Les Cahiers du GERAD ISSN: 0711–2440

Optimal planning of buffer sizes and inspection station positions

M. Ouzineb, F. Mhada,
R. Pellerin, I. El Hallaoui

G–2014–53

August 2014

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Optimal planning of buffer sizes and inspection station positions

Mohamed Ouzineb[a]
Fatima Mhada[b]
Robert Pellerin[c]
Issmail El Hallaoui[d]

[a] *GERAD & Institut National de Statistique et d'Économie Appliquée, Rabat, Maroc*

[b] *GERAD & Université Mohammed V/École Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Maroc*

[c] *CIRRELT & Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7*

[d] *GERAD & Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7*

mohamed.ouzineb@gerad.ca
f.mhada@um5s.net.ma
robert.pellerin@polymtl.ca
issmail.elhallaoui@gerad.ca

August 2014

Les Cahiers du GERAD
G–2014–53

Abstract: The buffer sizing problem in unreliable production lines is a complex combinatorial optimization problem. In the formulation of the problem, the system consists of n machines and n fixed-size buffers in series. These machines produce parts with two different quality levels: conforming and non-conforming parts. The production line can contain inspection stations whose job is to reject the non-conforming parts from the line. Thus, the design goal is to select the optimal combination of buffer sizes and inspection stations to meet the demand for conforming parts while minimizing total investment cost. In this paper, we propose an exact algorithm that not only finds an optimal location of the inspection stations but also optimizes the number of needed inspection stations. We present new theoretical results on buffer sizes bounds and stationarity and cost function convexity. These results help reducing the complexity of the algorithm. With this algorithm, we are able to solve to optimality large problems with 20 machines.

Key Words: Inspection, production lines, quality, combinatorial optimization, meta-heuristics.

1 Introduction

Quality and quantity modeling of manufacturing systems have been separately studied for long time, even though are highly coupled. In the literature, we assume that product quality is perfect so that the effect of the quality failure in the quantity modeling of production line is ignored. Both quantity and quality modeling have the same objectives: minimizing production cost and maximizing productivity. However, high quality requires early inspection of quality failure, which implies a reduction of inventory in the line. Inversely, low inventory may make the production line more vulnerable to machine failures and decrease the productivity of the line. Therefore, there is a need to integrate both quantity and quality modeling to optimize the performance of production lines.

Despite these issues being highly coupled, they have been studied separately for long time as stated by Inman et al. (2013). Indeed, researchers used to assume that product quality is perfect, so the effect of the quality failure in the quantity modeling of the production line has been ignored. However, high quality requires early inspection of quality failure, which implies a reduction of inventory in the line. Inversely, low inventory may make the production line more vulnerable to machine failures and decrease the productivity of the line. Therefore, there is a need to integrate both quantity and quality modeling to optimize the design of production lines.

Fortunately, a number of researchers have in recent years studied several issues situated in the interface between production and quality system design, including statistical process control design (Hajji et al. (2010)), equipment selection (Chincholkar and Herremann (2008)), tolerance design (Abdul-Kader et al. (2010)), and production control. For instance, Kim and Gershwin (2005, 2008) studied the relationship between quality and productivity by assuming that machines can enter a quality failure mode which is absorbing until proper maintenance is carried out.

In a similar manner, Colledani and Tolio (2005, 2006, 2009, 2011) considered a production system composed of unreliable manufacturing and inspection stations with different failure modes. Statistical quality control charts are introduced at inspection stations and act as noisy measurements on the quality state of the machines. Decomposition methods for the integrated production/quality performance studies of the line were developed.

Considering serial production lines consisting of producing and inspection machines that follow Bernoulli reliability and quality assumptions, Meerkov and Zhang (2010) provided important insights into the nature of both production and quality bottlenecks. Such systems are encountered in automotive assembly and painting operations where the downtime is relatively short and the defects are a result of uncorrelated random events.

In the same area, Mhada et al. (2014) proposed an analytical model for the integrated quality and quantity control of an unreliable production line with machines producing either conforming or defective parts. Solving even a small instance of 10 machines and one inspection station using a direct dynamic programming method takes three hours. When increasing the number of possible inspection stations, the dynamic programming approach becomes inefficient. Recognizing this limitation, this paper aims to optimize a generalization of this model using a different approach. An exact method is used to optimize the assignment of buffer sizes for a given location of inspection stations. An exhaustive method is then used to locate the extra inspection stations and give an optimal plan for the buffer sizes and inspection station positions problem. We also present new theoretical results on buffer sizes bounds and stationarity and cost function convexity. These results help reducing the complexity of the algorithm. With this algorithm, we are able to solve to optimality large problems with 20 machines and present some interesting properties of the problem on the basis of our empirical results.

The remainder of the paper is organized as follows. The problem formulation is first given in Section 2. Some theoretical results and the proposed resolution approach are presented and discussed in Section 3. Numerical results are reported in Section 4 and some conclusions are drawn in Section 5.

2 Problem formulation

2.1 Notation

Figure 1 represents a production line processing one part type. The line consists of processing machine M_i, denoted by square, a finite capacity storage buffer B_i for work in processes inventory, by a circle and an inspection station SI_i, denoted by circle in the middle of the square.

Let n be the number of the machines in the line. We assume that all the machines in the line are unreliable, and let $\alpha_i(t)$ denote the state of the machine M_i, respectively 1 if the machine is operational and 0 if it failed. The transition rates matrix of the stochastic processes $\alpha_i(t)$ are denoted by Q_i and expressed as follows: $Q_i = \begin{pmatrix} -p_i & p_i \\ r_i & -r_i \end{pmatrix}$ with p_i is the failure rate and r_i the repair rate.

The manufacturing system being studied consists of a synchronous tandem flow line that produces one part type with two types of qualities: "good quality" and "bad quality". The fraction between the bad parts and good parts is a constant β_i (depending on the machine). The inventory is stored between machines in finite buffers.

The production line can contain inspection stations located at the exit of buffer B_i and the provisioning point for machine M_{i+1}. The presence or absence of these stations is captured by a binary variable λ_i respectively 1 if there is an inspection station at the exit of the buffer B_i and 0 if not.

