	2812.67*
$+25\%\beta$	15.025	2567.88	2	0.211	2562.88	1	52	0.366	2550.67*
$-25\%\beta$	17.324	2767.20	2	0.321	2740.72*	2	110	0.327	2740.94
$+25\%t_0$	15.561	2655.20	2	0.261	2662.03	1	39	0.501	2650.72*
$-25\%t_0$	15.561	2655.20	1	0.312	2622.30	1	69	0.339	2619.10*
$+25\%c_{cun}$	12.951	3084.56	1	0.274	3057.02	1	64	0.307	3052.03*
$-25\%c_{cun}$	19.675	2198.60	1	0.468	2209.14	0	4	5.144	2196.37*
$+25\%c_{ins}$	15.561	2655.20	1	0.350	2656.68	1	46	0.443	2647.51*
$-25\%c_{ins}$	15.561	2655.20	2	0.232	2626.66	2	101	0.243	2624.92*
$+25\%c_{PM}$	15.940	2702.78	1	0.344	2692.31	1	52	0.412	2685.04*
$-25\%c_{PM}$	15.186	2607.25	1	0.332	2595.92	1	54	0.392	2589.35*
$+25\%c_{lp}$	18.347	2817.97	2	0.290	2810.09	1	46	0.512	2803.98*
$-25\%c_{lp}$	12.638	2471.14	1	0.274	2461.77	1	58	0.319	2456.30*
$+25\%T_{NS}$	16.013	2670.73	2	0.257	2663.68	1	48	0.444	2656.81*
$-25\%T_{NS}$	15.098	2638.94	1	0.317	2621.69	1	58	0.366	2616.01*
$+25\%c_{CM}$	15.346	2687.39	1	0.330	2677.91	1	52	0.398	2670.59*
$-25\%c_{CM}$	15.770	2622.89	2	0.250	2610.11	1	54	0.404	2603.75*
$+25\%t_{CM}$	15.092	2723.88	1	0.321	2716.21	1	48	0.411	2708.69*
$-25\%t_{CM}$	16.013	2583.71	2	0.256	2567.72	1	54	0.416	2562.67*
$+25\%p_0$	17.176	2966.08	2	0.312	2977.68	0	4	4.454	2963.73*
$-25\%p_0$	14.595	2340.34	1	0.268	2297.52	1	84	0.291	2293.74*
$+25\%p_1$	12.548	2763.99	2	0.207	2692.32	1	77	0.299	2690.90*
$-25\%p_1$	25.528	2486.41*	1	0.683	2520.64	0	2	12.92	2487.12
$+25\%c_{rej}$	15.561	2655.20	1	0.338	2644.35	1	53	0.404	2637.42*
$-25\%c_{rej}$	15.561	2655.20	1	0.338	2643.98	1	52	0.406	2637.12*
$+25\%c_{COM}$	15.544	2657.89	1	0.338	2646.99	1	53	0.402	2640.05*
$-25\%c_{COM}$	15.579	2652.51	1	0.339	2641.35	1	53	0.403	2634.47*
$+25\%T_S$	15.536	2659.02	1	0.337	2648.19	1	54	0.405	2641.24*
$-25\%T_S$	15.587	2651.37	1	0.339	2640.14	1	53	0.403	2633.26*
$+25\%c_0$	15.642	2657.98	1	0.342	2647.93	1	55	0.396	2640.86*
$-25\%c_0$	15.481	2652.40	1	0.334	2640.37	1	54	0.395	2633.68*

Noticeable changes in the decision variables have been observed when parameters such as η, β, T_0, μ, c_{cun}, c_{lp}, c_{ins}, p_0, and p_1 are changed. Some of those parameters have an obvious impact on all policies. For instance, m increases from 12.55 to 25.53 as p_1 decreases from 0.125 (+25%) to 0.075 (-25%) causing $LRCR_m$ to approach $LRCR$. For the same setup, h jumps to 12.92 and 0.68 in the integrated and sampling policies, respectively. Since p_1=0.075 becomes close to p_0=0.05, a smaller number of nonconforming units are produced, and hence, a frequent inspection may be unnecessary. Some other parameters have more influence on certain policies than others. For example, the maintenance policy is not affected by changing c_{ins} since there is no inspection carried out in this policy whereas, under the other two policies, a small c_{ins} allows n and r to increase, and h to decrease so more units can be inspected. The changes in c_{rej}, c_{COM}, T_S, c_{PM} and c_0 don't make a noticeable influence on m, r, n, and h in all policies. Further sensitivity analysis is performed on some parameters as illustrated as follows.

5.2.1. Effect of p_0

Table 3.6 and Figure 3.9.a depict the effect of p_0 on the optimal solutions of the three models. When p_0 is very small relative to p_1, n becomes very large relative to r, and therefore, the probability that the inspection cycle ends with a false alarm or a true alarm increases. This enables the process to stay more in the in-control state to avoid excessive costs of operating in the out-of-control state and benefits from the reduced maintenance cost of a false alarm. Since there is no inspection in the maintenance policy, the process may enter the out-of-control state earlier, and $LRCR_m$ becomes much higher than $LRCR$ and $LRCR_s$. For instance, $LRCR_m$ is 6.3% higher than $LRCR$ for p_0= 0.01875. Moreover, m reduces to avoid running longer in the out-of-control time. When p_0 increases and approaches p_1, the benefit of carrying out inspection reduces, and hence

m increases. For instance, when p_0= 0.8125, $h \approx m$ =25.2, and the integrated policy reduces to the maintenance policy since p_0 is very close to p_1, and there is no need for inspection.

