

Production control in a failure-prone manufacturing network using discrete event simulation and automated response surface methodology

Seyed Mojtaba Sajadi · Mir Mehdi Seyed Esfahani · Kenneth Sörensen

Received: 30 January 2010 / Accepted: 23 June 2010 / Published online: 21 July 2010
© Springer-Verlag London Limited 2010

Abstract In this paper, a system consisting of a network of machines with random breakdown and repair times is considered. The machines in this system can be in one of four states: operational, in repair, starved, and blocked. Failure and repair times of the machines are exponentially distributed. Previous research on multi-machine failure-prone manufacturing systems (FPMS) has focused on systems consisting of machines in series or in parallel. This paper considers a *network* of machines with relationship constraints. Additionally, the system under study models work in process for multiple products, intermediate and final buffers and one type of final product. The demand rate for the final commodity is constant and unmet demand is either backlogged or lost. The objective of this control problem is to find the production rates and policies of the different machines so as to minimize the long run average inventory and backlog cost. The applied control policy is the hedging point policy that is determined by factors representing the level of buffer inventory for each machine. Obtaining analytical solutions is generally impossible for such complex systems. To simultaneously control the production rates of the machines we have therefore

developed a method based on a combination of stochastic optimal control theory, discrete event simulation, experimental design and automated response surface methodology (RSM). The application of an automated RSM for Network FPMS is another contribution of this paper. The model can be extended easily to systems with age-dependent failure rates, a preventive repair maintenance policy and non-exponentially distributed up and down times.

Keywords Failure-prone manufacturing system · Discrete event simulation · Experimental design · Automated response surface methodology

1 Introduction and literature review

Failure-prone manufacturing systems (FPMS) have been formulated as a stochastic optimal control problem by Older and Souri [1] and later motivated by the pioneering work of Kimemia [2] and Kimemia and Greshwen [3]. Based on Rishel's work [4], the authors showed that the optimal control policy of a FPMS has a special structure called the hedging point policy (HPP). In such a policy, an optimal inventory level is maintained when excess production capacity is available in order to hedge against future capacity shortage caused by machine failures. If the current inventory level exceeds the optimal inventory level, the machine should not produce at all; if it is less, it should produce at the maximum rate. If the inventory level is exactly equal to the optimal level, the machine should produce just enough to meet demand. Akella and Kumar [5] find the exact analytical solution for a simple FPMS with only one machine, one product and exponentially

S. M. Sajadi · M. M. Seyed Esfahani (✉)
Department of Industrial Engineering,
Amirkabir University of Technology, Tehran, Iran
e-mail: msesfahani@aut.ac.ir

S. M. Sajadi
e-mail: msajjadi@aut.ac.ir

K. Sörensen
University of Antwerp, Faculty of Applied Economics,
Antwerp, Belgium
e-mail: kenneth.sorensen@ua.ac.be

distributed failure and repair times. The objective of the proposed problem is the expected discounted cost. Bielecki and Kumar [6] present the optimal production policy for such a system so as to minimize the long run average expected cost incurred per unit of time. They show that the optimal inventory level can be zero under special conditions. Sharifnia [7] solves the single product FPMS with an arbitrary number of machine states (failure modes). Xie [8] considers discrete-time FPMS and obtains an analytical solution when the production capacity is exactly equal to twice the demand rate. Kenne et al. [9] present—for the first time—a combination of an analytical and a simulation approach. Their paper deals with stochastic demand and lot size, and age-dependent production and failure rate of the machine. Later a modified version of the HPP called age-dependent hedging point policy is presented by Kenne and Gharbi [10]. Experimental design is used to prove the significance of the control variables with respect to the dependent variable or the incurred cost. Mourania et al. [11] address the optimization of a continuous-flow model of a single-stage, single-product manufacturing system with constant demand and transportation delay from the machine to the inventory. Kenne et al. [12] provide optimality conditions for simple FPMS in which all unmet demand is lost when the safety stock is consumed during machine breakdown. Their new control policy for preventive maintenance and related age-dependent repair is called multi threshold levels HPP. All of the papers mentioned above present simple models of FPMS that have only one machine, different objective functions, continuous or discrete time and Markovian or Non-Markovian distribution functions for failure and repair times.

A second group of papers has focussed on extensions of these models, with FPMS that have multiple machines and/or products. For two machines and one product, Boukas and Young [9] extend the approach in which the failure rate of a machine depends on its age. They present a numerical scheme restricted to small size systems and show that the HPP remains valid. Presman et al. [14] consider a two-machine flow shop with the objective to minimize the long run average inventory/backlog and production cost. They develop a theory of dynamic programming in terms of the method. Ching [15] studies a discrete inventory model for FPMS in tandem (m series machine). In this model, one type of product is produced and demand is assumed to be Poisson distributed. The production time for one unit of product in each machine is exponentially distributed. A Markov Modulated Poisson Process is used

and the system is solved using a numerical approximation method. The system with m parallel non-identical machines producing n different part types is simulated by Gharbi and Kenne [16]. Further information about this work can be found in Gharbi and Kenne [17].

In all of the former multi-machine FPMS, only machines in series or in parallel are considered. In this paper a *network* of multiple non-identical machines with relationship constraints between them is presented. Each intermediate machine produces one kind of product (WIP) assembling and/or operating on units of previous work-in-process. Production is strictly forward-flowing and discrete. One machine produces the final commodity to meet the constant demand rate facing the system. The stochastic nature of the system is due to machines that are subject to random breakdowns and repairs. Time between failures (TBF) and time to repairs (TTR) are assumed to be exponentially distributed variates with different parameters. Whenever a breakdown occurs, corrective maintenance is performed during a random amount of time (TTR) to restore the machine to its operational mode. When a machine is operational, its demand is met from the safety stock of the intermediate buffers that precede it. Demand for the final commodity is met from the final buffer or backlogged. However, when the backlog amount exceeds a given amount, demand is lost. Intermediate WIP cannot be backlogged, and when demand for an intermediate product cannot be met, the machine that requires this product is starved.