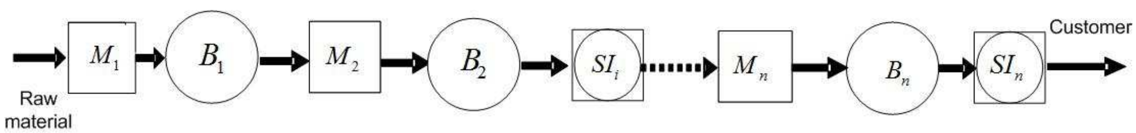

Figure 1: The production line

The following notations and assumption are used in the rest of this article:

- z_i: the inventory level (critical thresholds) for B_i, $1 \le i \le n$,
- k_i : the machine M_i maximal production rate,
- $u_i(t)$: the machine M_i production rate, $0 \le u_i(t) \le k_i$, $1 \le i \le n$,,
- q_i: the fraction of bad parts from good parts in B_i:

$$q_{i+1} = (1 - \lambda_i)\, q_i\, (1 + \beta_{i+1}) + \beta_{i+1} \tag{1}$$

  It is initialized by $q_1 = \beta_1$,
- $x_i^1(t)$: inventory level of the good parts on B_i,
- $x_i^2(t)$: inventory level of the bad parts on B_i,
- the level $x_i(t)$ of B_i consists of the sum of $x_i^1(t)$ and $x_i^2(t)$, with:

$$x_i^2(t) = q_i\, x_i^1(t) \tag{2}$$
$$x_i(t) = x_i^1(t) + x_i^2(t) = (1 + q_i)\, x_i^1(t) \tag{3}$$
$$x_i(t) = \frac{(1 + q_i)}{q_i}\, x_i^2(t) \tag{4}$$

- c_p: the storage cost per time unit and per part,
- c_I: the inspection cost per pulled part,
- machine M_1 is never starved,
- d: the demand rate that the line must satisfy from the stock of good parts x_{n_1} , which means that it must satisfy $(1 + q_n)\, d$ from the total stock x_n,

- $\lambda_n = 1$ means that one inspection station is placed at the end of the transfer line to guarantee that parts delivered to the customer are all conforming,
- $\sum_{i=1}^{n-1} \lambda_i = m$, $m = 0, 1, \ldots, (n-1)$: we are considering the problem of adding another m inspection stations within the transfer line and witch the placements of the m extra inspection stations are unknown,
- $\tilde{d}_i$, $i = 1, \ldots, n$, is the long term average number of parts pulled per unit time from the total stock $x_i(t)$, with:

$$\tilde{d}_i = d \prod_{i=j}^{n} (1 + \lambda_j q_j) \tag{5}$$

- a_n^{des} is the total parts wip availability coefficient at buffer B_n and it is a given desired coefficient of availability of conforming finished parts.

The optimization problem is that of the joint placement of the extra m inspection stations and the minimizing of the long term per unit time average global cost of storage, production shortages, and inspection. Meaning, the cost to be minimized is:

$$J_{T\{a_i,\lambda\}}(x_0^T, \alpha_0^T) = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{n} E\left[\int_0^T (c_p\, x_i(t) + c_I\, \tilde{d}_i\, \lambda_i) dt / (x_i(0), \alpha_i(0))\right] \tag{6}$$

under conditions: $\sum_{i=1}^{n-1} \lambda_i = m$, $m = 0, 1, \ldots, (n-1)$ and $\lambda_n = 1$.

2.2 Problem formulation

The decomposition/aggregation methodology of Sadr and Malhamé (2004b) allows us to define a virtual machine $\tilde{M}_i$, an aggregate representation of the complete transfer line up to machine M_{i+1}, as it appears viewed from the rest of the transfer line downstream. $\tilde{M}_i$ is a machine with a discrete state $\tilde{\alpha}_i$ (0 if failed and 1 if operational) with respectively $\tilde{r}_i$ the repair rate and $\tilde{p}_i$ the failure rate, for $i = 2 \ldots n$

$$ps_{i-1} = \tilde{r}_{i-1} \frac{1 - a_{i-1}}{a_{i-1}}, \tag{7}$$

$$\tilde{r}_i = \frac{(ps_{i-1} + p_i)\, \tilde{r}_{i-1}\, r_i}{p_i\, \tilde{r}_{i-1} + ps_{i-1}\, r_i}, \tag{8}$$

$$\tilde{p}_i = \left(\frac{(ps_{i-1} + \tilde{r}_{i-1})\,(p_i + r_i)}{\tilde{r}_{i-1}\, r_i} - 1\right) \tilde{r}_i, \tag{9}$$

with: $\tilde{r}_1 = r_1$ and $\tilde{p}_1 = p_1$.

By applying the averaging principle (Sadr and Malhamé (2004b)) with a corrected demand, the machine $\tilde{M}_i$ is subject to a constant demand $\frac{\tilde{d}_i}{a_i}$, $i = 1 \ldots n-1$ with shortage not permitted.

a_i is an unknown total parts wip availability coefficient at buffer B_i, $1 \le i \le (n-1)$ i.e, a_i is the probability that the wip $x_i(t)$ is *active*.

There is a relationship between a_i and z_i:

$$a_i = 1 - \frac{\tilde{p}_i}{\tilde{p}_i + \tilde{r}_i} \frac{1 - \frac{\tilde{r}_i (k_i - \frac{\tilde{d}_i}{a_i})}{\tilde{p}_i\, \frac{\tilde{d}_i}{a_i}}}{1 - \left(\frac{\tilde{r}_i (k_i - \frac{\tilde{d}_i}{a_i})}{\tilde{p}_i\, \frac{\tilde{d}_i}{a_i}} \exp\left(\frac{k_i \tilde{r}_i - \frac{\tilde{d}_i}{a_i}(\tilde{r}_i + \tilde{p}_i)}{(k_i - \frac{\tilde{d}_i}{a_i}) \frac{\tilde{d}_i}{a_i}}\, z_i\right)\right)}, \; i = 1 \ldots n-1$$

a_i must be between a lower bound defined to allow the feasibility of the demand $\tilde{d}_i$ by the pseudo machine $\tilde{M}_i$. So we have the following constraint of feasibility: $\tilde{d}_i < a_{i-1}\, k_i \frac{r_i}{(r_i + p_i)}$ i.e: $\frac{(r_i + p_i)\, \tilde{d}_i}{r_i\, k_i} < a_{i-1}$ and since a_{i-1} is a probability then the upper bound is 1.