Table 3.6: Effect of p_0 on the optimal solutions of all models.

p_0	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
0.01875	13.69	1864.12	1	0.1819	1756.22	1	150	0.1906	1753.18*
0.025	13.94	2023.24	1	0.2091	1939.04	1	130	0.2168	1936.37*
0.03125	14.24	2182.00	1	0.2379	2119.71	1	104	0.2513	2116.69*
0.0375	14.60	2340.34	1	0.2683	2297.52	1	84	0.2906	2293.74*
0.04375	15.03	2498.13	1	0.3012	2472.36	1	67	0.3388	2467.31*
0.05	15.56	2655.20	1	0.3382	2644.17	1	53	0.4016	2637.26*
0.05625	16.25	2811.31	2	0.2740	2812.71	1	41	0.4956	2803.40*
0.0625	17.18	2966.08	2	0.3117	2977.68	0	4	4.4539	2963.73*
0.06875	18.51	3118.84	1	0.5114	3138.68	0	3	6.3297	3117.91*
0.075	20.66	3268.31	1	0.6378	3293.39	0	2	10.4782	3268.26*
0.08125	25.16	3411.33*	1	0.9830	3436.23	0	1	25.2338	3411.94

5.2.2. Effect of p_1

As shown in Table 3.7 and Figure 3.9.b, when p_1 is slightly higher than p_0, the process almost runs with no considerable shift, and hence, m tends to be large and inspection may become unnecessary. When p_1 is 0.075 or 0.0875, $LRCR_m$ approximately equals $LRCR$. Even though h in the sampling policy increases at those values of p_1 in order to reduce inspection frequency and get some reduction in $LRCR_s$, $LRCR_s$ is still higher than the corresponding $LRCR$ and $LRCR_m$. On the other hand, as p_1 increases, inspection becomes more important, and it can be carried out either by the integrated or the sapling policies. For $0.075 \leq p_1 \leq 0.125$, it is more economical to carry out inspection by the integrated policy to benefit from the scheduled maintenance, while beyond p_1 =0.125, either the integrated or the sampling policies can be used. With r=2 and small h in both integrated and sampling policies, the process runs with frequent inspection and without a signal for a longer time. It is worth mentioning that the maintenance policy performs poorly with high p_1. For instance, $LRCR_m$ at p_1 =0.2,0.1875, 0.175, and 0.1625 is higher than $LRCR$ by 8.3%, 7.5%, 6.6%, and 5.7%, respectively.

Table 3.7: Effect of p_1 on the optimal solutions of all models.

p_1	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
0.075	25.53	2486.41*	1	0.6827	2520.64	0	2	12.9221	2487.12
0.0875	18.50	2583.61	1	0.4064	2600.94	0	4	4.7768	2582.11*
0.1	15.56	2655.20	1	0.3382	2644.17	1	53	0.4016	2637.26*
0.1125	13.79	2713.76	2	0.2208	2672.41	1	69	0.3247	2668.96*
0.125	12.55	2763.99	2	0.2069	2692.32	1	77	0.2990	2690.90*
0.1375	11.62	2808.32	2	0.1976	2707.42	1	85	0.2807	2707.16*
0.15	10.88	2848.17	2	0.1909	2719.26	2	139	0.1922	2718.34*
0.1625	10.28	2884.50	2	0.1859	2728.79	2	142	0.1869	2728.54*
0.175	9.77	2917.96	2	0.1820	2736.61	2	150	0.1826	2736.41*
0.1875	9.34	2949.02	2	0.1759	2743.19*	2	159	0.1683	2744.09
0.2	8.96	2978.05	2	0.1763	2748.59*	2	160	0.1674	2749.30

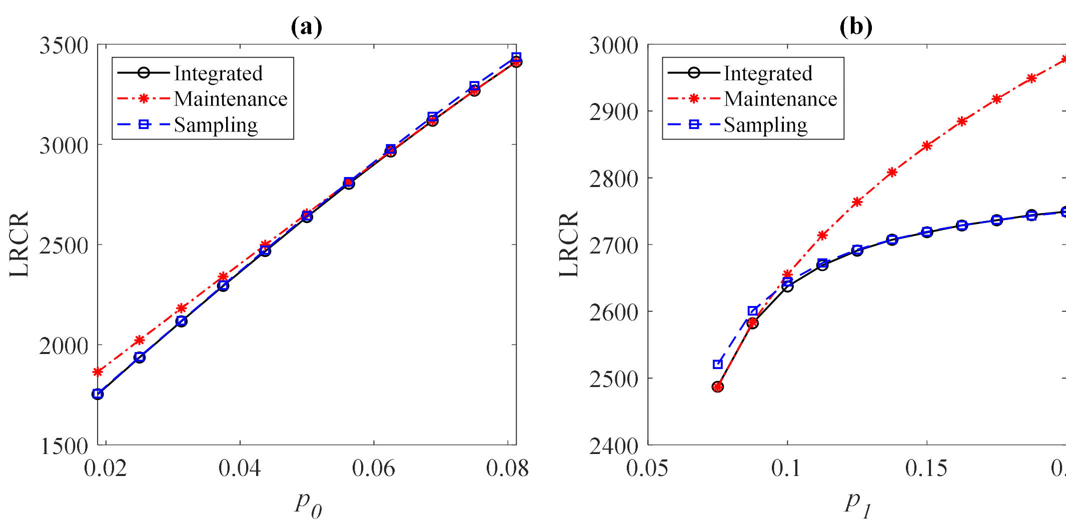

Figure 3.9. Effects of p_0 and p_1 on $LRCR$ of all policies.

5.2.3. Effect of η

Table 3.8 and Figure 3.10.a illustrate the influence η on the performance of the three policies.

Table 3.8: Effect of η on the optimal solutions of all models.

η	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
9.375	9.8817	3296.86*	1	0.3219	3349.05	0	2	5.0175	3298.12
12.5	10.8624	3083.35	1	0.2803	3119.56	0	3	3.717	3082.94*
15.625	12.0649	2931.42	1	0.287	2951.4	0	4	3.1292	2929.09*
18.75	13.268	2817.17	1	0.3021	2824.19	0	6	2.3437	2812.67*
21.875	14.4354	2727.62	1	0.3197	2724.53	1	47	0.4008	2715.89*
25	15.5614	2655.2	1	0.3382	2644.17	1	53	0.4016	2637.26*
28.125	16.6467	2595.2	2	0.258	2577.51	1	58	0.4098	2572.19*
31.25	17.6946	2544.52	2	0.2709	2521.02	1	61	0.4253	2517.36*
34.375	18.7082	2501.04	2	0.2842	2472.86	1	65	0.4365	2470.49*
37.5	19.6909	2463.24	2	0.2973	2431.25	1	68	0.4508	2429.9*
40.625	20.6452	2430.02	2	0.3101	2394.9	2	113	0.3195	2394.04*

As known, η refers to the characteristic life of a component, i.e., for the Weibull distribution, the time at which the CDF value equals 0.6321. Therefore, a smaller η means that the process takes a shorter time before a shift occurs since the failure rate becomes higher. Although inspection can help in detecting a shift, it might be subject to type II error. Hence, performing maintenance at shorter scheduled times with or without inspection is more economical for small values of η as shown for η ≤15.625 where $LRCR \leq LRCR_m < LRCR_s$.

