Reliability engineering of large jit production systems

Jose Faria[1], **Eusebio Nunes**[2]

[1] FEUP – Faculty of Engineering, University of Porto
email: jfaria@fe.up.pt
[2] School of Engineering, University of Minho
email: enunes@dps.uminho.pt

Jose Faria holds a degree in electrical engineering from the Faculty of Engineering of the University of Porto, Portugal (1983) and a Ph.D. also from FEUP (1996). The Ph.D. focused on reliability modelling, analysis and evaluation of the integrated management information system of a large European car manufacturer. He is currently an Assistant Professor at FEUP and his research interests include reliability and quality analysis and business operation systems.

Eusebio Nunes holds a degree in production engineering from the School of Engineering of the University of Minho, Portugal (1991), a MsC from the Technical University of Lisbon (1996), and a PhD from the Faculty of Engineering of the University of Porto, Portugal (2005). His PhD thesis addressed reliability evaluation of non-Markovian system with fuzzy parameters. His research interests include reliability analysis methods and quality management systems.

Corresponding author:
Prof. Jose A. Faria
FEUP-DEIG
Rua Roberto Frias, s/n
4200-465 Porto, Portugal
phone: +351.22.508.1831
email: jfaria@fe.up.pt

Reliability engineering of large just-in-time production systems

Keywords: manufacturing system, just-in-time, reliability, modelling, evaluation, case study

Abstract: This paper introduces the rationale and the fundamental elements and algorithms of a reliability engineering methodology, and discusses its application to the design of a large, multi-cell and heterogeneous production system with just-in-time (JIT) deliveries. The failure analysis and the non-reliability costs assessment of such systems is a complex task. In order to cope with such complexity, a two level hierarchical modelling and evaluation framework was developed. According to this framework, the internal behaviour of each manufacturing cell and the overall flow of materials are described, respectively, by local and global models. Local models are firstly obtained from the failure and repair processes of the manufacturing equipment. Then, these models are combined with the failure propagation delays introduced by the work-in-process buffers in order to obtain the system level model. The second part of the paper addresses several design issues of the production system that directly impact the reliability of the deliveries, such as the layout of the plant, the redundancy of the manufacturing equipment and the capacity of the work-in-process buffers. A distinctive feature of the reliability evaluation algorithm resides on the ability to deal with reliability models containing stochastic processes with generalized distributions. This fundamental requirement comes from the fact that repair and failure propagation processes typically present hyper-exponential distributions, e.g., lognormal distributions, that can't be assessed using the conventional reliability techniques. The paper will also explain how the behavioural and structural characteristics of JIT production systems were explored in order to implement effective evaluation algorithms that fit the requirements of this class of systems.

Notation and acronyms

Reliability costs

D_F: delivery failures penalties
T_x: extra working time cost
B_w: work-in-process buffers cost
R_d: redundant equipment cost.
nR: non-reliability cost (sum of D_F and T_x)
iR: reliability improvement cost (sum of B_w and R_d)
cR: total reliability costs (sum of nR and iR)
uc: standard unit of cost

Solutions under study

S_1: wip buffers at the output of the manufacturing cells
S_{2a}: unique buffer at the output of the production system, and unitary operation
S_{2b}: similar to S_{2a} but for batch operation

Buffers

B_b: capacity of buffer b
I_b: average inventory of buffer b
C_b: total cost of the buffer b
Ls_u: lot size at manufacturing unit u (u = m for manufacturing cell, u = a for assembly line)
Ss_u: safety stock at manufacturing unit u

Cost drivers

α^d: non-reliability cost rate (proportional to the duration of the failures)
α^f: non-reliability impulse cost (proportional to the frequency of the failures)
α_b^{Cp}: reliability improvement cost component (proportional to the capacity C_p of buffer b)
α_b^I: reliability improvement cost component (proportional to the average inventory I of buffer b)
$\alpha^x{}_u$: extra work cost rate (proportional to the duration of the extra work at manufacturing unit u)

Stochastic processes

p_ϖ: stochastic process ϖ
$\varpi(t)$: probability density function of process ϖ
m_ϖ: mean "time-to-occur" of process ϖ
r_ϖ: rate of occurrence of process ϖ (exponential processes only)
Δ: deterministic delay
λ: failure process
μ: repair process
ξ: reconfiguration process
γ_b: delay process of buffer b
h(t): Heaviside function

Canonical Models

M_u^x: canonical model of manufacturing unit u (with x = i for internal, x = o for output of cell, and x = b for output of buffer)
Λ_u^x: equivalent failure rate of M_u^x
$\rho_u^x(t)$: probability density function of M_u^x
s_{up}: normal operating state
s_d: down operating state
P_{up}: probability of s_{up}

Numerical application models

H_y: number of working hours per year
d_l: loading failure state
d_d: delivery failure state
γ_l: loading delay
γ_d: delivery delay

1. Introduction

Low production costs and strict compliance with delivery schedules are the two main pillars of competitiveness for companies that operate in the context of just-in-time (JIT) supply chains. The constant market demand for lower lead times and production costs has pushed manufacturing companies to adopt JIT techniques and to implement aggressive inventory reduction programmes. Many of those programmes have been successful in reducing production costs, but they have also had a more negative result as far as delivery reliability is concerned, once the flow of materials becomes much more sensitive to disturbances such as equipment failures and raw material shortages. Manufacturing companies are now looking for a different balance between production costs and delivery reliability, and recognize that work-in-process (wip) and finished products buffers are an indispensable element to guarantee the required reliability level, as shown in [1] and [2]. This idea is reinforced by the results of the survey presented in [3], which show that there has not been any statistically significant change in the inventory to sales ratio after the implementation of JIT techniques.

With this in mind, this paper presents a reliability engineering methodology intended to support the analysis and assessment of heterogeneous multi-cell manufacturing systems, and to help system planners and managers obtain the optimal system design. The optimization criteria is the minimization of the costs that directly depend on the reliability of the manufacturing system: the penalties due to failures on deliveries to the client, D_F; the extra working time costs required to compensate equipment breakdowns, T_x; the wip buffers, B_w; and the redundant equipment, R_d. The sum of the first two cost components is denoted as *nR*, i.e. the *non-reliability cost* of the production system, as these costs come from manufacturing equipment failures. The sum of the two other components is denoted as *iR*, or *reliability improvement cost*, and the sum of *nR* and *iR* is denoted as the production system reliability cost, *cR*.

For large production systems, comprising multiple pieces of heterogeneous equipment submitted to random failure processes, the determination of the optimal design is a complex task that demands effective methodologies and tools. Existing tools for performance analysis and evaluation often impose severe restrictions on the structure and behaviour of the production systems under study, which limit their application to relatively simple production systems.

Analytical models of production systems often impose idealized operating conditions and restrictive assumptions that undermine their application scope and practical usefulness. For example, [4] considers the optimization of the safety stock for a single-part type, single-unreliable machine production system; [5] investigates optimal production control for a tandem of two machines; and [6] analyses an unreliable bottleneck, assuming constant production and demand rate, constant restoration time and exponential failure processes.

Another major limitation of many tools is the assumption that all the stochastic processes have exponential distributions. . Homogeneity is a reasonable assumption for failure processes, but not for repair and buffer processes, which are typically hyper-exponential. As a typical example, consider a buffer whose inventory remains constant in normal operating conditions. That buffer will introduce a fixed (deterministic) delay in the propagation of a failure initiated in a upstream machine. On the other hand, if the buffer stays at the output of a batch cell and its inventory changes overtime according to a saw tooth pattern, it will introduce a uniformly distributed delay that can be modelled by a step distribution.

Very often, process homogeneity is adopted “lightly”, i.e., without a clear estimate of the error that this assumption will introduce in the calculations. However, as discussed in [7], when a reliability model contains deterministic or quasi-deterministic processes, the reliability and performance indices are highly sensitive to the shape of the probability distributions. This means that the use of a non-Markovian approach turns out to be mandatory in the assessment of production systems because, as it will be shown, repair, reconfiguration and propagation processes often present a quasi-deterministic behaviour.

In much of the literature that considers non-exponential time distributions, only very specific classes of problems are addressed, as it is the case in [8] where a control policy is discussed for a two-product and one-machine manufacturing system. Finally, most of the existing tools are oriented towards the evaluation of internal reliability and performance indices, such as availability and productivity. However, the important point for system planners is the global performance of the system from a business perspective, that is, the reliability of deliveries and the production costs. In [9] and [10], two cost models are proposed but, they are once again oriented towards specific classes of problems: the planning of regular preventive maintenance and the assessment of alternative delivery strategies.

1.1. The proposed approach

In order to overcome these shortcomings, this paper proposes a new approach. A hierarchical two-level modelling framework was developed to cope with the structural complexity of large manufacturing systems, At *local level*, models represent the internal behaviour of the cells, whereas at the *global level*, models represent the overall structure of the system and the flow of materials. Local level models are state diagrams describing the possible states of the cells in terms of their ability to meet production schedules (normal, halted, etc…), as well as the processes that govern the transitions between those states (failure, repair, reconfiguration, etc…).

Manufacturing cells may have different configurations and comprise heterogeneous equipment, but from the point of view of their consumers (or downstream subsystems), their behaviour may be described in terms of a standard two-states model, denoted as the *canonical model*. In the paper, two alternative approaches will be investigated to obtain the cells' canonical models, one based on the derivation of analytical expression, and the other based on simulation. In the second stage, these models are combined in accordance with the flow of materials represented in the global modelling level, in order to obtain an analytical model at the output of the manufacturing system. Next, the above mentioned reliability cost components are evaluated: buffer and extra work costs are assessed for each cell from the corresponding canonical model, while delivery penalties are assessed from the global canonical model.

Another distinctive feature of the proposed approach is the ability to assess stochastic models containing concurrent processes with generalized distributions. As it will be seen, during failure propagation, hyper-exponential repair and buffer processes are simultaneously active and remain active for several consecutive states without being reinitialized when a new state is entered, which is a behaviour pattern that corresponds to the pre-emptive resume age policy described in [11].

Despite the significant progress achieved in the last two decades, mostly based on stochastic Petri nets such as reported in [12] and [13], the assessment of stochastic models containing multiple generalized processes remains a largely open issue in reliability analysis.

In the paper, it will be shown how the behavioural and structural characteristics of JIT production systems can be explored in order to implement effective evaluation algorithms that fit the requirements of this class of systems. These algorithms may be seen a straightforward alternative to other well-established solutions to the analysis of non-Markovian systems, as those presented in [14], [15] and [16], or to the techniques based in Monte Carlo simulation as the one reported in [17].

1.2. Organization of the paper

The paper is organized as follows: after the introduction, Section 2 presents the manufacturing system of an industrial company in the automotive sector, the AutoParts Company, which is composed of three heterogeneous manufacturing cells and an assembly line. This section will cover the company's business context, the manufacturing system organization and the internal behaviour of each cell. The alternative solutions for the system implementation will also be introduced here. The next two sections present the main elements of the methodology. The two level hierarchical modelling framework will be introduced in Section 3, followed by an investigation of the algorithmic tools for the evaluation of reliability costs based on the canonical model concept in Section 4. Section 5 presents the practical application of the methodology, namely the canonical models of the AutoParts system and the design of the system obtained from these models, using total reliability costs as the optimization criteria. The final section presents some concluding remarks and perspectives for further work, including the extension of the methodology to other engineering domains. Annex 1 introduces complementary algorithms related to more complex behaviour patterns.