The average long term combined storage, shortage costs and inspection costs in (6) can be written after development and calculation (see article Mhada et al. (2014)) as:

$$J(a,\lambda) = \sum_{i=1}^{n-1} T^{(i)}(a,\lambda) + T_F(a,\lambda) + c_I \sum_{i=1}^{n} \lambda_i \, \tilde{d}_i \tag{10}$$

$$\begin{aligned} T^{(i)}(a,\lambda) = c_p \Bigg(& \frac{k_i\,\tilde{p}_i}{\sigma_i(k_i - \frac{\tilde{d}_i}{a_i})\,(\tilde{r}_i + \tilde{p}_i)} - \frac{k_i\,(1-a_i)}{\sigma_i(k_i - \frac{\tilde{d}_i}{a_i})} - \left[\frac{1}{\sigma_i} - \frac{(1-a_i)\,(\tilde{r}_i+\tilde{p}_i)}{\sigma_i^2(k_i - \frac{\tilde{d}_i}{a_i})}\right] \\ & ln\left[\frac{\tilde{p}_i\,\frac{\tilde{d}_i}{a_i}}{\tilde{r}_i\,(k_i - \frac{\tilde{d}_i}{a_i})} - \frac{\sigma_i\,\tilde{p}_i\,\frac{\tilde{d}_i}{a_i}}{(\tilde{r}_i+\tilde{p}_i)\,\tilde{r}_i\,(1-a_i)}\right]\Bigg), \; i = 1, \ldots, n-1 \end{aligned} \tag{11}$$

with : $\sigma_i = \frac{(\tilde{p}_i+\tilde{r}_i)\,\frac{\tilde{d}_i}{a_i} - k_i\,\tilde{r}_i}{(k_i - \frac{\tilde{d}_i}{a_i})\,\frac{\tilde{d}_i}{a_i}}$ and

$T_F(a,\lambda) =$

$$\frac{\rho_n\,[c_p\,(\frac{k_n\,(1-\exp(-\mu_n(1-\rho_n)\,z_n(a_n^{des})))}{1-\rho_n}}{(\tilde{p}_n+\tilde{r}_n)(1-\rho_n\,\exp(-\mu_n(1-\rho_n)\,z_n(a_n^{des})))} \frac{-(\tilde{p}_n+\tilde{r}_n)\,z_n(a_n^{des})\,\exp(-\mu_n(1-\rho_n)\,z_n(a_n^{des})))]}{(\tilde{p}_n+\tilde{r}_n)(1-\rho_n\,\exp(-\mu_n(1-\rho_n)\,z_n(a_n^{des})))} \tag{12}$$

with: $\rho_n = \frac{\tilde{r}_n(k_n - \frac{\tilde{d}_n}{a_n^{des}})}{\tilde{p}_n\,\frac{\tilde{d}_n}{a_n^{des}}}$, $\mu_n = \frac{\tilde{p}_n}{(k_n - \frac{\tilde{d}_n}{a_n^{des}})}$ where the expression of $z_n(a_n^{des})$ is given by

$$z_n(a_n^{des}) = -\frac{\ln\left[\frac{1}{\rho_n}\left(1 - \frac{(1-\rho_n)}{(1-a_n^{des})(\frac{(\tilde{p}_n+\tilde{r}_n)}{\tilde{p}_n})}\right)\right]}{\mu_n(1-\rho_n)} \tag{13}$$

The problem can be formulated as follows: find the minimal average global cost system structure (a,λ) that satisfies both constraints. That is,

$$\begin{aligned} \text{minimize} \quad & J(a,\lambda) && && (14) \\ \text{subject to} \quad & && && \\ & \sum_{i=1}^{n-1} \lambda_i = m && \forall m, \;\; 1 \le m \le n-1, && (15) \\ & \lambda_i \in \{0,1\} && \forall i, \;\; 1 \le i \le n-1. && (16) \\ & a_i \ge 0 && \forall i, \;\; 1 \le i \le n, && (17) \end{aligned}$$

This problem is a Mixed Integer Nonlinear Programming (MINLP) problem which refers to a mathematical programming with continuous and discrete variables and nonlinearities. This problem simultaneously optimizes the system structure: the location of the inspection stations (discrete) and parameters and the buffer sizing (continuous).

The variables a_i, $1 \le i \le n$ and $\lambda_i \in \{0,1\}$ $1 \le i \le n-1$ are the decision variables.

The vector of binary variables ((15),(16)) is connected to the line structure.

N.B: The theoretical results in Section 3 and the numerical results of Section 4 are for the homogeneous and partially homogeneous production lines: $r_1 = r_2 = \ldots = r$ and $\beta_1 = \beta_2 = \ldots = \beta$. i.e:

$$\tilde{r}_i = r \tag{18}$$

$$\tilde{p}_i = \frac{(p_i + r(1-a_{i-1}))}{a_{i-1}} \qquad \forall i, \;\; 1 \le i \le n \tag{19}$$

3 Solution methodology

The objective is to minimize the average long term combined storage and shortage costs, while also specifying the optimal location of inspection stations that meets a specified final conditions desired by the customer. More formally, we have to find an optimal solution to the problem defined by (14–17). As mentioned above, this problem is an MINLP problem with continuous and discrete variables and nonlinearities. Solving this problem by standard dynamic programming that can handle nonlinearities or directly by a MINLP solver takes hours for a small instance with one inspection station.

