A METHODOLOGY FOR THE DESIGN OF SIMULATION CAMPAIGNS BASED ON POPULATION VARIANCE CHARACTERIZATION

Andrea Grassi [a], Elisa Gebennini [b], Giuseppe Perrica [c], Cesare Fantuzzi [d], Bianca Rimini [e]

Dipartimento di Scienze e Metodi dell'Ingegneria, Università degli Studi di Modena e Reggio Emilia
Via Amendola 2 – Pad. Morselli, 42122 Reggio Emilia, Italy.

[a]andrea.grassi@unimore.it, [b]elisa.gebennini@unimore.it, [c]giuseppe.perrica@unimore.it,
[d]cesare.fantuzzi@unimore.it, [e]bianca.rimini@unimore.it,

ABSTRACT

Manufacturing industrial systems are complex systems whose performance is characterized by interactions among different parts of the system as well as by stochastic phenomena affecting the operation of the parts themselves.

A key aspect in studying a complex system is the ability to model its evolution over time and, as a consequence, to identify, from a statistical point of view, the trend of the performance measures (i.e. productivity) over time. Discrete event system simulation (DESS) is certainly the widespread technique adopted to this aim.

In this paper, a methodology to characterize the trend of the variance of the population for a flow-line production system is developed. The knowledge of the relation between the variance of the population and the system run time allows the analyst to better design simulation campaigns and define warm-up period. Moreover, this result is also useful when in-field tests have to be designed to certify performances of a newly deployed system.

Keywords: simulation campaign design, transitory analysis, variance estimation, flow-line systems

1. INTRODUCTION

Manufacturing industrial systems are complex systems whose behavior is characterized by the operation of several parts (i.e. machines performing processes on products) and the interactions among the parts themselves. Frequently, stochastic phenomena affect the operational state of the machines, and, as a consequence, disruptions of production flows are stochastically propagated all over the manufacturing system.

As an example, if we refer to a production line, operative conditions of machines positioned along the line is determined by failures and repairs of the process each very machine is executing on products, while disruptive interactions between the machines are due to interruptions of the production flow.

The optimal design of a manufacturing system is related to the definition of performance targets on machines (e.g. nominal capacity, reliability parameters, etc.) and of the structure of the system (e.g. buffer location and size), so as to reach a desired performance of the whole system.

Since the behavior of the system is influenced by stochastic phenomena, performance parameters have to be computed by means of probabilistic models, aiming at providing their steady state value, or estimated by adopting methodologies able to reproduce the evolution of the system over a limited run time. Discrete event system simulation (DESS) is the widely adopted approach for that latter case.

Adopting such an approach to asses system performances, and considering that it is impossible to execute a unique infinite run, we face with the need to execute several simulation runs reproducing different histories of the system, then running the same system with different sequences of random events. Hence, the performance measured in each run represents a different individual withdrew from the population of the individuals constituting all the possible values of the system performance measured at the specified run time. The set of the individuals is then a sample withdrew from the population.

Hence, once the sample is obtained, statistical analysis methodologies have to be used to obtain an estimation of the system performance in terms of confidence interval (Law, 1983). The wider the confidence interval, the lower the precision in assessing the true value of the performance measure is.

The variance of the sample is a key factor directly influencing the amplitude of the confidence interval (Montgomery and Runger, 2007). Moreover, the variance of the population is guessed with respect to the sample variance, thus producing a further extension of the confidence interval.

If we were able to directly compute the variance of the population with respect to some characteristics of the system, we would be able to restrict the confidence interval thus obtaining better estimation. Furthermore, knowing the trend of the population variance over run time will allow us to *a priori* define the right combination of simulation run length and number of runs to execute given a desired confidence interval. This aspect becomes very important when in-filed tests have to be deployed to certifies the performance of the system. In such a situation, system runs are executed in real time, thus there is an implicit need to reduce to a bare minimum the total time required for the test.

This paper develops a methodology to determine the trend of the population variance over run time for flow-line structured manufacturing systems. The methodology is based on some analytical considerations and is supported by experimental evidence. It is also shown how this information can be used to design simulation campaigns or in-field tests guaranteeing a specified confidence interval of the performance measure estimation.