In the existing literature two kind of objective functions are commonly found. Some authors use expected discounted cost (e.g. Akella and Kumar [5]) while others consider expected long run average inventory/backlog cost (e.g. Bielecki and Kumar [6]). The objective of the control problem in our research is to find the optimal production rates of all machines so as to minimize the long run average total inventory/backlog cost in all of the intermediate and final buffers. Production planning of FPMS is a complex stochastic control problem. The optimal control policy for such a system has been proved to be the solution of a set of coupled Hamilton–Jacobi–Bellman (HJB) equations [4] which are too complex to solve analytically. A possible way of coping with this difficulty is to develop heuristic methods based on a reduction of the size of the considered control problem. In the existing literature two different approaches can be commonly found [16]:

- The hedging point policy
- The singular perturbation method

For further details on these concepts we refer the readers to Boukas, and Hauri [18] and Kokotovic et al. [19]. In this paper we will first define the structure of the optimal control policy. Based on this structure, we will extend the production rates control model (HPP) presented in Kenne et al. [10] and Xiaolan Xie [8] in order to determine the control policy in a network FPMS including multiple non-identical machines with relationship constraints and discrete production flow. The resulting structure is described through a set of inventory levels (Input Factors). These inventory levels are used as inputs for the simulation model. For each combination of parameters the total cost incurred is obtained. By combining this discrete simulation model with statistical optimal control theory, experimental design and response surface methodology (RSM), an approximation of the optimal control policy and of the optimal values of the input factors is determined. The application of such an approach is motivated by the works of Gharbi and Kenne [17] and Kenne and Gharbi [12]. Stochastic objective functions with unknown variance, like the ones encountered in our problem, need a different approach in order to find good solutions. We have applied for the first time the framework of RSM for FPMS that is presented by Neddermeijer et al. [20], including restart procedures to improve it. The results of this method show that considerable improvement over the proposed setting in the existing RSM literatures can be obtained [21]. The remainder of the paper is organized as follows. In Section 2, we present the notation and the problem statement of the considered production planning problem for a network of FPMS. The properties of the objective function and the approximation of the optimal control policy for suitable values of the parameters are given in Section 3. The control approach and logic of the simulation model is described in Sections 4 and 5. In Section 6, the experimental design approach and automated response surface methodology are outlined. A numerical example is presented in Section 7 while Section 8 concludes the paper.

2 Notation and problem statement

Throughout this article, we use the following abbreviations, acronyms, and notations:

FPMS	Failure-prone manufacturing system
TBF	Time between failures
TTR	Time to repair

HPP	Hedging point policy
ARSM	Automated response surface methodology
CCD	Central composite design
M_i	Machine i
B_i	Buffer next to (following) machine i
m	Number of machines/products
d	Constant demand rate for final commodity (in number of units per time unit)
C_i^+	Inventory holding cost per unit of time for product i , $1 \leq i \leq m$
C_i^-	Penalty cost per unit of unmet demand for product i , $i = m$
μ_i	Mean time to breakdown of machine i
λ_i	Mean time to corrective maintenance of machine i
$u_i(t)$	Production rate of machine i , $1 \leq i \leq m$
$\mathbf{u}(t)$	Control policy vector
$\zeta_i(t)$	Stochastic capacity process of the machine i at time t
$x_i(t)$	Inventory level/backlog for buffer i at time t
K	Threshold of lost sales for final commodity
$J(t)$	Expected long run average cost (Inventory/Backlog) function
L_{ij}	Number of items of product i that is consumed by machine j to produce one unit of product j ($i < j$)
Z_i	Inventory threshold level parameter of buffer i
$\lambda_{\alpha\beta}^i$	Transition rate of machine i from state α to β

As illustrated in Fig. 1, the system under study consists of a network of m non-identical machines producing m part types comprising $m - 1$ work-in-process types and one final commodity. Each machine j creates one kind of part by operating and/or assembling $L_{(i)(j)}$ number of parts produced by machine i . Each machine has a maximum production rate u_i^{\max} with

$$\begin{aligned} u_i^{\max} &> \sum L_{ij} u_j^{\max} \quad \forall j > i, i \neq m \\ u_m^{\max} &> d \end{aligned} \tag{1}$$

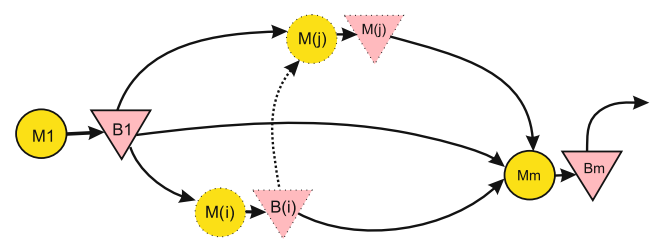


Fig. 1 Network failure-prone manufacturing system



Let $\mathbf{u}(t)$ be the vector of production rates. The set of feasible control policies of process $\zeta_i(t)$ is such that $\mathbf{u}(t) \in K(\alpha)$. This set is given by Eq. 2

$$K(\alpha) = \{\mathbf{u}(t) \in R^m, 0 \leq u_i(t) \leq u_i^{\max}, 1 \leq i \leq m\} \quad (2)$$

There are $m - 1$ intermediate buffers ($B_i, 1 \leq i \leq m - 1$), one buffer following each machine and finished goods are put into a final output stock B_m . Production is discrete and strictly follows a forward direction ($L_{ji} = 0$ for all $i < j$). Machine m produces the final commodity to meet the constant demand rate d facing the system. We assume that the first machine (M_1) is never starved and the last machine (M_m) is never blocked. Let $x_i(t)$ be the inventory/backlog of produced parts i and let $\zeta_i(t)$ be the stochastic Markov process which describes the state of machine M_i at time t . The dynamics of the system can then be described in terms of the states of the machines and the stock levels.