Observe that the problem (14–17) is equivalent to $min_\lambda(min_a J(a,\lambda))$ where a and λ satisfy constraints (16) and (17). It is interesting to see that by fixing λ, we obtain an objective function that is separable by variables a_i ($T^{(i)}(a,\lambda) = T^{(i)}(a_i, a_{i-1}, \lambda)$). The idea is to reformulate $min_a J(a,\lambda)$, for each fixed location λ, as a shortest path problem defined on a network (described below) and efficiently solve it by a standard shortest path algorithm to find an optimal assignment of buffer sizes. This fact is well stated in Proposition 3.1.

Consider the connected network $G(E,V)$ consisting of a set of nodes E and a set of links V as depicted in Figure 2. Each column i corresponds to a set of buffer availability possibilities for a machine i. Each node is connected to all nodes in the next column. To each arc is associated some real number $c_{i,i+1}$ that corresponds to the shortage and inspection costs, such that: $c_{0,1} = 1$, $c_{i,i+1} = T^i(a_i, a_{i+1}, \lambda) + c_I \times \lambda_i \times d_i$, $i = 1 \dots n-2$ and $c_{n-1,n} = T_F(a_{n-1}, a_n, \lambda) + c_I \times \lambda_n \times d_n$.

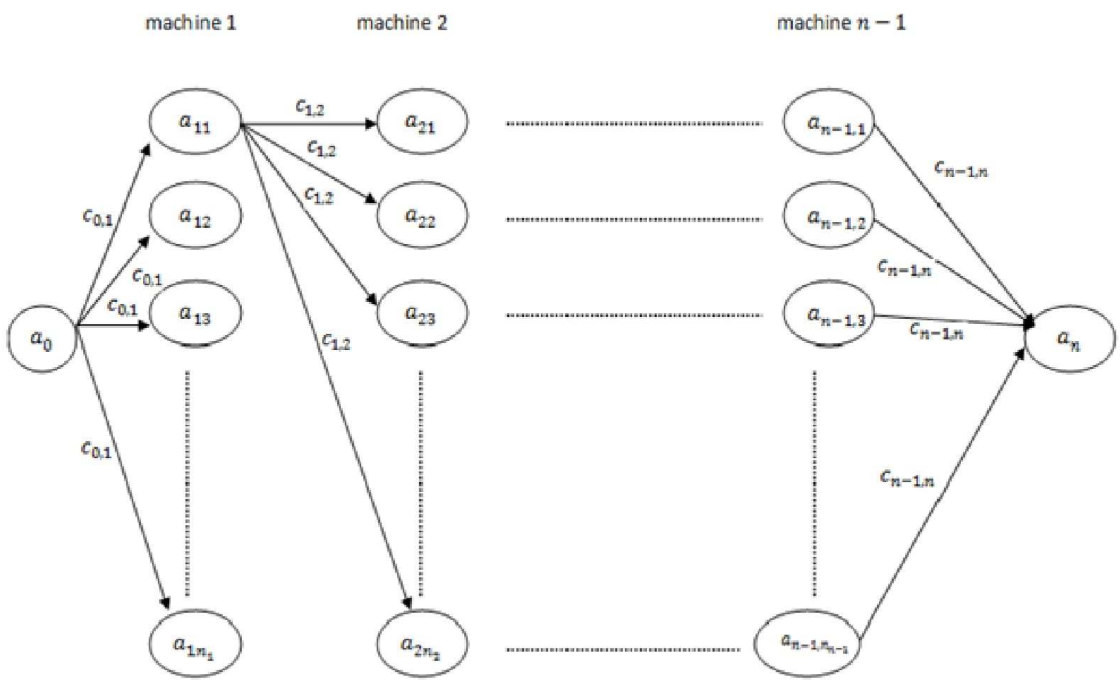

Figure 2: The network flow problem

Proposition 3.1 *For each fixed location λ, the minimum total storage and inspection costs is the sum of arc costs on the shortest path in G (between a_0 and a_n).*

Proof. For each fixed location $\lambda = \lambda_0$,

$$\begin{aligned} argmin\, J(a,\lambda) &= argmin\,(\sum_{i=1}^{n-1} T^{(i)}(a,\lambda) + T_F(a,\lambda) + c_I \sum_{i=1}^{n} \lambda_i\, \tilde{d}_i) \\ &= argmin\,(\sum_{i=1}^{n-1} T^i(a_i, a_{i+1}, \lambda_0) + T_F(a_{n-1}, a_n, \lambda_0)) \end{aligned} \tag{20}$$

From eq. (20), the minimum total storage and inspection costs problem can be considered as a single-source shortest path problem. So, to find its minimum, we need to find the shortest path (i.e.the path with lowest cost) from a_0 and a_n. □

From Proposition 3.1, we can deduce the following algorithm (a pseudocode is given in Algorithm 1). This algorithm can be significantly improved in two ways, either by restricting the possible values of the buffer availability coefficient (a_i), leading to a reduction in the network size and ultimately to a reduction of the time per iteration, or restricting possible positions of the inspection stations, hence reducing the number of iterations itself.

Algorithm 1 Optimal buffer sizing and inspection stations location algorithm

```
Z* = ∞, k = 0.
for all λ do
  Construct the network G.
  Find a shortest path in G (the buffer size vector a) and set Z to its cost.
  if Z < Z* then
    Z* = Z, a* = a, λ* = λ.
  end if
  k = k + 1.
end for
return (Z*, a*, λ*).
```

In order to reduce the number of nodes and arcs in the network $G(E, V)$, we use some characteristics of the production line to prove interesting theoretical results on the possible values of a_i.

Lemma 3.2 *The coefficient of availability a_i of the buffer B_i is constrained by the value of a_{i-1}.*

Proof. The coefficient a_i is at its minimum when the buffer B_{i-1} size is zero, because the buffer size is always larger than zero. This value is equal to $P[x_{i-i} \geq 0, M_i\ is\ ON] = \frac{r}{r+p_i}a_{i-1}$. Consequently, a_i must be larger than $\frac{r}{r+p_i}a_{i-1}$. □

Proposition 3.3 *The buffer size a_i is upperbounded by* $\min\left(\prod_{j=i}^{n}\left[\frac{r+p_j}{r}\right]a_n^{des}, 1\right)$.

