Production quality performance in manufacturing systems processing deteriorating products

Colledani, Marcello; Horvath, Andras; Angius, Alessio

Production quality performance in manufacturing systems processing deteriorating products

Marcello Colledani (2)[a,*], Andras Horvath [b], Alessio Angius [b]

[a] Politecnico di Milano, Department of Mechanical Engineering, Via la Masa, 1, 20156, Milan, Italy
[b] Università di Torino, Department of Computer Science, Corso Svizzera 185, Torino, Italy

1. Introduction, motivation and objectives

Production quality has been proposed recently as an emerging paradigm to achieve desired service levels of conforming products in advanced manufacturing systems, by simultaneously considering quality and productivity requirements [1]. With respect to this background, the importance of an integrated analysis of production logistics, product quality and equipment maintenance to achieve balanced manufacturing system solutions has been pointed out. This problem is particularly relevant in manufacturing systems producing deteriorating products.

Product quality and value deterioration due to excessive residence times (or lead times) during production is a significant phenomenon in several technology intensive industries, including automotive, food manufacturing, semiconductor and electronics manufacturing and in polymer forming. For example, in automotive paint shops a car body that is affected by prolonged exposure to the air in the shop floor caused by excessive lead times between operations, is prone to particle contamination, leading to unacceptable quality of the output of the painting process. Moreover, food production is pervaded by strict requirements on hygiene and delivery precision requiring a maximum allowed storage time before packaging. If the production lead time exceeds this limit, the product has to be considered as defective and cannot be delivered to the customer. In these systems, higher inventory increases the system throughput but also increases the production lead times, thus increasing the probability of producing defective items. Therefore, a relevant trade-off is generated between production logistics and quality performance that requires advanced engineering methods to be profitably addressed.

In spite of the relevance of this phenomenon in industry, the analysis of production quality performance under product deterioration has received relatively low attention in the literature. The manufacturing system is considered in a highly aggregate way in advanced Economic Production Quantity (EPQ) models considering quality deterioration [2]. In these works, the quality deterioration due to the parts residence time along the stages of the manufacturing system is neglected. Other works considered supply chain coordination mechanisms in presence of product obsolescence [3]. Furthermore, production control policies based on *WIP* [4] and part release [5,6] regulation could support the achievement of improved production quality performance under product deterioration, although they do not provide mechanisms to directly control production lead times.

The first model considering this interaction is proposed in [7] that analyzed un-buffered systems where the material under processing is scrapped after long machine failures. Moreover, the performance of serial lines with product deterioration is considered under Bernoulli reliability models of production stages in [8], and in small two-machine lines in [9]. While all these works are important to shed light on the problem, they do not provide methods to predict and control production lead time distributions under realistic manufacturing system features. As a matter of fact, a methodology to support the design of manufacturing systems under lead time dependent product deterioration that integrates quality and production logistics implications has never been proposed. Important questions like "What is the impact of buffers on the production rate of conforming products with product quality deterioration?" remain unsolved, resulting in sub-performing system configurations.

To overcome these limitations, in this paper an integrated model of manufacturing systems affected by product deterioration and a new method for the prediction of the production lead time distribution and the throughput of conforming products in these

systems are developed for the first time. This approach allows setting inventory levels to achieve desired production quality performance in real systems.

2. System description

The considered system is formed by K manufacturing stages and $K - 1$ buffers of finite capacity, configured in serial layout (Fig. 1). Stages are denoted as M_k, with $k = 1, \ldots, K$, and buffers are denoted as B_k, with $k = 1, \ldots, K - 1$. The capacity of buffer B_k is N_k, that is an integer number. Finite capacity buffers can either model physical conveyors or the implementation of token-based production control rules, such as kanban, regulating the material flow release at each stage [10].
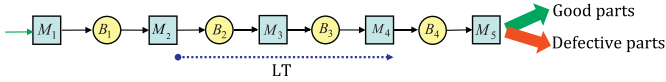

Fig. 1. Representation of the analyzed manufacturing system.

Single stage model. The dynamics of each stage is modeled by a discrete-time and discrete-state Markov chain of general complexity. This setup allows to analyze a wide set of different stage models within a unique framework. For example, stages with unreliable machines characterized by generally distributed up and down times and also stages with non-identical processing times can be considered within the same framework, thus making the proposed approach applicable to a wide set of real manufacturing systems.