A number of assumptions were adopted in the case study for the sake of simplicity. For example, identical failure and repair processes were assigned to every machine in the manufacturing cells. Despite this, the case study is representative of a broad range of systems.

2. The JIT manufacturing system

This subsection presents the manufacturing system of the AutoParts Company, a typical parts supplier for the automotive industry, which performs three main technological processes: metalworking, metal forming and assembly. AutoParts has to fulfil a fairly strict delivery plan. Every 4 hours a truck should leave the plant to go to the client facility, and a relatively short time frame (1 hour) is assigned for its loading at the dock station. When AutoParts is not able to complete the loading within the assigned time frame, it incurs a penalty proportional to the additional time spent at the plant. When the loading delay exceeds 5 hours, the operation in the destination plant is disturbed and AutoParts suffers a far more severe penalty. Table 1 shows the main service specifications agreed with the client, and the penalties applied to AutoParts when a delivery failure occurs. The penalties, as well as all other cost-related data presented in this paper, are expressed as a standard unit of cost, denoted by *uc* (typically, *uc* will range from 2000 to 10000 €. The subassemblies produced by the AutoParts manufacturing system are made up of three main components. After a preliminary analysis, process engineers agreed that the production system should be structured as sketched in Figure 1: the components are produced in three manufacturing cells ($cell_1$, $cell_2$ and $cell_3$ in Figure 1), and the final product is prepared on the assembly line.

Table 1 – Service penalties

Service specification	Penalty
loading time frame: 1 hr	per hour of delay: 3 uc h^{-1}
maximum delay: 5 hr	per occurrence of the delay: 30 uc

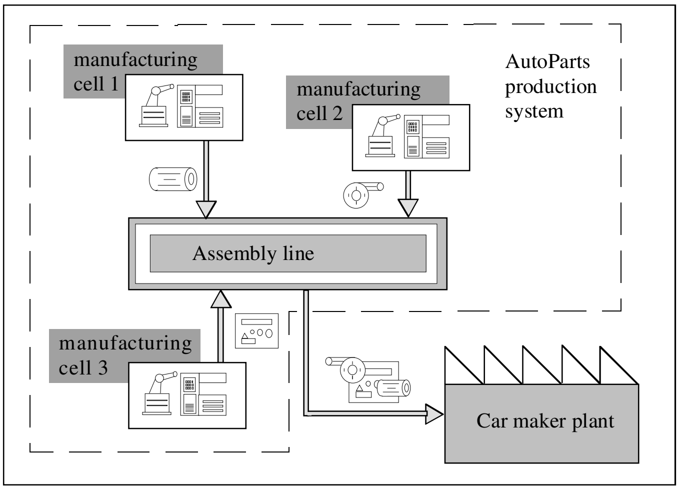


Figure 1 - The manufacturing system

The next subsection introduces the solutions for the AutoParts production system that will be analysed in this paper. The subsequent subsections will present the data required for their reliability analysis, which are the global organization and flow of materials within the manufacturing system, the internal behaviour of the manufacturing cells and the cost drivers for both non-reliability and improvement costs. In order to avoid data overload that could obscure the main ideas to be presented, it is assumed that the three manufacturing cells are identical. Even so, the case study deals with a rich set of structural and behavioural patterns, thus making it representative of a large number of practical systems. At this stage, it should be noted that AutoParts does not correspond to one existing company in particular. Instead, it is a synthesized company, whose organization and behaviour have been specified primarily to represent what is typical in the automotive industry[1].

2.1. The design problem

Despite intensive efforts by AutoParts to reduce its inventory levels, the existence of work-in-process (wip) buffers is still recognized as an indispensable element for the smooth and effective operation of the production system. Buffers filter the imbalance of manufacturing cells operating at different production rates. They also prevent the propagation of disturbances such as equipment failures and non-conforming lots, to the downstream units, thus improving the global throughput of the manufacturing system and the reliability of deliveries to the client. However, as buffers may represent significant additional costs, their design should be based on an economic analysis, balancing implementation costs (e.g. occupied area on the shop floor and inventory costs) against the productivity improvement they give [18]. The two solutions represented in Figure 2 may be analysed in the light of this consideration. In solution S_1, there is a single (and expensive) buffer at the output of the assembly line, whereas in solution S_2, there is a wip buffer at the output of each manufacturing cell, and the assembly line has a redundant implementation. The two solutions will be analysed and compared in Section 5, to determine the system design that minimizes global reliability costs.

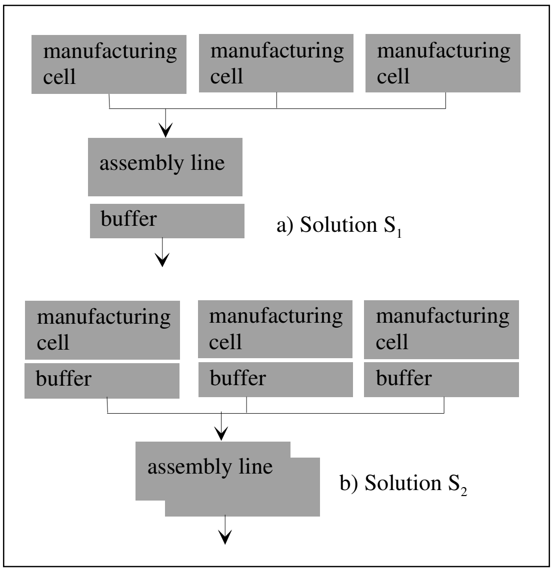


Figure 2 – The two solutions under study

Two important issues in this analysis are the propagation delays and the cost drivers associated with the wip buffers. The delays depend on the way the content of the buffers is managed: if the content remains almost constant (which is typically the case when there is a unitary flow of parts between the manufacturing cells), the propagation delay density function, $\gamma(t)$, will be close to the Dirac function: $\gamma(t) = \delta(t - \Delta)$, as shown in Figure 3.a. If the content varies according to the instantaneous production imbalance between input and output cells (which will typically be the case for batch operation) the density function will be close to the step function: $\gamma(t) = [\, h(t) - h(t- \Delta)\,] / \Delta$ (Figure 3.b). These are

[1] Many elements in the AutoParts manufacturing system correspond to ones that we have encountered among metalworking and plastic parts suppliers on several occasions, in both France and Portugal.

two typical but extreme situations. For intermediate situations, the density function of the propagation delay may be described by an n-order Erlang function (3.c):

$$\gamma(t) = \frac{(n/\Delta)^n \, t^{(n-1)} \, e^{-nt/\Delta}}{(n-1)!} \qquad (1)$$

As far as implementation costs are concerned, two cost components are to be considered for each buffer: an installation cost related to its capacity, and an inventory cost proportional to its average content:

$$C_b = \alpha_b^{Cp} B_b + \alpha_b^{I} I_b \qquad (2)$$

where B_b is the nominal capacity of b, I_b is its average content and α_b^{Cp} and α_b^{I} are the two cost drivers. The next two subsections will present the cost drivers for the buffers as well as the qualitative and quantitative data relating to the manufacturing cells and the assembly line, all of which are required for the analysis of the two solutions under study.

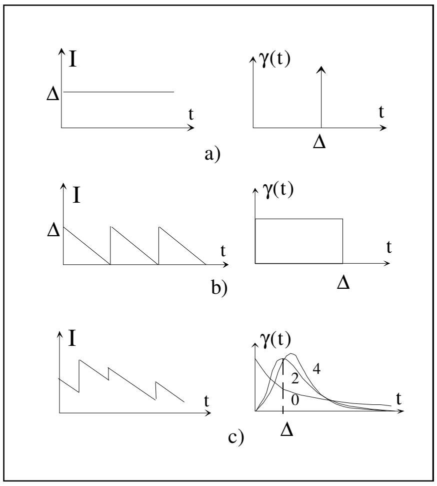
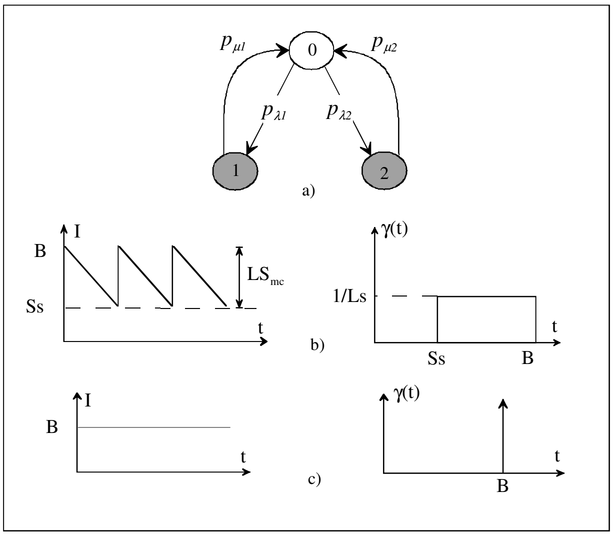


Figure 3- Types of probability density functions

2.2. Manufacturing cells

The internal model of the manufacturing cells represented in Figure 4.a includes two pairs of failure/repair processes, one corresponding to in-house repairs ($p_{\lambda 1}$, $p_{\mu 1}$) and the other corresponding to repairs requiring external resources ($p_{\lambda 2}$, $p_{\mu 2}$). Indeed, following a failure modes and effects analysis [19], manufacturing engineers concluded that equipment failures could be grouped into two main types: those solved by the internal maintenance service using in-house resources, and those requiring the spare parts to be ordered from an outside supplier. In the first case, time-to-repair will span from very short periods, when the machine operator is able to perform the repair by himself, to relatively long periods, when the intervention of a skilled technician is required. To model the execution time of these processes ($p_{\mu 1}$), the exponential distribution will be used. For external repair processes, AutoParts has settled maintenance contracts with external suppliers that guarantee a fixed lead-time (typically 10 hours). Given that time-to-repair is almost constant in this case, a 3rd order Erlang distribution will be used to model $p_{\mu 2}$. (It should be noted that other distributions could also have been chosen, as the evaluation algorithm is able to deal with any distribution.)

When a cell halts its operation due to a failure, extra working time will be needed in order to stay within the production plan. This time should be taken into account in the evaluation of the reliability costs, since different cost rates apply to normal and extra work. For the manufacturing cells, AutoParts' industrial engineering services have agreed to extra time overcharges at 0.3 uc h^{-1}.



a) Internal b) Batch c) Unitary

Figure 4 – Manufacturing cells behaviour

The propagation delays and the cost drivers for the buffers located at the output of the manufacturing cells depend on the operation mode of the cells. Therefore, two sub-solutions will be considered, S_{2a} and S_{2b}, corresponding respectively to unitary and batch operations at the manufacturing cells. Figures 4.b and 4.c show the evolution of the buffers for the two situations, where *Ss* is the safety stock, *Ls* is the lot size and *B* is the capacity of the buffer ($B = Ss + Ls$). For batch operation, the density function of the buffer delay is:

$$\gamma(t) = [h(t - S_s) - h(t - S_s - L_s)] / L_s \qquad (3)$$

In S_{2a}, the content of the buffer will be nearly constant (Figure 5.c). Therefore, the density function of the buffer propagation process will be close to the Dirac function:

$$\gamma(t) = \delta(t - S_s) \qquad (4)$$

Finally, the implementation costs of these buffers are:

$$B_w = \alpha_b^{Cp} S_s, \text{ for solution } S_{2a} \qquad (5)$$

$$B_w = \alpha_b^{Cp} (S_s + L_s) + \alpha_b^{I} (S_s + L_s/2) \text{ for solution S2b} \qquad (6)$$

where the following values were assigned to the cost drivers: $\alpha_b^{Cp} = 1\ uc\ h^{-1}$ $\alpha^{C} = 1\ uc\ h^{-1}$ and $\alpha_b^{I} = 0.5\ uc\ h^{-1}$.

