OPERATIONS RESEARCH CENTER

Working Paper

Integrating Quality and Quantity Modelling in a Production Line

by

Andrea Poffe
Stanley B. Gershwin

OR 377-05 **June 2005**

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Integrating Quality and Quantity Modelling in a Production Line

Andrea Poffe and Stanley B. Gershwin

June 29, 2005

Abstract

This paper investigates the behavior of automated production lines subject to quality inspection. Machines are unreliable and can fail for different reasons: operational failures stop the machine without involving quality issues; quality failures lead machines to produce defective parts without stopping. We analyze a 2-machine-1-buffer line where machines are modelled by Markov chains with discrete states and separated by a buffer of finite capacity. Using such models, we analyze how production system design, quality, and productivity are interrelated in production systems. We show how inventory capacity can influence system yield and productivity. The approach considers control charts, buffers of finite capacity and a delay in quality information due to remote inspection. Numerical results and comparisons with simulation are reported.

1 Introduction

1.1 Motivation

Production system design is a complex task with many different aspects. It starts with the analysis of all the manufacturing processes required for the realization of the specified product to find the best utilization of processing machines, inspection stations and storage space. Production system design has been driven for long time by two separate fields of study. On the one hand Manufacturing Systems Engineering has developed methods for understanding the behavior of production systems and has investigated techniques to design efficient factories. It has been focusing on quantity related issues, like estimating productivity and work in process (WIP). On the other hand, research on quality improvement like Statistical Quality Control (SQC), Total Quality Maintenance (TQM) and Six Sigma have investigated methods to better control system processes to increase product quality. Production system design has a significant impact on product quality in the way that the performance of the quality control system is affected by the architecture of the production system. However, there is little literature that deals with both these two fields together and that considers quality in a system context. An important class of problem to be investigated is the optimal allocation of inspection devices and of buffer storages. For this reason, the need for research in this direction [6] has been expressed in the past few years.

Manufacturers influenced by the success of popular techniques like Toyota Production System, Lean Manufacturing etc. are working to achieve the highest productivity and the highest quality while reducing WIP. These popular techniques are, however, qualitative in nature and mainly based on experience that lack a sound scientific quantitative foundation. For example, Toyota recently changed their view on inventory and are trying to re-adjust their interoperational storages [4]. Therefore, the behavior of automated production lines subject to quality inspection needs to be investigated with quantitative techniques to suggest clear and reliable advice to factory designers.

We develop in this paper a model of production line similar in spirit to earlier quantity-oriented models [5] [1], in that machines are modelled by Markov chains with discrete states and are separated by buffers of finite capacity. Here, however, the up states have quality information associated with them, e.g., the yield conditioned on the machine being in each state. Machines are subject to two different types of failures: operational failures (e.g. motor burnout) and quality failures (e.g. tool damage). These two types of failures are different in nature and have different mean times to repair (MTTR). For this reason they cannot be grouped into a single down state but must be modelled independently.

We develop a 2-machine-1-buffer (2M1B) model in which the first machine has both operational and quality failures and the second machine has only operational failures. The inspection station is placed either within the first machine or at the end of the line. On either case, it monitors the behavior of the first machine. We adopt a continuous time, continuous material model because it can handle deterministic but different operation times at each station. Using such models, we analyze how production system design, quality, and productivity are interrelated in production systems. We show how inventory capacity can influence system yield and productivity. The approach considers control charts, buffers of finite capacity and delays in quality information. Numerical results and comparisons with simulation are reported to show the accuracy of the proposed method.

1.2 Quality policy

Machines are unreliable and can fail for different reasons. An operational failure stops the machine without involving quality issues. Quality failures lead machines to an out-of-control state; in this state machines are operational but produce defective parts. We make the assumption that once a defective part has been produced, all the subsequent parts will be bad until the machine is repaired. This kind of failure happens after changes occur in the machines. For this reason this kind of failure is called *persistent-type quality failure* [10]. It is very important to catch defective parts and stop the machine very quickly to minimize the production of bad parts and the waste of downstream capacity. Inspection stations monitor the behavior of machines by the use of control charts telling whether a machine is in control or out of control. Quality control charts can work either on the data obtained from the inspections of produced parts or on the data obtained from the measurement of process parameters [10]. In this paper only quality control based on inspected parts is considered. When the control chart identifies an out-of-control situation, the machine that produced the feature that has been monitored is stopped so that an operator can investigate and possibly fix the problems that drove it out of control. Different inspection policies can be performed. For example in this paper all the parts that are processed in the line are measured. Alternatively it is possible to measure only a fraction of the processed parts. The first policy is performed when the cost of nonconforming parts is high and the cost of inspection is low. The sampling inspection policy entails a longer response time but is cheaper than the 100% inspection.

1.3 Literature review

Analytical tools in Manufacturing System Engineering have been developed by Buzacott [2], Gershwin [5] and others. In the field of SQC applied to production systems, Montgomery [10] contributed in the diffusion of statistical process control theory and Raz [12] dealt with the problem of the optimal allocation of inspection stations in multistage production lines. Tempelmeier and Burger [13] and Helber [7] proposed analytical methods for studying production lines with quality stations and scrap and rework policies.

Only few papers consider the intersection between quality control and system dynamics. Kim and Gershwin proposed a method for the evaluation of the performance of a production line considering the quality control issues and dealing with the delay in the quality information. Colledani and Tolio [3] proposed an approximate method for the analysis of production lines, in which SQC techniques are applied, which takes into accent scrap and rework policies.

In this paper we extend the model presented in [8] by modelling the first machine of the 2M1B line having separate down states for the two types of failures, therefore allowing a better characterization of the machine behavior. Numerical results are compared with those of the previous model.

1.4 Outline

The paper is structured as follows: in the next section we define the model assumptions and we investigate the behavior of a machine in isolation. In Section 3 we present the model of a 2-machine-

1-buffer line and in Section 4 we show the solution technique and its validation. Discussions on the behavior of production lines with different inspection policies, based on numerical experiments, are presented in Section 5. Section 6 provides summary of the contribution of this paper and investigate future research.

2 Mathematical models

2.1 Modelling assumptions

In this section, we specify the assumptions used in this work to model a production line with quality failure. *A manufacturing flow line* (also called *transfer line* or *production line*) is a system with a very special structure. It consists of work stations (machines M_1, M_2, ..., M_k) and finite storage areas (also called *buffers* B_1, B_2, ..., B_{k-1}) arranged in a linear network. Material flows from outside the system to M_1, then to B_1, then to M_2, and so forth until it reaches M_k after which it leaves. Figure 1 depicts a flow line. The squares represent machines and the circles represent buffers.

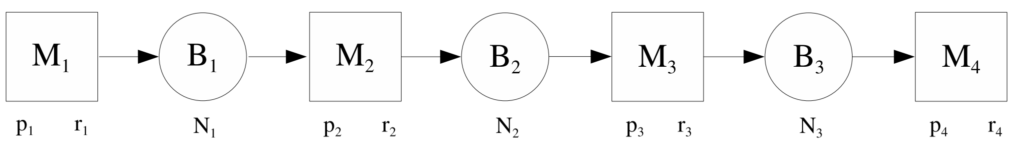

Figure 1: Example of Flow Line

We use the assumption of *independence of events*: events in the future are only contingent on the present state of the system and otherwise independent of each other and the past events. This independence assumption allows the system to be modelled as a *Markov Process*. In production models, this means that the time between failures of a given machine are independent of previous failure times and repair times. Other assumptions made here are:

- The line is modelled as a continuous system which has deterministic cycle times and blockage and starvation.
- The first machine is never starved and the last machine is never blocked.
- The buffer transit time is zero.
- Material flow is conserved: defective parts are reworked or scrapped later elsewhere.

