Moments of Cumulated Output and Completion Time of Unreliable General Markovian Machines

A. Angius * **A. Horváth** * **M. Colledani** **

* *Department of Computer Science, Università di Torino, Turin, Italy, (e-mail: angius@di.unito.it and horvath@di.unito.it).*
** *Mechanical Engineering Department, Politecnico di Milano, Milan, Italy, (e-mail: marcello.colledani@polimi.it).*

Abstract: Performance evaluation models are used by companies to design, adapt, manage and control their production systems. In the literature, most of the effort has been dedicated to the development of efficient methodologies to estimate the first moment performance measures of production systems, such as the expected production rate, the buffer levels and the mean completion time. However, there is industrial evidence that the variability of the production output may drastically impact on the capability of managing the system operations, causing the observed system performance to be highly different from what expected. This paper presents a general theory and a methodology to analyze the cumulated output and the lot completion time variability of unreliable machines and systems characterized by general Markovian models. Both discrete models and continuous reward models are considered. We then discuss two simple examples that show how the theory developed in this paper can be applied to analyse the dependency of the output variability on the system parameters.

Keywords: Performance Evaluation, Manufacturing Systems, Output Variability, General Markovian Machines, Markov Reward Models.

1. INTRODUCTION

Manufacturing System Engineering methods have been developed in the last decades for investigating the dynamic behaviour of manufacturing systems, for estimating their performance and for supporting their efficient design, improvement and reconfiguration. Among these, simulation and analytical tools are the most commonly adopted instruments. Typically, these approaches are focused on the first order performance measures of manufacturing systems, such as the average throughput, the average work in progress and the average system flow time. During the system configuration/reconfiguration phase, these tools are used to select system solutions that profitably exploit the trade-offs between these first order performance measures. Higher order performance measures are generally difficult to analyze and are rarely considered.

However, in the presence of random events and disturbances in the production, higher order performance measures are relevant to correctly predict the system output. Indeed, due to the production variability, the observed performance can be highly different from the average performance. This variability makes it difficult to meet customer orders on time and to assure the required service level of the system. This problem may directly corrupt the profitability of those systems designed only by considering the mean performance of the system, that are not robust to disturbances. Low output variance indicates stability of the output of the production line, less unforeseen delays and small fraction of escaped orders, which translates to lower costs. Symmetrically, high variance means instability of the output, i.e. significant differences in production quantity observed on a day by day basis.

Typical sources of variability in the production system behaviour are random failure occurrences and durations. A real case in the automotive sector (Colledani et al. (2010)) reports that the weekly output of the production system composed of 22 machines affected by the occurrences of 144 different failure modes, has a coefficient of variation, estimated from the available field data of three months, equal to 0.263. Thus, it is highly probable that the weekly demand will not be met if the system is designed only considering its average performance.

In spite of the relevance of this issue in industry, the number of papers discussing the variability of the output in production lines is limited and the underlying assumptions of the available methods are simplistic, thus preventing their wide application in industry. In Miltenburg (1987) a numerical method to calculate the variance of the number of parts produced by the system at a given time T is proposed. The author considers multi-stage buffered production lines featuring unreliable machines with geometric failures and repair times, and deterministic processing times. The approach is based on the state-space representation of the system. This makes the approach viable only for small and simple systems. This method was later improved and extended to other type of production lines and machines in Tan (1997), Tan (1999a) and Tan (1999b). In Ou and Gershwin (1989) the closed form expressions

of the variance of the lead time in a two machine line in which machines may fail in single mode is obtained. Gershwin also proposes a method for the calculation of the variance of the output of an isolated machine with a single geometric failure mode in closed form (Gershwin (1993)). His methods are based on the solution of the difference equation describing the system dynamics. In Carrascosa (1995) this method is extended to the case of the isolated machine with geometric multiple failure modes. Hendrics (1992) presents an approach, based on the structural properties of Markov chains, to estimate the asymptotic variance rate of interdeparture times in production lines with exponential processing times, perfectly reliable machines and finite buffer capacities. Li and Meerkov (2000) study the variance of the output for production lines composed of unreliable machines and finite buffers. The most limiting assumption to the application of this method is the Bernoulli reliability model, which assumes repair time equal to the cycle time of the machines. Recently, in He et al. (2007) the output variance of long production lines is studied using Markov Arrival Process (MAP). The limiting assumption in this paper is that only reliable machines with exponential processing times are considered.