2.3. Assembly line

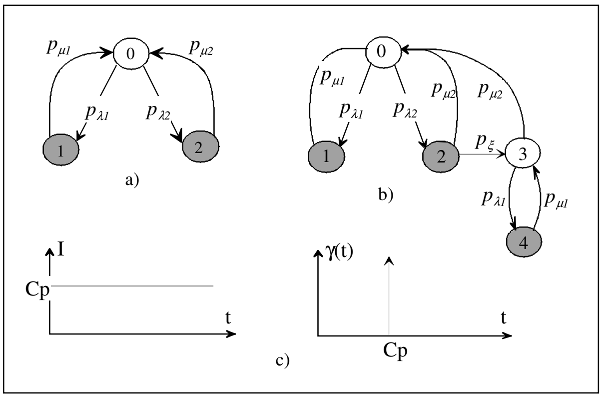
The implementation of the assembly line differs according to the solution under study, as represented in Figure 5. In the first solution, the line has a non-redundant implementation and an output buffer, whereas in solution S_2, there is no such buffer, but instead redundant equipment is added to improve the line inherent reliability. The additional implementation cost of this equipment is estimated at 30 uc. For the non-redundant solution, the internal model of the assembly line is similar to those of the manufacturing cells (Figure 5.a). The model corresponding to the redundant solution is shown in Figure 5.b. In this case, it is assumed that the reconfiguration process ξ that sets the redundant equipment into operation is triggered only for the longest failure/repair processes ($p_{\lambda 1}$, $p_{\mu 1}$). The assembly line has a unitary mode of operation, so the content of its output buffer will be nearly constant:

$$\gamma(t) = \delta(t - S_s) \qquad (7)$$

The cost drivers for this buffer are $\alpha^C = 5.2\, uc\, h^{-1}$ and $\alpha^I = 2.6\, uc\, h^{-1}$ so that:

$$B_w = \alpha_b^{Cp} S_s \qquad (8)$$

As may be expected, the cost drivers for this buffer are much higher than those of the manufacturing buffers, because the finished products also have a much higher added value than the parts produced at the manufacturing cells. AutoParts' engineering services assigned a value of 0.5 *uc* h^{-1} for the extra time surcharge of this unit.



a) Internal behaviour for S_1 b) Internal behaviour for S_2
c) Buffer behaviour

Figure 5 – Assembly line

3. Hierarchical modeling framework

The previous sections have highlighted that, in order to be useful for system planners, the reliability analysis of a manufacturing system should provide for the economic damages caused by failures, to make it possible to balance them against reliability improvement costs, that is, to balance *nR* against *iR*. The methodology to be presented thus provides a hierarchical modelling framework that enables the representation of the internal behaviour of the manufacturing cells and the flow of materials between the cells. It also provides a set of algorithmic tools that enable the evaluation of the indices driving the non-reliability costs, namely the probability and frequency of the failure states. The hierarchical modelling framework is introduced in this section, while the algorithmic tools for the evaluation of the reliability costs will be introduced in the next section.

3.1. Modelling levels

A production system may be seen as a network arrangement of two types of units – cells and buffers - interacting according to a producer/consumer scheme. The cells take their inputs from the upstream units, process them, and send them to the downstream units (note that the manufacturing cells and the assembly line of the AutoParts system are both *cells*). Each cell comprises a set of manufacturing equipment, whose behaviour is determined by processes such as failure, repair and reconfiguration. The output of a manufacturing cell may be linked directly to the input of one or more downstream cells. Alternatively, an intermediate wip buffer may exist between the producer and the consumer cells. (The solutions for the AutoParts system represented in Figure 2 provide examples of both possibilities.) In order to perform a reliability analysis, both the internal behaviour of each cell and the global structure of the production system must be known. To capture this data, a two-level modelling framework was adopted (Figure 6). At the *local level*, models represent the internal behaviour of the cells, whereas *global-level* models represent the overall structure of the production systems. For each modelling level, a conceptual model was defined describing the modelling entities and their properties and relations, as discussed below.

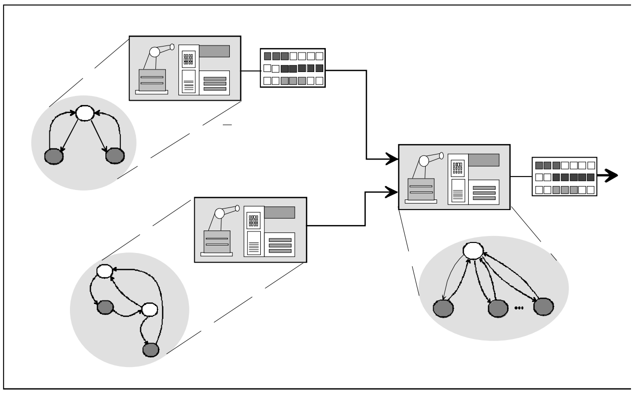


Figure 6 – The two modelling levels

3.2. Global level

A global-level model represents the structure of a production system through an oriented graph, where the *nodes* correspond to the manufacturing units (cells and buffers) and the *links* correspond to the flow of materials between the units. The conceptual model for the global level is represented in Figure 7.a, using UML notation [20]. According to this model, the following constraints apply to the structure of the production systems:

- a cell may be supplied by several input

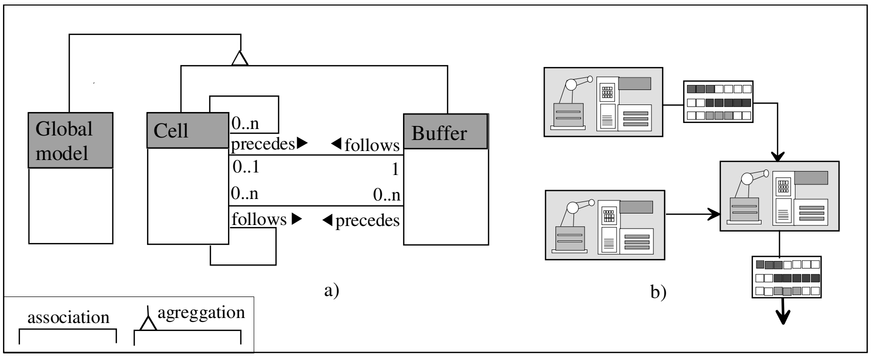


Figure 7 – Conceptual model for the global level

buffers and cells (cell *follows 0..n* buffers, and cell *follows 0..n* cells);

- a buffer is always supplied by a single cell (buffer *follows 1* cell);
- a cell may not supply more than one buffer (cell *precedes 0..1* buffer), but may directly supply more than one cell (cell *precedes 0..n* cells);
- a buffer may supply 0 or more cells (buffer *precedes 0..n* cells);

Two sets of *inherent* and *calculated* attributes are assigned to each class of the conceptual model. The values of the inherent attributes are assigned during the modelling process, whereas the calculated attributes are determined during the evaluation process. Table 2 shows the attributes for the two global-level classes, where the statements of the inherent attributes are preceded by an *. The use of these attributes will be considered in Section 4.2, together with the critical examination of the evaluation algorithms.

Table 2: Attributes of the global level classes

Cell
* precedes: Buffer
* follows[1..n]: Buffer, Cell
* M^i, M^o: $\{\Lambda,\rho(t)\}$
* α^d, α^f: real

Buffer
* precedes: Cell
* follows: Cell
* M^b: $\{\Lambda,\rho(t)\}$
* α^d, α^f: real
* $\gamma(t)$: function(t)

3.3. Local level

While the global level is oriented towards the system structure, the local level is oriented towards the internal behaviour of the manufacturing cells, which may be rather complex (see Annex 1 for an example). In the modelling framework, the internal behaviour of each cell is represented through a state diagram describing the possible states of the cell in terms of their ability to meet production schedules (normal, halted, etc…), along with the processes managing the transitions between states (failure, repair, reconfiguration, etc…). Figure 8.a shows the conceptual model for the local modelling level. According to this, a local model contains two types of entities: *states*, that represent the possible situations of the system being modelled, and *transitions*, that represent the possible transitions (*from* / *to*) between states. A process is a physical mechanism that causes a transition to occur. The same process may be active in several states and assigned to more than one transition. Consider the model in Figure 8.b as an example, in which:

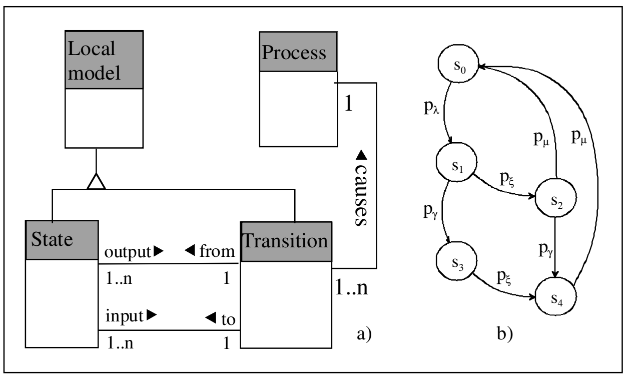


Figure 8 – Local-level conceptual model

- process p_ξ may cause the transitions $s_1 \rightarrow s_2$, and $s_3 \rightarrow s_4$;
- s_1 has two output transitions to s_2 and s_3, which are caused by processes p_ξ and p_γ, respectively;

- s_4 has two input transitions from s_2 and s_3, which are caused by processes p_γ and p_ξ, respectively.

As shown in Table 3, several inherent and calculated attributes are also assigned to the classes of this conceptual model.

Table 3: Attributes of the local level classes

State
* inputs[1..n]: Transition
* outputs[1..n]: Transition
* α^d, α^f: real
active: boolean
total_time: real
nº_occurrences: integer
probability: real
rate: real

Transition
* to, from: State
* process: Process
exec_time: real

Process
* f(t): function(t)
exec_time: real
active: boolean

4. Evaluation tools

The second main element of the methodology is the reliability evaluation tool, which enables the non-reliability cost components D_F and T_x to be obtained from the local and the global models. When an equipment failure occurs, a shortage of materials arises at the output of its manufacturing cell. As the shortages may propagate to the downstream cells (Figure 9), they are classified as *endogenous* if they were caused by an internal equipment failure, or as *exogenous* when caused by equipment belonging to an upstream cell. The occurrence of a materials shortage causes an economic loss.

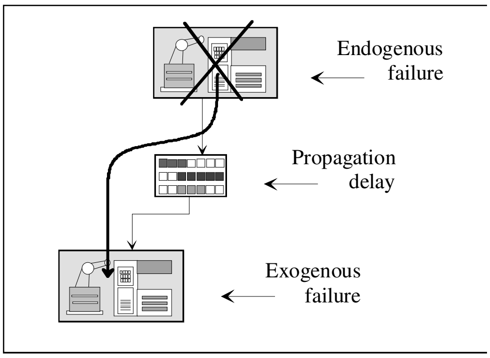


Figure 9 – Propagation of the failures

A *loss driver* is a ratio between the occurrence of economic damage and a reliability index. Two loss drivers will be considered for each manufacturing unit, one associated with the duration of the shortages (α^d cost driver) and the other associated with the frequency of the shortages (α^f cost driver). The non-reliability cost component coming from the extra working time, T_x, will typically be an α^d cost, whereas the penalties due to the failures on the deliveries to the client, D_F, will typically be an α^f cost.