Machine states The operating mode of machine i at time t , can be described by variable $\zeta_i(t)$ with values in $B = \{0, 1\}$ where

$$\zeta_i(t) = \begin{cases} 0 & \text{if machine } i \text{ is under corrective maintenance} \\ 1 & \text{if machine } i \text{ is in use or idle (starved or blocked)} \end{cases} \quad (3)$$

Let $\lambda_{\alpha\beta}^i$ be the transition rate of machine i from state α to state β . The following equations describe statistically the dynamics of the machine state.

$$P[\zeta_i(t + \delta t) = \beta | \zeta_i(t) = \alpha] = \begin{cases} \lambda_{\alpha\beta}^i \delta t + O(\delta t) & \text{if } \alpha \neq \beta \\ 1 + \lambda_{\alpha\alpha}^i \delta t + O(\delta t) & \text{if } \alpha = \beta \end{cases} \quad (4)$$

with

$$\lambda_{\alpha\beta}^i \geq 0, \quad \lambda_{\alpha\alpha}^i = - \sum_{\beta} \lambda_{\alpha\beta}^i, \quad \beta \neq \alpha \in B$$

and

$$\lim_{t \rightarrow 0} \frac{O(\delta t)}{\delta t}$$

We assume that the transition rates $\lambda_{\alpha\beta}^i$ are constant.

Stock dynamics The state equations of the stock levels are given by:

$$\begin{aligned} x_i(t) &= \int_t (u_i(t) - \sum_{i < j} L_{(i)(j)} u_j(t)) dt & x_i(0) &= x_i^0 \\ x_m(t) &= \int_t (u_m(t) - d(x_m(t))) dt & x_m(0) &= x_m^0 \\ d(x_m(t)) &= \begin{cases} d & \text{if } x_m(t) \geq -K \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

Where x_i^0 is a given initial stock and $x_i(t)$ represents the (integer) stock level of buffer i at time t and

$$\begin{aligned} -K &\leq x_m(t) \leq Z_m \\ 0 &\leq x_i(t) \leq Z_i \quad i \neq m \end{aligned} \quad (6)$$

Each machine has a capacity constraint ensuring that the production rate does not exceed a given value u_i^{\max} when the machine i is operational. We then have:

$$u_i(t) \in \begin{cases} \{0\} & \text{if } \zeta_i(t) = 0 \\ [0, u_i^{\max}] & \text{if } \zeta_i(t) = 1 \end{cases} \quad (7)$$

Our objective is to control the production rates so as to minimize the expected long run average cost including holding cost and shortage cost. The cost function is given by

$$\begin{aligned} J(\alpha, x, u) &= \sum_{i=1}^m \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} E \int_0^T [C_i^+(t) x_i^+(t) + C_i^-(t) x_i^-(t)] dt \right\} \\ \mathbf{X}(0) &= \mathbf{X}_0 \quad \zeta(0) = \mathbf{1} \end{aligned} \quad (8)$$

subject to the constraints in Eqs. 1–7 where C_i^+ and C_i^- represent the holding cost per unit for positive inventory or the backlog cost per unit, respectively, $x_i^+(t) = \max(x_i(t), 0), x_i^-(t) = \max(-x_i(t), 0)$. The value function of such a problem is

$$v(\alpha, x) = \min_{u \in K(\alpha)} J(\alpha, x) \quad (9)$$

The properties of the value function and also an approximation of the optimal control policy are given in next section.

3 Approximate control policy

The combination of an experimental design approach and response surface methodology using simulation experiments has been shown to be a useful tool to control FPMS [12]. The purpose of this paper is to combine such simulation-based models with the modified HPP as described by Xie [8] for the case of a discrete-time

network of FPMS, one-machine and single commodity with constant demand rate. In the following we extend the modified HPP presented by Xie [8] to a network of FPMS with multiple non-identical machines and connections between them. The machines in our model are non-identical in the sense that the production/transition rates are different. The HJB equations describe optimality conditions of the control problem for a failure-prone manufacturing system like our model. In this problem $x_i(t)$ is a discrete stochastic variable, which causes an additional increase of the complexity of these equations. The complexity of the proposed approach is well illustrated in a situation where the optimal production control problem can be solved analytically only for very simple system like one-machine, one-type of product [5, 8]. Given the complexity of the HJB equations for the proposed model, the objective of this paper is not to solve them analytically but to experimentally determine the parameters of the modified hedging point policy which gives the best approximation of the value for $\mathbf{u}(t)$. The optimal control policy of the production planning problem formulated above can be approximated by a single threshold surplus level Z_i such that for the intermediate buffer

$$u_{i \neq m}(t) = \begin{cases} 0 & \text{if } x_i(t) > Z_i^* \\ u_i^{\max} & \text{if } x_i(t) + u_i^{\max} < Z_i^* \\ Z_i - x_i(t) & \text{otherwise} \end{cases} \tag{10}$$

and for the final commodity

$$u_m(t) = \begin{cases} 0 & \text{if } x_m(t) - d > Z_m^* \\ u_m^{\max} & \text{if } x_m(t) + u_m^{\max} - d < Z_m^* \\ Z_m - x_m(t) + d & \text{otherwise} \end{cases} \tag{11}$$

We extend the concept to multiple non-identical machines in a network. We modify the production rates for each machine M_i in discrete time after they have produced one unit of part i or at the end of a constant period $(u_i^{\max})^{-1}$. The modified hedging point policy defined by Eqs. 10 and 11 is completely determined by the values of Z_i for $i = 1, 2, \dots, m$, which we therefore call the design factors. To find the optimal control policy, we need to determine the values of Z_i that result in the best value of $v(\alpha, x)$ given by Eq. 9. To achieve this, simulation experiments and response surface methodology are used to approximate the relationship between $v(\alpha, x)$ and Z_i . It is from this estimated function that the corresponding optimum values of Z_i are obtained.

This leads to an m -factor problem. The experimental design problem with m factors, each one having three levels, becomes difficult to solve for large m Gharbi and Kenne [17]. For a one-order model a 2^m design (instead of 3^m) augmented with center points is an excellent way to obtain an approximation of the response function [22]. It allows one to keep the size and complexity of the design low and simultaneously obtain some protection against curvature. In addition the two-level design can be augmented with axial runs so as to obtain a central composite design (CCD) for the second-order in the RSM, for which curvature is important. This sequential strategy of experimentation is far more efficient than running a 3^m factorial design with quantitative factors [22]. The control approach and the simulation model logic is described in Section 4.

4 Control approach

The results from traditional methods of planning in the environment of the FPMS are not sufficient to reach a comfortable level of desired performance. To improve these methods, we augment the descriptive capacities of conventional simulation model by using both analytical and simulation methods. The resulting structure is depicted in Fig. 2.