In Tan (2000), the author reviews the papers focused on the analysis of the output variability in production lines and proposes a classification of the existing methods based on the considered system layout, the considered system parameters and their distributions, the type of solution adopted, exact or approximate, and the complexity of the system that can be analyzed. According to this analysis, available methods only consider the assumptions of exponentially or, in the discrete time domain, geometrically distributed machine failure and repair times. However, while in real systems the times to machine failures can often be modeled using exponential or geometrical distributions with acceptable accuracy, given the mechanical and electronic nature of failures, the times to repair are rarely observed to follow exponential distributions (Inman (1999)). This paper considers general Markovian machines and systems in the analysis, thus enabling to revise this critical assumption of the existing methods.

Specifically, the objective of this paper is to develop a general theory and a methodology to analyze the moments of the cumulated output and the moments of the completion time in manufacturing systems modeled as arbitrarily complex Markov chains. The generality of the proposed approach allows modeling and studying the output variability under many different system configurations within a unique framework, also including several previously uninvestigated system layouts. Among these, unbuffered multi-stage serial-parallel systems with limited repair capacity and buffered two-machine lines with degrading machines are modeled and analyzed as examples in this paper, within the proposed framework. Moreover, the impact of the main system and machine parameters on the variability of the output can be investigated, with the objective of deriving new system design rules for reducing the variability and meeting the due-time performance of the system. Indeed, very little is known on how to manage production systems to reduce the variability of their output. Important questions like "What is the due date to be quoted for a given order?" and "What is the probability of delivering a given order on time, under a particular system configuration?" still remain unsolved.

The paper is organized as follows. In Sect. 2 we discuss Markov reward models reporting existing results and deriving new ones. In Sect. 3 we illusrate the application of the proposed framework. Conclusions are drawn in Sect. 4.

2. MARKOV REWARD MODELS

In a Markov reward model (MRM) an underlying Markov chain modulates the reward rate and a sojourn of length u in an up-state accumulates a quantity of reward proportional to u while in a down-state the production is zero. The generality of modeling machines with MRMs lies in the facts that: **(i)** MRMs allow us to model general failure/repair mechanisms with phase type distributed durations and **(ii)** the set of phase type distributions is dense in the field of all positive-valued distributions (Neuts (1981)), i.e., it can be used to approximate any failure/repair distribution.

While to compute the distribution of the accumulated reward and the distribution of completion time is computationally heavy, there exist efficient methods for the calculation of the moments of these distributions. In Sect. 2.1 we report these results for continuous time models with continuous reward and in Sect. 2.2 we derive the counterpart for discrete time models with discrete reward. The corresponding quantities will be denoted by the same symbols in the continuous and the discrete case.

2.1 Continuous Markov Reward Model

A continuous time Markov reward model is defined by the infinitesimal generator of the underlying continuous time Markov chain (CTMC), denoted by $\mathbf{Q}$, and the reward rates associated with the states denoted by r_i for state i and organized into a diagonal matrix denoted by $\mathbf{R}$. During a sojourn of length u in state i the system accumulates ur_i amount of reward. Denoting by $X(t)$ the state of the chain at time t and by $Z(t)$ the accumulated reward (the production) by time t we have $Z(t) = \int_{s=0}^{t} r_{X(s)} ds$. The completion time is defined as $C(x) = \min[t \geq 0 : Z(t) = x]$. The quantities describing the accumulated reward and the completion time are

$$F_{ij}(t,x) = P\{X(t) = j, Z(t) < x | X(0) = i\} \quad (1)$$

$$G_{ij}(t,x) = P\{X(C(x)) = j, C(x) < t | X(0) = i\} \quad (2)$$

where $F_{ij}(t,x)$ is the joint distribution of the accumulated reward and the state of the underlying CTMC supposing that the initial state is i and $G_{ij}(t,x)$ is the joint probability of the completion time and the state at completion supposing that the initial state is i. The matrices formed of the above quantities, $\mathbf{F}(t,x) = [F_{ij}(t,x)]$ and $\mathbf{G}(t,x) = [G_{ij}(t,x)]$, can be described in double transform domain as (Kulkarni et al. (1986))

$$\mathbf{F}^{**}(s,v) = \frac{1}{v}(s\mathbf{I} + v\mathbf{R} - \mathbf{Q})^{-1} \quad (3)$$

$$\mathbf{G}^{**}(s,v) = \frac{1}{s}(s\mathbf{I} + v\mathbf{R} - \mathbf{Q})^{-1}\mathbf{R} \quad (4)$$

where $f^*(s) = \int_{t=0}^{\infty} f(t)e^{-st}dt$ denotes the Laplace transform which is applied both according to time ($t \to s$) and

accumulated reward ($x \to v$) and $\mathbf{I}$ denotes the identity matrix. In theory, both the accumulated reward and the completion time can be analyzed based on (3) and (4) by numerical or symbolic inverse Laplace transformation. In practice, this approach works only for small and/or specially structured MRMs.