The evaluation algorithm is based on the *canonical model* concept, an equivalent representation of a manufacturing cell or set of cells from the point of view of the downstream

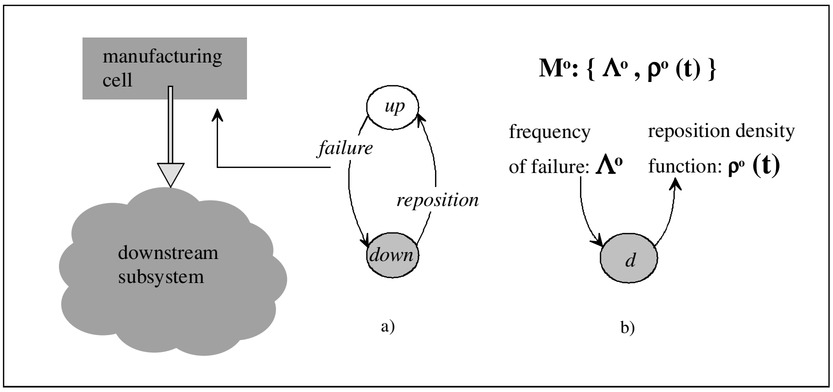


Figure 10 –Canonical model of a cell

subsystem. This concept is illustrated by the manufacturing cell in Figure 10. The behaviour at its output can be described in terms of two states - *up* and *down* - corresponding to the situations where (i) the cell is producing according to its schedule and (ii) the cell is halted and the normal flow of materials at its output has been interrupted (Figure 10.a). If the frequency of arrival to the down state, Λ^o, and the distribution of the reposition process, $\rho^o(t)$, are known, then the non-reliability costs at the output of the cell can readily be evaluated from:

$$nR = \alpha^d \, \overline{A}^o + \alpha^f \, \Lambda^o \qquad (9)$$

where $\overline{A}^o$ is the probability of the down state, given by:

$$\overline{A}^o = \Lambda^o \int_0^\infty t \, \rho^o(t) \, dt \qquad (10)$$

The couplet $\{\Lambda^o, \rho^o(t)\}$ will hereafter be designated as the *output canonical model* of cell (10.b):

$$M^o = \{\Lambda^o, \rho^o(t)\} \qquad (11)$$

Once the canonical model of a manufacturing unit is known, therefore, the corresponding non-reliability costs can readily be calculated.

Similar models can also be used to describe the internal behaviour of a cell, and the behaviour at the output of a buffer. Canonical models may actually be used in three different situations: (i) modelling of the internal behaviour of a cell, (ii) modelling of the behaviour at the output of a cell and (iii) modelling of the behaviour at the output of a buffer. In the first case, the canonical model will show the frequency of failure and the reposition process at the output of a cell, when only the endogenous failure processes of that cell are being considered. In the second situation, the down state of the model represents the situations where the cell halts its operation, due to an endogenous or to an exogenous failure in an upstream cell. In the latter case, the failure state will correspond to the situations where the buffer is empty and unable to supply the downstream cells. To distinguish these three models for a particular *cell c*, they will be designated as M^i, M^o and M^b, respectively (Figure 11). The next subsection introduces the procedures for the determination of the canonical models for a production system.

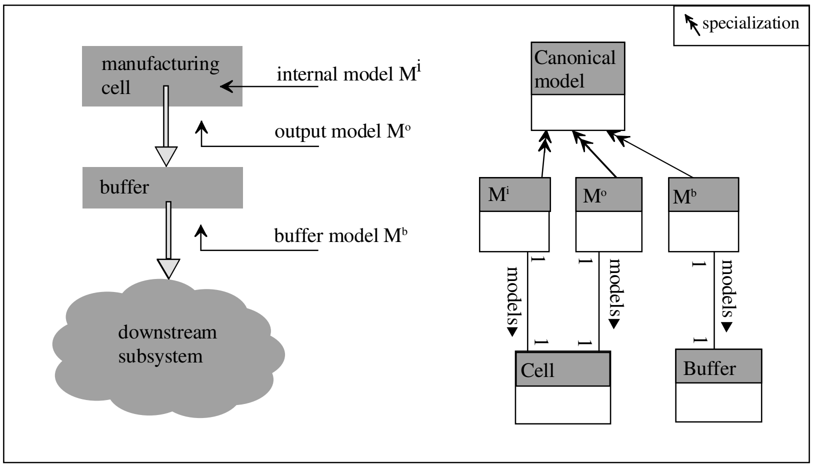


Figure 11 – Internal and external canonical models

4.1. Determination of the canonical model

Two main approaches may be followed in the determination of the canonical models: an *analytical approach* that provides expressions for Λ and $f_\rho(t)$, and a *simulation-based approach*, that provides a numerical estimation for Λ and a histogram for $f_\rho(t)$. Both approaches will be presented below, with a set of practical examples covering the situations present in the AutoParts manufacturing system. The evaluation algorithm will then be introduced.

4.1.1. Analytical approach

Consider the internal canonical model of a cell *c* composed by two non-redundant machines, whose behaviour is represented in Figure 12.a. The probability of the normal operation state, P_{up}, is given by:

$$P_{up} = \frac{1}{1 + r_{\lambda 1} m_{\mu 1} + r_{\lambda 2} m_{\mu 2}} \qquad (12)$$

where r_p and m_p denote, respectively, the rate of occurrence and the mean "time-to-occur" of process *p*. The first parameter of the internal canonical model, i.e., the frequency of failure, is given by:

$$\Lambda^i = \Lambda_1 + \Lambda_2 \qquad (13)$$

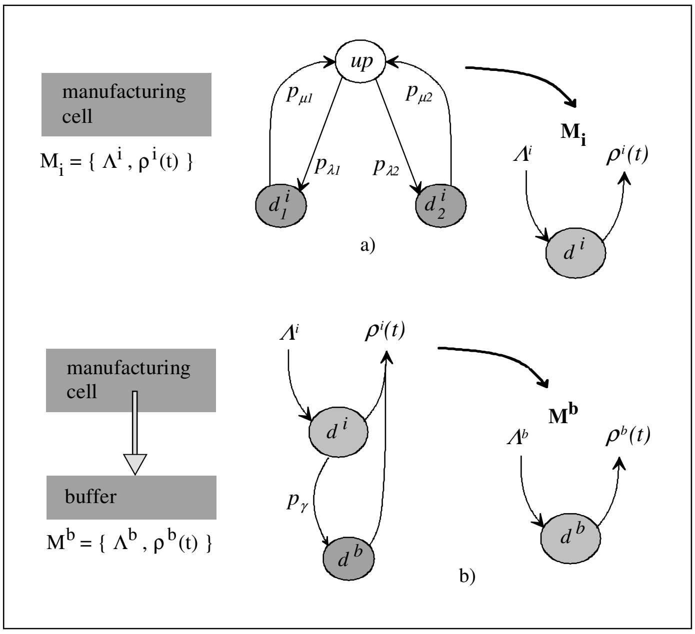


Figure 12 –Canonical model of a cell with 2 non-redundant machines

where Λ_n denotes the frequency of arrival to the down state d_n^i due to process $p_{\lambda n}$ ($\Lambda_n = r_{\lambda n} P_{up}$). The second parameter of the model, i.e. the probability density function of the reposition process, is given by:

$$\rho^i(t) = \frac{\Lambda_1}{\Lambda^i} f_{\mu 1}(t) + \frac{\Lambda_2}{\Lambda^i} f_{\mu 2}(t) \qquad (14)$$

Now, suppose that an output buffer is added to the cell (Figure 12.b) and p_γ denotes the corresponding propagation process. The canonical model at the output of the buffer, M^b, is obtained as follows: the failure rate Λ^b comes from the product of the frequency of arrival to state d^i, and the probability of transition $d^i \rightarrow d^b$:

$$\Lambda^b = \Lambda^i \int_0^\infty \gamma(t_1) \int_{t_1}^\infty \rho^i(t_2)\, dt_2\, dt_1 \qquad (15)$$

Function $\rho^i(t)$ comes from the ratio between the density function of residence time in state d_b, given that the system has arrived to that state, that is:

$$\int_0^\infty \gamma(t_1)\,\rho^i(t+t_1)\,dt_1 \qquad (16)$$

and the probability of transition $d^i \rightarrow d^b$. Thus:

$$\rho^b(t) = \frac{\int_0^\infty \gamma(t_1)\,\rho^i(t+t_1)\,dt_1}{\int_0^\infty \gamma(t_1)\int_{t_1}^\infty \rho^i(t_2)\,dt_2\,dt_1} \quad \text{or} \quad \rho^b(t) = \frac{\int_0^\infty \gamma(t_1)\,\rho^i(t+t_1)\,dt_1}{\Lambda^b / \Lambda^i} \qquad (17)$$

A systematic method is presented in [21], making it possible to obtain analytical expressions for Λ and $\rho(t)$. For the general case of a cell made by n non-redundant machines, the equivalent internal canonical model is given by the following expressions, where λ_j and μ_j are the failure and repair processes of machine j:

$$P_{up} = \frac{1}{1+\sum_{j=1}^{n} r_{\lambda j} m_{\mu j}} \qquad (18)$$

$$\Lambda^i = \sum_{j=1}^{n} r_{\lambda j}\, P_{up} \qquad (19)$$

$$\rho^i(t) = \sum_{j=1}^{n} \frac{r_{\lambda j}\, P_{up}}{\Lambda^i} \mu_j(t) \qquad (20)$$

If a buffer is then added to this cell, expressions (15) and (17) may be employed again to obtain the canonical model at its output, M^b. These results will be employed in the numerical analysis of the AutoParts system presented in Section 5: the canonical models for the manufacturing cells and for the assembly line will be determined using the above procedure.

4.1.2. Multiple cells

An important consideration is the fact that the canonical model at the output of a cell, M^o, can be obtained by the combination of the internal canonical model of the cell, M^i, and the canonical models of the upstream buffers M^b. Moreover, the canonical model equivalent to a set S of manufacturing cells can be obtained by successively combining the internal models of the cells of S. To introduce the corresponding procedure, consider $cell_3$ in Figure 13, and suppose that:

- its internal model, M_3^i, and the models at its inputs, M_1^b and M_2^b, have already been determined;

- the canonical model at its output, M_3^o, is to be determined.