1. The control problem statement of the network-FPMS, as described in Section 2, consists of the representation of the production planning through a stochastic optimal control model based on control theory. Hence, the problem of determining the optimal values of the control variables is described in this first step, which contains a specification of the objective of the study. That objective is to find the

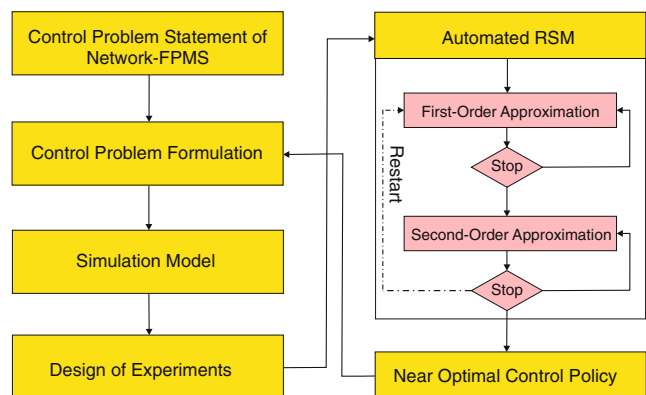


Fig. 2 Proposed control approach

control variables $u_i(t)$ called the production rates in order to improve the related output (i.e. the incurred cost).

- The aim of the control problem formulation block is to develop a mathematical representation of the system based on some simplified hypotheses. By applying an analytical approach, the structure of a feedback control policy is derived. Such a policy is parameterized by several factors and is considered as an input of the simulation model. The cost related to each entry, given by values of input factors, is defined as the output of the simulation model.
- The simulation model presented in Section 5 uses the near optimal control policy defined in the previous step as input in order to evaluate the performance of the manufacturing system.
- In this step we design experiments to define how the control factors can be varied in order to determine the effects of the main factors and their interactions on the cost through a minimal set of simulation experiments. For this purpose we use a regression approach in an analysis of variance (ANOVA) framework [22]. Adequate regression models (first-/second-order models) are later used to apply an Automated RSM method as presented by Neddermeijer et al. [20] and Nicolai et al. [21].
- The automated response surface methodology is then used to obtain an approximation of the relationship between the incurred cost and the significant main factors and their interactions. The obtained regression model is optimized in order to determine the best values of the production rates Z_i . Because the RSM should be precise and fast, we use the Automated RSM including building blocks, strategic moves, stopping rules and a restarting algorithm.
- From this estimated relation, the optimal values of the input factors, called Z_i^* , are determined by performing a canonical analysis. The hedging point policy $\mathbf{u}(Z^*)$ determined in this way can be applied to the Network-FPMS.

5 Simulation model

A discrete event simulation model that describes the dynamics of the system is developed using the Visual SLAM language [23]. In order to obtain the cost of the system for a given set of input factors Z_i the behavior of the system is simulated following the simulation model described in Fig. 3 (for a Network-FPMS with

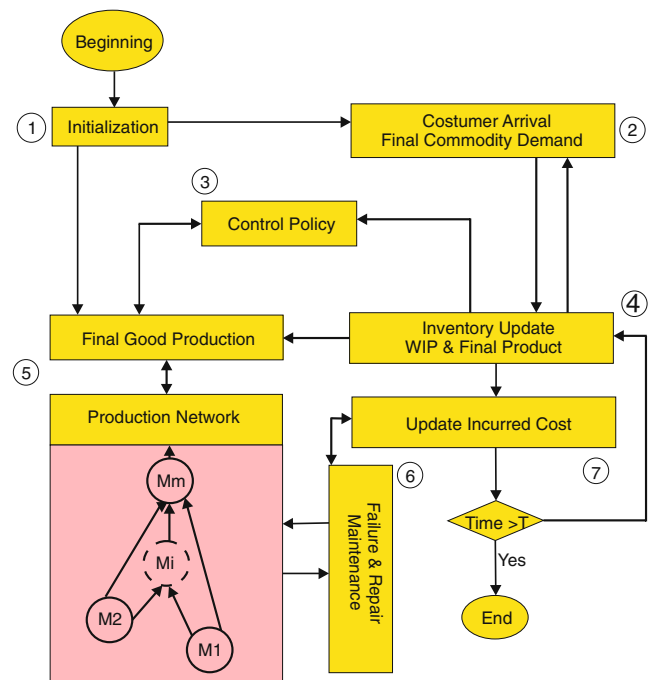


Fig. 3 Diagram of the simulation model

$m = 4$ machines). The different building blocks of the simulation model are the following:

- The *initialization* block sets the variables (production machine rates, holding and shortage inventory cost, production coefficient, demand rate, etc.) to their initial values
- The *customer arrival* block generates the arrival of a customer (demand) for a unit of the final product each d^{-1} unit of time.
- The *control policy* block is defined in Section 3 (see Eqs. 10 and 11). The control policy is given by the output of the inventory update block. This block permanently sends signals to verify the variation in

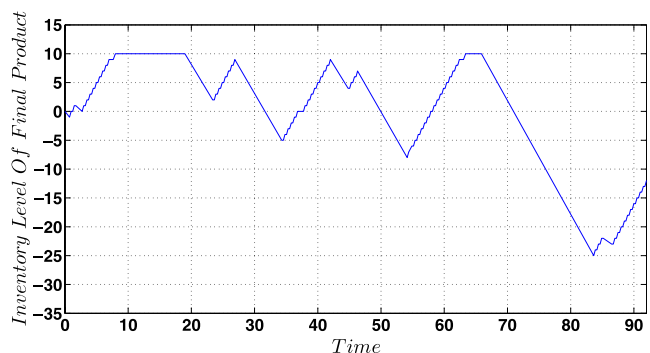


Fig. 4 Trajectory of stock of final product (negative stock represents a backlog)

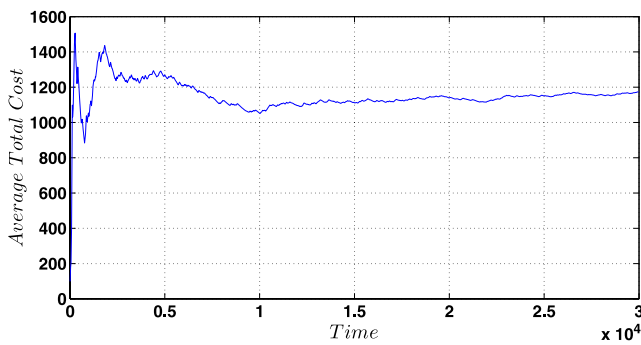


Fig. 5 Average total cost

the stock/shortage levels $x_i(t)$. If the final commodity backlog is equal to K , the demand is lost.