For larger MRMs, as it was proposed in Telek and Rácz (1999), the moments of the quantity defined in (1) can be computed efficiently. To this purpose define the moments of the state dependent accumulated reward as

$$K_{ij}^{(n)}(t) = \int_{x=0}^{\infty} x^n dF_{ij}(t,x),\ n = 0,1,2,... \qquad (5)$$

such that $K_{ij}^{(n)}(t)$ is the nth moment of the accumulated reward after t time units multiplied by the probability that the state at time t is j supposing that the initial state is i. Note that if $n = 0$ then (5) corresponds simply to the transient probabilities of the underlying CTMC. Applying (5), the mean of the accumulated reward after t time units given that the initial state is i is given by $\sum_j K_{ij}^{(1)}(t)$. It was shown in Telek et al. (2004) that the matrices $\mathbf{K}^{(n)}(t) = [K_{ij}^{(n)}(t)], n = 0,1,2,...$, satisfy the differential equations

$$\frac{d\mathbf{K}^{(n)}(t)}{dt} = n\mathbf{K}^{(n-1)}(t)\mathbf{R} + \mathbf{K}^{(n)}(t)\mathbf{Q}. \qquad (6)$$

In Telek and Rácz (1999) the authors proposed randomization based methods for the calculation of $\mathbf{K}^{(n)}(t)$ for $n = 1,...,N$ whose time complexity is roughly $\sum_{i=0}^{N}(i+1)$ times higher than the complexity of the transient analysis of the CTMC underlying the MRM. The space complexity is instead $N+1$ higher than for the transient analysis of the CTMC. The same authors proposed a method for the calculation of the moments of the completion time as well but this method is efficient only if there are not states with zero reward rate in the model.

2.2 Discrete Markov Reward Model

In the discrete case a discrete time Markov chain (DTMC) modulates the accumulation of reward and a discrete amount of reward is gained in each time slot (see, e.g., Mullubhatla and Pattipati (2000)). Let $\mathbf{P}$ denote the transition probability matrix of the DTMC and $r_i, 0 \le r_i \le c$, the integer reward rate associated with state i where c is the maximum reward per step. A sojourn of length l in state i provides lr_i reward. Denoting by X_n the state of the chain after the nth transition and by Z_n the reward accumulated in the first n steps, we have $Z_n = \sum_{i=0}^{n-1} r_{X_i}$ assuming that $Z_0 = 0$. Further, let $S_i, 0 \le i \le c$, denote the set of states with reward rate i and define the matrices $\mathbf{P}^{(i)}, 0 \le i \le c$, with entries: $\mathbf{P}_{kl}^{(i)} = \mathbf{P}_{kl}$ if $k \in S_i$ and $\mathbf{P}_{kl}^{(i)} = 0$ otherwise.

Our aim is to characterise the joint probability of the accumulated reward and the background state defined for $n = 0,1,2,...$ and $k = 0,1,2,...$ as $F_{ij}(n,k) = Pr\{Z_n = k, X_n = j | X_0 = i\}$. It is easy to see that the following recursion holds for $n = 1,2,3,...$ and $k = 0,1,2,...$

$$F_{ij}(n,k) = \sum_{m=0}^{\min(k,c)} \sum_{l \in S_m} F_{il}(n-1,k-m)\mathbf{P}_{lj}$$

which, introducing the matrix notation $\mathbf{F}(n,k) = [F_{ij}(n,k)]$, becomes

$$\mathbf{F}(n,k) = \sum_{m=0}^{\min(k,c)} \mathbf{F}(n-1,k-m)\mathbf{P}^{(m)} \qquad (7)$$

The following theorem provides the discrete counterpart of (3).

Theorem *1.* The double z-transform of $\mathbf{F}(n,k)$ is given by

$$\mathbf{F}^{**}(z_1,z_2) = \left(\mathbf{I} - z_1 \sum_{m=0}^{c} z_2^m \mathbf{P}^{(m)}\right)^{-1} \qquad (8)$$