Proof. From Lemma 3.2, we have:

$$a_i \geq \frac{r}{r+p_i}a_{i-1} \Rightarrow a_{i-1} \leq \frac{r+p_i}{r}a_i.$$

The recurrence of this equation makes it possible to conclude that:

$$\begin{aligned} a_{i-1} &\leq \prod_{j=i}^{n}\left[\frac{r+p_j}{r}\right]a_n && (21)\\ &\leq \prod_{j=i}^{n}\left[\frac{r+p_j}{r}\right]a_n^{des}; && (22)\end{aligned}$$

□

Lemma 3.4 *The coefficient of availability a_i of the buffer B_i is constrained by the value of demand d.*

Proof. The second characteristic of the production lines is that each machine M_i must be able to meet the demand d. So, the production rate of the machine M_i, when it is functional and has enough parts to process, must be superior to d *i.e.,*

$$k_i\ P[x_{i-i} \geq 0, M_i\ is\ ON] \geq d \Rightarrow k_i \frac{r}{r+p_i} a_{i-1} \geq d \Rightarrow a_{i-1} \geq \frac{r+p_i}{r\ k_i} d.$$

□

Proposition 3.5 *The buffer size a_i is lower bounded by* $\max\left(\prod_{j=1}^{i-1}\left[\frac{r}{r+p_j}\right], \frac{r+p_i}{r\ k_i}d\right)$.

Proof. From Lemma 3.2 , we have:

$$a_i \geq \frac{r}{r+p_i} a_{i-1}.$$

We suppose that the recurrence of this equation makes it possible to conclude that:

$$a_{i-1} \geq \prod_{j=1}^{i-1}\left[\frac{r}{r+p_j}\right] a_0 \tag{23}$$

$$\geq \prod_{j=1}^{i-1}\left[\frac{r}{r+p_j}\right]; \tag{24}$$

From Lemma 3.4, we have:

$$a_{i-1} \geq \frac{r+p_i}{rk_i} d.$$

And,

$$a_{i-1} \geq \max\left(\prod_{j=1}^{i-1}\left[\frac{r}{r+p_j}\right], \frac{r+p_i}{rk_i} d\right)$$

□

Bounding by Propositions 3.3 and 3.5 immediately affects the number of nodes and arcs in G. Actually, each column i contains only nodes, such that: $a_{i1} = 100 \times \max\left(\prod_{j=1}^{i-1}\left(\frac{r}{r+p_j}\right), \frac{d}{k_i}\frac{r+p_i}{r}\right)$ and $a_{in_i} = 100 \times \min\left(\prod_{j=i}^{n}\left[\frac{r+p_j}{r}\right] a_n^{des}, 1\right)$. The nodes corresponding to outranged values are removed. Obviously, a "quadratic" number of arcs between removed nodes can be eliminated.

A stationary characteristic of homogeneous lines can be used to reduce the solution space. Indeed, we establish that:

Proposition 3.6 *For a perfectly homogeneous transfer line (with identical r_i, p_i, k_i, β_i) with a large number of machines n goes to infinity and a fixed $\lambda = \lambda_0$, there is a stationary feedback control law that is optimal. This state policy contains some constant availability coefficients $\bar{a}$ for the machines considered "sufficiently far" from the first machine upstream, meaning that the optimal profile would be flat at the fixed level $\bar{a}$, independent of the boundary condition on a_0 and a_n.*

Proof. Let

$$P_s: \quad min\ J(a,\lambda_0) = \sum_{i=1}^{n-1} T^{(i)}(a_{i-1}, a_i, \lambda_0)$$

with:

$$T^{(i)}(a_{i-1},a_i,\lambda_0)=c_p\left(\frac{k\,\frac{p+r(1-a_{i-1})}{a_{i-1}}\,\frac{\tilde{d}_i}{a_i}}{\left((\frac{p+r}{a_{i-1}})\,\frac{\tilde{d}_i}{a_i}-k\,r\right)(\frac{p+r}{a_{i-1}})}-\frac{k\,(1-a_i)\frac{\tilde{d}_i}{a_i}}{\left((\frac{p+r}{a_{i-1}})\,\frac{\tilde{d}_i}{a_i}-k\,r\right)}-\left[\frac{(k-\frac{\tilde{d}_i}{a_i})\,\frac{\tilde{d}_i}{a_i}}{(\frac{p+r}{a_{i-1}})\,\frac{\tilde{d}_i}{a_i}-k\,r}-\frac{(1-a_i)\,(\frac{p+r}{a_{i-1}})(k-\frac{\tilde{d}_i}{a_i})\left(\frac{\tilde{d}_i}{a_i}\right)^2}{\left((\frac{p+r}{a_{i-1}})\,\frac{\tilde{d}_i}{a_i}-k\,r\right)^2}\right]\right.$$
$$\left.ln\left(\frac{\frac{p+r(1-a_{i-1})}{a_{i-1}}\,\frac{\tilde{d}_i}{a_i}}{r\,(k-\frac{\tilde{d}_i}{a_i})}-\frac{\left((\frac{p+r}{a_{i-1}})\,\frac{\tilde{d}_i}{a_i}-k\,r\right)\left(\frac{p+r(1-a_{i-1})}{a_{i-1}}\right)}{(\frac{p+r}{a_{i-1}})\,r(1-a_i)(k-\frac{\tilde{d}_i}{a_i})}\right)\right),\ i=1\ldots n-1$$

The idea is to apply a theoretical result for a homogeneous line without any inspection station ($m=0$) to lines with m inspection stations ($m>0$) placed in the positions: $e_1,\,e_2,\,\ldots,\,e_m$ with $e_m<n$.