4. As its name suggests, the *inventory update* block (work in process or final product) updates inventories when a unit is produced or when a unit of demand for the final product occurs. The results of the simulation model for a Network-FPMS with four machines, three types of work in process, one type of final good, and using the control policy described by Eqs. 10 and 11 for $Z_1 = 60$, $Z_2 = 20$, $Z_3 = 10$, $Z_4 = 10$, $K = 30$, $l_{12} = 3$, $l_{23} = l_{14} = 2$, $l_{13} = l_{24} = l_{34} = 1$ are illustrated in Fig. 4 for the final product stock trajectory. The inventory level decreases during repair or lack of raw material (here as WIP) caused by breakdowns of intermediate machines.
5. The *final good production* block performs the production of finish goods with support of production network block. In fact a chain of connected machines comprise multi-intermediate buffers and multi-type parts, produce a final good to meet the demand.
6. The *failure–repair* block performs two functions. It defines the TBF and TTR for each machine independently. As mentioned, the breakdown and repair times are exponentially distributed with mean μ_i and λ_i , respectively.
7. The block *update the incurred cost* calculates the cost of inventory and backlogs for all of the buffers. A warmup period T was defined to ensure that the total cost is measured in steady state. The value of T was chosen after a pilot study (Fig. 5) that showed that $T = 11,000$ is sufficiently large. We perform five replications of the simulation model.

6 Experimental design and Automated RSM

The objectives of the experimental design and Automated RSM are to determine whether the input factors

affect the response factor and to estimate the relationship between the cost and significant input factors using design of experiments (DOE) and Automated RSM. RSM is an optimization tool that was introduced in the early 1950s by Box and Wilson [24]. It is a collection of mathematical and statistical techniques that is useful to approximate and optimize stochastic functions. Using regression analysis based on a number of observations of the stochastic objective function, the best local solution is determined together with a search direction for possible improvement. To this end, the stochastic function is evaluated in an arrangement of points referred to as an experimental design. In a manual setting the user can interfere in the optimization process according to his/her personal intuition and insight. In an Automated RSM optimization exercise the settings of the algorithm are fixed in a systematic manner. The algorithm reads the input, performs a systematic search for a (local) optimum and reports the optimum back to the user. Designs for fitting a response surface are called response surface designs. These designs will be discussed in the following steps. Automated RSM is a sequential procedure. Often we are at a point on the response surface that is remote from the optimum, there is little curvature in the system and the first-order regression model will be appropriate. Our objective here is to go rapidly and efficiently along a path of improvement toward the general vicinity of the optimum. Once the region of the optimum has been found, a more elaborate second-order model is employed. In the following we set the best settings for such an Automated RSM based on work of Nicolai et al. [21] and Neddermeijer [20]. The statistical analysis of the simulation data is done by using the statistical software application, SAS, to provide coefficients of regression models and adequate statistical tests of these models.

The following steps outline the automated RSM algorithm that we use to determine the optimal values of the control parameters.

Step 1. Approximate the response surface function locally by a first-order model given by

$$y = \beta_0 + \sum_{i=1}^m \beta_i Z_i + \varepsilon \quad (12)$$

This design is used to estimate coefficient β_i and is evaluated once in the 2^m points of a two-level full factorial design and five times in the center point of the current region of interest. For the advantages of this design we refer the readers to the Nicolai et al. [21] and references therein. Go to step 2.

- Step 2.** Test the first-order model for adequacy. A test for lack of fit and a test for significance of regression are performed. The test for lack of fit is a joint test for interaction between factors as well as for curvature. All statistical tests are performed at a significance level of 5%. If the first-order model is inadequate, go to step 3 otherwise, go to step 4.
- Step 3.** Solve the inadequacy of the first-order model. If the first-order model is inadequate, there is some evidence of curvature or interaction between the factors in the current region of interest, or the regression coefficients are all cannot be discerned from zero. If this occurs during the first optimization exercise, shrink the region of interest with 50% and go to step 1 otherwise, go directly to step 5 and approximate the response surface function by a second-order model.
- Step 4.** Perform a line search in the direction of steepest descent. If the first-order model is accepted, then it is used to determine the direction in which most improvement of the simulation response is expected. The steepest descent direction is given by $\mathbf{b} = (b_1, b_2, \dots, b_m)$. Parameters b_i are estimators for β_i obtained in step 1 and are significantly different from zero. Increments $(\Delta Z_1, \Delta Z_2, \dots, \Delta Z_m)$ along the direction of steepest descent are chosen with

$$\Delta Z_i = \frac{b_i}{b_j} \quad b_j = \max |b_i| \quad (13)$$

End the line search when 3 consecutive observed values of the simulation response functions are higher than the preceding observation (3-in-a-row stopping rule). From the previous points in the research line that one has minimum value function set as a new center point of the current region of interest, then go to step 1.

- Step 5.** Approximately, the response surface function locally by a second-order model. The second-order model is given by

$$y = \beta_0 + \sum_{i=1}^m \beta_i Z_i + \sum_{i=1}^m \beta_{i,i} Z_i^2 + \sum_{i=1}^m \sum_{j>i}^m \beta_{ij} Z_i Z_j + \varepsilon \quad (14)$$

The regression coefficients of the second-order model are again determined by using regression analysis applied to observations performed in an experimental design. The class of second-order designs is the CCD. Generally, the CCD consist of a 2^m factorial (or fractional factorial resolution V), $2m$ axial runs and $2n_c$ center runs. This design can be easily constructed by augmenting the factorial design that was used for estimating the first-order model in step 1 (Montgomery [22]). Go to step 6.