In detail, each stage M_k is represented by I_k states, and thus the state indicator α_k assumes values in $\{1, \ldots, I_k\}$. The set containing all the states of M_k is called S_k. The dynamics of each stage in visiting its states is captured by the transition probability matrix λ_k, that is a square matrix of size I_k. Moreover, a quantity reward vector $\mu_\mathbf{k}$ is considered, with I_k binary entries. While in the generic state i, M_k produces $\mu_{k,i}$ parts per time unit. Therefore, state i with $\mu_{k,i} = 1$ can be considered as an operational state for stage M_k, while state i with $\mu_{k,i} = 0$ is a down state for stage M_k.

Material flow dynamics. A discrete flow of parts is considered in the system. Stage M_k is blocked if the buffer B_k is full. Stage M_k is starved if the buffer B_{k-1} is empty. Stage M_K is never blocked, i.e., infinite amount of space is available to store finished products. Stage M_1 is never starved, i.e., unlimited supply of raw parts is assumed. Operational Dependent Transitions are considered, i.e., a machine cannot make transitions to other states if it is starved or blocked. State transitions take place at the beginning of the time unit and buffer levels are updated at the end of the time unit.

Part quality deterioration. The quality of parts deteriorates with the time parts spend in a *critical portion of system*, denoted by two integers, e and q with $1 \le e < q \le K$, and composed of those buffers that are between stages M_e and M_q. The lead time of a part is the time spent in buffers $B_e, B_{e+1}, \ldots, B_{q-1}$. For example, for the system in Fig. 1, $e = 2$ and $q = 4$. If $e = 1$ and $q = K$, then the time spent in the whole system is considered. The probability that a part is defective at the end of the line is a non-decreasing function of its lead time. The function $g(h)$ indicates the probability that a part released by the system is defective given that it spent h time units in the critical portion of system. Defective parts are scrapped at the end of the line.

Performance measures. The main performance measures of interest are:

- Average total production rate of the system, denoted by E^{Tot},
- Probability that the lead time, LT, is equal to a given number of time units, h, i.e., $P(LT = h)$,
- Average effective production rate of conforming parts, E^{Eff}, which is given by:

$$E^{Eff} = E^{Tot} \cdot \sum_{h=1}^{\infty} P(LT = h)[1 - g(h)] \tag{1}$$

- System yield, Y^{system}, i.e. fraction of conforming parts: (E^{Eff}/E^{Tot}).

- Total average inventory of the system, *WIP*.

3. Lead time distribution evaluation

In this section, an efficient and exact analytical method to compute the distribution of the lead time in the critical portion of the system is described. The rationale of the approach is explained in the following. Firstly, the probability that, at the moment when a randomly selected part enters buffer B_e, the system is in a given state is derived. Secondly, the probability that, given that state of the system as initial condition, the last part in buffer B_e crosses the critical portion of the system in h time units is analyzed. By taking the product of these two quantities and summing for all possible states, the distribution of the lead time in the critical portion of the system can be reconstructed. In the following, the steps of this procedure are briefly described.

Starting from the parameters of the stages and according to the modeling assumptions, a discrete time Markov chain, denoted by $\mathfrak{D}_1$, can be determined that characterizes the overall behavior of the system. A state of $\mathfrak{D}_1$ is described by a vector $\mathbf{s_1} = (n_1, n_2, \ldots, n_{K-1}, \alpha_1, \alpha_2, \ldots, \alpha_K)$ where n_k is the number of parts in buffer B_k and α_k is the state of machine M_k. The set of all the system states is denoted by Ω_1 and the transition probability matrix by $\mathbf{P_1}$. The entries of $\mathbf{P_1}$ can be obtained as shown in [11]. The row vector of the steady state probabilities, denoted as $\pi_\mathbf{1}$, and the total production rate, E^{Tot}, can be calculated as in [11,12].