- Only M_1 can have both operational failures and quality failures and these failures are operation dependent (ODF). M_2 has only operational failures (ODF).

- All the failures and repairs are uncorrelated.

- Inspections are nondestructive, operation dependent and have *Type II errors* only. (H_0 is not rejected although it is false, where H_0 is the hypothesis that the machine is producing non-defective parts.)

- μ_i is the speed at which Machine i processes material while it is operating and not constrained by the other machine or the buffer.

- p_i is the the reciprocal of the *Mean Time To Operational Failure (MTTF)* of M_i.

- r_i and r_i^Q are the the reciprocal of the *Mean Time To Repair (MTTR)* of operational failures and quality failures of M_i, respectively.

- g_i is the reciprocal of the *Mean Time To Quality Failure (MTQF)* of M_i. A more stable operation leads to a larger *MTQF* and a smaller g_i.

- h_i is the reciprocal of the *Mean Time To Detect* (MTTD) of M_i. A more reliable inspection leads to a shorter MTTD and a larger h_i.

- All the indicated transition times are assumed to follow exponential distributions.

2.2 Single machine model

There are many possible ways to characterize a machine for the purpose of simultaneously studying quality and quantity issues. Kim and Gerhwin [8] proposed a three-state machine model. In their paper, a machine produces good parts in state 1 and produces bad parts due to a quality failure in state -1. When the machine is under repair (state 0), an operator can not tell whether the machine is down due to a quality failure or an operational failure. Therefore, whenever a machine is under repair, the operator fixes the machine completely so that the machine goes back to state 1. As a result, the repair rates of the operational and quality down states (r and r^Q) are identical. This is not always realistic because it means that every repair is a quality repair, even if the machine was in state 1.

In this paper we model a machine as a discrete state, continuous time Markov process as represented in Figure 2. The model is an improvement of [8] because the machine has five states, therefore allowing operational and quality failures to have separate down states. This means that an operator can perform different repairs associated with different failures; also we can specify different repair rates r and r^Q associated with the two failures. In the previous model, once the machine was down, the operator checked and repaired every source of failure because he didn't know the reason of the failure. The five states are:

- State 1: The machine is operating and producing good parts.
- State D_1: The machine is not operating due to an operational failure that occurred when the machine was in state 1. When this failure is repaired, the machine returns to state 1.
- State -1: The machine is operating and producing bad parts, but the operator does not know this yet.
- State D_{-1}: The machine is not operating due to an operational failure that occurred when the machine was in state 1. When this failure is repaired, the machine returns to state -1.
- State D_Q: The machine is not operating due to a quality failure. When this failure is repaired, the machine returns to state 1.

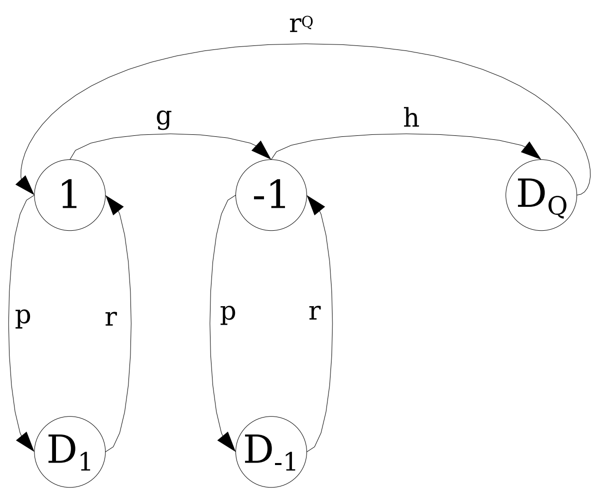

Figure 2: Proposed five-state machine model

2.3 Isolated machine behavior

When a machine is in state 1, it can fail due to an operational failure such as a motor or fuse burnout. In that case it goes to state D_1 with probability rate p. After that an operator fixes it, and the machine goes back to state 1 with probability rate r. Sometimes, due to an assignable cause, the machine goes out of control and begins to produce bad parts, so there is a transition from state 1 to state -1 with a probability rate of g. The machine, when it is in state -1, can be stopped for two reasons: it may experience the same kind of operational failure as it does when it is in state 1; or the operator may stop it for repair when he learns that it is producing bad parts. In the first case it goes in state D_{-1} at rate p. Then an operator fixes the machine without knowing it was making bad parts. Therefore, only the operational failure is repaired and when becoming operational the machine will still make bad parts. It returns to state -1 at rate r. In the second case the transition from state -1 to state D_Q occurs at probability rate h. In this case the operator fixes the quality failure and the machine goes back to state 1 with probability rate r^Q. Operational failure rates do not depend upon whether the machine is in state 1 or in state -1. The transition rate between -1 and D_{-1} is the same as that between 1 and D_1.

To determine the production rate of a single machine, we determine the steady-state probability distribution, calculated based on the probability balance principle: in steady state, the probability rate of leaving a state is the same as the probability rate of entering that state. We have

$$(g+p)P(1) = rP(D_1) + r^Q P(D_Q) \tag{1}$$

$$(h+p)P(-1) = rP(D_{-1}) + gP(1) \tag{2}$$

$$r^Q P(D_Q) = hP(-1) \tag{3}$$

$$rP(D_1) = pP(1) \tag{4}$$

$$rP(D_{-1}) = pP(-1) \tag{5}$$

The probabilities must also satisfy the normalization equation:

$$P(1) + P(-1) + P(D_1) + P(D_{-1}) + P(D_Q) = 1 \tag{6}$$

The solution of (1)-(6) is

$$P(1) = \frac{hrr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{7}$$

$$P(-1) = \frac{grr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{8}$$

$$P(D_1) = \frac{hpr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{9}$$

$$P(D_{-1}) = \frac{pgr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{10}$$

$$P(D_Q) = \frac{phr}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{11}$$

The *total production rate*, including good and bad parts, is

$$P_T = \mu(P(1) + P(-1)) = \mu \frac{(h+g)rr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{12}$$

The *effective production rate*, the production rate of good parts only, is

$$P_E = \mu P(1) = \mu \frac{hrr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{13}$$

The *yield* is

$$Y = \frac{P_E}{P_T} = \frac{P(1)}{P(1) + P(-1)} = \frac{h}{h+g} \tag{14}$$

2.4 Simplified two-machine-one-buffer (2M1B) model

2.4.1 State aggregation: from five-state to two-state machine model

In production lines, machines are stopped either because an operational or a quality failure occurred, or because they are starved or blocked by other machines. The simplest non-trivial model of a production line is a two-machine-one buffer line (2M1B). The 2M1B model is particularly useful in the study of long lines because it is used as a building-block in the decomposition technique [5]. In this paper, only the first machine has quality failures and needs a five-state model. The second machine having only operational failures can be described with the same two-state model as the one proposed by Gershwin [5]. We assume that neither the first machine nor its operator, when producing, can distinguish between good and bad parts; from this point of view each part is identical to the others. The control station will then determine if a part is defective or not. Because of the assumption that each machine works on different features, quality failures at an operation do not influence the quality of other operations. This mean that if a machine produces a bad part, other machines will not see any difference between this part and a good part. Buffers also will not see any difference in the flow of parts that pass through them, when one of the machines starts making bad parts. This allows us to treat machines as having only one up state and several down states. Therefore, in first approximation we can add together the different down states and we end up with a two-state model.