On the contrary, as η increases, the failure rate slows down, and the process stays in control for a longer time. This permits to perform multiple inspections that could speed up shift detection or generate a false alarm. However, to maximize the benefit from operating in the in-control period, r increases to delay the occurrence of a false alarm as illustrated for η ≥28.125 in the sampling policy and for η ≥40.625 in the integrated policy.

Generally, $LRCR < LRCR_s < LRCR_m$ for 18.75≤ $\eta \leq 37.5$ since the integrated policy benefits from both the scheduled maintenance time and the reduced cost of false alarm. The integrated policy doesn't benefit much from the scheduled maintenance as η ≥40.625 since the failure rate becomes very slow. Overall, the maintenance or the integrated policies can be used for very low η, the integrated policy should be used for medium η, and the sampling or the integrated policies can be used for large η as seen in Figure 3.10.a.

5.2.4. Effect of β

In Table 3.9 and Figure 3.10.b, all policies are examined for different values of β.

Table 3.9: Effect of β on the optimal solutions of all models.

β	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
1.125	17.32	2767.20	2	0.3208	2740.72*	2	110	0.3268	2740.94
1.3125	16.21	2707.97	1	0.3797	2691.56*	2	112	0.2392	2695.06
1.5	15.56	2655.20	1	0.3382	2644.17	1	53	0.4016	2637.26*
1.6875	15.20	2608.70	1	0.3113	2601.36	1	52	0.3815	2591.37*
1.875	15.03	2567.88	2	0.2105	2562.88	1	52	0.3663	2550.67*
2.0625	14.96	2532.05	2	0.2021	2528.39	1	49	0.3712	2514.82*
2.25	14.98	2500.51	2	0.1967	2498.37	1	49	0.3665	2483.40*
2.4375	15.04	2472.65	2	0.1933	2472.17	1	47	0.3772	2455.81*
2.625	15.14	2447.93	2	0.1912	2449.30	1	46	0.3799	2431.45*
2.8125	15.25	2425.89	2	0.1900	2429.29	1	45	1.6179	2409.88*
3	15.39	2406.16	2	0.1892	2411.72	1	45	0.3881	2390.69*

As explained above in the effect of η, the sampling and the integrated policies are more economical than the maintenance policy when the failure rate is slow, and it also applies for β too. For instance, at time equals to η, the failure rate is β/η. Since β generally is small relative to η, the failure rate increases slowly with the increase of β. This justifies why sampling and integrated policies are preferable for a wide range of β. For instance, for very small β=1.125, the failure rate is very low, and it may not be necessary to have scheduled maintenance. Basically, for $\beta \leq$2.4375, $LRCR < LRCR_s < LRCR_m$ whereas, for β >2.4375, $LRCR < LRCR_m < LRCR_s$ as shown in Figure 3.10.b.

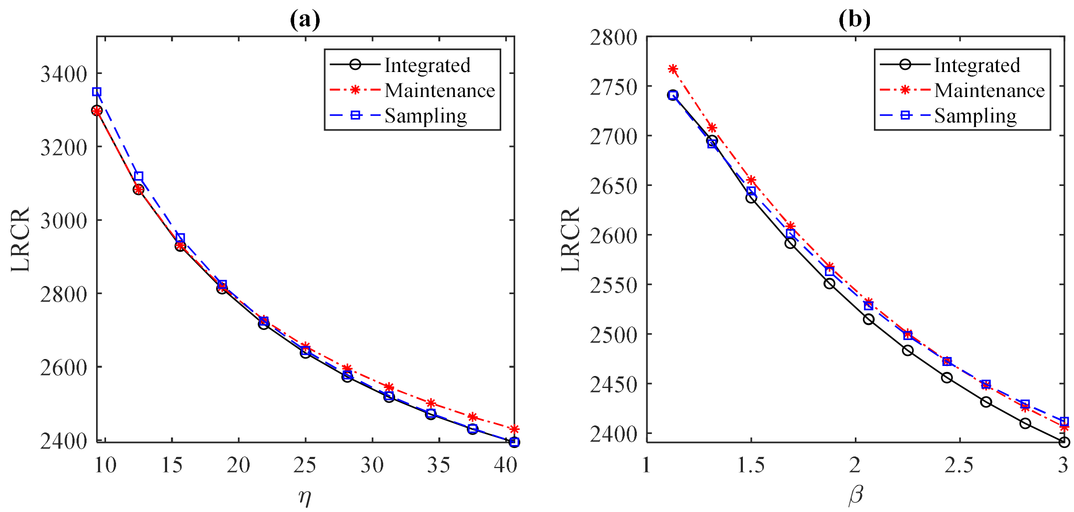

Figure 3.10. Effects of η and β on $LRCR$ of all policies.

5.2.5. Effect of c_{cun}

Table 3.10 and Figure 3.11.a show how c_{cun} affects the performance of all policies.

The expected total cost of the inspected units produced without inspection in one cycle depends on how many units produced with p_0 and p_1. The necessity of inspection increases as c_{cun} increases. While the maintenance policy is economical for low c_{cun}, the integrated policy dominates the other two policies for 21.875≤ c_{cun}. The flexibility of the integrated policy allows n to increase and h to decrease for higher c_{cun}, and hence, an alarm could be alerted earlier to avoid further losses. At the same time, the maximum cycle length is minimized so scheduled maintenance can be expedited. For clarification, the maximum cycle length when c_{cun}=48.125 is 70*0.2730=19.11 whereas, it is 73*0.2497=18.23 when c_{cun}=52.50.

Table 3.10: Effect of c_{cun} on the optimal solutions of all models.

c_{cun}	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
17.5	32.53	1691.47*	1	1.3224	1714.21	0	1	32.6330	1692.00
21.875	23.71	1954.67	1	0.6261	1975.85	0	2	12.0282	1954.33*
26.25	19.67	2198.60	1	0.4677	2209.14	0	4	5.1435	2196.37*
30.625	17.37	2431.05	1	0.3885	2430.36	1	45	0.4937	2421.90*
35	15.56	2655.20	1	0.3382	2644.17	1	53	0.4016	2637.26*
39.375	14.13	2872.69	1	0.3013	2852.68	1	58	0.3494	2846.87*
43.75	12.95	3084.56	1	0.2740	3057.02	1	64	0.3068	3052.03*
48.125	11.94	3291.45	1	0.2512	3257.91	1	70	0.2730	3253.59*
52.5	11.04	3493.82	1	0.2320	3455.80	1	73	0.2497	3451.99*
56.875	10.24	3691.96	1	0.2154	3650.99	1	75	0.2323	3647.60*