2. METHODOLOGY DESCRIPTION

Provided that the behavior of the system is affected by stochastic phenomena, a generic performance measure is described by means of a random variable (i.e. characterized by a mean value, a variance, and a probability density function) for any finite time. All of the mean value, variance and probability density function of the performance measure vary over time. Nevertheless, we can state that, if the system is ergodic, as time tends to infinite the mean value tends to the steady state asymptotic value of the performance measure. Consequently, the variance tends to zero and the random variable degenerates to a deterministic variable.

A number of studies were carried out in years to analytically derive trends of mean value and variance over time of performance measures of some canonical systems (Kelton and Law, 1985; Li and Meerkov, 2000; Tan, 1999).

Starting from the simplest case of a single machine operating with a Bernoulli like production process (Li and Meerkov, 2000) the mean and the variance of the number of units produced after t time steps $N(t)$ can be expressed as

$$\mathrm{E}[N(t)] = t\,p \tag{1}$$

$$\mathrm{Var}[N(t)] = t\,p\,(1-p) \tag{2}$$

Defining the productivity of the machine as

$$P(t) = \frac{1}{t} N(t)\,, \tag{3}$$

we can state that the mean productivity at t is

$$\mathrm{E}[P(t)] = \frac{1}{t}\mathrm{E}[N(t)]\,, \tag{4}$$

while its variance is

$$\mathrm{Var}[P(t)] = \frac{1}{t^2}\mathrm{Var}[\mathrm{N(t)}]\,. \tag{5}$$

Hence, by substituting eq. 1 and 2 in eq. 4 and 5 we can obtain

$$\mathrm{E}[P(t)] = p\,, \tag{6}$$

$$\mathrm{Var}[P(t)] = \frac{p\,(1-p)}{t}\,. \tag{7}$$

Let's now consider the case of a machine characterized by a deterministic production rate but affected by failure and repair phenomena represented by continuous time Markovian processes. Tan (1999) determined the number of parts produced at time t, $N(t)$, and its variance. Considering the case in which the machine has a probability $r/p+r$ to be operational at time zero, such a variance is equal to

$$\begin{aligned}\mathrm{Var}[N(t)] = {} & \frac{p\,r\,(2-p-r)}{(p+r)^3}\,t + \\ & + \frac{2\,p\,r\,(1-p-r)}{(p+r)^4}\left((1-p-r)^t - 1\right),\end{aligned} \tag{8}$$

where p and r represent the failure rate and the repair rate, respectively.

By substituting eq. 8 in eq. 5 we obtain

$$\begin{aligned}\mathrm{Var}[P(t)] = {} & \frac{p\,r\,(2-p-r)}{(p+r)^3}\frac{1}{t} + \\ & + \frac{2\,p\,r\,(1-p-r)}{(p+r)^4}\frac{\left((1-p-r)^t - 1\right)}{t^2}.\end{aligned} \tag{9}$$

Hence, we can observe that the variance of the productivity of a machine has the following form

$$\mathrm{Var}[P(t)] = \frac{K}{t} + \circ(t)\,, \tag{10}$$

where K is a parameter to be determined. Having an exact mathematical solution for K allows us to exactly know the variance of the population, otherwise K could also be determined in an experimental way by fitting the variance of a large sample with the relation

$$\overline{\mathrm{Var}}[P(t)] = \frac{K}{t}\,. \tag{11}$$

Experimental observations, realized by means of simulation campaigns, showed that eq. 10 represents a good approximation also when longer lines (i.e. lines with several machines decoupled with buffers) are considered.

Another important aspect addressed is the identification of the mass probability function of the performance measure considered, in our case $P(t)$, that changes over time. The shape of such a mass probability function is certainly not Gaussian since the value of the productivity is limited between 0 and a maximum value depending by the characteristics of the system and of the machines.