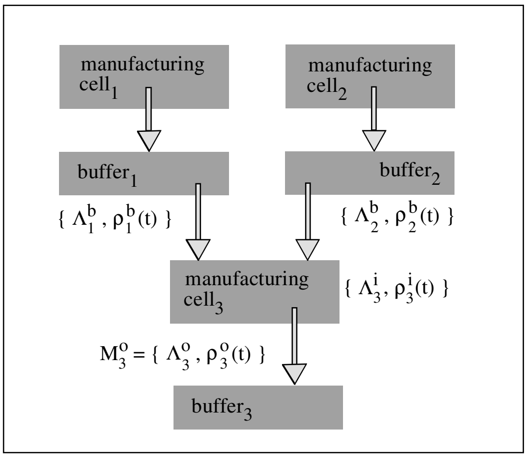


Figure 13 – Canonical model equivalent to a set of cells

The operations in *cell$_3$* may stop due to an internal failure, or due to a shortage of materials at its inputs, i.e., an exogenous failure in *cell$_1$* or in *cell$_2$*. In a typical situation, the failure rate at the output of *cell$_3$* will be close to the sum of the endogenous and exogenous failure rates:

$$\Lambda_3^o = \Lambda_1^b + \Lambda_2^b + \sigma \, \Lambda_3^i \qquad (21)$$

In fact, in a JIT manufacturing system, the global availability of the system is normally well above 90%. As the individual availability of each manufacturing cell is significantly higher, the probability of simultaneous failures is small enough to be overlooked (as an example, the manufacturing cells of the AutoParts system present an endogenous availability of about 95%, therefore, the probability of simultaneous failures is below 1%). Finally, note that in (21) the internal failure rate of *cell$_3$* is affected by a factor σ given by:

$$\sigma = \frac{P_{up3} - \Lambda_1^b m_{\rho_1^b} - \Lambda_2^b m_{\rho_2^b}}{P_{up3}} \qquad (22)$$

The reason is that Λ_3^i was determined considering only the internal failure processes, whereas, in the global model, *cell$_3$* is also submitted to the exogenous processes. Finally, as far as the distribution of the reposition process at the output of *cell$_3$* is concerned, it will come from the weighted average of the three reposition processes involved:

$$\rho_3^o(t) = \frac{\Lambda_3^i}{\Lambda_3^o} \rho_3^i(t) + \frac{\Lambda_1^b}{\Lambda_3^o} \rho_1^b(t) + \frac{\Lambda_2^b}{\Lambda_3^o} \rho_2^b(t) \qquad (23)$$

The canonical model equivalent to any subset of a manufacturing system can be obtained by successively incorporating new cells in a global model, starting from the upstream cells. This same approach is implemented in the algorithm presented in subsection 4.2.

4.1.3. Simulation approach

The analytical approach presented so far is not effective for models presenting complex behaviour patterns, nor for large models. As an example, consider the system introduced in Annex 1. If the processes p_μ and p_η do not present exponential distributions, the analytical approach leads to rather complex expressions. In such situations, an alternative approach based on Monte Carlo simulation proves to be more effective. It makes it possible to obtain the *rate of arrival* and the *histogram for the time of residence* in the failure states. Using statistical techniques, it is then possible to obtain the parameters of a distribution (typically, a Weibull or the Erlang distribution) that closely fits the empirical distribution of the reposition process $\rho(t)$. This function can then be used to model the behaviour of the cell in the context of a larger model.

The simulation algorithm[2] is sketched in Table 4. The histograms obtained with this algorithm for the system described in Annex 1 are shown in Figure 14.a. This shows the histogram for the duration of failure at the output of the cell, while 14.b shows the histogram at the output of the buffer. The first histogram presents a very high frequency at 0.5 hours, due to the reconfiguration process. Figure 14.c shows the same histogram without this process.

Figure 14.d shows the Weibull distribution that best fits the density function of reposition process at the output of the

Table 4 – Simulation algorithm

```
//Global variables declaration
current_state: State
sim_time: Real
next_transition: Transition
// 1. Initialize
current_state = s0
sim_time  =  0
// 2. Simulation cycle
while (sim_time < sim_horizon) {
   // 2.1. Determine next transition to occur
    next_transition.exec_time =  ∞
    for each t in current_state.outputs {
       if t.process.active = false then {
       t.process.exec_time = sim_time +
                     + random_generator(t.process.fdp(t))
       t.exec_time = t.process.exec_time }
       if t.exec_time < next_transition.time then
          next_transition = t }

    // 2.2. Execute transition
    // 2.2.1. Update histograms
    current_state.total_time +=
                     next_transition.exec_time - sim_time
    current_state.nº_occurrences +=  1
    // 2.2.2. Prepare next simulation cycle
    current_state = next_transition.to
    sim_time = next_transition.exec_time
end while  //end of simulation cycle
// 3. Calculate indices
for each s in State
    s.Probability = s.total_time / sim_horizon
    s.Rate = s.nº_occurrences / sim_horizon
}
```

[2] Note that this algorithm makes use of the attributes of the local-level conceptual model introduced in 3.3.

buffer, obtained using a general purpose mathematical tool [22]:

$$\rho^b(t)=\beta\,\alpha^{-\beta}\,t^{\,(\beta-1)}\,e^{-(\frac{t}{\alpha})^\beta}$$

with α = 14.931 and β = 1.0869. For the reposition process at the output of the cell, the density function comes from the combination of a Dirac pulse and a Weibull distribution. Once the number of reconfiguration failures represents 40% of the total number of failures, this can be shown as:

$$\rho^b(t)=0.4\,\delta(t\text{-}\Delta)+0.6\,\beta\,\alpha^{-\beta}\,t^{\,(\beta\text{-}1)}\,e^{-(\frac{t}{\alpha})^\beta}$$

with Δ = 0.5, α = 14.36 β = 1.2667. As before, the practical application of the concepts and results presented here will be considered using numerical analysis in Section 5.

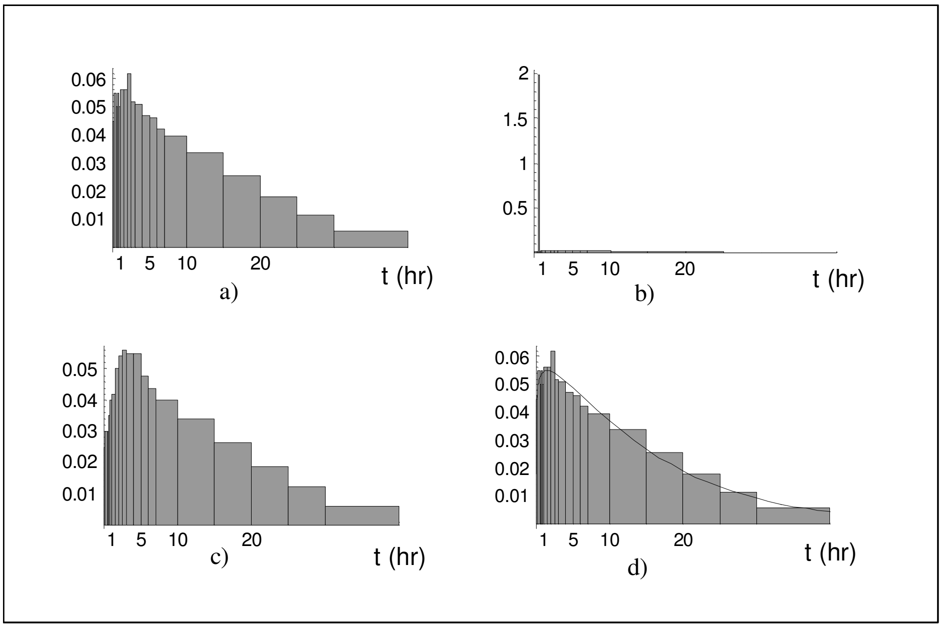


Figure 14 – Histograms of the duration of the failures

The simulation approach is insensitive to the dimension of the model, and to the shape of the density functions of the underlying behaviour processes. On the other hand, it will normally demand more processing power than the analytical approach, especially when the reliability analysis involves sensitivity analysis, because a full evaluation of the system has to be performed for each set of values of the parameters. In the analytical approach, once the expressions for the cost model are obtained, they just have to be re-evaluated for each set of parameters.

4.2. Evaluation algorithm

The evaluation algorithm that makes it possible to obtain the non-reliability costs is sketched in Table 5. It involves three main steps:

1. For each node of the global model → obtain the internal canonical model.
2. For each cell of the global model, starting from the upstream cells and proceeding according to the flow of materials → obtain the output and the buffer canonical models.
3. For each cell where relevant loss occurs → evaluate the non-reliability costs.

Table 5 –Evaluation algorithm

```
//Global declarations
type Canonical_model {
        Λ: real
        ρ(t): function(t)
        }
        M_u: Canonical_model
        nR: Real

function evaluate_nR(model): real {
// This function determines the non-reliability cost for
// all the manufacturing units in model

        nR = 0
   //step 1: Obtain the internal model of each manuf. cell
  for each c in model
                        c.M^i = obtainM^i(c)

  // for each manufacturing unit (cell or buffer) in model,
  // determine the non-reliability loss
  for each u in model such that  u.α^d ≠ 0 or u.α^f ≠ 0 {
      // step 2: Obtain the model at the output of u
      M^o_u = M^i_u
      if u is a buffer then
                        M^b_u = obtainM^b(u)
      if u is a cell then
                        M^o_u = obtainM^o(u)

      // step 3: Evaluate production losses at the output of u
      u.nR = M^o_u.Λ [ u.α^d_n ∫_0^∞ t M^o_u.ρ(t) dt + u.α^f_n ]
      nR += u.nR
      }
  return(nR)
}
```

The first step of the algorithm was presented in the previous subsection, while the two others will be discussed now. The second step is based on a recursive procedure whose core is implemented by the functions *obtainM*^{o} and *obtainM*^{b}. Table 6 sketches the algorithms of these two functions. To illustrate the recursive procedure, consider again the system sketched in Figure 13.

To evaluate the system non-reliability costs, the canonical model at the output of *buffer*$_3$, M_{b3}, must be known. So, function *obtainM*^{b} is invoked with input argument *buffer*$_3$. As the determination of M_3^b requires M_3^o to be known, function *obtainM*^{o} will be invoked for *cell*$_3$. The model M_3^o is obtained from the combination of M_3^i, M_1^b and M_2^b (as was seen in 4.1.2). Function *obtainM*^{b} will then be invoked twice, in order to determine the models at the outputs of *buffer*$_1$ and *buffer*$_2$. The recursive invocation of *obtainM*^{b} and *obtainM*^{o} will continue until the upstream cells are reached. For this example M_1^o and M_2^o will be determined first, then M_1^b, M_2^b and M_3^o, and finally M_3^b. The third step of the evaluation algorithm consists of the assessment of the non-reliability costs from the relevant canonical models.

Table 6 – Obtaining the canonical model of a subsystem

function obtainM^b(b): Canonical model {
// obtains the canonical model of buffer b
// $\gamma(t)$ is the density function for the delay of b

// Declarations
M^o, M^b: {$\Lambda, \rho(t)$}

// **1.** Obtain M^o for the preceding cell
M^o = obtainM^o(b.input)

// **2.** Obtain the model at the output of b

$$M^b.\Lambda = M^o.\Lambda \int_0^\infty b.\gamma(t) \int_t^\infty M^o.\rho(\tau)\, d\tau\, dt$$

$$M^b.f\,(t) = \frac{\int_0^\infty b.\gamma(t_1)\, M^o.\rho(t+t_1)\, dt_1}{\int_0^\infty b.\gamma(t_1) \int_{t_1}^\infty M^o.\rho(t_2)\, dt_2\, dt_1}$$

return(M^b)
}

function obtainM^o(c): Canonical model {
// obtains the output canonical model of cell c

// Declarations
M^b: array of Canonical model

// 1. Obtain M^b for each b in c.inputs
M^b[b] = obtainM^b(b)

// 2. Determine M^o for cell c
// 2.1. Endogenous failures
Λ_Σ = c.M^i.Λ
$\rho(t)$ = c.M^i.ρ

// 2.2. Exogenous failures
for each buffer b in c.inputs
Λ_Σ += M^b[b].Λ
$\rho(t)$ += M^b[b]. $\rho(t)$ x M^b[b].Λ

$M^o.\Lambda = \Lambda_\Sigma$
$M^o.\rho(t) = \rho(t) / \Lambda_\Sigma$

return(M^o)
}

The main cost components will typically come from shortages of materials that directly impact on deliveries to the clients, i.e., the shortages at the production system output. However, shortages at internal cells may also cause significant losses when extra work becomes necessary to fit in with production schedules. For the general case of a manufacturing system *S* having n cells, therefore, the non-reliability cost will be obtained from:

$$nR_S = H_y \sum_n \Lambda_n^o \left[\alpha_n^d \int_0^\infty t\, \rho_n^o(t)\, dt + \alpha_n^f \right] \qquad (24)$$

where H_y is the number of working hours per year (a typical value is 5,000 hours); and α_n^d and α_n^f are the cost drivers for cell *n*.