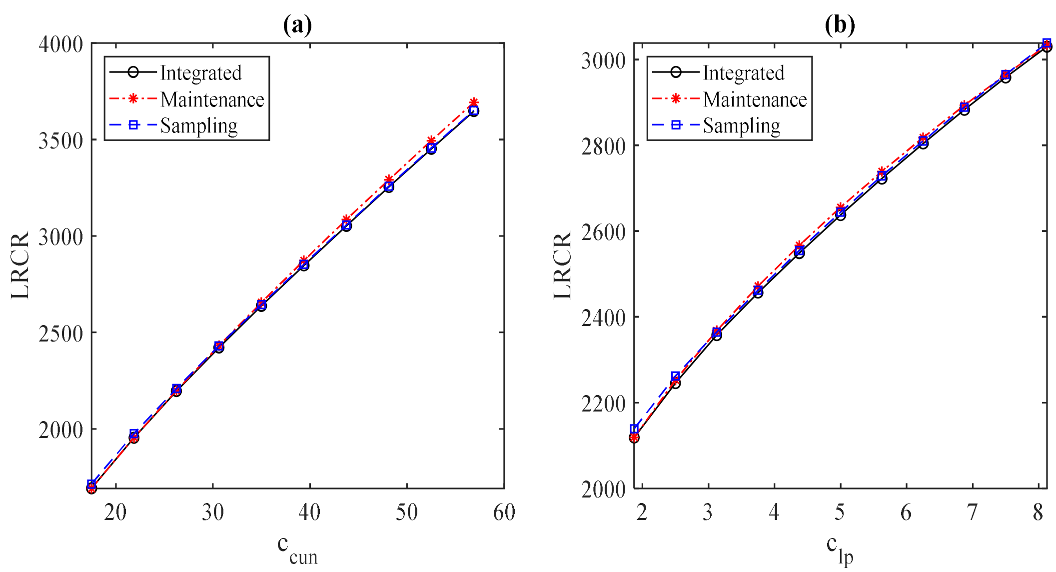

Figure 3.11. Effects of c_{cun} and c_{lp} on $LRCR$ of all policies.

5.2.6. Effect of c_{lp}

In Table 3.11 and Figure 3.11.b, the influence of c_{cun} on all policies is investigated.

Table 3.11: Effect of c_{lp} on the optimal solutions of all models.

c_{lp}	Maintenance model		Sampling model			Integrated model			
	m	$LRCR_m$	r	h	$LRCR_s$	r	n	h	$LRCR$
1.875	7.06	2119.20	0	0.2642	2139.22	0	8	0.9833	2118.48*
2.5	9.23	2252.66	1	0.2119	2261.99	0	17	0.6694	2245.53*
3.125	11.03	2367.67	1	0.2428	2364.48	0	17	0.7907	2357.22*
3.75	12.64	2471.14	1	0.2744	2461.77	1	58	0.3187	2456.30*
4.375	14.13	2566.37	1	0.3057	2554.81	1	55	0.3611	2548.74*
5	15.56	2655.20	1	0.3382	2644.17	1	53	0.4016	2637.26*
5.625	16.96	2738.81	2	0.2662	2728.75	1	50	0.4508	2722.22*
6.25	18.35	2817.97	2	0.2904	2810.09	1	46	0.5118	2803.98*
6.875	19.75	2893.24	2	0.3169	2888.84	1	41	0.5934	2882.49*
7.5	21.19	2965.00	2	0.3469	2964.98	1	36	0.6980	2957.81*
8.125	22.69	3033.53	2	0.3812	3038.46	1	30	0.8719	3029.93*

The amount of lost production depends on the production rate and the expected maintenance time. Surprisingly, the increase in c_{lp} causes m to increase which automatically increases the expected time and cost of maintenance. Such a trend might be due to that the increase in c_{lp} results in longer operational time in order to absorb the increase in the expected total cycle cost since $LRCR_m = E[CC_m]/E[CL_m]$. For the integrated policy, the increase in c_{lp} from 1.875 to 3.75 causes n to increase and h to decrease to end the cycle earlier by alerting an alarm. For c_{lp} from 4.375 to 8.125, n decreases and h increases in order to maximize the cycle length and benefit from operating with longer cycle length. The same applies to the sampling policy as r increases from 1 to 2. For the very low and the very high c_{lp}, the integrated policy dominates the other two policies mainly because of the scheduled maintenance whereas, for the middle values of c_{lp}, the integrated policy benefits more from inspection. This justifies why the sampling policy is less economical for low and high values of c_{lp} and why the integrated policy dominates the other two policies for the whole range of c_{lp}.

6. Conclusion and future work

In this chapter, an integrated model for attributes sampling and proactive maintenance is introduced with a special consideration on maintenance upon a false alarm. Since the inspection cycle may end with scheduled maintenance, a true alarm, or the first false alarm, the developed model is built with sampling based on binomial and truncated negative binomial distributions. The analysis demonstrates that the integrated model generally outperforms the separate models for maintenance and sampling and having multiple maintenance opportunities makes the integrated model more flexible than the separate alternatives. For instance, for $p_0 \leq 0.05$, the sampling model performs better than the maintenance model whereas, for $p_0 > 0.05$, the maintenance model is better. However, for the full range, the integrated model may be used instead. The sensitivity analysis shows the benefit of taking the maintenance opportunity upon a false alarm. Indeed, the integrated model benefits from the discounted false alarm maintenance more than the scheduled maintenance in some situations such as when p_0 is small relative to p_1 or when η is large.

This work can be extended in different ways for future research. First, the integrated model can be further expanded to involve production schedules and delays, and inventory. Second, gradual performance deterioration and complete failure of the production unit can be considered. Third, multiple shifts are often encountered in practice and thus are worth further investigation. Last, since the process shift may cause both the proportion of nonconforming and the production rate to deteriorate, advanced models need to be developed to handle the increased complexity.

7. References

Ardakan, M. A., Hamadani, A. Z., Sima, M. and Reihaneh, M. (2016) A hybrid model for economic design of MEWMA control chart under maintenance policies. *International Journal of Advanced Manufacturing Technology*, **83**(9–12), 2101–2110.

Barlow, R. E. and Hunter, L.C. (1960) Optimum preventive maintenance policies. *Operations Research*, **8**(1), 90-100.

Ben-Daya, M. and Duffuaa, S. O. (2003) Integration of Taguchi's loss function approach in the economic design of $\bar{x}$-chart. *Iinternational Journal of Quality & Reliability Management,* **20**(5), 607-619.