In reality in a production system every machine has quality failures. Therefore each machine has a behavior that can be captured with a five-state model. However, for systems with finite buffers, studying models with more than one machine with five states leads to a very complicated system of differential equations. For example, in a two-machine line with both machines having five states, we must solve a system of 25 internal transition equations and many boundary equations. The transformation of a five-state machine into a two-state machine, allows us to study lines where each machine has quality failures and where there is a finite buffer.

Here, we derive a relationship between a five-state and a two-state model: the two up states of the five-state-machine (state 1 and state -1) are consolidated into the up state of the two-state-model, as depicted in Figure 3. The three down states of the five-state model are consolidated into the down state of the two machine model. We refer to the five-state model with the superscript 5 and to the two-state model with the superscript 2. The parameters of the two-state model are p' and r'. From equations (7)-(11) we have:

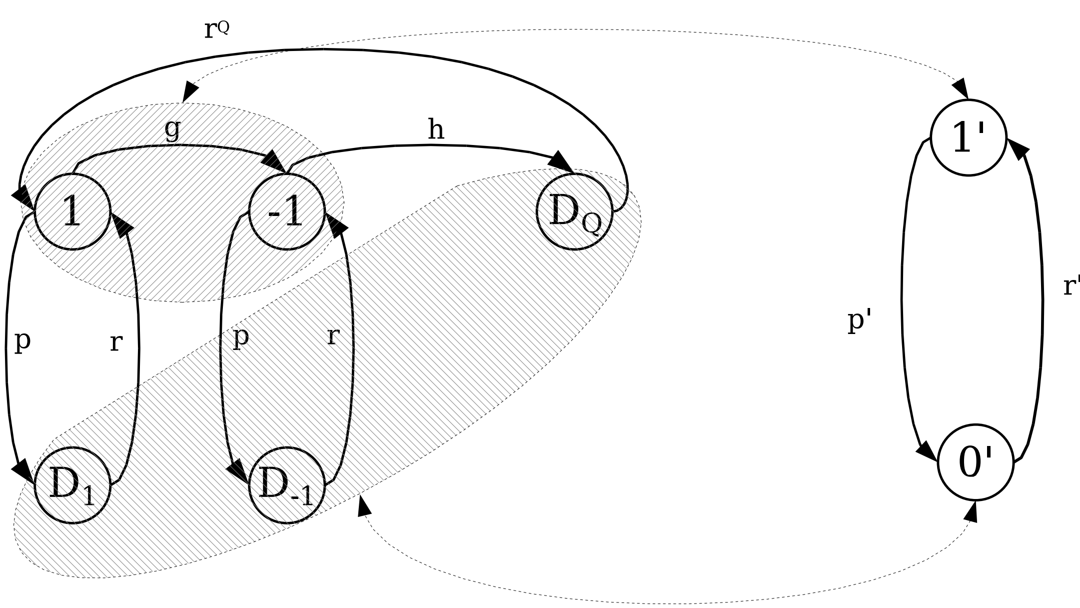

Figure 3: Transformation of a five-state model into a two-state model

$$P^5(1) = \frac{hrr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{15}$$

$$P^5(-1) = \frac{grr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{16}$$

$$P^5(D_1) = \frac{hpr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{17}$$

$$P^5(D_{-1}) = \frac{pgr^Q}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{18}$$

$$P^5(D_Q) = \frac{phr}{hrr^Q + grr^Q + phr^Q + pgr^Q + ghr} \tag{19}$$

For a two-state machine in isolation, the probability of the machine being in each state is:

$$P^2(1') = \frac{r'}{p' + r'} \tag{20}$$

$$P^2(0') = \frac{p'}{p' + r'} \tag{21}$$

The probabilities of the states of the five-state model are calculated as follows:

$$P^5(1) + P^5(-1) = P^2(1') \tag{22}$$

$$P^5(D_1) + P^5(D_{-1}) + P^5(D_Q) = P^2(0') \tag{23}$$

For the failure and repair parameter of the two-state machine we proceed as follows: we add together the failure rates p and p^Q

$$p' = p + p^Q \tag{24}$$

where p^Q is obtained by considering D_Q. The probability rate of exiting state D_Q must be equal to the probability rate of entering it from state $1'$. This is an approximation because in the five-state model there is no transition from state 1 to state D_Q. Then

$$p^Q = \frac{r^Q P^5(D_Q)}{P^2(1')} \tag{25}$$

The repair rate of the two-state model r' is obtained with a weighted average of operational and quality failure:

$$\frac{1}{r'} = \frac{p}{p'}\frac{1}{r} + \frac{p^Q}{p'}\frac{1}{r^Q} \tag{26}$$

2.4.2 Infinite buffer case

Before we consider the general 2M1B line with a finite buffer we analyze two extreme situations. In the first, the storage space between the two machines is infinite. In this case the first machine (M_1) never suffers from blockage. In the second case, there is no buffer space between the machines. This is the other extreme where blockage and starvation take place most frequently. In zero-buffer lines whenever one of the machines stops, the other one also stops. In addition, if the machines have different operation rates, when both of them are working, the production rate is $min[\mu_1, \mu_2]$. To derive expressions for the total production rate and the effective production rate, we observe that when there is infinite buffer capacity between the two machines (M_1, M_2), the total production rate of the 2M1B system is a minimum of the total production rates of M_1 and M_2. The total production rate of machine i is given by (12), so the total production rate of the 2M1B system is

$$P_T^{\infty} = \min\left[\mu_1 \frac{(h_1+g_1)r_1 r_1^Q}{h_1 r_1 r_1^Q + g_1 r_1 r_1^Q + p_1 h_1 r^Q + p_1 g_1 r_1^Q + g_1 h_1 r_1},\ \mu_2 \frac{r_2}{p_2+r_2}\right] \tag{27}$$

The probability that machine M_1 does not add non-conformities is the same as (14). The probability that machine M_2 does not add non-conformities is 1 since the second machine does not have quality failures. Since there is no scrap and rework in the system, the system yield is

$$Y_{sys} = \frac{h_1}{h_1+g_1} \tag{28}$$

As a result, the effective production rate is

$$P_E^{\infty} = Y_{sys} P_T^{\infty} \tag{29}$$

2.4.3 Zero buffer case

To calculate the production rate we follow the method of Kim and Gershwin [8]; for a detailed explanation refer to [11]. The total production rate is:

$$P_T^0 = \frac{Min[\mu_1, \mu_2]}{1 + \frac{p_1^b}{r_1^b} + \frac{h_1^b g_1^b}{r_1^b(h_1^b+g_1^b)} + \frac{p_2^b}{r_2^b}} \tag{30}$$

where the superscript b refers to the reduction of the probability rates for the fastest machine. The effective production rate is

$$P_E^0 = \frac{h_1^b}{h_1^b+g_1^b} P_T^0 \tag{31}$$

In the next Section we study 2M1B model with finite storage capacity. For a each set of parameters, the results of the 2M1B line with buffer of finite capacity must lie between the two extreme cases.

3 Solution method for a Finite-Buffer System

In this chapter we study a 2M1B line in which the first machine is represented by a five-state model and the second machine by a two-state model. We present the model, a solution technique and a validation of the proposed 2M1B line model.