Proof. Multiplying the left hand side of (7) by $z_1^n z_2^k$ and summing for $n = 1,2,\ldots$ and $k = 0,1,\ldots$ gives

$$\sum_{n=1}^{\infty}\sum_{k=0}^{\infty} z_1^n z_2^k \mathbf{F}(n,k) = \mathbf{F}^{**}(z_1,z_2) - \sum_{k=0}^{\infty} z_2^k \mathbf{F}(0,k) = \mathbf{F}^{**}(z_1,z_2) - \mathbf{I} \qquad (9)$$

because $\mathbf{F}(0,0) = \mathbf{I}$ and $\mathbf{F}(0,k) = \mathbf{0}$ for $k = 1,2,3,....$ By the same operation on the right hand side of (7) we have

$$\sum_{n=1}^{\infty}\sum_{k=0}^{\infty} z_1^n z_2^k \sum_{m=0}^{\min(k,c)} \mathbf{F}(n-1,k-m)\mathbf{P}^{(m)} =$$
$$z_1 \sum_{m=0}^{c} z_2^m \sum_{n=1}^{\infty}\sum_{k=m}^{\infty} z_1^{n-1} z_2^{k-m} \mathbf{F}(n-1,k-m)\mathbf{P}^{(m)} =$$
$$z_1 \sum_{m=0}^{c} z_2^m \mathbf{F}^{**}(z_1,z_2)\mathbf{P}^{(m)}\ . \qquad (10)$$

From the fact that (9) equals (10), simple rearrangements provide the theorem. □

The joint distribution of the completion time and the state is defined for $n = 1,2,3,...$ and $k = 1,2,3,...$ as

$$G_{ij}(n,k) = Pr\{Z_n \ge k, Z_{n-1} < k, X_n = j | X_0 = i\}.$$

It is easy to see that the following relation holds for $n = 1,2,3,...$ and $k = 1,2,3,...$

$$G_{ij}(n,k) = \sum_{m=1}^{\min(k,c)} \sum_{h=m}^{c} \sum_{l \in S_h} F_{il}(n-1,k-m)\mathbf{P}_{lj}$$

which, by the matrix notation $\mathbf{G}(n,k) = [G_{ij}(n,k)]$, becomes

$$\mathbf{G}(n,k) = \sum_{m=1}^{\min(k,c)} \sum_{h=m}^{c} \mathbf{F}(n-1,k-m)\mathbf{P}^{(h)}\ . \qquad (11)$$

The following theorem is the discrete counterpart of (4).

Theorem *2.* The double z-transform of $\mathbf{G}(n,k)$ is

$$\mathbf{G}^{**}(z_1,z_2) = \sum_{m=1}^{c}\sum_{h=m}^{c} z_1 z_2^m \left(\mathbf{I} - z_1 \sum_{m=0}^{c} z_2^m \mathbf{P}^{(m)}\right)^{-1} \mathbf{P}^{(h)} \qquad (12)$$

Proof. Multiplying the left hand side of (11) by $z_1^n z_2^k$ and summing for $n = 1,2,\ldots$ and $k = 0,1,\ldots$ gives

$$\sum_{n=1}^{\infty}\sum_{k=0}^{\infty} z_1^n z_2^k \mathbf{G}(n,k) = \mathbf{G}^{**}(z_1,z_2) - \sum_{k=0}^{\infty} \mathbf{G}(0,k) = \mathbf{G}^{**}(z_1,z_2) - \mathbf{I}$$

because $\mathbf{G}(0,0) = \mathbf{I}$ and $\mathbf{G}(0,k) = \mathbf{0}$ for $k = 1, 2, 3, ...$ By the same operations on the right hand side of (11) we have

$$\sum_{n=1}^{\infty}\sum_{k=0}^{\infty} z_1^n z_2^k \sum_{m=1}^{\min(k,c)}\sum_{h=m}^{c} \mathbf{F}(n-1,k-m)\mathbf{P}^{(h)} = \sum_{m=1}^{c}\sum_{h=m}^{c} z_1 z_2^m \mathbf{F}^{**}(z_1,z_2)\mathbf{P}^{(h)}$$

from which by applying (8) the theorem follows. □

As in the continuous case, in theory, the accumulated reward and the completion time can be analysed based on (8) and (12) by applying double inverse z-transform. This approach is viable for small or highly structured DTMCs. The following theorem, which is the discrete counterpart of (6), provides an efficient approach for the calculation of the moments of the accumulated reward defined as $\mathbf{K}^{(l)}(n) = \sum_{k=0}^{\infty} k^l \mathbf{F}(n,k), k = 0, 1, 2, ...$

Theorem *3.* The moments of the accumulated reward, $\mathbf{K}^{(l)}(n), k = 0, 1, 2, ...,$ satisfy the following recursive relation

$$\mathbf{K}^{(l)}(n) = \sum_{i=0}^{l} \binom{l}{i} \mathbf{K}^{(l-i)}(n-1) \sum_{m=0}^{c} m^i \mathbf{P}^{(m)}. \qquad (13)$$