5. Analysis and Evaluation

Now that the underlying concepts and algorithms of the methodology have been presented, this section will discuss their practical application to the reliability analysis of the AutoParts manufacturing system. According to the rationale presented in the first part of the paper, the aim of the analysis is the minimization of the manufacturing system global reliability costs as they are defined in (1).

The analysis is organized as follows. The first step involves identifying the relevant cost components and investigating their relationship to the canonical models of the manufacturing system. Next, the relevant canonical models will be determined using the analytical approach presented in 4.1.1. The third step corresponds to the numerical evaluation of the reliability costs for the solutions being studied, and considers different wip and output buffers. The optimal solution for the manufacturing system will be based on the analysis of these results. The analysis also includes a comparison with the results that would be obtained using the conventional Markov approach. This will confirm the occurrence of significant errors in calculations using that approach that lead system planners to non-optimal solutions. The following table summarizes the data that is relevant to the calculations previously introduced in the paper.

Table 7 – Input data for the reliability analysis process

Processes	
In-house maintenance	
Failure	$r_{\lambda_1} e^{-r_{\lambda_1} t}$, $r_{\lambda_1} = 0.005\ h^{-1}$
Repair	$r_{\mu_1} e^{-r_{\mu_1} t}$ $\mu_1 e^{-\mu 1 t}$, $r_{\mu_1} = 0.5\ h^{-1}$
External maintenance	
Failure	$r_{\lambda_2} e^{-r_{\lambda_2} t}$, $r_{\lambda_2} = 0.001\ h^{-1}$
Repair	$\frac{1}{2}(1/m_{\mu_2})^3 t^2 e^{-t/m_{\mu_2}}$, $m_{\mu_2} = 0.3\ h^{-}$
Reconfiguration	
assembly line (S_2 only)	$\delta(t - \Delta_\xi)$, with $\Delta_\xi = 0.5$ h
Propagation	
assembly line (S_1 only)	$\delta(t - Ss_a)$
manufacturing cells (unitary)	$\delta(t - Ss_m)$
manufacturing cells (batch)	$[h(t-Ss_m) - h(t-Ss_m-Ls_m)] / Ls_m$

Cost drivers	
Delivery	
Loading delay (1hr)	$\alpha^d = 2\ h^{-1}$
Delivery delay (5 hr)	$\alpha^f = 20$
Extra work	
assembly line	$\alpha^x = 1\ h^{-1}$
manufacturing cells	$\alpha^x = 0.5\ h^{-1}$
Buffers	
assembly line	$\alpha^{Cp} = 5.2\ h^{-1}$, $\alpha^I = 2.6\ h^{-1}$
manufacturing cells	$\alpha^{Cp} = 1.3\ h^{-1}$, $\alpha^I = 0.7\ h^{-1}$
Redundancy	
assembly line	$R_d = 30$

5.1. Cost components analysis

In the first section of the paper, four cost components were identified. With regard to reliability improvement costs, B_w may be evaluated from the formula and the cost drives introduced in paragraphs 2.1 and 2.2 respectively. However, the cost of the redundant equipment, R_d, considered in solution S_2, is fixed and estimated at 30 *uc*. The two other cost components – extra working time and delivery failure penalties – depend on the manufacturing equipment failures. This subsection shows how these costs may be evaluated using the canonical models of the manufacturing system and the associated Formulae.

The cost of extra time, T_x, will be evaluated from the canonical models at the output of the manufacturing cells, and at the output of the assembly line, M_m^o and M_a^o, using expression (11). For the delivery failures, according to the service agreement with the client (Table 1), two types of penalties are to be considered: one that is proportional to the length of delay at the loading station (loading failure); and the other that is proportional to the number of delivery failures (delays longer than 5 hr). If M_a^o is known, the penalties can be evaluated from the models presented in Figure 15, where:

- $p_{\gamma a}$ is the propagation process of buffer at the output of the assembly line, with $\gamma_a(t)= \delta(t - B_a)$;
- p_l is the delay process corresponding to the loading-time frame of 60 min., with $l(t)= \delta(t - \Delta_l)$;
- d_l is the down state corresponding to a loading failure;
- p_d is the delay process corresponding to the maximum delay, with $d(t)= \delta(t \; \Delta_d)$;
- d_d is the down state corresponding to a delivery failure.

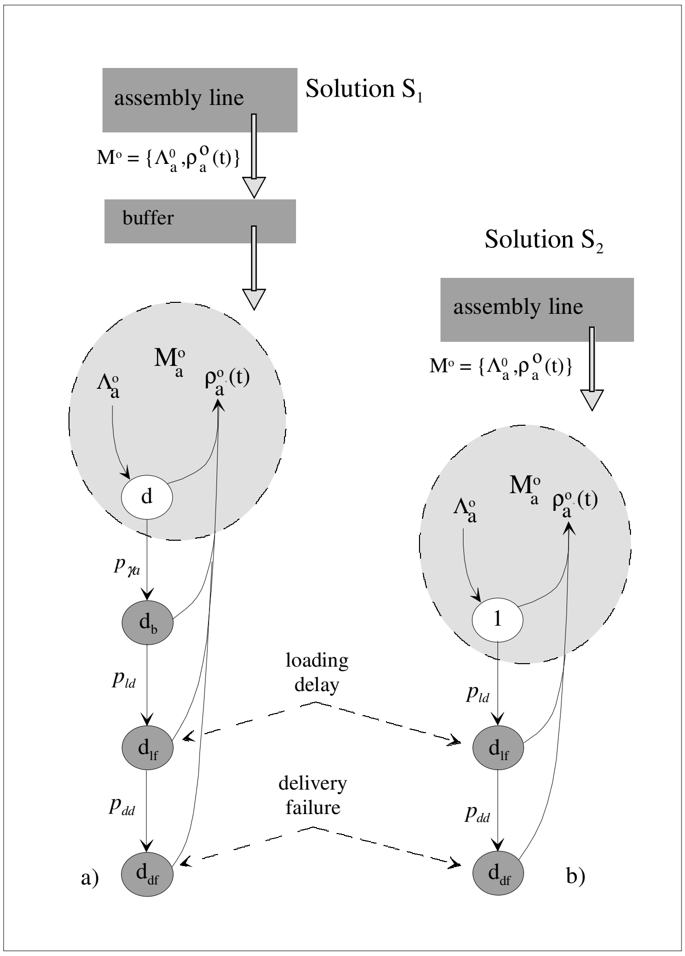


Figure 15 – Models for the evaluation of the delivery penalties

The penalties due to loading delays are proportional to the sum of the probabilities of states d_l and d_d, $P(d_l) + P(d_f)$, whereas the penalties due to delivery failures are proportional to the arrival frequency of state d_d, $\Lambda(d_d)$. Once the canonical model at the output of the assembly line is known, these reliability indices may be readily evaluated from:

$$P(d_l) + P(d_d) = \frac{\Lambda_a^o \int_{\Delta_l}^{\infty} (t - Cp_a + \Delta_l)\ \rho_a^o(t)\,dt}{1 + \Lambda_a^o \int_0^{\infty} t\ \rho_a^o(t)\,dt} \tag{25}$$

$$\Lambda(d_d) = \frac{\Lambda_a^o \int_{\Delta_l + \Delta_d}^{\infty} \rho_a^o(t)\,dt}{1 + \Lambda_a^o \int_0^{\infty} t\ \rho_a^o(t)\,dt} \tag{26}$$

5.2. Determination of the canonical models

To assess the non-reliability costs of the AutoParts manufacturing system, its canonical models must first be determined. Tables 8 and 9 show the relevant canonical models, which were obtained using the procedures presented in Section 4.1.1 (for M_m^i, M_m^o, M_m^b and M_a^i in solution S_1), in Annex 1 (for M_a^i in solution S_2) and in Section 4.1.2 (for M_a^o in solution S_2) .

Table 8 – The canonical models for the manufacturing cells

M_m^i	M_m^o
$\Lambda_m^i = (r_{\lambda_1} + r_{\lambda_2})\frac{1}{k}$, with $k = 1 + r_{\lambda_1} m_{\mu_1} + r_{\lambda_2} m_{\mu_2}$	$\Lambda_m^o = \Lambda_m^i$
$\rho_m^i(t) = \frac{r_{\lambda_1}}{\Lambda_m^i}\mu_1(t) + \frac{r_{\lambda_2}}{\Lambda_m^i}\mu_2(t)$	$\rho_m^o(t) = \rho_m^i(t)$
M_m^b, batch operation	**M_m^b, unitary operation**
$\Lambda_m^b = \Lambda_m^o \int_{Ss}^{Ss+Ls} 1/Ls \int_{t_1}^{\infty} \rho_m^o(t_2)\,dt_2\,dt_1$	$\Lambda_m^b = \Lambda_m^o \int_0^{\infty} ((1/m_{\mu_2})^3\, t^2\, e^{-t_1/m_{\mu_2}}/2) \int_{t_1}^{\infty} \rho_m^o(t_2)\,dt_2\,dt_1$
$\rho_m^b(t) = \frac{\Lambda_m^o}{\Lambda_m^b} \int_{Ss}^{Ss+Ls} 1/Ls\ \rho_m^o(t_1 + t)\,dt_1$	$\rho_m^b(t) = \frac{\Lambda_m^o}{\Lambda_m^b} \int_0^{\infty} ((1/m_{\mu_2})^3\, t^2\, e^{-t_1/m_{\mu_2}}/2)\ \rho_m^o(t_1 + t)\,dt_1$

Table 9 – The canonical models for the assembly line

M_a^i, **solution** S_1	M_a^i, **solution** S_2
$\Lambda_a^i = (r_{\lambda_1} + r_{\lambda_2})\frac{1}{k}$	$\Lambda_a^i = r_{\lambda_1} + r_{\lambda_2}\,(1+\int_0^\infty r_{\lambda_1}\, e^{-t/r_{\lambda_1}} \int_{t_1}^\infty \mu_2(t_2)\, dt_2\, dt_1)$
$\rho_a^i(t) = \frac{r_{\lambda_1}}{\Lambda_a^i} m_1(t) + \frac{r_{\lambda_1}}{\Lambda_a^i} m_2(t)$	$\rho_a^i(t) = \frac{r_{\lambda_1}}{\Lambda_a^i}\mu_1(t) + \frac{r_{\lambda_2}}{\Lambda_a^i}\xi(t) + \frac{r_{\lambda_2}}{\Lambda_a^i}\int_0^t r_{\lambda_1}\, e^{-t_1/r_{\lambda_1}} \mu_2(t-t_1)\,dt_1$
M_a^o, **solutions** S_1 **and** S_2	M_a^b, **solution** S_1
$\Lambda_a^o = 3\Lambda_m^b + \frac{k}{k+3(\Lambda_m^b\ m_{\rho_c^m})}\Lambda_a^i$	$\Lambda_a^b = \Lambda_a^o \int_{\Delta_b}^\infty \rho_a^o(t)\, dt$
$\rho_a^o(t) = 3\frac{\Lambda_m^b}{\Lambda_a^o}\rho_m^b(t) + \varphi\frac{\Lambda_a^i}{\Lambda_a^o} m_2(t)$	$\rho_a^b(t) = \frac{\Lambda_a^o}{\Lambda_a^b}\rho_a^o(t+\Delta_a)$

5.3. Numerical results

The following figures show a number of results that were obtained using the canonical models above, for the solutions of the AutoParts manufacturing system that are under consideration, including:

(i) a single buffer at the output of the assembly line and no redundancy (solution S_1);

(ii) a wip buffer at the output of each manufacturing cell, a redundant assembly line and unitary operation at the manufacturing cells (solution S_{2a});

(iii) identical to S_{2a} but for batch mode operation at the manufacturing cells (solutions S_{2b}).