Ben-Daya, M. and Rahim, M. A. (2000) Effect of maintenance on the economic design of $\bar{x}$-control chart. *European Journal of Operational Research*, **120**(1), 131–143.

Cassady, R. C., Bowden, R. O., Liew, L. and Pohl, E. A. (2000) Combining preventive maintenance and statistical process control: a preliminary investigation. *IIE Transactions*, **32**(6), 471–478.

Charongrattanasakul, P. and Pongpullponsak, A. (2011) Minimizing the cost of integrated systems approach to process control and maintenance model by EWMA control chart using genetic algorithm. *Expert Systems with Applications,* **38**(5), 5178-5186.

Christopher, C. A., Kros, J. F. and Said, S. E. (2010) Economic design of $\overline{X}$ control charts with continuously variable sampling intervals. *Quality and Reliability Engineering International*, **26**(3), 235-245.

Das, T. K. and Sarkar, S. (1999) Optimal preventive maintenance in a production inventory. *IIE Transactions*, **31**(6), 537–551.

Dodge, H. F. (1943) A sampling inspection plan for continuous production. *The Annals of Mathematical Statisticst,* **14**(3), 264-279.

Duarte, B. P. M. and Saraiva, P. M. (2008) An optimization-based approach for designing attribute acceptance sampling plans. *International Journal of Quality & Reliability Management,* **25**(8), 824-841.

Duncan, A. J. (1956) The economic design of $\overline{X}$ charts used to maintain current control of a process. *Journal of the American Statistical Association*, **51**(274), 228-242.

El-Ferik, S. and Ben-Daya, M. (2006) Age-based hybrid model for imperfect preventive maintenance. *IIE Transactions,* **38**, 365–375.

Huang, Y.-S., Lin, Y.- J. and Ho, J.-W. (2013) A study on negative binomial inspection for imperfect production systems. *Computers & Industrial Engineering*, **65**, 605–613.

Huang, Y.-S., Lo, H.-C. and Ho, J.-W. (2008) Design of effective inspection schema for imperfect production systems. *International Journal of Production Research*, **46**(16), 4537–4551.

Huynh, K. T., Castro, I. T., Barros, A. and Bérenguer, C. (2012) Modeling age-based maintenance strategies with minimal repairs for systems subject to competing failure modes due to degradation and shocks. *European Journal of Operational Research*, **218**, 140–151.

Kaya, I. (2009) A genetic algorithm approach to determine the sample size for attribute control charts. *Information Sciences,* **179**(10), 1552-1566.

Lee, B. H. and Rahim, M. A. (2001) An integrated economic design model for quality control, replacement, and maintenance. *Quality Engineering*, **13**(4), 581–593.

Lee, J., Ni, J., Djurdjanovic, D., Qiu, H. and Liao, H. (2006) Intelligent prognostics tools and e-maintenance. *Computers in Industry*, **57**(6), 476-489.

Li, C., Su, Q. and Xie, M. (2016) Economic modelling for statistical process control subject to a general quality deterioration. *International Journal of Production Research*, **54**(6), 1753–1770.

Liao, W., Pan, E. and Xi, L. (2010) Preventive maintenance scheduling for repairable system with deterioration. *Journal of Intelligent Manufacturing,* **21**(6), 875-884.

Lin, Z.-L. and Huang, Y.-S. (2010) Nonperiodic Preventive Maintenance for Repairable Systems. *Naval Research Logistics*, **57**(7), 615-625.

Linderman, K., McKone-Sweet, K. E. and Anderson, J. C. (2005) An integrated system approach to process control and maintenance. *European Journal of Operational Research,***164**(2), 324-340.

Liu, L., Jiang, L. and Zhang, Di. (2017) An integrated model of statistical process control and condition-based maintenance for deteriorating systems. *Journal of the Operational Research Society*, **68**(11), 1452-1460.

Lorenzen, T.J. and Vance, L.C. (1986) The economic design of control charts: A unified approach. *Technometrics,* **28**(1), 3-10.

MATLAB & Simulink. (2019) Mixed Integer Optimization. https://www.mathworks.com/help/gads/mixed-integer-optimization.html#bs1cifg. Accessed October 22, 2019.

Mehrafrooz, Z. and Noorossan, R. (2011) An integrated model based on statistical process control and maintenance. *Computers & Industrial Engineering,* **61**(4),1245-1255.

Moghaddam, K. S. and Usher, J. S. (2010) Optimal preventive maintenance and replacement schedules with variable improvement factor. *Journal of Quality in Maintenance Engineering*, **16**(3), 271-287.

Montgomery,D.C.(2009) *Introduction to Statistical Quality Control*. 6th ed. Wiley: Hoboken, N.J.

Mujahid, S. N. and Rahim, M. A. (2010) Optimal preventive maintenance warranty policy for repairable products with periodically increasing failure rate. *International Journal of Operational Research*, **9**(2), 227-240.

Panagiotidou, S. and Tagaras, G. (2007) Optimal preventive maintenance for equipment with two quality states and general failure time distributions. *European Journal of Operational Research*, **180**(1), 329-353.

Pandey, D., Kulkarni, M. S. and Vrat, P. (2012) A methodology for simultaneous optimisation of design parameters for the preventive maintenance and quality policy incorporating Taguchi loss function. *International Journal of Production Research*, **50**(7), 2030–2045.

Radhoui, M., Rezg, N. and 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.

Rahim, M. A. (1993) Economic design of $\overline{\mathrm{X}}$ control charts assuming Weibull in-control times. *Journal of Quality Technology*, (25), 296-305.

Rahim, M. A. and Ben-Daya, M. (1998) A generalized economic model for joint determination of production run, inspection schedule and control chart design. *International Journal of Production Research*, **36**(1), 277–289.

Rasay, H., Fallahnezhad, M. S. and Mehrjerdi,Y. Z. (2018) An integrated model for economic design of chi-square control chart and maintenance planning. *Communications in Statistics - Theory and Methods*, **47**(12), 1-16.

Ross, S. M. (2003) *Introduction to Probability Models*. 8th ed. Academic Press, San Diego.

Safaei, A. S., Kazimzadeh, R. B. and Gan, H.-S. (2015) Robust economic-statistical design of $\overline{\mathrm{X}}$ control chart. *International Journal of Production Research*, **53**(14), 4446-4458.