Proof. The moments $\mathbf{K}^{(l)}(n)$ can be derived starting from (7) by multiplying both sides by k^l and summing up for $k = 0, 1, 2, ...$ which leads to

$$\mathbf{K}^{(l)}(n) = \sum_{k=0}^{\infty} k^l \sum_{m=0}^{\min(k,c)} \mathbf{F}(n-1,k-m)\mathbf{P}^{(m)}$$

from which by changing the order of the summation and applying $k^l = \sum_{i=0}^{l} \binom{l}{i}(k-m)^{l-i}m^i$ we get $\mathbf{K}^{(l)}(n) =$

$$\sum_{m=0}^{c}\sum_{k=m}^{\infty}\sum_{i=0}^{l} \binom{l}{i}(k-m)^{l-i}m^i \mathbf{F}(n-1,k-m)\mathbf{P}^{(m)} = \sum_{m=0}^{c}\sum_{i=0}^{l} \binom{l}{i} \mathbf{K}^{(l-i)}(n-1) m^i \mathbf{P}^{(m)}$$

which yields the expression given in (13). □

The expression given in Theorem 3 directly provides a procedure for the calculation of the moments. The time complexity of calculating $\mathbf{K}^{(l)}(n)$ for $l = 1, ..., N$ is roughly $\sum_{i=0}^{N}(i+1)$ times the complexity of the transient analysis of the underlying DTMC. The space complexity is instead $N+1$ higher than for the transient analysis of the DTMC.

3. ANALYSIS OF MACHINES BY MARKOV REWARD MODELS

Due to the generality of the proposed approach, several system architectures and machine models can be analyzed within the framework described in the previous sections. However, due to space limitations, in this paper we report the detailed analysis and results only for two system architectures, including serial-parallel systems with limited repair capacity and buffered two-machine lines with degrading machines. Such complex systems have never been analyzed previously in the literature.

3.1 Parallel Machines with Limited Repair Capacity

Consider a manufacturing system or a portion of it composed by N identical parallel machines, each one characterized by failure rate λ, repair rate μ and attended by $K \leq N$ repairmen. The associated continuous MRM has the following infinitesimal generator (where, due to space limitations, we omit the diagonal entries)

$$\mathbf{Q} = \begin{vmatrix} \bullet & N\lambda & & & & & \\ \mu & \bullet & (N-1)\lambda & & & & \\ & 2\mu & \bullet & (N-2)\lambda & & & \\ & & & \ddots & & & \\ & & & & K\mu & \bullet & \lambda \\ & & & & & K\mu & \bullet \end{vmatrix}$$

and the matrix composed of the reward rates is

$$\mathbf{R} = \begin{vmatrix} N & & & & \\ & N-1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 0 \end{vmatrix}$$

The matrices $\mathbf{Q}$ and $\mathbf{R}$ of this simple model are such that the expressions given in (3) and (4) can be computed in a symbolic manner for any value of N and K. This gives the possibility of analysing the production as function of λ and μ. Based on the properties of the Laplace transform, the Laplace transform of the moments of the accumulated reward (i.e., the Laplace transform of $\mathbf{K}^{(n)}(t)$) can be obtained from $\mathbf{F}^{**}(s,v)$ as

$$\mathbf{K}^{(n)*}(s) = \int_{t=0}^{\infty} e^{-st}\mathbf{K}^{(n)}(t)dt = \lim_{v\to 0}(-1)^n \frac{d^n v\mathbf{F}^{**}(s,v)}{dv^n}$$

In general it is possible to obtain the inverse Laplace transform of the entries of $\mathbf{K}^{(n)*}(s)$ only if the numerical values of λ, μ, N and K are known. For example, with $\lambda = 1, \mu = 2, N = 3, K = 2$ the mean accumulated reward at time t given that the system is started with all machines up (obtained by summing up the entries of the first row of $\mathbf{K}^{(1)*}(s)$ and applying inverse Laplace transform) is

$$0.3709 + 1.963t + 0.001644e^{-8.372t} - 0.03e^{-5.t} - 0.3425e^{-2.627t}$$

while the second moment (obtained the same way from $\mathbf{K}^{(2)*}(s)$) is

$$-0.2429 + 2.0009t + 3.8558t^2 + 0.004623e^{-8.372t}t - 0.084e^{-5.t}t$$
$$-0.8434e^{-2.627t}t + 0.003189e^{-8.372t} + 0.1776e^{-5.t} + 0.06217e^{-2.627t}.$$

In Figure 1 and 2 we have depicted, respectively, the mean and the index of dispersion of the production as function of time. While the cases $K = 2$ and $K = 3$ are almost identical for what concerns the mean, they somewhat differ for the variability of the production.