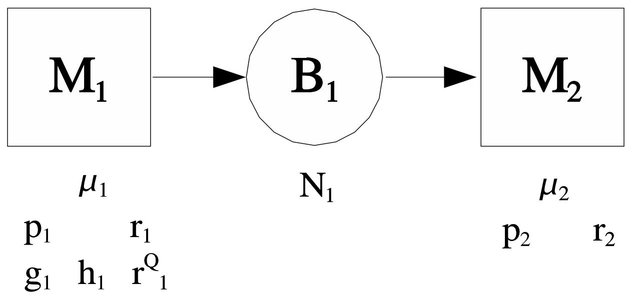

Figure 4: 2M1B line

The state of the 2M1B line illustrated in Figure 4 is (x, α_1, α_2) where:

- x: the amount of material in buffer B with $0 < x < N$
- α_1: the state of machine M_1 ($\alpha_1 = 1,\ -1,\ D_1,\ D_{-1}\ or\ D_Q$)
- α_2: the state of machine M_2 ($\alpha_1 = 1\ or\ 0$)

The parameters of machine M_1 are $\mu_1, p_1, r_1, g_1, h_1, r_1^Q$. The parameters of machine M_2 are μ_2, p_2, r_2. The buffer size is N. The probabilistic behavior of the 2M1B is described by probability density functions ($f(x, \alpha_1, \alpha_2)$) when buffer B is neither empty nor full, and by probability masses ($P(0, \alpha_1, \alpha_2)$ and $P(N, \alpha_1, \alpha_2)$) when the buffer is empty or full. If we find all the probability density functions and the probability masses, we can calculate the performance measures of the 2M1B line. The probability density functions and probability masses are to be found by solving the internal transition equations and the boundary transition equations presented below.

3.1 Internal Transition Equations

When the buffer is neither empty nor full its level can rise or fall depending on the states of the machines. Since it can change only a small amount during a short time interval, it is reasonable to use a continuous probability density $f(x, \alpha_1, \alpha_2)$ and differential equations to describe its behavior. For example, the probability of finding both machines operational with the buffer level between x and $x + \delta x$ at time $t + \delta t$ is given by $f(x, 1, 1, t + \delta t)\delta x$, where

$$f(x,1,1,t+\delta t) = -(g_1+p_1+p_2)\delta t f(x+(\mu_2-\mu_1)\delta t, 1, 1) + r_1 \delta t f(x+\mu_2\delta t, D_1, 1) + \\ + r_1^Q \delta t f(x+\mu_2\delta t, D_Q, 1) + r_2\delta t f(x-\mu_1\delta t, 1, 0) + o(\delta t)$$

The first term, except for a factor of δx, is the probability of transition from between $(x + (\mu_2 - \mu_1)\delta t, 1, 1)$ and $(x + (\mu_2 - \mu_1)\delta t + \delta x, 1, 1)$ at time t,to between $(x, 1, 1)$ and $(x + \delta x, 1, 1)$ at time $t + \delta t$. After linearizing and letting $\delta t \to 0$, this equation becomes

$$f(x,1,1) = (\mu_2-\mu_1)\frac{\partial f(x,1,1)}{\partial x} - (g_1+p_1+p_2) f(x,1,1) + r_1 f(x,D_1,1) + r_1^Q f(x,D_Q,1) + r_2 f(x,1,0)$$

In steady state $\frac{\partial f}{\partial t} = 0$. Therefore we have

$$(\mu_2-\mu_1)\frac{df(x,1,1)}{dx} - (p_1+g_1+p_2) f(x,1,1) + r_1 f(x,D_1,1) + r_1^Q f(x,D_Q,1) + r_2 f(x,1,0) = 0 \quad (32)$$

In the same way we derive the other nine internal transition equations:

$$-\mu_1\frac{df(x,1,0)}{dx} - (p_1+g_1+r_2) f(x,1,0) + r_1 f(x,D_1,0) + r_1^Q f(x,D_Q,0) + p_2 f(x,1,1) = 0 \quad (33)$$

$$(\mu_2-\mu_1)\frac{df(x,-1,1)}{dx} - (p_1+h_1+p_2) f(x,-1,1) + r_1 f(x,D_{-1},1) + g_1 f(x,1,1) + r_2 f(x,-1,0) = 0 \quad (34)$$

$$-\mu_1\frac{df(x,-1,0)}{dx} - (p_1+h_1+r_2) f(x,-1,0) + r_1 f(x,D_{-1},0) + g_1 f(x,1,0) + p_2 f(x,-1,1) = 0 \quad (35)$$

$$-(r_1+r_2) f(x,D_1,0) + p_2 f(x,D_1,1) + p_1 f(x,1,0) = 0 \quad (36)$$

$$\mu_2\frac{df(x,D_1,1)}{dx} - (r_1+p_2) f(x,D_1,1) + r_2 f(x,D_1,0) + p_1 f(x,1,1) = 0 \quad (37)$$

$$-(r_1+r_2) f(x,D_{-1},0) + p_2 f(x,D_{-1},1) + p_1 f(x,-1,0) = 0 \quad (38)$$

$$\mu_2 \frac{df(x, D_{-1}, 1)}{dx} - (r_1 + p_2) f(x, D_{-1}, 1) + r_2 f(x, D_{-1}, 0) + p_1 f(x, -1, 1) = 0 \tag{39}$$

$$-(r_1^Q + r_2) f(x, D_Q, 0) + p_2 f(x, D_Q, 1) + h_1 f(x, -1, 0) = 0 \tag{40}$$

$$\mu_2 \frac{df(x, D_Q, 1)}{dx} - (r_1^Q + p_2) f(x, D_Q, 1) + r_2 f(x, D_Q, 0) + h_1 f(x, -1, 1) = 0 \tag{41}$$

3.2 Boundary Transition Equations

The boundary transition equations depend on the relative speeds of machines M_1 and M_2. We have three different cases that we must analyze separately. Within each case we have a set of lower boundary equations (when the buffer is empty) and upper boundary equations (when the buffer is full). We report here only the case of the two machines having equal speeds ($\mu_1 = \mu_2$). The other cases are similar and are discussed in detail in [11]. For the lower boundary we first determine the transient states, that is, they have zero steady state probability:

$$P(0, 1, 0) = P(0, -1, 0) = 0 \tag{42}$$

These states are transient because $P(0, 1, 0)$ cannot be reached from any state and $P(0, -1, 0)$ can only be reached from $P(0, 1, 0)$.