Samrout, M., Chatelet, E., Kouta, R. and Chebbo, N. (2009) Optimization of maintenance policy using the proportional hazard model. *Reliability Engineering & System Safety*, **94**(1), 44–52.

Saniga, E. M. (1989) Economic statistical control-chart designs with an application to $\overline{\mathrm{X}}$ and R charts. *Technometrics*, **31**, 313–320.

Seif, A., Faraz, A. and Saniga, E. (2015) Economic statistical design of the VP $\overline{\mathrm{X}}$ control charts for monitoring a process under non-normality. *International Journal of Production Research*, 53(14), 4218-4230.

Serel, D. A. (2009) Economic design of EWMA control charts based on loss function. *Mathematical and Computer Modeling*, **49**, 745–759.

Shrivastava, D., Kulkarni, M. S. and Vrat, P. (2016) Integrated design of preventive maintenance and quality control policy parameters with CUSUM chart. *International Journal of Advanced Manufacturing Technology*, **82**(9–12), 2101–2112.

Tagaras, G. (1988) An Integrated Cost Model for the Joint Optimization of Process Control and Maintenance. *Journal of the Operational Research Society*, **39**(8), 757–766.

Xiang, Y. (2013) Joint optimization of $\overline{\mathrm{X}}$ control chart and preventive maintenance policies: A Discrete-Time Markov Chain approach. *European Journal of Operational Research,* **229**(2), 382-390.

Yeong, W. C., Khoo M. B. C., Lee, M. H. and Rahim, M. A. (2013) Economic and Economic Statistical Designs of the Synthetic $\overline{\mathrm{X}}$ Chart using Loss Functions. *European Journal of Operational Research*, **228**(3), 571–581.

Yeung, T. G., Cassady, C. R. and Schneider, K. (2008) Simultaneous optimization of $\overline{\mathrm{X}}$ control chart and age-based preventive maintenance policies under an economic objective. *IIE Transactions*, **40**(2), 147–159.

Yin, H., Zhang, G., Zhu, H., Deng, Y. and He, F. (2015) An integrated model of statistical process control and maintenance based on the delayed monitoring. *Reliability Engineering and System Safety*, **133**, 323–333.

Zhong, J. and Ma, Y. (2017) An integrated model based on statistical process control and maintenance for two-stage dependent processes.*Communications in Statistics: Simulation and Computation,* **46**(1), 106-126.

Zhou, W.-H. and Zhu, G.-L. (2008) Economic design of integrated model of control chart and maintenance management. *Mathematical and Computer Modelling*, **47**(11–12), 1389–1395.

Zong, S., Chai, G., Zhang, Z. G. and Zhao, L. (2013) Optimal replacement policy for a deteriorating system with increasing repair times.*Applied Mathematical Modlling,* **37**(23), 9768–9775.

Chapter 4 : Conclusions and Future Research

Quality and maintenance are two key operational areas in any production system. Sampling is an SPC approach that is used to alert for unusual variations that may occur in production systems. The aims of sampling are to control the quality of produced products and to alert for maintenance if a shift in a process has occurred.

This dissertation focuses on modelling sampling schemes with attributes in production systems that produce discrete products. Two studies are conducted in Chapters 2 and 3. While a sampling model in a multistage system with multiple assignable causes is developed in Chapter 2, an integrated model of sampling and scheduled maintenance is proposed in Chapter 3.

1. Conclusions

In Chapter 1, a multistage system of two unreliable machines is studied with one assignable cause that could occur on each machine. The study mainly aims at modelling such systems where a propagating shift can occur downstream. Therefore, a stochastic methodology based on the competencies of shifts is constructed to model all possible scenarios. This methodology forms the base that could be extended for more than two-machine systems. The sampling model is built to compromise between the quantity and the quality of the produced products. An economic-statistical design with constraints on availability, effective production rate, and average time to signal is built. Results show the applicability and usability of the model. Moreover, sensitivity analyses on some parameters and decision variables provide managerial insights that could be beneficial for practitioners.

In Chapter 2, an integrated model of sampling and scheduled maintenance is proposed. In this model, the scheduled maintenance and the parameters of sampling are determined jointly. Multiple maintenance opportunities are offered by this model. In addition to the traditional maintenance opportunities at the times of true alarm and scheduled maintenance, a preventive maintenance opportunity at the time of the false alarm is included. The maintenance time of this opportunity is assumed to increase with time, and hence, the maintenance time is expressed as a function of the Weibull distribution CDF. Results show that the integrated model generally outperforms the separate models of maintenance and sampling. The flexibility of the integrated model makes it used instead of the separate models for wide ranges of model parameters. Furthermore, the benefit of having a false alarm opportunity is illustrated in the sensitivity analysis. Last, the developed model can be used for sampling with the truncated negative binomial distribution.

2. Future research

Three topics are proposed for future research. The first topic aims at developing an integrated model of selective maintenance and sampling on a single machine. An integrated model of selective maintenance and sampling is explained in the second topic. Those two topics are proposed because of all selective maintenance model focus on maximizing a quantitative measure and ignore the quality of produced products. The third topic extends the study investigated in Chapter 2 by considering more than two competing shifts.

2.1. An integrated model of sampling and selective maintenance on a single machine

Selective maintenance is a new topic in maintenance that is first introduced by Rice et al. (1998). Some systems are required to perform a sequence of operations (or missions) with finite breaks between each operation. It may be impossible to perform all desirable maintenance activities within the maintenance break and before the beginning of the next mission due to limitations on maintenance resources (Cassady, 2001). Figure 4.1 illustrates the basic selective maintenance model.

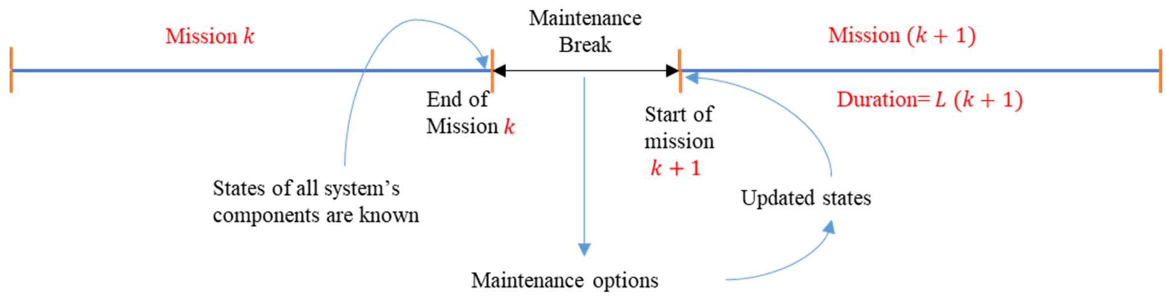

Figure 4.1: The basic selective maintenance model.