In general it is not possible to obtain expressions for the moments of the production as function of time and the parameters of the model (λ, μ, N, K) but a symbolic expression for the asymptotic behaviour of these functions can be obtained. This can be done as follows. The non-vanishing terms of the mean are of the form $a_{10} + a_{11}t$ while for the second moment it is $a_{20} + a_{21}t + a_{22}t^2$. Denoting by $m_1(t)$ and $m_2(t)$ the mean and the second moment of the accumulated reward at time t and by $m_1^*(s)$ and $m_2^*(s)$ their Laplace transforms (which can be easily obtained by appropriate summations of the entries

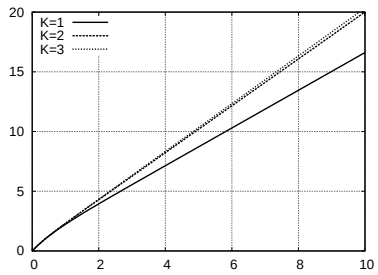

Fig. 1. Mean accumulated reward as function of time for identical machines with limited repair capacity with $\lambda = 1, \mu = 2, N = 3$ for different values of K

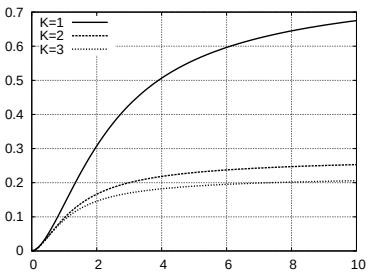

Fig. 2. Index of dispersion of reward as function of time for identical machines with limited repair capacity with $\lambda = 1, \mu = 2, N = 3$ for different values of K

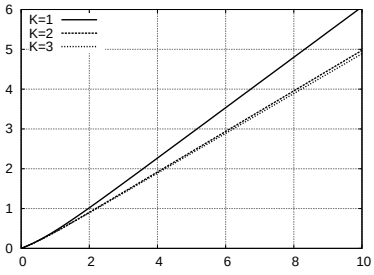

Fig. 3. Mean completion time as function of the required quantity for identical machines with limited repair capacity with $\lambda = 1, \mu = 2, N = 3$ for different values of K

of $\mathbf{K}^{(1)*}(s)$ and $\mathbf{K}^{(2)*}(s)$), the coefficients describing the non-vanishing terms can be obtained as

$$a_{11} = \lim_{s\to 0} s^2 m_1^*(s), \quad a_{10} = \lim_{s\to 0} s\left(m_1^*(s) - \frac{a_{11}}{s^2}\right)$$
$$a_{22} = \frac{1}{2}\lim_{s\to 0} s^3 m_2^*(s), \quad a_{21} = \lim_{s\to 0} s^2\left(m_2^*(s) - \frac{2a_{22}}{s^3}\right)$$
$$a_{20} = \lim_{s\to 0} s\left(m_2^*(s) - \frac{2a_{22}}{s^3} - \frac{a_{21}}{s^2}\right).$$

Having obtained the above coefficients, the limiting value of the index of dispersion is $I = (a_{21} - 2a_{11}a_{10})/a_{11}$. The above procedure provides symbolic expressions for any value of N and K. For example, with $N = 3$ we have that the limiting values of the index of dispersion is

$$\frac{2\lambda\left(72\lambda^6 + 216\mu\lambda^5 + 324\mu^2\lambda^4 + 240\mu^3\lambda^3 + 94\mu^4\lambda^2 + 16\mu^5\lambda + \mu^6\right)}{\left(2\lambda^2 + 2\mu\lambda + \mu^2\right)\left(6\lambda^3 + 6\mu\lambda^2 + 3\mu^2\lambda + \mu^3\right)^2},$$
$$\frac{\lambda\left(18\lambda^6 + 108\mu\lambda^5 + 234\mu^2\lambda^4 + 249\mu^3\lambda^3 + 146\mu^4\lambda^2 + 48\mu^5\lambda + 8\mu^6\right)}{(\lambda+\mu)^2\left(3\lambda^3 + 6\mu\lambda^2 + 6\mu^2\lambda + 2\mu^3\right)^2},$$

and $\frac{2\lambda}{(\lambda+\mu)^2}$ for $K = 1, 2$ and 3, respectively, where the simplicity of the last expression is due to the fact that with $K = 3$ the model simplifies to the sum of three independent machines.