This theoretical result is based on Proposition 2 (Sadr and Malhamé (2004a)) :

> "For a perfectly homogeneous transfer line (identical r_i, p_i, k_i, β_i) as the number of machines n goes to infinity, the problem P_s admits a stationary state feedback control policy which is optimal. This state feedback control policy contains some constant availability coefficients $\bar{a}=\arg\min$ $T^{(i)}(a,a,0)$, subject to: $\left(\max\left[\left(\frac{r}{r+p}\right)^{(i)},\ \frac{\tilde{d}_i}{k}\,\frac{r+p}{r}\right]\right)<a<1.$"

So, we divided the line into $m+1$ homogeneous lines separated by the m inspections stations. The parameters of these lines are identical except for the pulled demand $\tilde{d}_i$ since:

$$\begin{cases} 1\ \le i\le e_1 & q_i=(1+\beta)^i-1 & \tilde{d}_i=d\,(1+\beta)^n \\ e_1<i\le e_2 & q_i=(1+\beta)^{i-e_1}-1 & \tilde{d}_i=d\,(1+\beta)^{n-e_1} \\ \vdots & \vdots & \vdots \\ e_{m-1}<i\le e_m & q_i=(1+\beta)^{i-e_{m-1}}-1 & \tilde{d}_i=d\,(1+\beta)^{n-e_{m-1}} \end{cases}$$

And then we apply Proposition 2 (of Sadr and Malhamé (2004a)) to each line segment with $\bar{a}=argminT^{(i)}$ (a,a,λ_0). This will allow us to calculate the values of the optimal availability coefficients in which each line segment will be flat. □

A numerical example is given in Subsection 4.4 to illustrate the theoretical result above.

Conjecturing on the convexity of the cost function may lead to reducing the number of iterations of the algorithm. For instance, the local minimum of a convex function is also a global minimum and there are many efficient specialized methods for optimizing convex functions. This reduction in the number of iterations will lead to a significant reduction in the overall solution time. Actually, Conjectures 3.7 and 3.8 emphasize the fact that the cost function is convex on the position of the inspection station and on the number of the located inspection stations. These conjectures are likely to be true as empirically shown by the numerical results in Section 4.

Conjecture 3.7 *The minimal total cost (storage and inspection costs) is a convex function of λ, i.e the location of the internal inspection station.*

Conjecture 3.8 *The minimal total cost (storage and inspection costs) is a convex function of the number of the internal inspection stations.*

4 Numerical results

To test different aspects of the algorithm, two instances are used: a sample with 10 machines and a larger test problem with 20 machines. The algorithm was implemented in C^{++}. The numerical tests were completed on an Intel Core i7 at 2.8 GHz with 8 Gbytes of RAM running on Linux.

4.1 Benchmark data

Table 1 shows the different parameters used for the numerical results. For machine M_i, $(i = 1, 2, \ldots, n)$, k_i is the maximal production rate, p_i the failure rate, r_i the repair rate, β_i the ratio of non-conforming parts to conforming parts, d the demand rate for good parts, c_p the storage cost per time unit and per part, c_I the inspection cost per pulled part, m the number of additional inspection stations, and a_n^{des} the required availability rate of conforming finished parts.

Table 1: System parameters

	n	p_i	r_i	k_i	β_i	d	c_p	c_I	a_n^{des}
10 machines	10	0,2	0,9	4	0,1	1	1	2	0.95
20 machines	20	0,2	0,9	9	0,1	1	0.1	0.2	0.95

4.2 Optimal location of the inspection station for $m = 1$

Figures 3 and 4 display the optimal cost as a function of the location of the internal inspection station λ_i, $i = 1, 2, \ldots, (n-1)$ for the test problems with 10 machines and with 20 machines. We observe that in our case (a homogeneous production line) the minimal total cost (storage and inspection costs) is a convex function of the location of the internal inspection station i, i.e. $\lambda_i = 1$. The optimal cost corresponds to the fourth position $\lambda_4 = 1$ for the two problems. This result makes sense since the station cannot be placed at the beginning of the line and so it loses a bit of its role in rejecting nonconformities. It also cannot be placed at the end of the line and so it increases the number of nonconformities in the system which translates into a high cost of storage.

4.3 Cost function according to the number of inspection stations

Figures 5 and 6 display the optimal cost as a function of the amount of inspection stations $(m = 1, 2, \ldots, (n-1))$ for the test problems with 10 machines and with 20 machines. We also observe that in this case the minimal total cost is a convex function of the number of inspection stations. The minimum cost corresponds to use one inspection station for the first test problem, to use 3 inspection stations for the second test problem, the optimal inspection stations positions in this case are:

- The second position $\lambda_2 = 1$.
- The sixth position $\lambda_7 = 1$.
- The fourteenth position $\lambda_{18} = 1$.

Figure 7 displays the optimal availability coefficients a_i for each buffer B_i, $i = 1 \ldots (n-1)$ corresponding to the optimal solution for the case of 20 machines, as seen previously.

4.4 The availability coefficient state space $a_i, i = 1 \ldots n-1$

In the 20 machines case, and with 4 inspection stations placed after M_2, M_7, M_{18} and M_{20}, we calculate $\bar{a}$ for each line segment located between two consecutive inspection stations and which can be considered as containing a large number of machines. We obtain:

- For the line segment composed by M_3 to M_7, $\bar{a} = 0.7649$.
- For the line segment composed by M_8 to M_{18}, $\bar{a} = 0.4758$.

We compare this $\bar{a}$'s with availability for the optimal solution found by the exhaustive method (Table 2 and Figure 7), and noted that with this calculation, it is possible to form an initial solution to improve the running time of the standard shortest path problem.