$$P(0, D_1, 0) = P(0, D_{-1}, 0) = P(0, D_Q, 0) = 0 \tag{43}$$

These states are transient because for example $P(0, D_1, 0)$ can be reached only from itself or $P(0, 1, 0)$. It cannot be reached from $P(0, D_1, 1)$ or $P(0, 1, 1)$ since the second machine cannot fail. Then we characterize the boundary equations:

$$-r_1 P(0, D_1, 1) + p_1 P(0, 1, 1) + \mu_2 f(0, D_1, 1) = 0 \tag{44}$$

$$-r_1 P(0, D_{-1}, 1) + p_1 P(0, -1, 1) + \mu_2 f(0, D_{-1}, 1) = 0 \tag{45}$$

$$-r_1^Q P(0, D_Q, 1) + h_1 P(0, -1, 1) + \mu_2 f(0, D_Q, 1) = 0 \tag{46}$$

$$-(p_1 + g_1 + p_2) P(0, 1, 1) + r_1 P(0, D_1, 1) + r_1^Q P(0, D_Q, 1) = 0 \tag{47}$$

$$-(p_1+h_1+p_2)P(0,-1,1)+r_1P(0,D_{-1},1)+g_1P(0,1,1)=0 \tag{48}$$

$$\mu_1 f(0,1,0)=p_2P(0,1,1) \tag{49}$$

$$\mu_1 f(0,-1,0)=p_2P(0,-1,1) \tag{50}$$

For the upper boundary region we first determine the transient states:

$$P(N,D_1,1)=P(N,D_{-1},1)=P(N,D_Q,1)=0 \tag{51}$$

$$P(N,D_1,0)=P(N,D_{-1},0)=P(N,D_Q,0)=0 \tag{52}$$

and then we characterize the boundary equations:

$$-r_2P(N,1,0)+p_2P(N,1,1)+\mu_1 f(N,1,0)=0 \tag{53}$$

$$-r_2P(N,-1,0)+p_2P(N,-1,1)+\mu_1 f(N,-1,0)=0 \tag{54}$$

$$-(p_1+g_1+p_2)P(N,1,1)+r_2P(N,1,0)=0 \tag{55}$$

$$-(p_1+h_1+p_2)P(N,-1,1)+r_2P(N,-1,0)+g_1P(N,1,1)=0 \tag{56}$$

$$\mu_2 f(N,D_1,1)=p_1P(N,1,1) \tag{57}$$

$$\mu_2 f(N,D_{-1},1)=p_1P(N,-1,1) \tag{58}$$

$$\mu_2 f(N,D_Q,1)=h_1P(N,-1,1) \tag{59}$$

3.3 Normalization equation

In addition to the internal and boundary equations, all the probability density function and masses must satisfy the normalization equation:

$$\sum_{\alpha_1=-1,1,D_1,D_{-1},D_Q} \sum_{\alpha_2=0,1} \left[\int_0^N f(x,\alpha_1,\alpha_2)dx + P(0,\alpha_1,\alpha_2) + P(N,\alpha_1,\alpha_2) \right] = 1 \tag{60}$$

4 Solution Technique

4.1 Solution of the Internal Equations

We are dealing with ordinary linear differential equations with constant coefficients. Therefore it is logical to assume an exponential form for the solution to the steady state density functions. This approach worked successfully in the continuous model with perfect quality [5] and with a simpler model of quality failure [8]. Therefore, we assume a solution of the form:

$$f(x,\alpha_1,\alpha_2) = e^{\lambda x} G_1(\alpha_1) G_2(\alpha_2) \tag{61}$$

in which we must determine $\lambda, G_1(1), G_1(-1), G_1(D_1), G_1(D_{-1}), G_1(D_Q), G_2(1), G_2(0)$, a total of 8 unknowns. This form satisfies the transition equations if all of the following equations are met. After substituting (61) into (32) - (41) we have:

$$\{\lambda(\mu_2-\mu_1)-(p_1+g_1+p_2)\}G_1(1)G_2(1)+r_1G_1(D_1)G_2(1)+r_1^QG_1(D_Q)G_2(1)+r_2G_1(1)G_2(0) = 0 \tag{62}$$

$$-\{\lambda\mu_1 + (p_1+g_1+r_2)\}G_1(1)G_2(0) + r_1G_1(D_1)G_2(0) + r_1^QG_1(D_Q)G_2(0) + p_2G_1(1)G_2(1) = 0 \tag{63}$$

$$\{(\lambda(\mu_2-\mu_1)-(p_1+h_1+p_2)\}G_1(-1)G_2(1)+r_1G_1(D_{-1})G_2(1)+g_1G_1(1)G_2(1)+r_2G_1(-1)G_2(0) = 0 \tag{64}$$

$$-\{\lambda\mu_1+(p_1+h_1+r_2)\}G_1(-1)G_2(0)+r_1G_1(D_{-1})G_2(0)+g_1G_1(1)G_2(0)+p_2G_1(-1)G_2(1) = 0 \tag{65}$$

$$-(r_1+r_2)G_1(D_1)G_2(0) + p_2G_1(D_1)G_2(1) + p_1G_1(1)G_2(0) = 0 \tag{66}$$

$$\{\lambda\mu_2 - (r_1+p_2)\}G_1(D_1)G_2(1) + r_2G_1(D_1)G_2(0) + p_1G_1(1)G_2(1) = 0 \tag{67}$$

$$-(r_1+r_2)G_1(D_{-1})G_2(0)+p_2G_1(D_{-1})G_2(1)+p_1G_1(-1)G_2(0)=0 \quad (68)$$

$$\{\lambda\mu_2-(r_1+p_2)\}G_1(D_{-1})G_2(1)+r_2G_1(D_{-1})G_2(0)+p_1G_1(-1)G_2(1)=0 \quad (69)$$

$$-(r_1^Q+r_2)G_1(D_Q)G_2(0)+p_2G_1(D_Q)G_2(1)+h_1G_1(-1)G_2(0)=0 \quad (70)$$

$$\{\lambda\mu_2-(r_1^Q+p_2)\}G_1(D_Q)G_2(1)+r_2G_1(D_Q)G_2(0)+h_1G_1(-1)G_2(1)=0 \quad (71)$$

Now we have 10 equations in 8 unknowns. Thus, there must be eight independent equations and two dependent ones in order for us to determine these quantities. To simplify the study of this system, we can divide each equation by the most frequent $G_1(\alpha_1)G_2(\alpha_2)$ within that equation. Therefore we divide equation (62) by $G_1(1)G_2(1)$, equation (63) by $G_1(1)G_2(0)$, equation (64) by $G_1(-1)G_2(1)$, equation (65) by $G_1(-1)G_2(0)$, equation (66) by $G_1(D_1)G_2(0)$, equation (67) by $G_1(D_1)G_2(1)$, equation (68) by $G_1(D_{-1})G_2(0)$, equation (69) by $G_1(D_{-1})G_2(1)$, equation (70) by $G_1(D_Q)G_2(0)$, equation (71) by $G_1(D_Q)G_2(1)$ and we can define new variables:

$$\frac{G_1(D_1)}{G_1(1)}=A_1 \qquad \frac{G_1(D_{-1})}{G_1(-1)}=A_2 \qquad \frac{G_1(D_Q)}{G_1(-1)}=A_3$$
$$\frac{G_1(D_Q)}{G_1(1)}=B_1 \qquad \frac{G_1(1)}{G_1(-1)}=B_2 \qquad \frac{G_2(1)}{G_2(0)}=C$$

If we rewrite equations (62) - (71) we have:

$$\lambda(\mu_2-\mu_1)-(p_1+g_1+p_2)+r_1A_1+r_1^QB_1+r_2\frac{1}{C}=0 \quad (72)$$

$$-\lambda\mu_1-(p_1+g_1+r_2)+r_1A_1+r_1^QB_1+p_2C=0 \quad (73)$$

$$\lambda(\mu_2-\mu_1)-(p_1+h_1+p_2)+r_1A_2+g_1B_2+r_2\frac{1}{C}=0 \quad (74)$$

$$-\lambda\mu_1-(p_1+h_1+r_2)+r_1A_2+g_1B_2+p_2C=0 \quad (75)$$

$$-(r_1+r_2)+p_2C+p_1\frac{1}{A_1}=0 \quad (76)$$

$$\lambda\mu_2-(r_1+p_2)+r_2\frac{1}{C}+p_1\frac{1}{A_1}=0 \quad (77)$$

$$-(r_1 + r_2) + p_2 C + p_1 \frac{1}{A_2} = 0 \tag{78}$$

$$\lambda \mu_2 - (r_1 + p_2) + r_2 \frac{1}{C} + p_1 \frac{1}{A_2} = 0 \tag{79}$$

$$-(r_1^Q + r_2) + p_2 C + h_1 \frac{1}{A_3} = 0 \tag{80}$$

$$\lambda \mu_2 - (r_1^Q + p_2) + r_2 \frac{1}{C} + h_1 \frac{1}{A_3} = 0 \tag{81}$$