The sub-system including only that portion of line that is downstream machine M_e is considered next. This sub-system comprises buffers $B_e, B_{e+1}, \ldots, B_{K-1}$ and machines $M_{e+1}, M_{e+2}, \ldots, M_K$. Machines $M_1, \ldots, M_e$ are not considered because, once a part is put in buffer B_e, these machines have no impact on its lead time. For this sub-system, a Markov chain, denoted by $\mathfrak{D}_2$, with initial probability vector $\pi_\mathbf{2}$ and transition probability matrix $\mathbf{P_2}$ is determined. $\mathfrak{D}_2$ can be considerably reduced with respect to $\mathfrak{D}_1$. A state of this model is given by a vector $\mathbf{s_2} = (n_e, n_{e+1}, \ldots, n_{K-1}, \alpha_{e+1}, \alpha_{e+2}, \ldots, \alpha_K)$. The set of states of $\mathfrak{D}_2$ will be denoted by Ω_2. Note that, since no machine feeds parts in the first buffer of this model, all parts will eventually leave and the system will get empty. In other words, $\mathfrak{D}_2$ is an absorbing Markov chain [12]. The subset of states in S_2 in which there are no parts in the buffers of the critical portion $B_e, B_{e+1}, \ldots, B_{q-1}$ will be denoted by Ω_0. The lead time in the critical portion of system can be expressed as follows:

$$P(LT \le h) = \sum_{s_2 \in \Omega_2} \pi_{\mathbf{2},s_2} f(s_2, h) \tag{2}$$

where π_{2,s_2} is the probability that a random part that enters in buffer B_e finds the system in state $\mathbf{s_2}$ and $f(\mathbf{s_2}, h)$ is the probability that a part that enters the critical portion in state $\mathbf{s_2}$ leaves machine M_q in at most h time units. The first term in Eq. (2) is obtained by properly mapping the states $\mathbf{s_1}$ in $\mathfrak{D}_1$ into the states $\mathbf{s_2}$ of the sub-system in $\mathfrak{D}_2$. For this reason the matrix $\mathbf{F}$ of binary entries is defined that links the states in $\mathbf{s_1}$ and $\mathbf{s_2}$ if the states of all the stages M_{j+1} and buffers B_j, with $j > e$, are the same. In other words, the initial probability of the state $\mathbf{s_2}$ in $\mathfrak{D}_2$, π_{2,s_2}, has to be equal to the sum of the probabilities that a part is put in the buffer B_e in the connected set of states $\mathbf{s_1}$ in $\mathfrak{D}_1$. Therefore:

$$\pi_\mathbf{2} = \frac{\pi_\mathbf{1} \mathbf{Q_1}}{E^{Tot}} F \tag{3}$$

where $\mathbf{Q_1}$ is identical to $\mathbf{P_1}$, except that those transitions that do not cause the release of a part in buffer B_e are set to zero.

The second term in (2) can be easily computed as the distribution of the first passage time to Ω_0 in the absorbing Markov chain $\mathfrak{D}_2$ [12]. The probability that the lead time is exactly h is given by $P(LT = h) = P(LT \le n) - P(LT \le n - 1)$. The other performance measures defined in Section 2 can be derived based on the lead time distribution and the total throughput.

4. Numerical results and system behavior

In the first experiment, the distribution of the lead time in multi-stage manufacturing systems is investigated. A production line with five stages is considered with parameters reported in Table 1. The stages correspond to machines with one operational state and one failure state, with failure probability $p = 1/MTTF$ and repair probability $r = 1/MTTR$ ($MTTF$ is Mean Time to Failure and $MTTR$ is Mean Time to Repair). The machines produce one part per time unit if operational. The results are reported in Fig. 2. The following considerations hold:

Table 1
Data of the systems analyzed in the experiments.

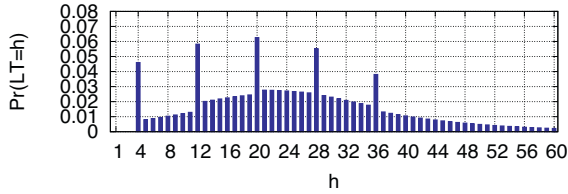| Exp. no.	System data
1	Balanced line: $\lambda_k = \begin{vmatrix} 1-p_k & p_k \\ r_k & 1-r_k \end{vmatrix}, \mu_k = [1,0]$
	with $p_k = 0.01$ and $r_k = 0.1$, $N_1 = N_2 = N_3 = N_4 = 10$. $e = 1$, $q = K = 5$.
2	Same machine model as exp. 1
	$p_1 = 0.01$; $r_1 = 0.05$; $p_2 = 0.02$; $r_2 = 0.2$. $N_1 = [5:50]$. $e = 1$, $q = K = 2$.
3	Same machine model as exp. 1
	$p_1 = 0.01$; $r_1 = 0.25$; $p_2 = 0.02$; $r_2 = [0.02:0.5]$. $e = 1$, $q = K = 2$.