A shown in Figure 4.1., the system is available for maintenance at the end of mission k. An essential assumption of all selective maintenance models is that the states of the system's components are exactly known at the end of mission k and before starting maintenance. A component at the end of mission k could be in the operational state or in the failure state (binary-state component).

A component could also have a nominal operational state, degraded operational states, and a failure state (multistate component). Due to the limited maintenance resources during the maintenance break, not all components may be maintained. Given the duration of the next mission $L(k+1)$, maintenance break time, and maintenance options for each component, a subset of components is selected for maintenance such that the performance of the system in the next mission $k+1$ is maximized. The system's reliability and availability are some examples of system performance.

The integrated design of sampling and selective maintenance aims at maximizing the availability of a production system. The production system (or production unit) operates for a definite period. At the end of this period, the production unit is available for maintenance. It is assumed that the production unit consists of different components, and the failure of any component causes the failure of the production unit.

During operation, the production process is monitored by a sampling plan by attributes to ensure the quality of products produced. A presence of an assignable cause causes the proportion of nonconforming (PON) products produced to increase. The shift in quality (an increase of PON) cannot be observed without inspecting products. The necessary adjustment to the process is made when a shift is detected by sampling.

Both the system failure and the shift in quality require to stop the system for restoration. It is assumed that if a system failure occurs during operation, minimal repair is performed to restore the system to the operational state. However, if a quality shift is detected, the assignable cause is removed. In both cases, the system is nonoperational (unavailable). Accordingly, the objective of the integrated design is to find the best setup of sampling parameters and determine the subset of components selected for maintenance such that the system's availability in the next operational period is maximized. As shown in Figure 4.2, the operational period consists of multiple sampling cycles where each cycle ends with shift removal. In each cycle, the system is stopped if a false or a true alarm is alerted. It also stopped for maintenance to remove the assignable cause.

In this model, the states of the components are expressed by their ages. A component that receives replacement starts the next production period with age zero. A component that is not selected for maintenance begins the next duration of the same age that ends within the previous duration. Imperfect maintenance reduces the age of a selected component to an age greater than 0 and less than the age it ends within the previous duration. Different models proposed in the literature can be used to model imperfect maintenance. Some of these models are the Kijima type I model (Kijima et al., 1988), Kijima type II model (Kijima,1989), improvement factor method (Malik, 1979), and hybrid imperfect model (Lin et al., 2000). In some cases, it is not necessary to maintain a component to its perfect state to achieve the desired system's performance in the next mission. Instead, an imperfect repair can be carried out with less time and cost. Therefore, an imperfect repair can reduce the overall maintenance cost/time.

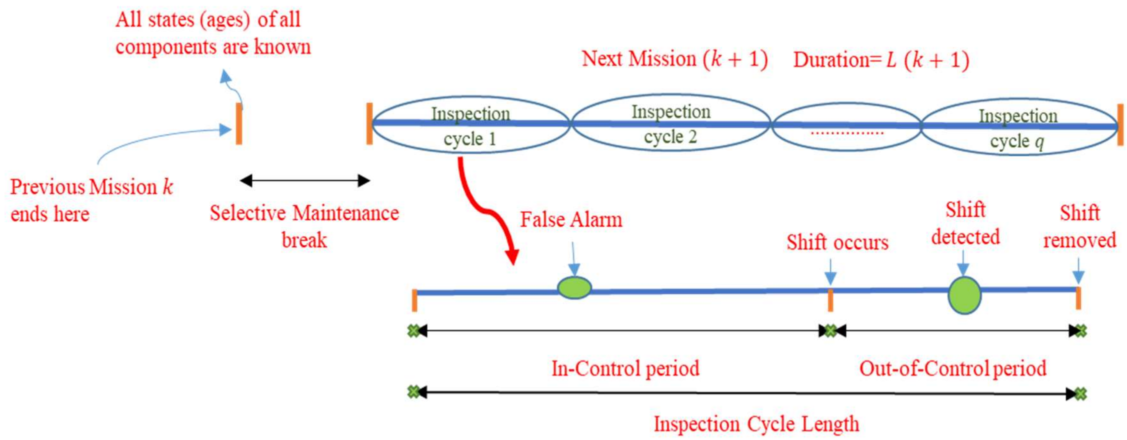

Figure 4.2. The integrated model of sampling and selective maintenance.

The above description of the integrated model assumes that the shift and the failure processes are independent. However, these processes could be dependent such that the increasing failure rate of a component could increase the shift rate. Also, the failures of all components are assumed to be stochastically independent. The failure dependency between components could be also considered.

2.2. An integrated model of sampling and selective maintenance for a multistage multistate system (MSS) with discrete states of the output

Many production lines consist of machines connected in a series-parallel configuration. The whole system consists of subsystems connected in series, and each subsystem consists of independent machines connected in parallel. Each machine has more than one discrete state of the output. Such systems are called multistage multistate systems (MSS). For clarification, consider the series-parallel system shown in Figure 4.3 below.

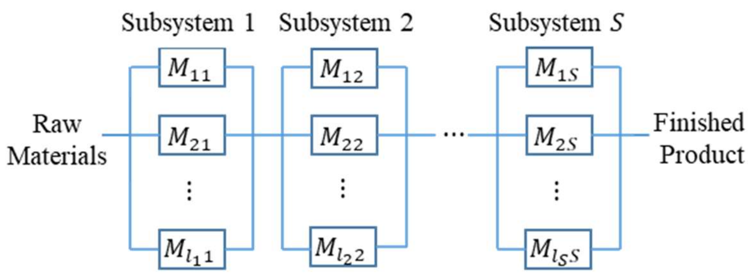

Figure 4.3. MSS with a series-parallel configuration.

The whole system consists of S series subsystems where production flows from subsystem 1 to subsystem 2 until a finished product is produced at subsystem S. Each subsystem consists of $l_s, s \in \{1, \cdots, S\}$ independent machines connected in parallel where a product can be processed at any of these machines. Each machine has multiple degradation states where each state corresponds to a discrete performance rate. Let i be the state of machine j in subsystem s at time t, then

$$i \in \{0, \cdots, m_{js}\}.$$

The corresponding performance rates of the above set are represented by the set

$$g_{ijs} \in \{g_{0js}, \cdots, g_{m_{js}js}\},$$

where $g_{m_{js}js}$ is the highest performance rate that corresponds to state m_{js} and g_{0js} is the lowest performance rate that corresponds to the lowest state 0. When a machine degrades from a higher state to a lower state, the performance rate decreases accordingly. Usually, the performance rate g_{ijs} is measured by the capacity, i.e., production rate.