- Step 6.** Test the second-order model for adequacy. A second-order model should be tested for the presence of lack-of-fit. If the second-order model is inadequate, go to step 7 otherwise, go to step 8.
- Step 7.** Solve the inadequacy of the second-order model. If the second-order model is found to be inadequate, we assume that the region of interest is too large. We then shrink the region of interest with 50% and go to step 5, otherwise go directly to step 8. In RSM it is not customary to fit a higher than second-order polynomial.

Table 1 Factorial design (2^4) and response values for fitting the first-order model

Z ₁	Z ₂	Z ₃	Z ₄	Coded variables				\bar{Y}
				X ₁	X ₂	X ₃	X ₄	
61	21	11	11	+1	+1	+1	+1	1,211.47
61	21	11	9	+1	+1	+1	-1	1,250.56
61	21	9	11	+1	+1	-1	+1	1,244.08
61	21	9	9	+1	+1	-1	-1	1,281.82
61	19	11	11	+1	-1	+1	+1	1,224.62
61	19	11	9	+1	-1	+1	-1	1,262.85
61	19	9	11	+1	-1	-1	+1	1,257.09
59	19	9	9	+1	-1	-1	-1	1,297.44
59	21	11	11	-1	+1	+1	+1	1,216.68
59	21	11	9	-1	+1	+1	-1	1,255.39
59	21	9	11	-1	+1	-1	+1	1,246.55
59	21	9	9	-1	+1	-1	-1	1,284.56
59	19	11	11	-1	-1	+1	+1	1,228.89
59	19	11	9	-1	-1	+1	-1	1,266.82
59	19	9	11	-1	-1	-1	+1	1,259.97
59	19	9	9	-1	-1	-1	-1	1,298.03
60	20	10	10	0	0	0	0	1,253.91
60	20	10	10	0	0	0	0	1,315.18
60	20	10	10	0	0	0	0	1,245.80
60	20	10	10	0	0	0	0	1,292.50
60	20	10	10	0	0	0	0	1,251.08

Table 2 ANOVA for the first-order model

Source	DF	Sum of squares	Mean squares	F-ratio	P value	N/S
Model	4	10,402.3374	2,600.5843	8.72	0.0006	Significant
Z ₁	1	56.0152	56.0152	0.19	0.6705	Not significant
Z ₂	1	644.9746	644.9746	2.16	0.1608	Not significant
Z ₃	1	3,882.8269	3,882.8269	13.02	0.0024	Significant
Z ₄	1	5,818.5208	5,818.5208	19.51	0.0004	Significant
Residuals	16	4,771.6399	298.2275			
Interaction	1	0.6982	0.6982	0.00	0.9599	Not significant
Pure quadratic	1	1,031.3583	1,031.3583	3.86	0.0696	Not significant
Pure error	14	3,739.5833	267.1131			
Total	20	15,173.9772				

Step 8. Perform canonical analysis. Canonical analysis is used to determine the location and the nature of the stationary point of the second-order model. The estimated second-order approximation can be written as follows:

$$\hat{y} = \beta_0 + \mathbf{X}'\mathbf{b} + \mathbf{X}'\mathbf{B}\mathbf{X} \tag{15}$$

Where

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12}/2 & \dots & b_{1m}/2 \\ & b_{22} & \dots & b_{2m}/2 \\ & & \ddots & \vdots \\ S_{ym} & & & b_{mm} \end{pmatrix} \tag{16}$$

The stationary point *S* of the second-order polynomial is determined by

$$\mathbf{X}_S^* = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b} \tag{17}$$

If all eigenvalues of **B** are positive, then the quadratic surface has a minimum at the stationary point **X_S**. In this state, it could be useful to explore a region around this minimum with a new second-order approximation, therefore go to step 5. If the eigenvalues are mixed in sign, then the stationary point *S* is a saddle point. It

means the optimum is still located far away from the current region of interest. In this case, approximating this region with a first-order model and consequently performing a line search would be preferable, so go to

Table 3 The CCD design and response values for fitting the second-order model

Z ₁	Z ₂	Z ₃	Z ₄	Coded variables				\bar{Y}
				X ₁	X ₂	X ₃	X ₄	
61	21	35.51	41	+1	+1	+1	+1	777.59
61	21	35.51	39	+1	+1	+1	-1	781.51
61	21	33.51	41	+1	+1	-1	+1	782.00
61	21	33.51	39	+1	+1	-1	-1	785.84
61	19	35.51	41	+1	-1	+1	+1	783.85
61	19	35.51	39	+1	-1	+1	-1	789.76
61	19	33.51	41	+1	-1	-1	+1	788.14
59	19	33.51	39	+1	-1	-1	-1	793.66
59	21	35.51	41	-1	+1	+1	+1	780.55
59	21	35.51	39	-1	+1	+1	-1	785.94
59	21	33.51	41	-1	+1	-1	+1	783.87
59	21	33.51	39	-1	+1	-1	-1	788.20
59	19	35.51	41	-1	-1	+1	+1	789.64
59	19	35.51	39	-1	-1	+1	-1	794.66
59	19	33.51	41	-1	-1	-1	+1	791.64
59	19	33.51	39	-1	-1	-1	-1	796.23
62	20	34.51	40	+2	0	0	0	782.93
58	20	34.51	40	-2	0	0	0	788.59
60	22	34.51	40	0	+2	0	0	779.61
60	18	34.51	40	0	-2	0	0	794.83
60	20	36.51	40	0	0	+2	0	783.43
60	20	32.51	40	0	0	-2	0	789.12
60	20	34.51	42	0	0	0	+2	782.16
60	20	34.51	38	0	0	0	-2	791.57
60	20	34.51	40	0	0	0	0	664.94
60	20	34.51	40	0	0	0	0	762.88
60	20	34.51	40	0	0	0	0	748.98
60	20	34.51	40	0	0	0	0	781.72
60	20	34.51	40	0	0	0	0	734.15
60	20	34.51	40	0	0	0	0	731.50
60	20	34.51	40	0	0	0	0	741.59