Now we turn our attention to the analysis of the completion time. The same steps we have applied to the double transform of the accumulated reward, $\mathbf{F}^{**}(s, v)$, can be applied to the double transform of the completion time, $\mathbf{G}^{**}(s, v)$. We report briefly the results that can be obtained. With $\lambda = 1, \mu = 2, N = 3, K = 2$, the mean of the completion time for x units of product is

$$-0.1183 + 0.5092x - 0.006899e^{-5.302x} + 0.1252e^{-1.697x}$$

while its second moment is

$$-0.05676 - 0.04862x + 0.2593x^2 - 0.01401e^{-5.302x}x + 0.1398e^{-1.697x}x - 0.005304e^{-5.302x} + 0.06207e^{-1.697x}.$$

Figure 3 and 4 provide, respectively, the mean and the index of dispersion of the completion time as function of the required quantity of the product. The same phenomenon can be observed as for the accumulated reward: the mean completion time is almost identical for the cases $K = 2$ and $K = 3$ but some difference occurs between these two cases for what concerns the variability of the completion time.

Symbolic expressions can be obtained for the limiting index of dispersion of the completion time and with $N = 3$ we have

$$\frac{2\lambda\left(72\lambda^6 + 216\mu\lambda^5 + 324\mu^2\lambda^4 + 240\mu^3\lambda^3 + 94\mu^4\lambda^2 + 16\mu^5\lambda + \mu^6\right)}{3\mu\left(2\lambda^2 + 2\mu\lambda + \mu^2\right)^2\left(6\lambda^3 + 6\mu\lambda^2 + 3\mu^2\lambda + \mu^3\right)},$$
$$\frac{\lambda\left(18\lambda^6 + 108\mu\lambda^5 + 234\mu^2\lambda^4 + 249\mu^3\lambda^3 + 146\mu^4\lambda^2 + 48\mu^5\lambda + 8\mu^6\right)}{6\mu(\lambda+\mu)^4\left(3\lambda^3 + 6\mu\lambda^2 + 6\mu^2\lambda + 2\mu^3\right)},$$

and $\frac{2\lambda}{3\mu(\lambda+\mu)}$ for $K = 1, 2$ and 3, respectively.

The procedure we have illustrated here for parallel machines with limited repair capacity with continuous time and continuous product can be performed with discrete time and discrete product as well by applying basic properties of z-transform.

3.2 Discrete Time Machine-Buffer-Machine Block

As second example, we analyse the output of the second machine of a machine-buffer-machine block. By assuming discrete time and production of discrete quantities, it is straightforward to build a DTMC accounting for both the states of the two machines and the number of parts present in the buffer. We assume that the two machines are identical and operate according to a degradation/reparation scheme. The underlying DTMC has the following transition probability matrix:

$$\mathbf{P} = \begin{vmatrix} 1-p & p & & & & & & \\ & 1-p & p & & & & & \\ & & \ddots & & & & & \\ & & & 1-p & p & & & \\ & & & & 1-q & q & & \\ & & & & & \ddots & & \\ q & & & & & & & 1-q \end{vmatrix}$$

with N_u up-states and N_d down-states. We assume that the quantity produced in the up-states is $N_u, N_u - 1, N_u - 2, ..., 2, 1$, i.e., the productivity of the machines is degrading as it is passing through the up-states. Naturally, there is no production in the down-states. We assume that in each time slot the first machine is putting parts in the buffer (if it is in an operational state) and then the second machine takes away parts from the buffer (if it is in an operational state) producing the final outcome of the machine-buffer-machine block. When the buffer is full the first machine is blocked and does not change state and the same happens to the second machine when it does not find parts in the buffer. When the buffer is close to being full (empty) then the first (second) machine is producing as much as allowed by the current buffer content.

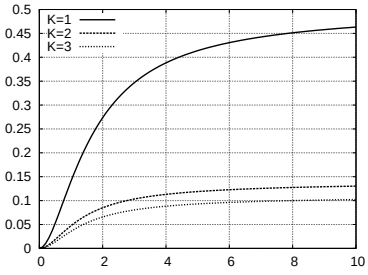

Fig. 4. Index of dispersion of the completion time as function of the required quantity of the for identical machines with limited repair capacity with $\lambda = 1, \mu = 2, N = 3$

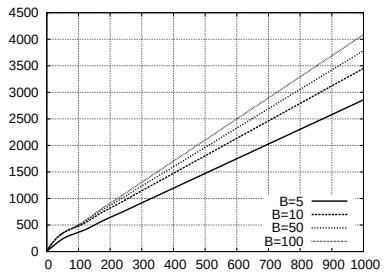

Fig. 5. Mean number of produced parts as function of time for the machine-buffer-machine block with different buffer sizes