We notice that equations (75), (77), (79) and (81) are linearly dependent on the others, therefore are eliminated. If we rearrange equations (76), (78) and (80) we obtain

$$A_1 = \frac{p_1}{r_1 + r_2 - p_2 C} \tag{82}$$

$$A_2 = \frac{p_1}{r_1 + r_2 - p_2 C} = A_1 \tag{83}$$

$$A_3 = \frac{h_1}{r_1^Q + r_2 - p_2 C} = B_1 B_2 \tag{84}$$

From equation (73) we derive λ

$$\lambda = \frac{1}{\mu_1} \left[-(p_1 + g_1 + r_2) + r_1 A_1 + r_1^Q B_1 + p_2 C \right] \tag{85}$$

We then substitute A_1, A_2, B_2, λ into equations (72) and (74)+(72). We denote $\frac{\mu_2 - \mu_1}{\mu_1} = \delta - 1$. After rearranging we have two equations in two unknowns (B_1 and C)

$$B_1 = f(C) = \frac{(p_1 + g_1 + r_2)}{r_1^Q} + \frac{(p_2 - r_2)}{r_1^Q \delta} - \frac{p_2(1 - 1/\delta)C}{r_1^Q} - \frac{r_1 p_1}{r_1^Q (r_1 + r_2 - p_2 C)} - \frac{r_2}{r_1^Q \delta C} \tag{86}$$

$$C = f(B_1) = \frac{r_1^Q + r_2}{p_2} - \frac{g_1 h_1}{p_2 \left[h_1 - g_1 + r_1^Q B_1 \delta \right] B_1} \tag{87}$$

If we plug equation (87) into (86) we get a single equation $g(B_1)$ with one unknown. In Figure 5 we display the plot of (88). It is very easy to locate the roots of the function with numerical tools and thus obtain the solution to the internal equations.

For a detailed explanation of the numerical techniques adopted here, refer to [11].

$$\begin{aligned} g(B_1) = & -(p_1+g_1+r_2)(\delta-1)-(p_1+g_1+p_2)+r_1^Q\delta B_1 \\ & +\frac{r_1 p_1 \delta(h_1-g_1+r_1^Q\delta B_1)B_1}{(r_1-r_1^Q)(h_1-g_1+r_1^Q\delta B_1)B_1+g_1h_1} \\ & +(\delta-1)\left[r_1^Q+r_2-\frac{g_1h_1}{(h_1-g_1+r_1^Q\delta B_1)B_1}\right] \\ & +\frac{p_2r_2(h_1-g_1+r_1^Q\delta B_1)B_1}{(r_1^Q+r_2)(h_1-g_1+r_1^Q\delta B_1)B_1-g_1h_1}=0 \end{aligned} \tag{88}$$

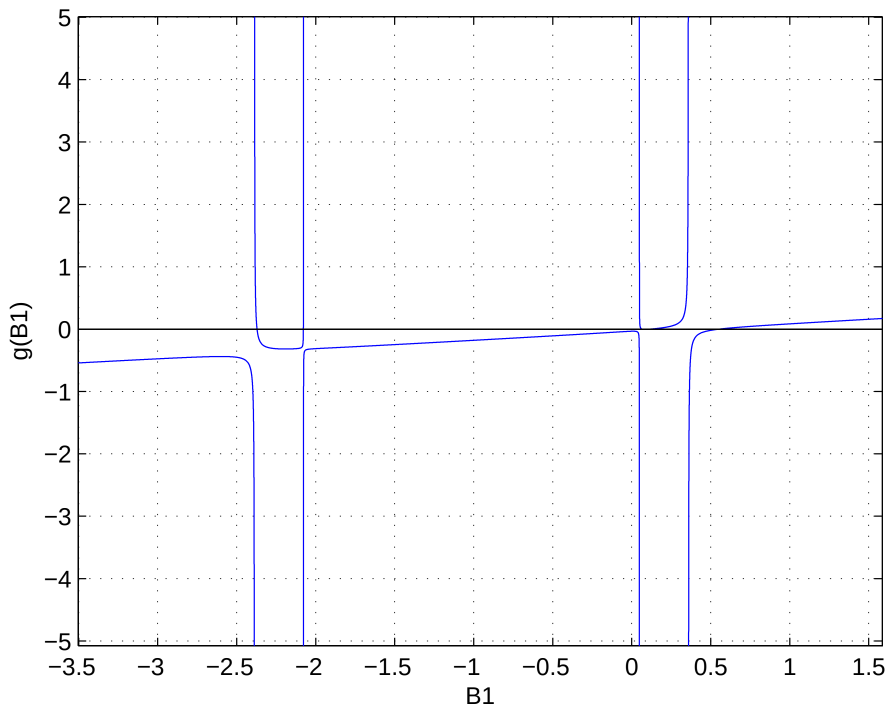

Figure 5: Plot of equation (88)

The other internal expression parameters are obtained from (72)-(88).

The general expression of the probability density function is

$$f(x,\alpha_1,\alpha_2)=\sum_{i=1}^{RN} c_i f_i(x,\alpha_1,\alpha_2) \tag{89}$$

where RN is the number of roots of equation (88).

The remaining unknowns, including coefficients c_i $i=1,2...,RN$ and probability masses at the boundaries, can be calculated by solving the boundary transition equations and the normalization equation.

4.2 Solution of the Boundary Equations

The boundary equations (43) – (59) are linear equations in which the unknowns are the probability masses and the coefficients in equation (89). Some of the probability masses are 0 according to the equations, and functions $f_i(x, \alpha_1, \alpha_2)$ are found by solving the internal transition equations in Section 3.1. The boundary equations can be simplified as follows:

- Eliminate the probability masses which are known to be 0.
- Temporarily set $P(0, 1, 1) = 1$.
- Substitute $f(x, \alpha_1, \alpha_2) = c_1 f_1(x, \alpha_1, \alpha_2) + c_2 f_2(x, \alpha_1, \alpha_2) + c_3 f_3(x, \alpha_1, \alpha_2) + c_4 f_4(x, \alpha_1, \alpha_2) + c_5 f_5(x, \alpha_1, \alpha_2)$
  where $f_i(x, \alpha_1, \alpha_2) = e^{\lambda_i x} G_1^i(\alpha_1) G_2^i(\alpha_2)$.

Then, we have an equation $AX = B$ where X is a vector of coefficients c_i and of probability masses. This equation is in matrix form and can be solved using a linear equation solver; all the unknowns are expressed as multiples of $P(0, 1, 1)$. Then, the value of $P(0, 1, 1)$ can be calculated from the normalization equation (60).