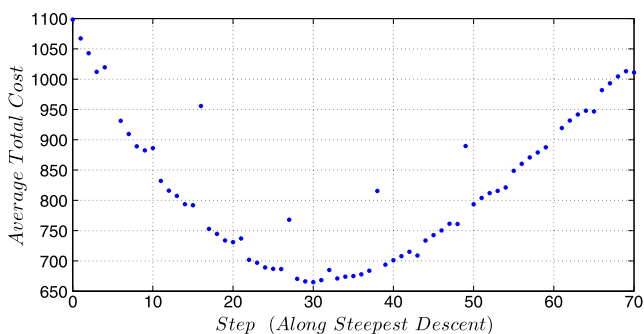


Fig. 6 System cost along the path of steepest descent

Table 4 ANOVA for the complete second-order model

Source	DF	Sum of squares	Mean squares	F-ratio	P value	N/S
Model	14	13,598.0492	971.29	1.93	0.1044	Not significant
Residuals	16	8,062.2094	503.89			
Total	30	21,660.2587				

the step 1. The Automated RSM is stopped at this step when one of the following stopping rules occurs:

1. The estimated response value does not improve sufficiently anymore
2. A fixed maximum number of simulations have been performed
3. The input factor values Z_i do not change anymore

To escape from a non-optimal region, restart the algorithm as soon as the algorithm is stopped by one of the above stopping criteria using the new starting point. Because the algorithm cannot escape from a non-optimal region when the size of the region of interest is too small, reset the size of the region to its initial value. In all optimization runs the best solution should be remembered.

7 Numerical example

For the numerical example experiment in this section, the following values are used: $d = 2$, $m = 4$, $u_1^{\max} = 48$, $u_2^{\max} = 12$, $u_3^{\max} = u_4^{\max} = 4$, $K = 30$, $l_{12} = 3$, $l_{23} = l_{14} = 2$, $l_{13} = l_{24} = l_{34} = 1$, $C_1^+ = 1$, $C_2^+ = 3$, $C_3^+ = 6$, $C_4^+ = 10$, $C_4^- = 100$, $\mu_i = 23$, $\lambda_i = 5$. The first center point of the current region of interest is $Z_1 = 60$, $Z_2 = 20$, $Z_3 = 10$, $Z_4 = 10$. We use a 2^4 factorial design augmented with five center points, presented in Table 1. Five replications (response evaluation) were conducted for each combination of the factors. Table 1 also shows the average of responses for these evaluations in the yield column. The process variables of interest are the inventory hedging points (Z_i). The total inventory cost of the FPMS at this point is **1271.69** units of money per unit of time. Statistical analysis was done using the SAS statistical software application. We use the

following first-order model to construct the path of steepest descent.

$$\hat{y} = 1,844.88 - 1.87Z_1 - 6.35Z_2 - 15.58Z_3 - 19.07Z_4 \quad (18)$$

Before exploring along the path of steepest descent, the adequacy of the first-order model is investigated. The analysis of variance for this model is summarized in Table 2. Both the interaction and quadratic curvature checks are not significant, whereas the F-test for overall regression is significant. At this point, we therefore have no reason to question the adequacy of the first-order model.

Table 2 shows that only regression coefficients b_3 and b_4 are significantly different from zero. Therefore the steepest descent direction is $(0, 0, 15.58, 19.07)$. To move away from the design center point along the path of steepest descent, we move $\Delta Z_3 = -15.57 / -19.07 = 0.82$ units in the Z_3 direction for every $\Delta Z_4 = -19.07 / -19.07 = 1$ units in the Z_4 direction while $\Delta Z_1 = \Delta Z_2 = 0$. We compute points along this path and observe the yields at these points until 3 consecutive observed values higher than the preceding observation are noted. Figure 6 plots the total inventory system cost along the path of steepest descent. Decreases in the response of the model are observed up to 30th step with coordinates of $(60, 20, 34.51, 40)$ and a response value of **644,94**, where 3 consecutive steps beyond this point result in an increase in yield. The response value decreases more than **52** percent at this new point compared to the starting point of **1271.69**.

Analysis of variance implies that the new first-order model around the new point $(60, 20, 34.51, 40)$ is not adequate. This means that the curvature and interaction are found to be quite evident and a new steepest decent procedure will certainly soon be truncated [25].

Table 5 ANOVA for the second-order model (without interaction effects)

Source	DF	Sum of squares	Mean squares	F-ratio	P value	N/S
Model	8	13,591.1471	1,698.89	4.63	0.0020	Significant
Residuals	22	8,069.1116	366.78			
Lack of fit	15	10.2411	0.68	0.00	1.0000	Not significant
Pure error	7	8,058.8705	1,151.27			
Total	30	21,660.2587				

At this point we are interested in finding optimum conditions through the use of a fitted second-order model. To apply a second-order model to this new region, we augment the $2^{(m=4)}$ factorial design with $n_c = 7$ central runs and $2(m = 4) = 8$ axial runs with $\alpha = \pm 2$. α is the distance of the axial points from the design center. Table 3 shows this design and related response values. This CCD is rotatable and provides equal precision of estimation in all directions [22, 25]. The results of ANOVA for a complete second-order model (Eq. 14) and a second-order model without interaction are summarized in Tables 4 and 5.

The overall regression F-test for the second-order model without interaction is significant whereas a lack-of-fit test is non-significant so the following regression model is adequate.

$$\hat{y} = 83,556.62 - 1,446.28Z_1 - 499.99Z_2 - 841.33Z_3 - 987.52Z_4 + 12.04Z_1^2 + 12.40Z_2^2 + 12.17Z_3^2 + 12.31Z_4^2 \tag{19}$$

From Eq. 17 the stationary point is

$$X_S^* = -\frac{1}{2}B^{-1}b = \begin{pmatrix} 60 \\ 20 \\ 35 \\ 40 \end{pmatrix} \tag{20}$$