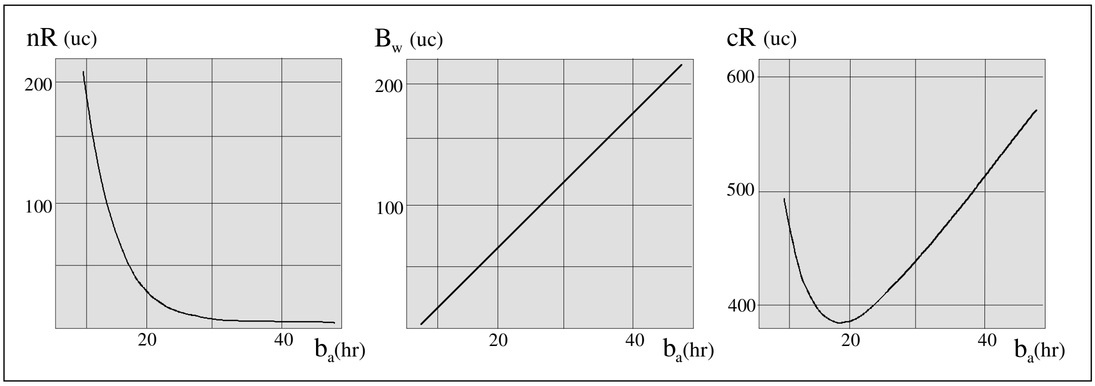
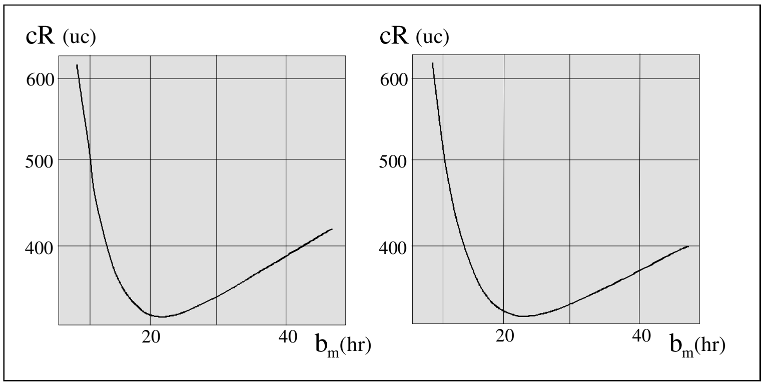


Figure 16 –Evaluation results for solution S_1

The graphs of Figure 16 show, for solution S_1, the evolution of the non-reliability costs (16.a), the cost of the output buffers (16.b) and the total reliability costs (16.c), versus the capacity of the buffer at the output of the assembly line, b_a. As may be expected, these curves show that there is a capacity of the buffer that minimizes the global cost of the system.

Figure 17 shows the evolution of the reliability costs, versus the capacity of the buffers located between the manufacturing cells and the assembly line, b_m, for solutions S_{2a} (17.a) and S_{2b} (17.b),. Based on these results (Table 10), it is possible to make the following statements:

- if the annual cost of redundant equipment is less than 62 uc, then S_1 will be the best solution. In this case, the optimal capacity of the buffer at the output of the assembly line is 18.4 hr.
- If the redundant equipment has a cost higher than 62 uc, S_2 becomes a better solution. In this case, the optimal design of the intermediate buffers, between the manufacturing cells and the assembly line, will be 20.4 hr and 21.3 hr respectively for S_{2a} and S_{2b}.
- The results also show that there is not a significant difference between solutions S_{2a} and S_{2b}, i.e., the operation mode of the manufacturing cells has a minor impact on the reliability costs of the AutoParts system.



a) Unitary operation b) Batch operation

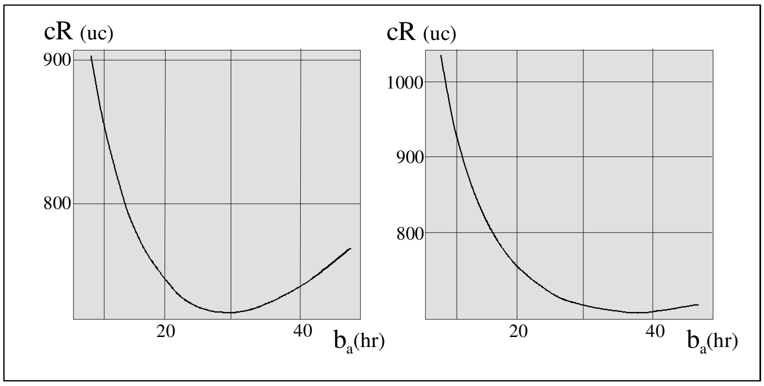
Figure 17 – Evaluation results for solution $\mathbf{S_2}$

Table 10 – Optimal buffer design

Solution	Total Cost (uc)	Buffer (h)
$\mathbf{S_1}$	375.2	18.4
$\mathbf{S_{2a}}$	313.5	20.4
$\mathbf{S_{2b}}$	317.5	21.3

5.4. Exponential versus non-exponential models

An important feature of the reliability methodology presented in this paper is the ability to deal with non-exponential distributions, thus avoiding the errors introduced in the calculation when a Markov model is misused. Therefore, it appears interesting to compare the results presented so far (obtained using empirical, non-exponential distributions for repair, reconfiguration and propagation processes), with those obtained when all the processes are supposed to have exponential distributions. Figure 18 shows such results for solutions S_{2a} and S_{2b}. Table 11 compares the values obtained for the optimal design of the buffers, from both the exponential and non-exponential models. These results reinforce the idea that the adoption of the Markovian hypothesis (exponential model) may introduce very significant errors in the calculations. In this system, the error reaches about 117% in the evaluation of the reliability costs, and 83% in the design of the buffer.



a) Unitary operation b) Batch operation

Figure 18 – Evaluation results for exponential processes

Suppose now, that the results obtained from the exponential model for solution S_1 were the basis for the design of the system. In this case, a buffer with a capacity of 30.0 hr would be implemented at the output of the assembly line (when the optimal capacity of the buffer is 18.4 hr). The reliability costs for such a buffer (obtained from the non-exponential model) are 383.3 uc. The comparison of this value with the minimum losses for S_1 (326.2 uc) shows that the exponential model would lead to a design of the system, that presents reliability costs 17.5% higher than those obtained from the "correct" non-exponential model. Similar conclusions could be drawn for solution S_2.

Table 11 – Comparison between exponential and non-exponential models

Solution	Reliability cost (uc)			Buffer capacity(h)		
	non-exp	exp	error	non-exp	exp	error
S_1	375.2	724.6	93%	18.4	30.0	63%
S_{2a}	313.5	690.2	120%	20.4	37.4	83%
S_{2b}	317.5	690.2	117%	21.3	37.4	76%

5.5. Non-instantaneous buffer replenishment

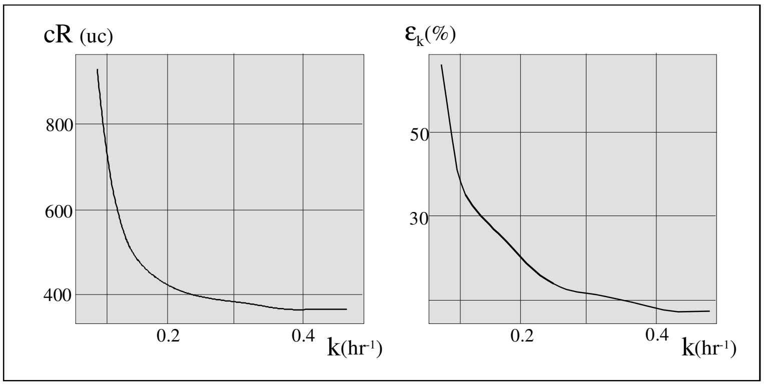


Figure 19 – Evaluation results for non-instantaneous replenishment

The replenishment process of the buffer can be taken into account in the evaluation of the system, using the simulation technique introduced in Section 4.1.3,. This is important when there is a significant probability that a new failure will occur before the buffer has recovered its nominal content, after the previous failure. By applying this technique to the AutoParts manufacturing system, it is possible to compare the results presented above with those obtained when the replenishment of the buffer is not ignored. For an output buffer corresponding to the optimal capacity determined before (18.4 hr), Figure 19.a shows the non-reliability costs versus replenishment rate k, defined as the reciprocal of the time needed to recover the nominal content after a shortage has occurred at the output of the buffer. These results were obtained using the simulation algorithm in Table 4, and show that, for low replenishment rates, there are significant differences, compared to previous results obtained for instantaneous replenishment ($k = \infty$). Figure 19.b shows the error ε_k in the evaluation of reliability costs when the replenishment of the buffer is ignored:

$$\varepsilon_k = \frac{nR_\infty - nR_k}{nR_k} 100\% \qquad (27)$$

For small values of k, the error becomes very significant (for example, reaching 37.2% for k=0.1), and the use of the non-instantaneous replenishment model becomes mandatory. Figure 20 shows the evolution of the reliability costs versus the capacity of the output buffer for different values of k. By analysing these curves it is possible to obtain the design of the buffer that minimizes the total cost, as shown in Table 12. However, for k greater than 0.5, the error is negligible (smaller than 5%) and the replenishment process of the buffer can consequently be ignored in the design of the system.

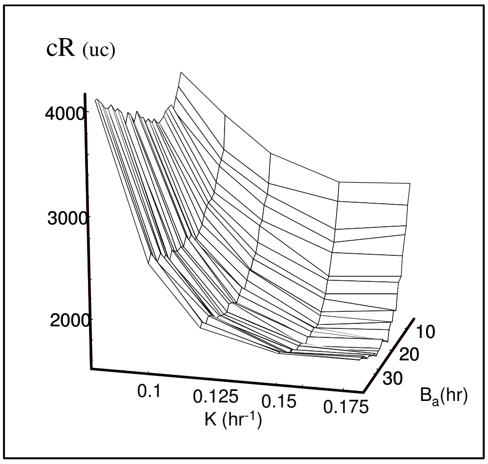


Figure 20 – Production losses vs buffer capacity and replenishment rate

Table 12 – Optimum buffer for non-instantaneous replenishment

k	Buffer (hr)	k	Buffer (hr)
0.1	14.5	0.15	27.4
0.125	17.2	0.175	33.5

6. Conclusions

Reliability is a major concern for planners and managers of JIT manufacturing systems, given that low reliability means increased production costs and lower compliance with delivery schedules. In a demanding (yet fairly common) business context like that of the AutoParts Company described in this paper, the reliability of the manufacturing system directly impacts with the overall performance of the integrated supply chain. In spite of this, a systematic reliability analysis such as the one presented here is often neglected during the design stage of manufacturing systems. Similar comments may be applied to other classes of distributed engineering systems, such as management information systems, extended logistical systems and electrical power systems. This situation is most likely due to the absence of ground engineering methodologies and tools to support planners and managers throughout the reliability analysis process.