The MSS performance rate $G(t)$ at any time t is a random variable that depends on the machines' states in all subsystems, and hence, $G(t)$ can be represented as

$$G(t) = \Phi\left(g_{011}(t), \cdots, g_{il_S S}(t)\right),$$

where $\Phi(\cdot, \cdots, \cdot)$ is a function of machines' performance rates at time t.

Let $F_s(t)$ be the performance of subsystem s at time t, then

$$F_s(t) = \sum_{j=1}^{l_s} g_{ijs}(t).$$

Following these definitions, $G(t)$ can be rewritten as

$$G(t) = \min_{S}\{F_s(t)\}.$$

At the end of the production period, all machines' states are known, and maintenance is carried out selectively within a limited time. With selective maintenance, a machine state could be upgraded to a better state or not. The objective of selective maintenance is to maximize the system reliability in the next production period within the available maintenance resources. The reliability of the system is defined as the probability that the output of the system exceeds a certain demand, and hence, system reliability R can be defined as the probability that the system successfully completes the next mission $k+1$, i.e, the system's performance rate exceeds the demand rate W_{k+1} at the end of mission $k+1$. In other words, R is given as

$$R = P\ (G(t) \geq W_{k+1}).$$

Constant transition rates are assumed, and a continuous-time Markov chain for each machine can be used to find the probabilities of states at the end of the production period. Such studies of the series-parallel system can be found in Pandey et al. (2013) and Liu and Huang (2010). These studies develop quantitative models of selective maintenance and don't consider the quality of the produced products. Therefore, the integrated model of sampling and selective maintenance for MSS aims at maximizing the system's reliability (or availability) in order to meet the demand rate with acceptable quality. Although sampling helps in controlling the quality of produced products if the quality degrades because of the occurrence of assignable causes, the sampling practice decreases the availability of the system due stoppages of false alarms and performing an adjustment to the process when a true alarm is confirmed. This directly influences the decisions of selective maintenance, and therefore, combining sampling and selective maintenance in an integrative design is required when both quality and quantity are of interest.

2.3. Sampling with propagating shifts in multistage systems with more than two machines and two states

In Chapter 2, sampling with propagating shifts in a two-machine serial system is presented. However, a system could consist of more than two machines. This makes the problem more complicated since the number of scenarios, due to the competencies among shifts, becomes larger. For instance, if there are three machines in the system, there will be one in-control state and seven out-of-control states. The number of a system's states increases exponentially as the number of machines increases.

As shown in Chapter 2, there are two competing shifts such that $T_1 > T_2$ or $T_2 > T_1$, whereas with three shifts, there are six competing scenarios such that $T_1 > T_2 > T_3$, $\cdots\cdots$, and $T_3 > T_2 > T_1$. Moreover, when three shifts are considered, Case II will consist of three subcases, and Case III will consist of four subcases as shown in Table 4.1. As illustrated in Table 4.1, for the scenario $T_1 > T_2 > T_3$, $(T_3T_2T_1)$ means that all three shifts occur in the same interval, (T_3, T_2, T_1) means that the three shifts occur at different intervals, and (T_3T_2, T_1) means that S_3 and S_2 occur in the same interval while S_1occurs in a different interval. Since there are 6 competing scenarios and each scenario has 8 possible subcases, the total number of instances to model is 48. Generally, if a denotes the number of shifts "number of machines", there is always $(a!)$ possible competing scenarios

Table 4.1: A scenario development in a three-machine system.

Scenario	Case I	Case II	Case III
$T_1 > T_2 > T_3$	$(T_3 T_2 T_1)$	(T_3, T_2, T_1)	(T_3, T_2, T_1) S_3 is detected
			(T_3, T_2, T_1) S_3 is not detected, but the propagated shift of S_3 and S_2 is detected
		$(T_3 T_2, T_1)$	$(T_3 T_2, T_1)$ the propagated shift of S_3 and S_2 is detected
		$(T_3, T_2 T_1)$	$(T_3, T_2 T_1)$ S_3 is detected

If there are four shifts, there are $4! = 24$ competing scenarios, and for each scenario, there are 7 subcases for Case II and 12 subcases under Case III with a total of 20 cases including Case I. This means that there are $24 \times 20 = 480$ possible instances for modeling. Therefore, there is a necessity to create a methodology or an algorithm to count these large numbers of scenarios and reduce the computational burden maybe by eliminating the least effect scenarios. Moreover, a simulation-based solution could be an option.

In the above design, it is always considered that there is only one in-control state and the rest are out-of-control states. Thus, another extension of the above design is to consider a set of in-control "desired" states and another set of out-of-control "undesired" states. This may help in reducing the total number of scenarios to be modeled.

3. References

Cassady, C. R., Pohl, E. A. and Murdock Jr, W. P. (2001) Selective maintenance modeling for industrial systems. *Journal of Quality in Maintenance Engineering,* **7**(2), 104-117.

Kijima, M., Morimura, H. and Suzuki, Y. (1988) Periodical replacement problem without assuming minimal repair. *European Journal of Operation Research,* **37**(2), 194-203.

Kijima, M. (1989) Some results for repairable systems with general repair. *Journal of Applied Probability,* **26**(1), 89-102.

Lin, D., Zuo, M. J. and Yam, R. C. M. (2000) General sequential imperfect preventive maintenance models. International Journal of Reliability. *Quality and Safety Engineering,* **7**(3), 253–266.

Liu, Y. and Huang, H. Z. (2010) Optimal selective maintenance strategy for multi-state systems under imperfect maintenance. *IEEE Transactions on Reliability,* **59**(2), 356–367.

Malik, M. A. K. (1979) Reliable preventive maintenance policy. *AIIE Transactions,* **11**(3), 221-228.

Pandey, M., Ming J. Z. and Moghaddass, R. (2013) Selective maintenance modeling for a multistate system with multistate components under imperfect maintenance. *IIE Transactions,* **45**(11), 1221–1234.

Rice, W. F., Cassady, C. R. and Nachlas, J. A. (1998) Optimal maintenance plans under limited maintenance time. *Proceedings of the Seventh Industrial Engineering Research Conference*.