4.3 Evaluation of performance

Throughput

The total production rate, the production of good and bad parts is:

$$P_T = P_T^1 = \sum_{\alpha_2=0,1} \mu_1 \left[\int_0^N (f(x, 1, \alpha_2) + f(x, -1, \alpha_2))dx + P(0, -1, \alpha_2) + P(0, 1, \alpha_2) \right] + \\ + \mu_2 \Big[P(N, -1, 1) + P(N, 1, 1) \Big] \tag{90}$$

The effective production rate of the first machine is:

$$P_E^1 = \sum_{\alpha_2=0,1} \mu_1 \left[\int_0^N f(x, 1, \alpha_2)dx + P(0, 1, \alpha_2) \right] + \mu_2 P(N, 1, 1) \tag{91}$$

The fraction of parts produced by the first machine that are good is $Y_1 = \frac{P_E^1}{P_T}$.

Average Inventory

The average number of parts in the buffer is:

$$\bar{x} = \sum_{\alpha_1=-1,1,D_1,D_{-1},D_Q} \sum_{\alpha_2=0,1} \left[\int_0^N x f(x, \alpha_1, \alpha_2)dx + NP(0, \alpha_1, \alpha_2) \right] \tag{92}$$

4.4 Numerical Results

The mathematical model for the two-machine-one-finite-buffer system has been solved. We compare analytical and simulation results in this section. But as we have indicated, we represent discrete parts in this model as a continuous fluid and time as a continuous variable. On the other hand, in simulation and in most real systems, both material and time are discrete. For simulation, a transient period of 100,000 time units and 1,000,000 time units of data collection period are used.

Figure 6 illustrates the comparison of the total production rate and the average inventory from the analytic model and the simulation respectively. By changing machine and buffer parameters, 30 cases are generated and % errors are plotted in the vertical axis. The parameters for these cases are given in [11] and are randomly chosen. The % errors in the production rates are calculated from

$$P_T \quad \%error = \frac{P_T(A) - P_T(S)}{P_T(S)} \times 100(\%)$$

where $P_T(A)$ and $P_T(S)$ are respectively the total production rate calculated from the analytical model and estimated from the simulation. We find the % error for the effective production rate P_E in a similar way. The % error in the average inventory is calculated from

$$Inv \quad \%error = \frac{Inv(A) - Inv(S)}{0.5 \times N} \times 100(\%) \tag{93}$$

where $Inv(A)$ and $Inv(S)$ are average inventory estimated from the analytical model and the simulation respectively and N is buffer size. This equation is an unbiased way to calculate the error in average inventory. The average absolute value of the % errors in the total production rate,

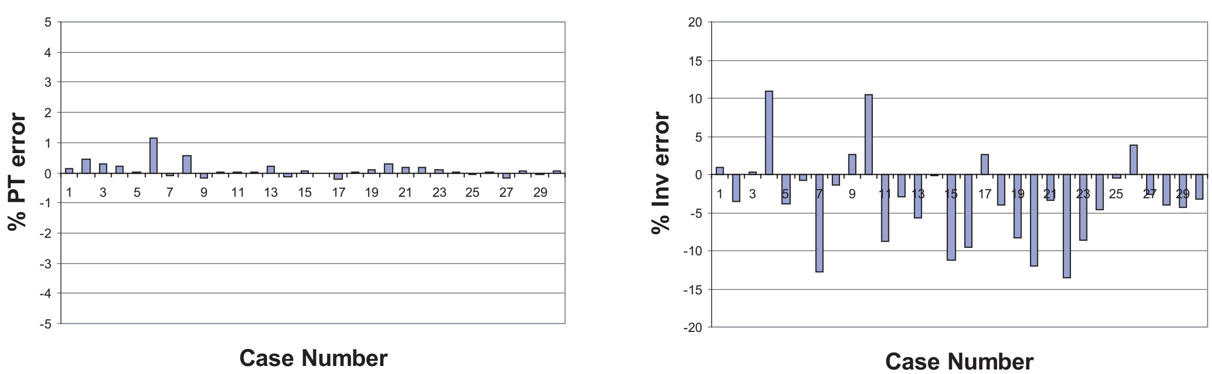

Figure 6: Numerical results

the effective production rate, and the average inventory are 0.14%, 0.22%, and 4.8% respectively.

The observation that the production rates estimates are better than average buffer level estimates is consistent with the rest of the literature [5] [8].

5 Quality information feedback

Sometimes inspection stations are designed to perform multiple inspections at the end of the line. When a bad part is detected, the machine that made that feature is informed of its out-of-control condition and stopped. (This is called *quality information feedback*.) An example of the quality information feedback in 2M1B systems is when M_1 produces defective features but does not have inspection, and M_2 has inspection and it can detect bad features made by M_1.

As detailed in [8], the mean time to detect a bad part is function of the size of the buffer and not only of the sampling policy adopted. In fact, the presence of parts in the buffer delays the inspection of that operation. To take into account the quality information feedback we adjust the transition rate h_1 of M_1 from state -1 to state D_Q, and we call it h_1^Q. We define

$$\chi_{21} = \frac{h_{21}}{\mu_2} \tag{94}$$

where $\frac{1}{h_{21}}$ is the mean time until the inspection at M_2 detects a bad part made by M_1 after M_2 receives the bad part. We call K_1^b the expected number of bad parts generated by M_1 before it is stopped by quality information feedback (from the time it enters state -1 until it enter state D_Q). It is given by

$$K_1^b = (w+1)\chi_{21} + (w+2)(1-\chi_{21})\chi_{21} + (w+3)(1-\chi_{21})^2\chi_{21} + \dots \tag{95}$$

where w is average inventory in the buffer B. After some mathematical manipulation we obtain

$$K_1^b = w + \frac{1}{\chi_{21}} \tag{96}$$

h_1^q is the inverse of the mean time needed for the second machine to detect a bad part produced by M_1.

$$h_1^q = \frac{\mu_1}{K_1^b} \tag{97}$$

Since the average inventory is a function of h_1^q, and h_1^q is dependent on the average inventory, an iterative method is used to get these values. We compare analytical and simulation results as done in the previous section. The average absolute value of the % errors in the total production rate, the effective production rate, and the average inventory are 0.21%, 0.54%, and 6.84% respectively. Figure 7 illustrates the comparison of the total production rate and of the effective production rate.

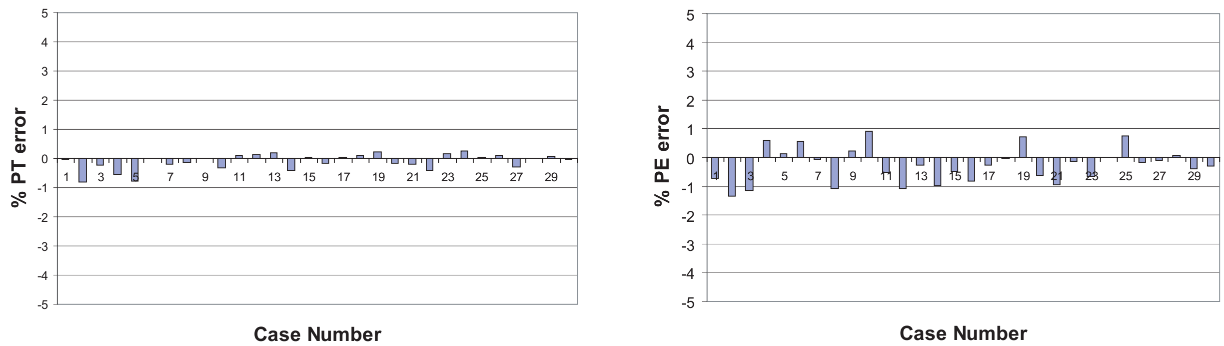

Figure 7: Numerical results for the quality information feedback case

In Figure 8 we let the buffer size increase from 1 to 50 and we compare two situations: we place the same inspection station first within the first machine and then at the end of the line to catch defective parts made by the first machine. The second case has quality information feedback. The total production rate with remote inspection is always higher than than with local inspection. This is intuitive because the machine is stopped less frequently when making defective parts. However, for this reason the system yield is monotonically decreasing as the buffer size increase. In this example, the effective production rate increases up to a certain point and then decreases monotonically as the buffer size increases. This is explained by the fact that a little increase in the buffer size, when it is very small, gives an higher increase in effective production rate than the decrease given by the delay of the quality information. It is very important to observe that the system yield is a function of the buffer size if there is quality information feedback. All these remarks are coherent with those presented by Kim and Gershwin [8].