Since the productivity derives from the sum of the independent random variables representing the units produced over time, according to the central limit theorem, the mass probability function of $P(t)$ asymptotically tends to a Gaussian. Hence, from a

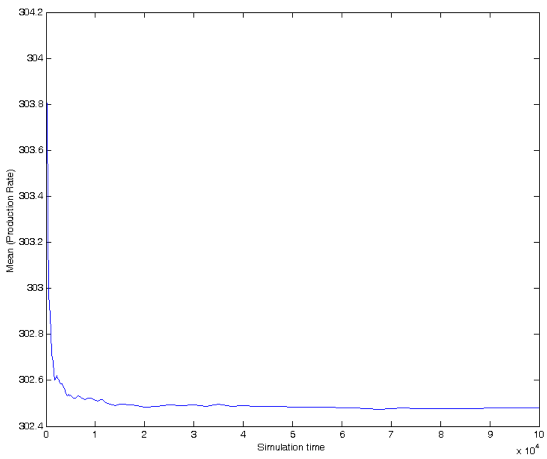

Figure 2: Mean trend.

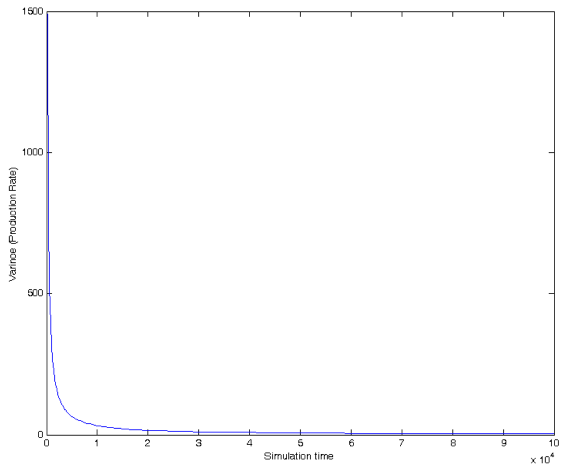

Figure 3: Variance trend.

certain time $t > t_0$ we can apply the inferential techniques explained in the next section.

3. ESTIMATION APPROACH

When dealing with simulation of stochastic processes, estimation of parameters of interest is provided in terms of confidence intervals. This approach states that the estimated range of value contains the true population parameter with a confidence of $(1-\alpha)\%$. That is, the method used to obtain this range, yields correct statements $(1-\alpha)\%$ of the time. The length of a confidence interval is a measure of precision of the estimation.

If $\overline{\vartheta}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100 \cdot (1-\alpha)\%$ confidence interval on ϑ is given by

$$\overline{\vartheta} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \vartheta \leq \overline{\vartheta} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \tag{12}$$

where n is the sample size, $1-\alpha$ is the confidence level and $z_{\alpha/2}$ is the upper $100 \cdot \alpha/2$ percentage point of the standard normal distribution. As it is possible to see in eq. 12, the larger the sample size, the narrower the interval is. Conversely, the larger the sample variability, the less the accuracy on the estimation is (Montgomery and Runger, 2007).

By substituting the approximate variance of eq. 11 in eq. 12 we obtain

$$\mathrm{E}[P(t)] - z_{\alpha/2}\sqrt{\frac{K}{t \cdot n}} \leq \vartheta \leq \mathrm{E}[P(t)] + z_{\alpha/2}\sqrt{\frac{K}{t \cdot n}}. \tag{13}$$

4. NUMERICAL EXAMPLE

An exemplificative case study has been carried out to show how the methodology presented in this paper can be used to organize and design simulation campaigns.

Let us consider a simple production line consisting of two machines decoupled by a finite buffer (Figure 1).

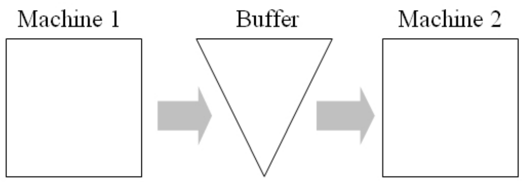

Figure 1: two-machine one-buffer line.

The input parameters of interest are as follows: production capacity (μ), failure rate (p) and repair rate (r) of each machine, and buffer size (N). The corresponding values adopted in the case study, derived from a real installation typical of the beverage/packaging field, are reported in Table 1 (where t.u. is for time unit).

Given the simulation model of such a system, a simulation campaign, consisting of 50000 runs of 10000000 time units [t.u.] length each, has been carried out. The performance measure of interest is the production rate $P(t)$ at any time instant t, i.e. the ratio between the number of produced items and the simulation time $t - t_0$. The production rate $P(t)$ has been captured at several time values over all the simulation length.