Fig. 2. Lead time distribution for a five machine line.

- The distribution of the lead time is multi-modal with a number of peaks equal to the number of production stages, K.
- The v-th peak, with $v = 1, \ldots, K$, is located in correspondence to the following number of time units, h_v:

$$h_v = K - v + \sum_{i=1}^{v-1}(N_i - 1) \tag{4}$$

- The peaks appear for the following reasons. The first peak corresponds to the situation where the system is empty and the parts, after being processed by stage M_1, cross the remaining $K - 1$ stages, each one spending one time unit for processing. The v-th peak, with $v = 2, \ldots, K$, corresponds to the situation where stage M_v failed for sufficiently long time to fill up all the upstream buffers and empty all the downstream buffers. After the stage is recovered from the failure, an incoming part has to wait $N_i - 1$ time units in the upstream buffers, with $i < v$, and 1 time unit in each of the $K-v$ downstream stages, if no other failure occurs. The knowledge of this behavior can be exploited to correctly operate and design the production line.

In the second experiment, the impact of the buffer size on the effective throughput, E^{Eff}, is investigated for a simple system with two machines and one buffer, with parameters reported in Table 1. The quality deterioration function $g(h)$ assumes value zero for $h = 1, \ldots, h^*$ and value 1 for $h > h^*$, with $h^* = 40$. In other words, if the parts spend more than 40 time units in the system, they turn into defective. In the experiment, the buffer capacity N is varied from 5 to 50. Results are reported in Fig. 3. The following considerations hold:

- The effective throughput curve is a concave function presenting a maximum corresponding to a buffer capacity equal to 25.
- This behavior is explained in the following. It is well known that the total throughput of the system is monotonically increasing with the buffer size, with a marginal gain that reduces while increasing the buffer size. At the same time, if the first machine is more reliable than the second, while increasing the buffer size also the average lead time increases, as the average inventory
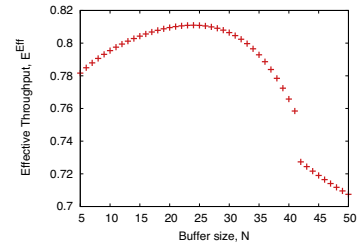

Fig. 3. Effect of the buffer size on the effective throughput.

increases almost linearly. Therefore, the scrap rate will increase as a function of the buffer size since the probability that the lead time exceeds the threshold increases. The result of this combined effect is the following. When the first phenomenon is more relevant (small buffers), the effective throughput curve increases with the buffer size. When the second phenomenon is more relevant (large buffers), the effective throughput curve starts decreasing, as the system produces more parts but also a higher fraction of defective parts. Therefore, a trade-off is generated that is optimized for a specific buffer size. This optimal buffer size could not be found without the proposed approach.

- It is worth to note that for $N > h^* + 1$ the curve drops abruptly in a discontinuous way and shows a different slope. This phenomenon is due to the fact that, according to equation (4), the lead time distribution curve in a two-machine line shows a probability peak at level $N-1$. Therefore, operating a system with $N-1 > h^*$ does not make sense in practical settings.

The third experiment shows the impact of the repair probability of the first machine on the effective throughput, for different buffer sizes in a line with two machines and one buffer. Results are reported in Fig. 4. The following considerations hold:
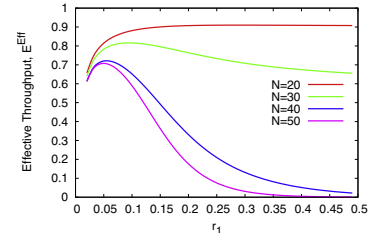

Fig. 4. Effect of the repair probability on the effective throughput.