In comparison with existing methods and tools for reliability analysis of production systems, the methodology presented here has a wider scope for application, and far less restrictive assumptions. Some of its main features are:

- the use of a hierarchical modelling framework, separating the endogenous and the exogenous behaviours of each unit, which is an effective approach to cope with the inherent complexity of large distributed systems;
- the orientation of the evaluation tools towards the assessment of economic damages caused by failures (i.e., the non-reliability costs), rather than towards the assessment of reliability indices that are of limited interest for system planners (e.g., availability and frequency of failure).

The capacity to deal with non-exponential processes is another fundamental feature, as error propagation delays will typically present a deterministic or quasi-deterministic behaviour. Furthermore, the assumption of exponential distributions would lead to wrong design decisions, as was shown in the final part of the case study.

The practical usefulness of the methodology and the kinds of results it can provide have been demonstrated through the detailed analysis and design of the manufacturing system presented in this paper. This case study has shown that the methodology may help planners of manufacturing systems to determine the most effective solutions for their systems in terms of: overall structure (installed production capacity, plant layout); number and type of equipment redundancy (active or stand by); number of maintenance resources (repairmen and spare parts); maintenance policy (responsibility for undertaking maintenance operations); and dimensioning of wip buffers (capacity and inventory level). The results of the case study also reinforce the notion that the so-called Markovian hypothesis often leads to dramatic errors that undermine design decisions.

The methodology presented in this paper was developed with industrial production systems in mind. However, it can be extended and adapted in order to accommodate the above mentioned engineering domains, which present a number of similar characteristics. All these systems can in fact be seen as large networks of units acting as producers and consumers of data, goods, power, etc. Each such unit will typically tolerate a temporary unavailability of the services delivered to it by the upstream subsystems[3]. In such conditions, propagation delays play a key role in the assessment of damages, which will typically have one component driven by the duration of the failures and another driven by their frequency.

References

[1] Salameh, M. and Ghattas, R. , "Optimal just-in-time buffer inventory for regular preventive maintenance", International Journal of Production Economics, v 74, nº 1-3, Dec. 2001, p 157-161

[2] Cheung, K. and Hausman, W., "Joint determination of preventive maintenance and safety stocks in an unreliable production environment", Naval Research Logistics, v 44, nº 3, Apr. 1997, p 257-272

[3] Biggart T. and Gargeya V., "Impact of JIT on inventory to sales ratios", Industrial Management & Data System, 102 (3-4), p 197-202, 2002

[4] Giordano, M., Martinelli, F., "Optimal safety stock for unreliable, finite buffer, single machine manufacturing systems", IEEE International Conference on Robotics and Automation, Washington, USA. Vol.3, p 2339-2344, May 2002.

[5] Van Ryzin, G., Lou, S., Gershwin, S. "Production control for tandem two-machine system", Transactions of the Institute of Electrical Engineers, Vol. 25, No. 5, p 5-20, Sep. 1993.

[6] Moinzadeh, K., Aggarwal, P., Management Science, Vol. 43, No. 11, p 1577-1588, Nov 1997.

[7] Nunes, E., Faria, J., Matos, M., "A comparative analysis of dependability assessment methodologies", Proceedings of the λμ13 ESREL Conference, Lyon, France, May 2002.

[8] Kenne, A. and Gharbi, A., "A simulation optimization based control policy for failure prone one-machine, two-product manufacturing systems", Computers & Industrial Engineering, v 46, nº 2, Apr. 2004, p 285-92

[9] Chelbi, A. and Ait-Kadi, D., "Analysis of a production/inventory system with randomly failing production unit submitted to regular preventive maintenance", European Journal of Operational Research, v 156, nº 3, Aug. 2004, p 712-18

[3] In the case of power distribution systems, regulation often specifies a tolerated unavailability of the service per year. In terms of loss evaluation, this is fairly similar to the manufacturing buffers that act as a temporary barrier to the propagation of the failures to clients.

[10] Kreng V. and Wang I., “Economical delivery strategies of products in a JIT system under a global supply chain”, International Journal of Advanced Manufacturing Technology 26 (11-12): p 1421-1428 Nov. 2005

[11] Puliafito, A., Scarpa, M., Trivedi, K., “Petri nets with k simultaneously enabled generally distributed timed transitions”, Performance Evaluation 32 (1): p 1-34, Feb 1998

[12] Balbo, G. ”Introduction to generalized stochastic Petri nets, Formal Methods for Performance Evaluation”, 7th International School on Formal Methods for the Design of Computer, Communication and Software Systems, SFM 2007. Advanced Lecture (Lecture Notes in Computer Science Vol.4486): p 83-131, 2007.

[13] Bobbio, A., Telek, M., “Non-exponential stochastic Petri nets: an overview of methods and techniques”, Computer Systems Science and Engineering 13 (6): 339-351, Nov 1998

[14] Carneiro, J.., Ferrarini, L.; “Reliability analysis of power system based on generalized stochastic Petri nets”, Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2008), 2008

[15] Haiyan, Z.; Shengqiang, L.; “Modeling and analysis of reverse supply chain based on generalized stochastic Petri nets”, 2009 International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII 2009): p 437-40, 2009

[16] Zhan, H.; Gu, J.; “Study of the normal generalized stochastic Petri nets and its application in testing system”, IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No. 06CH37714C), 2006

[17] Windebank, E.; “A Monte Carlo simulation method versus a general analytical method for determining reliability measures of repairable systems” Reliability Engineering, v 5, n 2: p 73-81, 1983

[18] Mahadevan, B., Narendran, T., “Buffer levels and choice of material handling device in Flexible Manufacturing Systems”, European Journal of Operational Research, p 166-176, 1993

[19]. Failure Mode and Effects Analysis in Manufacturing and Assembly Processes Reference Manual, Chrysler, Ford and General Motors under the sponsorship of USCAR, SAE, Warrendale, USA, 2002

[20] Rumbaugh, J., Jacobson, I., Booch, G., The Unified Modelling Language Reference Manual, Addison-Wesley, USA, 1998

[21] Faria, J., Matos, M., “An analytical methodology for the dependability evaluation of non-Markovian systems with multiple components”, Journal of Reliability Engineering and System Safety, 74, p 193-210, 2001.

[22] Wolfram, S., Mathematic: A system for doing Mathematics by Computer, 2nd Edition, Reading, MA Addison-Wesley, 1991.

Annex 1. Non-instantaneous buffer replenishment

For complex behaviour patterns, the determination of the canonical models using the analytical approach becomes ineffective, because the expressions for Λ and $\rho(t)$ are too complex to be of practical interest. In these situations, Monte Carlo simulation proves to be a better solution. As a typical example of application, consider the model in Figure A.2 which corresponds to a manufacturing cell comprising two machines in passive redundancy (M_1 and M_2); an automatic guided vehicle

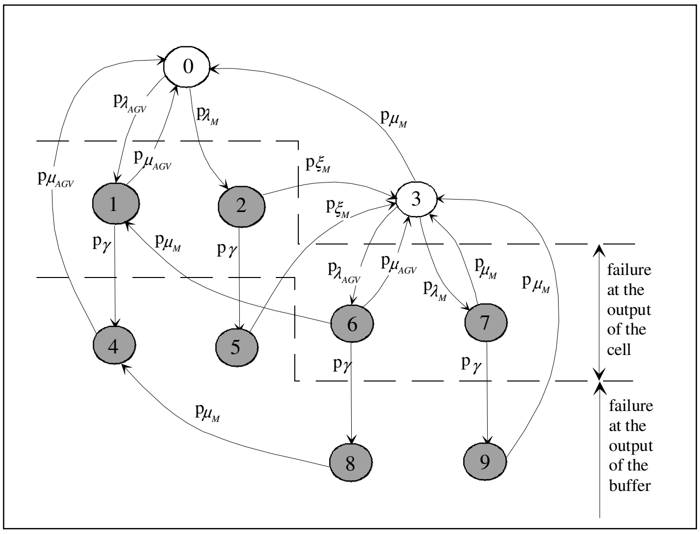


Figure A.2 – Redundant cell with output buffer

(AGV) for materials handling; and a single repairman. The behaviour of the cell is as follows: if the AGV fails ($p_{\lambda_{AGV}}$), the cell immediately stops its operation, until its repair takes place ($p_{\mu_{AGV}}$). If the first redundant machine fails ($p_{\lambda_{M1}}$), a reconfiguration is undertaken ($p_{\xi M}$) in order to put the redundant machine into operation. If the latter also fails ($p_{\lambda_{M2}}$), the cell stops its operation. Processes $p_{\mu_{M1}}$ and $p_{\mu_{M2}}$ model the repair processes of the two machines, (an operational procedure states that when the two machines are simultaneously down, the repair priority is assigned to the last machine to fail). The sets of states {1, 2, 6, 7} and {4, 5, 8, 9} correspond to the situations in which there is a shortage of material at the output of the cell and at the output of the buffer respectively.

Suppose now that there is a significant probability that a new failure will occur before the buffer has recovered its nominal content. In this condition the delay introduced by the buffer (equal to its content when the cell enters a failure state) will depend on the previous states occupied by the system.

$$\gamma(t) = \delta (t - I_{in}) \qquad \text{(A.12)}$$

The value of a delay for a failure state n can be determined as shown in Table A.1, where : I_n denotes the nominal content of the buffer; I_n^{in} and I_n^{out} denote the content of the

Table A.1 – Expressions for the determination of I_n^{in} and I_n^{out}

$I_1^{in} = I_0^{out}$	$I_1^{out} = I_1^{in} - (t_1^{out} - t_1^{in})$
$I_2^{in} = I_0^{out}$	$I_2^{out} = I_2^{in} - (t_2^{out} - t_2^{in})$
$I_3^{in} = I_2^{out}$ (for transition 2 → 3) $I_3^{in} = 0$ (for transition 5 → 3)	$I_3^{out} = \min (I^n , I_3^{in} + k (t_3^{out} - t_3^{in}))$
$I_4^{in} = 0$	$I_4^{out} = 0$
…	

buffer at instants t_n^{in} and t_n^{out} respectively, that is, when state n is entered and when it is left; and k denotes the replenishment rate of the buffer, defined as the reciprocal of the time needed by the buffer to recover its nominal content after a shortage has occurred at its output.

For such complex behaviour patterns, the simulation approach is the only effective one. The simulation algorithm is discussed in Section 4.1.3, together with a practical example relating to the cell presented here. It assumes that (i) the failure processes are exponential with $r_{\lambda_{AGV}}$ = 0.01 h^{-1} and $r_{\lambda_{M1}}$ = $r_{\lambda_{M2}}$ = 0.05 h^{-1}; (ii) the three repair processes present 3rd order Erlang density functions with $m_{\mu_{AGV}}$ = 4 hr and $m_{\mu_{M1}} = m_{\mu_{M2}}$ = 20 hr; (iii) the reconfiguration process p_{ξ} presents a Dirac density function with m_{ξ} = 0.5 hr. This same behaviour pattern, non-instantaneous buffer replenishment, is considered in the numerical analysis of the AutoParts system presented in Section 5.