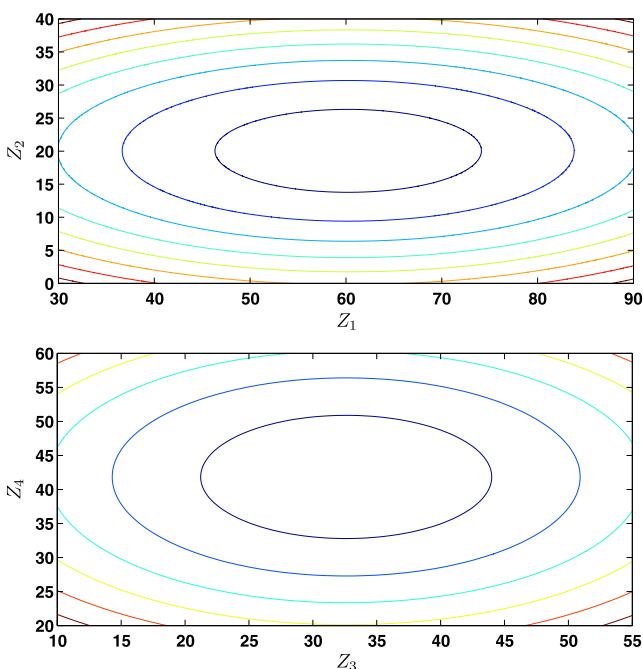


Fig. 7 Contour plots in terms of the control variables

The canonical analysis characterizes the response surface. The eigenvalues of **B** related to Eq. 19 are

$$\text{Eigenvalues} = \begin{pmatrix} 12.04 \\ 12.40 \\ 12.17 \\ 12.31 \end{pmatrix} \tag{21}$$

Because all of the eigenvalues of **B** are positive, we conclude that the stationary point (60, 20, 35, 40) is a minimum. The expected long run average cost at the minimum point is $J^* = 665$. We restart the algorithm for the different starting points and present the best response in all of optimization runs as a numerical example. Figure 7 shows contour plots in terms of control variables Z_i .

8 Summary and concluding remarks

In this paper we have extended the concept of HPP to a network-FPMS consisting of multiple machines, multiple products and multiple buffers with restrictions on the feeding materials. The proposed approach for solving this more complex problem is based on a combination of analytical experiments, experimental design and automated RSM. In the automated RSM we have used a two-level factorial design, augmented with center and axial runs (CCD). This design can be easily constructed by augmenting the factorial design that was used to estimate the first-order model. This design allowed to reduce the number of simulation runs required. In the future, we plan to extend the results of this paper to the control of network-FPMS with age dependent failure rates, preventive repair maintenance policy and non-exponential distribution break down/repair times.

Acknowledgements The authors would like to thank professor Peter Goos for his invaluable help and Mrs. Maryam Hassanzadehavval for her kind encouragements.

References

1. Older GJ, Suri R (1980) Time operational part-routing in manufacturing system with failure prone machines. In: Proc. 19th conf. decis. control, Albuquerque, NW
2. Kimemia JG (1982) Hierarchical control of production in flexible manufacturing systems. PhD thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science
3. Kimemia JG, Gershwin SB (1983) An algorithm for the computer control of production in felxible manufacturing system. IIE Trans 15:353–362

4. Rishel R (1975) Control of systems with jump Markov disturbances. *IEEE Trans Automat Contr* 20:241–244
5. Akella PR, Kumar PR (1986) Optimal control of production rate in a failure prone manufacturing system. *IEEE Trans Automat Contr AC-31(2)*:116–126
6. Bielecki T, Kumar PR (1988) Optimality of zero-inventory policies for unreliable manufacturing systems, operations research archives. 36(4):532–541
7. Sharifnia A (1988) Production control of manufacturing system with multiple machine states. *IEEE Trans Automat Contr* 33(7)
8. Xie X (1989) Optimal control in a failure prone manufacturing system, a research for INRIA-LORRAINE Institute
9. Boukas EK, Yang H (1996) Manufacturing flow control and preventive maintenance: a stochastic control approach. *IEEE Trans Automat Contr* 41(6):881–885
10. Kenne JP, Gharbi A, Boukas EK (1997) Control policy simulation based on machine age in a failure prone one-machine, one-product manufacturing system. *Int J Prod Res* 35(5):1431–1445
11. Mourania I, Hennequina S, Xie X (2006) Simulation-based optimization of a single-stage failure-prone manufacturing system with transportation delay. *International of Production Economics*
12. Kenne JP, Gharbi A (1999) Experimental design in the production and maintenance control problem of a single machine, single product manufacturing system. *Int J Prod Res* 37(3):621–637
13. Kenne JP, Gharbi A, Beit M (2007) Age-dependent production planning and maintenance strategies in unreliable manufacturing systems with lost sale. *Eur J Oper Res* 178: 408–420
14. Presman EL, Seti SP, Zhang H (2000) Average cost optimality for an unreliable two-machine flowshop with limited internal buffer. *Ann Oper Res* 98:333–351
15. Ching WK (2001) Markovian approximation for manufacturing systems of unreliable machines in tandem. *Nav Res Logist* 48:65–78
16. Gharbi A, Kenne JP (2005) Maintenance scheduling and production control of multiple-machine manufacturing systems. *Comput Ind Eng* 48:693–707
17. Gharbi A, Kenne JP (2000) Production and preventive maintenance rates control for manufacturing system: an experimental design approach. *Int J Prod Econ* 65:275–287
18. Boukas EK, Haurie A (1990) Manufacturing flow control and preventive maintenance: a stochastic control approach. *IEEE Trans Automat Contr* 33(9):1024–1031
19. Kokotovic P, Khalil HP, O'Reilly J (1986) Singular perturbation methods in control: analysis and design. Academic Press, New York
20. Neddermeijer HG, Gerrit J, Oortmarssen V, Piersma N, Dekker R (2000) A frame work for response surface methodology for simulation optimization. In: Proceedings of the 32nd conference on winter simulation, Orlando, Florida, 10–13 December 2000
21. Nicolai RP, Dekker R, Piersma N, Oortmarssen JG (2004) Automated response surface methodology for stochastic optimization models with unknown variance. In: Proceedings of the 36th winter simulation conference, Washington, D.C., 5–8 December 2004
22. Montgomery DC (2005) Design and analysis of experiments, 6th edn. Wiley, New York
23. Pritsker AAB, O'Reilly JJ, LaVal DK (1999) Simulation with visual SLAM and AweSim. Wiley, New York
24. Box, GEP, Wilson KB (1951) On the experimental attainment of optimum conditions. *J R Stat Soc B13*:1–38, discussion, 38–45
25. Myers RH, Montgomery DC (2002) Response surface methodology: process and product optimization using designed experiments, 2nd edn. Wiley, USA

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.