- For all the cases, it is possible to notice that faster repair operations are not always beneficial in terms of effective throughput of the system. Indeed, a faster repair of the first machine increases the total throughput of the system but also increases the amount of inventory stored in the buffer, as the buffer will be more frequently full. This phenomenon increases the lead time and, consequently, the scrap rate.
- This behavior is more prominent for large buffers. Indeed, in these cases, the positive effect of shorter repair time on the total throughput is less significant than the negative effect on the lead time increase.
- This analysis suggests that improvement actions that are beneficial on the total throughput of the system (faster repair operations) can be detrimental on the effective throughput, if the products are perishable. Thus, neglecting this aspect can lead to sub-performing system configurations.

5. Real case study

The proposed approach has been applied to the production of micro-catheters as high value medical products for the aging society in the medical technology sector at ENKI s.r.l. [13]. The manufacturing process is composed of three main phases: (i)

material compound preparation and control, the (ii) micro-extrusion of the micro-tubes and (iii) final micro-catheter assembly. Defects are mainly geometrical, generated within the micro-extrusion process. The above defects lead to extremely high defect rates (up to 70% in standard production).

During the phase of preparation of the raw material, in forms of granules, the moisture level is measured and if the value exceeds a fixed limit, the granules are dried to reduce the moisture level before the downstream micro-extrusion process. The lead time between the drying and extrusion processes should not exceed a certain limit to avoid increase of the moisture level by exposure to the air. A maximal moisture level of 300 ppm is usually the accepted limit. Excessive moisture level affects the material viscosity. This results in fluctuations of the material shear rate and instability in the extrusion process, thus affecting the quality of the tube section. The line under analysis is composed of three main stages. Stage M_1 is the compounding stage provides additives to the material and makes the size of the pellets homogeneous. Stage M_2 is the drying stage, which reduces the moisture level of the material. The residence time of the material in the dryer depends on the material type and properties. For example, for PEBA pellets the drying time is usually 4 h. After the drying process, the material is loaded in the input hopper of the extruder at stage M_3. The extrusion speed selection is one of the key competences of the company, directly affecting the output quality of the tube. It is strongly dependent on the material type and tube section complexity, usually ranging between 3 m/s and 20 m/s. After the extrusion, each tube undergoes a set of very severe EOL (End of Line) inspections and non-conforming tubes are scrapped. However, the company's strategic goal is to avoid the generation of defects by improving the control of the material properties before the extrusion process. For this reason, the lead time between the end of the drying process and the feeding of the granules in the extruder hoppers must be limited to avoid gradual increase in the humidity on the granules. After a deep study of the material behavior in extrusion, the company assessed that the maximal residence time of the material between these stages is 45 s. Therefore, the main objective of the analysis is to properly design the in-process buffer included between the drying stage and the extrusion stage in order to decrease the fraction of defects due to material deterioration in extrusion, thus increasing the effective production rate levels.

The micro-tube forming line under the current configuration has been modeled. Two types of tubes have been considered with part-type dependent process parameters. The number of lumens ranges from 4 to 26 and the external diameters ranges from 2 mm to 5 mm. Time-consuming set-up operations are required while switching between part types; therefore different production settings have been considered, one for each tube type. The operational and down times were provided by the company. The performance of the system was evaluated and the results were validated through comparison with historical production data, showing deviations always below 5%.

The approach proposed in this paper has been applied to design a different production control policy to manage the available hopper capacity before the extrusion machine. In detail, for each tube type, the hopper capacity that maximizes the effective throughput of the system, under a constraint on the maximum hopper capacity currently available in the line (1500 g), has been investigated. Results are reported in Table 2. Similarly to what was shown for experiment 2 in Section 4, for each case, the effective throughput is maximized for a hopper capacity that is smaller than the currently adopted capacity. With the proposed optimal policy, in all the cases only 1% of the tubes is defective due to excessive lead time, against the 6% and 17%, respectively, for each tube type, in the current configuration (a reduction of more then 80% for both types). Since defective tubes cannot be repaired and are sent to recycling, this solution entails a large environmental benefit. Moreover, benefits are achieved in terms of effective throughput (3.47% and 16.8%), while also reducing the work in progress in the

Table 2
Micro-tube production line performance under optimized buffer size (the time unit is omitted for confidentiality reasons).