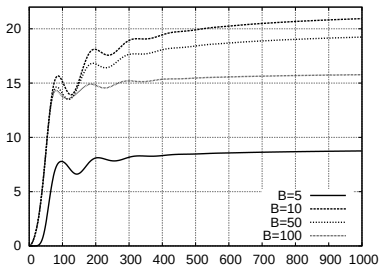

Fig. 6. Index of dispersion of the number of produced parts as function of time for the machine-buffer-machine block with different buffer sizes

We calculated, having empty buffer and fully operational machines as initial state, the mean and the index of dispersion of the number of parts produced by the second machine applying the recursion given in (13) for $N_u = 10, N_d = 5, p = 0.1, q = 0.5$ varying the size of the buffer B. The results are depicted in Figure 5 and 6. As expected, the mean amount of production is increasing as the buffer is enlarged. The variability of the production instead is low for low buffer size ($B = 5$), it has a sharp increase for medium buffer size ($B = 10$) and then it decreases as the buffer is enlarged ($B = 50, 100$). The oscillatory behaviour of the variability is due to the fact that both the time to complete failure and the time to repair of the machines are of low variance. The size of the state space is $(N_u + N_d)^2(B+1)$ which is 22725 in our case for $B = 100$. Note that the method is applicable for much larger state spaces as well.

4. CONCLUSIONS

In this paper we proposed a methodology for the analysis of the moments of the cumulated output and the completion time of unreliable machines. The framework is based on Markov reward models which allows for modeling general failure/repair mechanisms and failure/repair durations with phase type distributions. The approach is particularly promising for the analysis of the variability of the production.

REFERENCES

Carrascosa, M. (1995). Variance of the output in a deterministic two-machine line. *M.S. Thesis, Massachusetts Institute of Technology, Cambridge MA*.

Colledani, M., Ekvall, M., Lundholm, T., Moriggi, P., Polato, A., and Tolio, T. (2010). Analytical methods to support continuous improvements at scania. *International Journal of Production Research*, 48, 1913–1945.

Gershwin, S.B. (1993). Variance of output of a tandem production system. In I.F.A. R. D. Onvural (ed.), *Queueing Networks with Finite Capacity*, 291–304. Elsevier Science Publishers, Amsterdam.

He, X.F., S.Wu, and Li, Q.L. (2007). Production variability of production lines. *International Journal of Production Economics*, 107, 78–87.

Hendrics, K. (1992). The output processes of serial production lines of exponential machines with finite buffers. *Operations Research*, 40, 1139–1147.

Inman, R. (1999). Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. *Production Operations Management*, 8, 409–432.

Kulkarni, V., Nicola, V., and Trivedi, K. (1986). On modeling the performance and reliability of multi-mode computer systems. *The Journal of Systems and Software*, 6, 175–183.

Li, J. and Meerkov, S. (2000). Production variability in manufacturing systems: Bernoulli reliability case. *Annals of Operations Research*, 93, 299–324.

Miltenburg, G. (1987). Variance of the number of units produced on a transfer line with buffer inventories during a period of length t. *Naval Research Logistics*, 34, 811–822.

Mullubhatla, R. and Pattipati, K. (2000). Discrete-time Markov reward models of automated manufacturing systems with multiple part types and random rewards. *IEEE Tr. on Robotics and Automation*, 16(5), 553–566.

Neuts, M. (1981). *Matrix Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore.

Ou, J. and Gershwin, S. (1989). The variance of the lead time of a two machine transfer line with a finite buffer. *Laboratory for Manufacturing and Productivity, MIT, Techical report*, LMP-90-028.

Tan, B. (1997). Variance of the throughput of an n-station production line with no intermediate buffers and time dependent failures. *European Journal of Operational Research*, 101, 560–576.

Tan, B. (1999a). Asymptotic variance rate of the output of a transfer line with no buffer storage and cycle dependent failures. *Mathematical and Computer Modelling*, 29, 97–112.

Tan, B. (1999b). Variance of the output as a function of time: Production line dynamics. *European Journal of Operational Research*, 177, 470–484.

Tan, B. (2000). Asymptotic variance rate of the output in production lines with finite buffers. *Annals of Operations Research*, 93, 385–403.

Telek, M., Horváth, A., and Horváth, G. (2004). Analysis of inhomogeneous Markov reward models. *Linear Algebra and Its Applications*, 386, 383–405.

Telek, M. and Rácz, S. (1999). Numerical analysis of large Markovian reward models. *Performance Evaluation*, 36&37, 95–114.