6 Conclusion

In this paper we have analyzed how production system design, quality and productivity are interrelated. Starting from the model presented in [8] we developed a new Markov process for machines with both quality and operational failures. The first machine is described by a five-state model that makes it possible to consider separately the two types of failures, thus allowing a better characterization of the real system. The second machine is described by a two-state model. The inspection station has been placed either within the first machine or at the end of the line and the two kinds of systems have been compared. When the inspection station is placed at the end of the line we

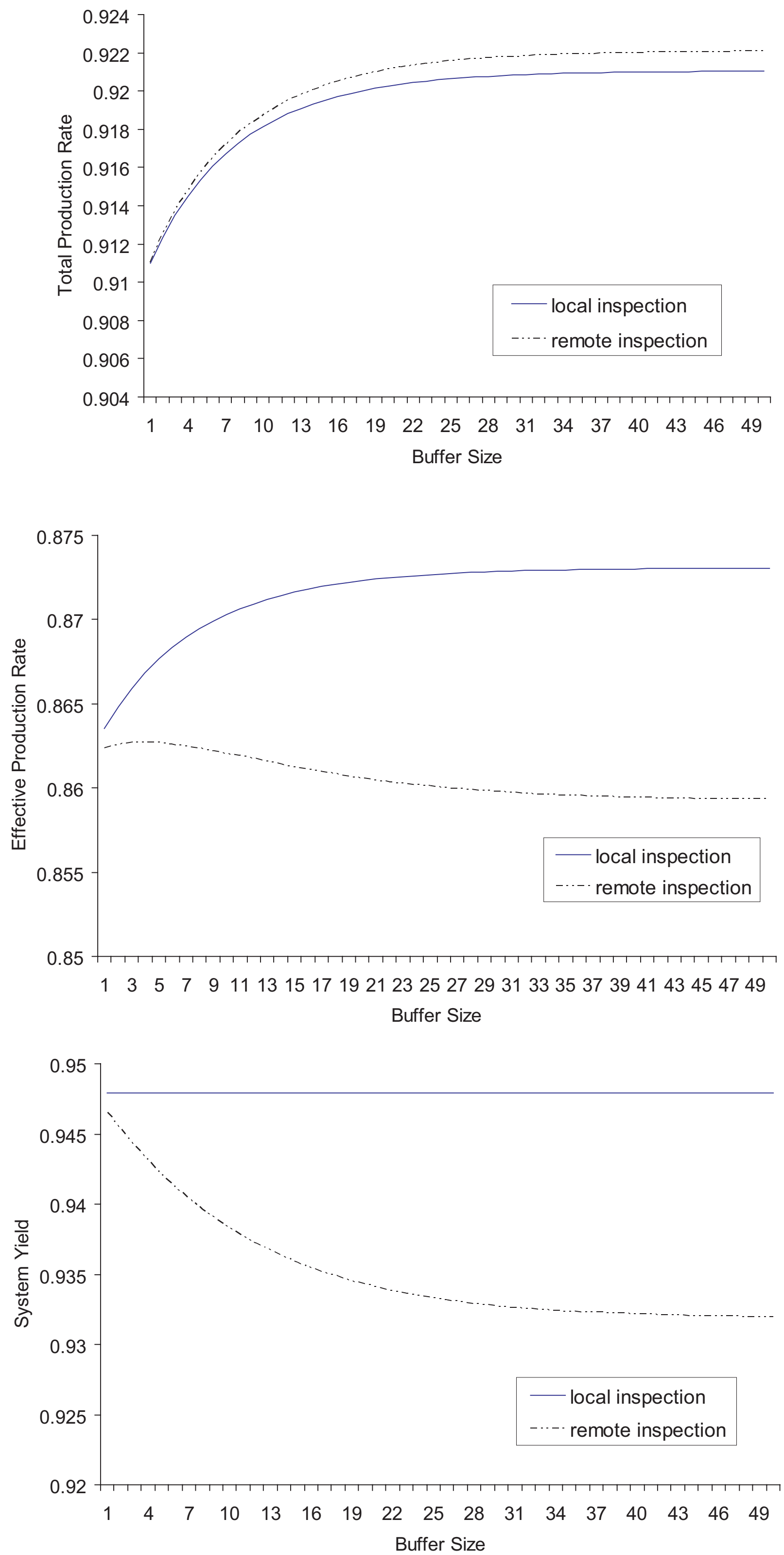

Figure 8: Performances comparison between models with local and remote inspection

analyzed how the buffer causes a delay in the quality information. This delay reduces the system yield. We presented analytic models, solution techniques, performance evaluations and validation of a 2M1B line. The present model allows system designers to investigate different configuration to reach a target effective production rate. The 2M1B line could be used in future as a building block in the decomposition techniques for the study of longer lines.

References

[1] Burman, M., *New results in flow line analysis*, Ph.D. thesis dissertation, Operations Research Center, Massachusetts Institute of Technology, 1995.

[2] Buzacott, J. A. and Shanthikumar, J. G., *Stochastic models of manufacturing systems*, Prentice-Hall, 1993

[3] Colledani, M. and Tolio, T. *Impact of SPC on the performance of production systems*, 5th International conference on Analysis of Manufacturing Systems - Production Management, pp. 77 - 84, May, 2005.

[4] Fujimoto, T., *The evolution of a manufacturing systems at Toyota*, Oxford University Press, 1999

[5] Gershwin, S. B., *Manufacturing systems engineering*, Prentice Hall, 1994

[6] Inman, R. R. Blumenfeld, D. E., Huang, N., and Li, J., *Designing production systems for quality: research opportunities from an automotive industry perspective*, International journal of production research Vol. 41, No. 9, pp. 1953 - 1971, 2003.

[7] Helber, S., *Performance analysis of flow lines with non-linear flow of material*, Springer, 1999.

[8] Kim, J. and Gershwin, S. B., *Integrated quality and quantity modeling of a production line*, OR Spectrum, Vol. 27, pp. 287-314, 2005.

[9] Monden, Y., *Toyota production system - An integrated approach to Just-In-Time*, EMP Books,1998.

[10] Montgomery, D. C., *Introduction to statistical quality control*, John Wiley & Sons, Inc, 1991.

[11] Poffe, A. and Gershwin, S. B., *Integrating Quality and Quantity Modelling in Manufacturing Systems Engineering*, Technical Report, LMP, Massachusetts Institute of Technology, 2005.

[12] Raz, T., *A survey of models for allocating inspection effort in multistage production systems*, Journal of quality technology, Vol. 18, No. 4, pp. 239 - 246, October 1986.

[13] Tempelmeier, H. and Burger, M., *Performance evaluation of unbalanced flow lines with general distributed processing times, failures and imperfect production*, IIE Transaction, Vol. 33, pp. 293 - 302, 2000.