Table 1: Input Parameters

	Machine 1	Machine 2	Buffer
Capacity [itmes/t.u.]	400	400	-
Failure rate [1/t.u.]	0.01667	0.06667	-
Repair rate [1/t.u.]	0.083333	0.4	-
Size [items]	-	-	1500

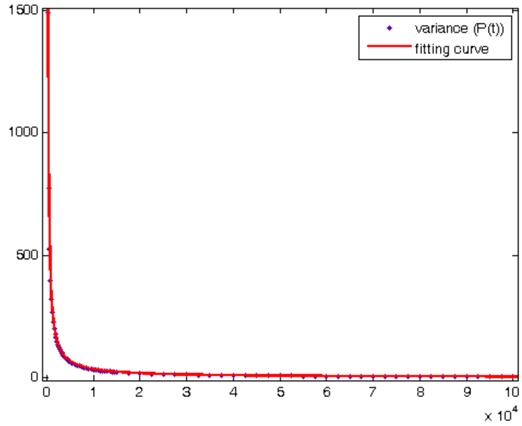

Figure 4: Fitting results.

At any capture time the sample mean $\mathrm{E}[P(t)]$ and variance $\mathrm{Var}[P(t)]$ were computed with respect to the simulated data. Figure 2 and Figure 3 depict the trend of $\mathrm{E}[P(t)]$ and $\mathrm{Var}[P(t)]$, respectively, against the simulation time *t*.

Note that the sample mean quickly approaches its steady state value. When analytical model for steady state performance computation are available, such the ones proposed in Gershwin (2002) and in Gebennini et al. (2009), the steady state value of the productivity can be conveniently *a priori* determined.

Focusing on the variance $\mathrm{Var}[P(t)]$, Figure 4 shows how the curve of $\mathrm{Var}[P(t)]$ can be properly fitted by the function reported in eq. 11. In this specific case, $K = 3 \cdot 10^5$.

In order to estimate the goodness of the fit, the R-square (i.e. the square of the multiple correlation coefficient and the coefficient of multiple determination) is computed. Specifically, its value is about 0.99, very close to 1 (i.e. the value corresponding to a perfect fit).

This result proves that the form of the approximated variance reported in eq. 11 holds also when complex lines are dealt with.

As can be seen, knowing *K* in advance for a specific system configuration allows to effectively represent the variance of the population. Practically, *K* can be determined whether by using specific analytical formula obtained in the literature, or by adopting heuristic methods (i.e. neural networks) to interpolate *K* values from a set of observations conducted with different values of system parameters.

Thus, we are provided with all the data necessary for computing the confidence interval (see eq. 13) in relation to a certain simulation time $t - t_0$ and a certain number *n* of simulation runs/repetitions. Figure 5 shows the surface representing the confidence interval half length (CI/2) for a significant range of simulation times and numbers of repetitions, i.e. the range [200,…,8000] t.u. for the simulation time and the range [1,…,30] for the number of runs. This is a useful result for dimensioning simulation campaigns by identifying isolines for any specified value of the confidence interval (Figure 6).

5. CONCLUSION

The paper presents a methodology for the design of simulation campaigns or in-field tests, specifically for identifying a trade-off between replication length and number of simulation runs to execute as a function of a specified estimation precision.

The method is based on the direct estimation of the population variance by means of a properly shaped fitting function. That shape is determined by some analytical analyses on the structure of the manufacturing system.

The trend of the population variance can be then expressed as a function of the system run time, and