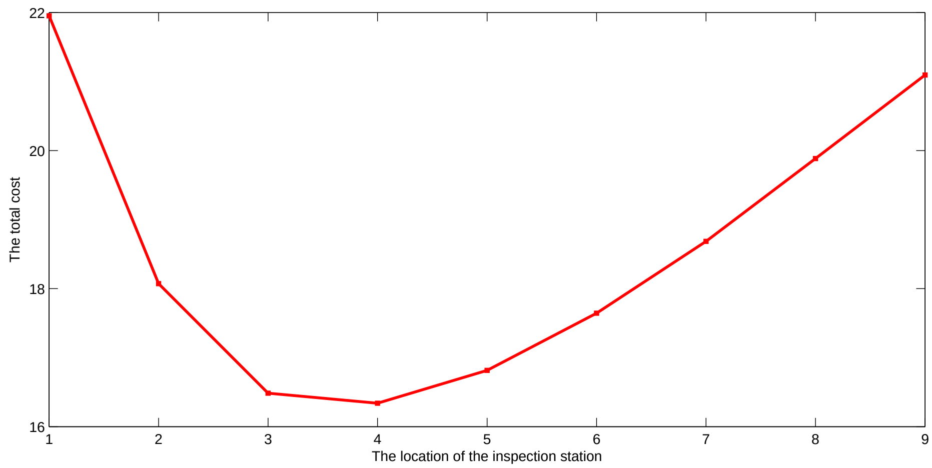

Figure 3: The optimal cost as a function of λ_i, $i = 1, 2, \ldots, 9$: the 10 machines case

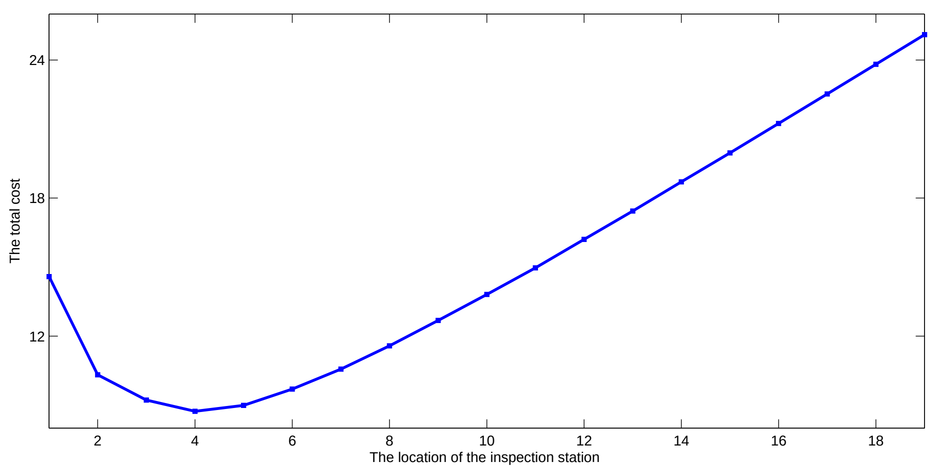

Figure 4: The optimal cost as a function of λ_i, $i = 1, 2, \ldots, 19$: the 20 machines case

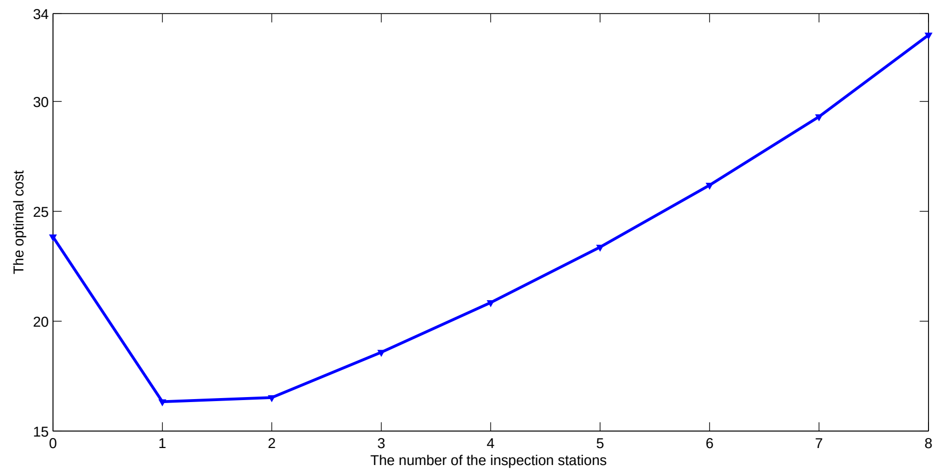

Figure 5: The optimal cost as a function of the number of inspection stations, $m = 1, 2, \ldots, 9$: the 10 machines case

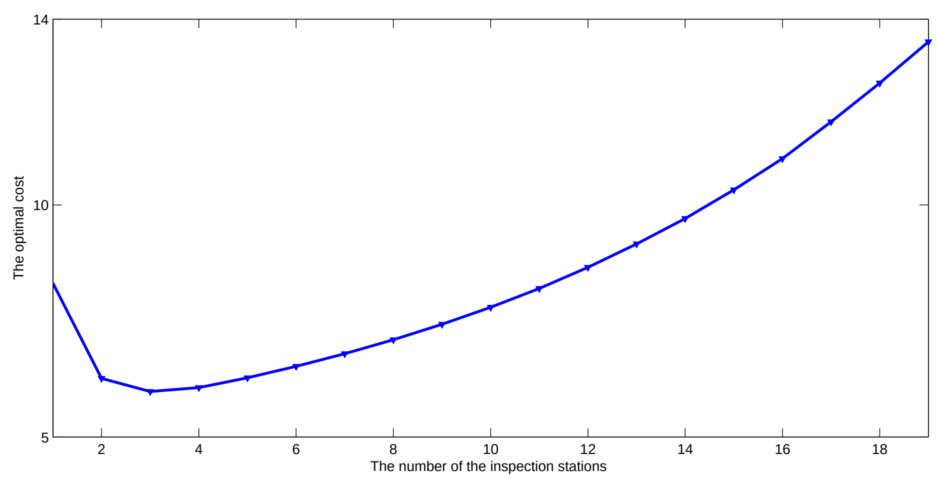

Figure 6: The optimal cost as a function of the number of inspection stations, $m = 1, 2, \ldots, 19$: the 20 machines case