Tube type	Buffer size (g)	E^{Tot}	Y^{system}	E^{Eff}	WIP (g)	Scrap rate
1	$N^{Current} = 1500$	0.862	0.94	0.812	884	0.06
	$N^{opt} = 783$	0.849	0.991	0.84	448	0.009
Difference Δ%		−1.5%	+5%	+3.47%	−49.3%	−82%
2	$N^{Current} = 1500$	0.868	0.829	0.719	901.7	0.17
	$N^{opt} = 591$	0.85	0.991	0.842	338	0.009
Difference Δ%		−2%	+19.3%	+16.8%	−62.5%	−93.9%

system (−50% as a minimum). This directly translates into a reduction of production costs, thus positively impacting on the company's competitiveness.

6. Conclusions

This paper proposes a modeling framework and a methodology to predict the lead time distribution in multi-stage manufacturing systems. The lead time distribution is used to compute the system production quality performance, when products deteriorate. The theoretical work developed in this paper opens the way for applications in several different fields. For example, supply chain management approaches can benefit from the availability of a solution to the problem of deriving the lead time distribution in unreliable manufacturing systems. Moreover, production planning approaches that take into account the exact distribution of the residence time of parts in the system can be developed. Finally, new production control policies to provide a direct control on the lead time can be elaborated from these results.

References

[1] Colledani M, Fischer A, Iung B, Lanza G, Schmitt R, Tolio T, Vancza J (2014) Design and Management of Manufacturing Systems for Production Quality. *CIRP Annals – Manufacturing Technology* 63(2):773–796.

[2] Subbaiah KV, Rao SVUM, Rao KS (2011) An Inventory Model for Perishable Items with Alternating Rate of Production. *International Journal of Advanced Operations Management* 3(1):66–87.

[3] Váncza J, Monostori L, Lutters D, Kumara SR, Tseng M, Valckenaers P, Van Brussel H (2011) Cooperative and Responsive Manufacturing Enterprises. *CIRP Annals – Manufacturing Technology* 60(2):797–820.

[4] Duffie NA, Shi L (2009) Maintaining Constant WIP-Regulation Dynamics in Production Networks with Autonomous Work Systems. *CIRP Annals – Manufacturing Technology* 58(1):399–402.

[5] Windt K, Knollmann M (2014) Consequences of Planned Lead Time Adaptions in Scope of the Lead Time Syndrome of Production Control. *CIRP Annals –Manufacturing Technology* 63(1):405–408.

[6] Bertsch S, Schmidt M, Nyhuis P (2014) Modeling of Lateness Distributions Depending on the Sequencing Method with Respect to Productivity Effects. *CIRP Annals – Manufacturing Technology* 63(1):429–432.

[7] Liberopoulos G, Kozanidis G, Tsarouhas P (2007) Performance Evaluation of an Automatic Transfer Line with WIP Scrapping during Long Failures. *Manufacturing and Service Operations* 9(1):62–83.

[8] Naebulharam R, Zhang L (2014) Bernoulli Serial Lines with Deteriorating Product Quality: Performance Evaluation and System-Theoretic Properties. *International Journal of Production Research* 52(5):1479–1494.

[9] Gershwin S-B, Shi C (2012) Part Waiting Time Distribution in a Two-Machine Line. *Proceedings of the 14th IFAC Incom*, May 23–25, Bucharest, Romania, 457–462.

[10] Colledani M, Tolio T (2011) Joint Design of Quality and Production Control in Manufacturing Systems. *CIRP Journal of Manufacturing Science and Technology* 4(3):281–289.

[11] Colledani M (2013) A Decomposition Method for the Analysis of Long Buffered Production Systems with Discrete General Markovian Machines. *Proceedings of the 7th IFAC Conference on Manufacturing Modelling, Management, and Control*, June 19–21, Saint Petersburg, Russia, 1644–1649.

[12] Stewart WJ (2009) *Probability, Markov Chains, Queues and Simulation: The Mathematical Basis of Performance Modeling*, Princeton University Press.

[13] EU, FP7, FoF.NMP. 2011-5, 2011, MuProD-Innovative Proactive Quality Control System for In-process Multi-stage Defect Reduction – Grant Agreement no. 285075.