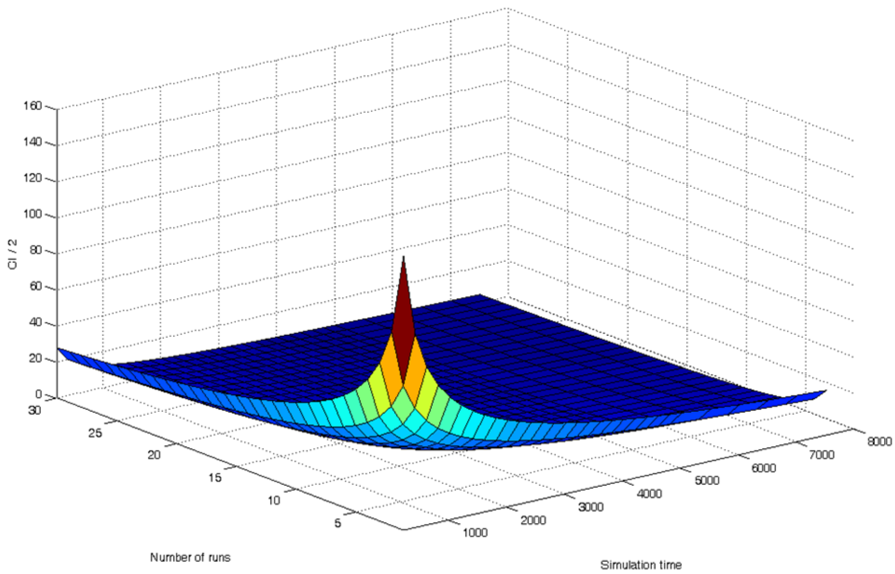

Figure 5: CI/2 surface.

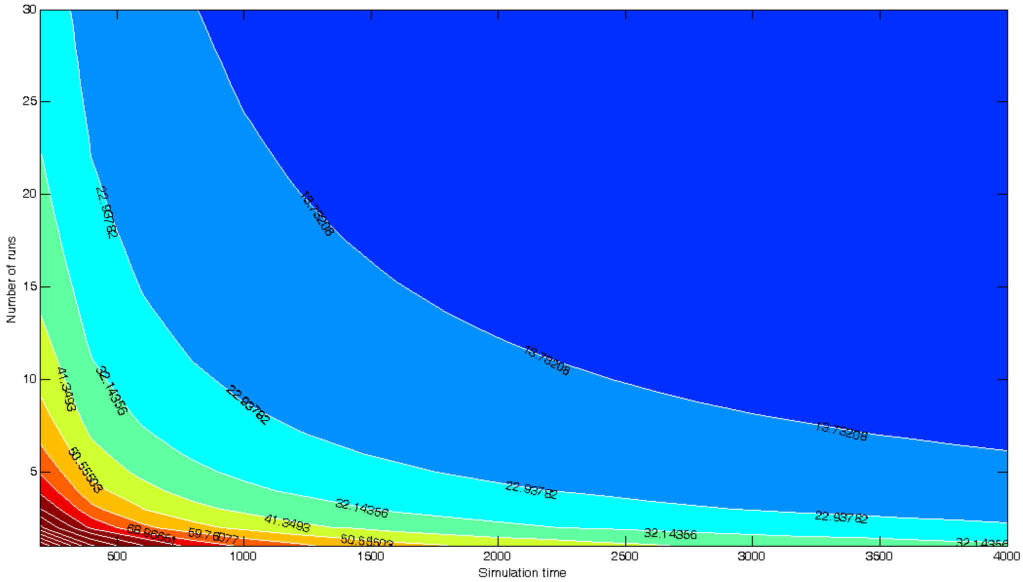

Figure 6: CI/2 contour plot.

confidence interval diagrams can be determined to find better trade-offs between the number of simulation to run and simulation length.

REFERENCES

Gebennini, E., Grassi, A., Fantuzzi, C., Gershwin, S.B., and Schick, I.C. (2009). On the introduction of a restart policy in the two-machine, one-buffer transfer line model. In *Proceedings of the 7th International Conference on Stochastic Models of Manufacturing and Service Operations*, 81–88. Ostuni, Italy.

Gershwin S.B., 2002. *Manufacturing Systems Engineering*, Massachusetts Institute of Technology. Second private printing, available via http://web.mit.edu/manuf-sys/www/gershwin.book.html

Kelton, W.D., and Law, A.M., 1985. The transient behaviour of the M/M/s queue, with implications for steady-state simulation, *Operations Research*, 33 (2), 378-396.

Law, A.M., 1983. Statistical analysis of simulation output data, *Operations Research*, 31(6), 983–1029.

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

Montgomery, D.C., and Runger, G.C., 2007. *Applied Statistics and Probability for Engineers*, 4th ed. USA: John Wiley & Sons.

Tan, B., 1999. Variance of the output as a function of time: Production line dynamics, *European Journal of Operational Research*, 117, 470-484.