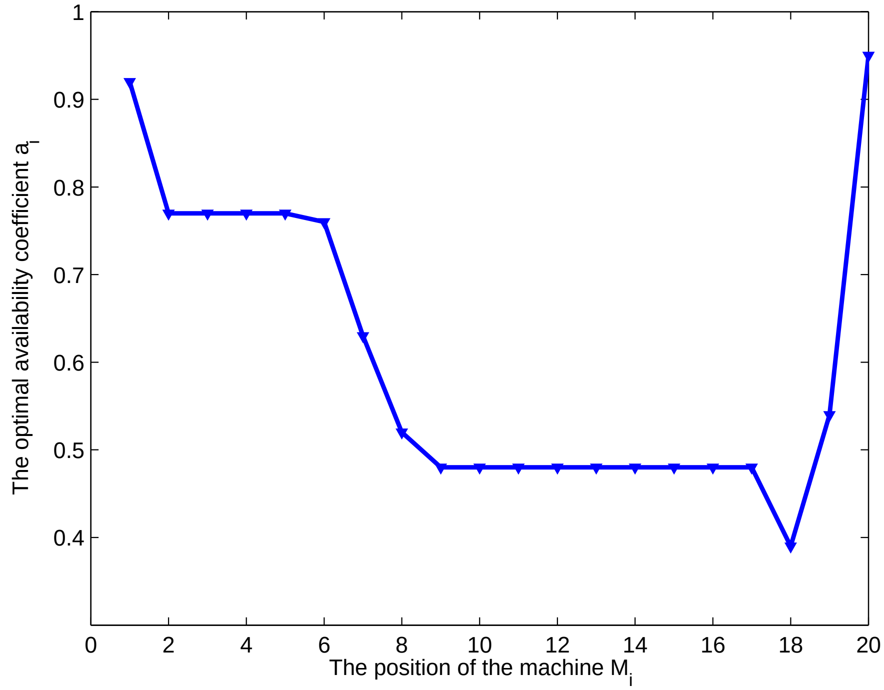

Figure 7: The optimal availability coefficient: the 20 machines case

Table 2: The optimal availability coefficient state space $a_i, i = 1 \dots n-1$: 20 machines case

i	1	2	3	4	5	6	7	8	9	10
a_i	0.92	**0.77**	**0.77**	**0.77**	**0.77**	**0.76**	0.63	0.52	**0.48**	**0.48**
i	11	12	13	14	15	16	17	18	19	20
a_i	**0.48**	**0.48**	**0.48**	**0.48**	**0.48**	**0.48**	**0.47**	0.39	0.54	0.95

The availability coefficients $a_i^0, i = 1 \dots n-1$ for this initial solution can be defined by:

$$a_i^0 = \max\left[\bar{a},\ \max\left(\left(\frac{r}{r+p}\right)^{(i)},\ \frac{\tilde{d}_i}{k}\frac{r+p}{r}\right) + 0.01\right]$$

5 Conclusions

This paper suggests solving a complex manufacturing problem formulated as a Mixed Integer NonLinear Problem (MINLP) with an exact method. Our manufacturing problem is a production line with intermediary stock and inspection stations positioned between machines. The objective is to design these inventories while optimizing the number of inspection stations throughout the line to guarantee a fixed rate of service to the final customer with minimal total cost.

This problem is considered a network flow problem which can be efficiently solved as a standard shortest path problem and optimal solutions are obtained from using this approach. We were able to develop some theoretical results (Section 3) which allow us not only to offer a shortest path but also to reduce the complexity. Empirically, we have shown some interesting properties of the problem. For instance, this problem is convex on the position of the inspection station and is also convex on the number of internal inspection stations.

In addition, we have confirmed that some properties found in the homogeneous line without inspection stations remain valid in our case (Section 3). This will allow us to improve our exact method by providing an initial solution close enough to the optimum by theoretical calculation. This will be treated in a future work where we will be increasing the size of the line up to 30 machines.

As a research perspective, it would be interesting to see whether the convexity (Figures 5 and 6) remain valid in the non homogeneous case and if it is possible to combine the exact method with a meta-heuristic to speed up the run time in the case of a line with more machines, without losing the performance of the exact method.

References

W. Abdul-Kader, O. Ganjavi, and A. Solaiman. An integrated model for optimisation of production and quality costs. International Journal of Production Research, 48(24):7357–7370, 2010.

M. Chincholkar and J.W. Herremann. Estimating manufacturing cycle time and throughput in flow shops with process drift and inspection. International Journal of Production Research, 46(24):7057–7072, 2008.

M. Colledani and T. Tolio. Performance evaluation of production systems monitored by statistical process control and off-line inspections. International Journal of Production Economics, 120(2):348 – 367, 2009.

M. Colledani and T. Tolio. Integrated analysis of quality and production logistics performance in manufacturing lines. International Journal of Production Research, 49(2):485 –518, 2011.

M. Colledani and T. Tolio. Impact of statistical process control (spc) on the performance of production systems-part 2 (large systems). Zakynthos Island, Greece, 2005.

M. Colledani and T. Tolio. Impact of quality control on production system performance. Annals of the CIRP, 55(1): 453–456, 2006.

A. Hajji, F. Mhada, A. Gharbi, R. Pellerin, and R. Malhame. Integrated product specifications and productivity decision making in unreliable manufacturing systems. International Journal of Production Economics, 129(1): 32–42, 2010.

R. Inman, D. Blumenfeld, N. Huang, and J. Li. Survey of recent advances on the interface between production system design and quality. IIE Transactions, 45(6):557–574, 2013.

J. Kim and S. Gershwin. Integrated quality and quantity modeling of a production line. OR Spectrum, 27:287–314, 2005.

J. Kim and S. Gershwin. Analysis of long flow lines with quality and operational failures. IIE Transactions, 40: 284–296, 2008.

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

F. Mhada, R. Malhamé, and R. Pellerin. Joint assignment of buffer sizes and inspection points in unreliable transfer lines with scrapping of defective parts. Production and Manufacturing Research, 1:79–101, 2014.

J. Sadr and R. Malhamé. Decomposition/aggregation-based dynamic programming optimization of partially homogeneous unreliable transfer lines. IEEE Transactions Automatic Control, 49:68–81, 2004a.

J. Sadr and R. Malhamé. Unreliable transfer lines: decomposition/ aggregation and optimisation. Annals of Operations Research, 125:167–190, 2004b.