Integrated Model for Capacity Planning for Manufacturing Systems

Andrés Ramos

September 1992

1 Introduction

1.1 Problem framework

In general, manufacturing production lines can be classified as:

- job shops

  These lines are characterized by production of complex elements requiring numerous operations to be performed. Elements are made in very small series and manufacturing time can be around a month. For example, a high voltage transformer or a thermal generator. The emphasis of these lines is on coordination and scheduling of the different operations.

- continuous production lines

  These lines are characterized by production of simple elements requiring few operations. The elements are made in big amounts and manufacturing time can be around one minute. As an example, we can think of bolts. The emphasis of these lines is on throughput and in-process inventory.

- flexible manufacturing systems (FMS)

  These lines lie in between the other two. They produce a variety of elements but in small quantities. These are intermittent production systems with the following characteristics: batch production, many products which share the workstations, products relative short-lived and uncertain demand. As an example, the High Mix Low Volume Production Line at the Hewlett-Packard facility in Puerto Rico [27] produces around 30 different board types at a rate of 250 boards per day. The production mix ranges from 1 to 1500 board per month depending on the type. The typical time in system is 2 hours. In these lines the emphasis is on minimizing the makespan (i.e., the time to achieve the prescribed orders) and other desirable attributes are minimize the difference with respect to due delivery dates, in-process inventory, machine and personnel utilization. Also it is a flow shop type because products are moving in a uniform direction. Each operation is done just by one machine/operator (there are few exceptions). There is no machines in parallel although there are shortcuts in the line because certain operations are not needed for any type of product.

The present model is aimed at solving the capacity planning of flexible manufacturing systems.

An integrated capacity planning model has to implement the different type of decisions addressed in manufacturing systems, borrowed from [15]:

- STRATEGIC PLANNING

  Process of deciding the resources used to attain the objectives of a manufacturing facility, policies governing the acquisition, use and disposition of these resources.

- TACTICAL PLANNING

  Efficient allocation of the available resources (e.g., machines, work force).

- OPERATIONS CONTROL

  Process of assuring that the specific tasks are carried out effectively and efficiently.

Moreover, this classification tries to match functions and people responsible for taking these decisions.

One has to include the complex interactions among decision levels while keeping a balance and coordination in modeling aspects in order to achieve minimum costs. Thus, an integrated approach is needed if one wants to avoid the problems of suboptimization. This approach suggests a hierarchical model linking all the decision levels in an effective manner. Decisions made at higher levels provide constraints for lower ones. While, detailed results provide the necessary feedback to evaluate upper level decisions. Because a capacity planning model is concentrated in solving the upper level decisions it will be necessary to simplify the lower levels while keeping their main implications.

The application of the previous three levels to decisions taken inside a manufacturing facility, where the examples shown below have been taken from an electronic memory board production line described in [27], will result in:

- CAPACITY PLANNING (FACILITY DESIGN) DECISIONS.

  Type of *decisions* addressed:

  - updating or replacement of existing equipment (stencil printer, infrared soldering) during or after its useful life
  - acquisition of new equipment
  - location and size (discrete by nature) of inventory buffers
  - replacement of support equipment (AGV, conveyors)
  - acquisition of new support equipment
  - equipment specification
  - reinforcement of the human activity (inspection personnel)
  - new configuration or layout of equipment

These decisions allow the firm to maintain their competitive capabilities. The driving forces for these decisions can be: increment in demand causing investments in new faster and more reliable machines, new type of products requiring new production technologies.

A capacity planning proposal will be a combination of the above decisions and can be modeled as substitution of old activities by new ones. Decisions are *discrete* (number of machines to instal) or *continuous* (processing speed) involving different equipment specifications for a production line.

Consequences: negotiating purchasing agreements with equipment manufacturers, new capital requirements (land, building, labor, machines).

- PRODUCTION SCHEDULING DECISIONS.

  The importance of the production scheduling function is related to the type of line and parts produced. This function can be less relevant if products are similar and a great amount of them are produced in one time period or can be important when existing interdependence among products (like in the case of sequence-dependent setup times).

  Besides, this decision level is specially required when appearing highly uncertain future demand where previous scheduling patterns are ever repeated.

  A master production scheduling plan controls and organizes the line operations after evaluating and deciding different policies (*decisions*) for dispatching orders to the shop floor: choosing batch sizes, designing rules for moving parts, sequencing of parts through the system. A schedule or plan is an ordering of the products and operations onto machines subject to priorities and routing relationships.

  Commonly used *criteria* for production scheduling decisions are: minimize work-in-process inventory or related costs, maximize completion of the products on time or minimize penalty costs, maximize utilization of the available capacity, minimize makespan or time to complete all the prescribed products.

  Consequences: delivery schedules, setup time reductions.

  Uniform scheduling is common practice in just-in-time environments. It is a planning method for resource allocation based on smooth, homogenized production flow, see [31]. This strategy is trying to decrease the time needed in setup changes and to maintain a constant rate in production delivery.

- OPERATIONAL (INTERMEDIATE RANGE PLANNING) DECISIONS.

  Operating *decisions* are: priority rules changes when certain events occur at the shop floor.

  An operations model reproduces the production line rules and performance giving *attributes* (or performance measures) such as:

  - utilization of personnel and equipment

  - productivity or throughput
  - time in system for parts
  - makespan (total time needed to produce certain number of parts)
  - inventory in system (is proportional to the time in system for parts)
  - timeliness of deliveries (proportion of late orders)
  - times that parts spend in queues, transport, waiting for transport
  - product reliability and quality

  Some of them can be difficult to measure or quantify. Furthermore, robustness of each attribute (dependence or variance with respect to stochastic factors) for any of the capacity planning proposals should be considered.

  Usually, the most important attributes are converted into costs according to some factors (e.g., penalty for deviation from a due time).

  Consequences: changes in equipment design and operation, identification of bottlenecks, product specifications.

The application of a model to a certain type of manufacturing lines conditions its design. The manufacturing lines addressed by this model are FMS.

1.2 Time frames

The useful life of the machines is up to 5-7 years, then the time frame of the investment decisions will be 1-3 years. Intrinsically, these decisions are time dependent (dynamic). One has to decide when to invest.

It is supposed that production is a continuous task that does not depend on the time (month, day) during the year. Therefore, any representative time unit (e.g., week or month) is convenient. It seems that a month is a natural time interval.

Then, the total annual demand estimate is split into monthly objectives being the production demand for each month. Monthly demand is scheduled along the month to balance production for each day and type of product. The orders are dispatched to the shop floor once a month and production begins as soon as an order arrives.

1.3 Uncertainty

One has to take decisions on new investments, replacements, process improvements that take into account and hedge against all the problem uncertainties.

On one hand, there is uncertainty in volume and mix (number) of each different type of product (memory boards, CPUs) to be demanded during the investment time frame. Besides, there is uncertainty in the production line itself because of the many stochastic factors involved in (equipment failure, human behavior). These uncertainties are considered within the model. In other planning environments (e.g., electric power systems planning) the optimization process can take into account both stochasticities. However, in the present model the last factors are considered into the operations model, while demand uncertainty is considered in an upper level. Synthetic demand series could be used to include the variation of the demand along the time horizon, obtained from previous demand patterns or from predictions about future evolution.

On the other hand, uncertainties associated with cost, reliability or delivery time of new equipment are also considered in the upper level.

2 Formulation

1. Investment decision variables (machine of certain type, certain inspection position, new layout). Using the simulation terminology they would be number of servers (machines, people) in each workstation (machine type, human activity). X are by nature integer, convertible to 0/1 variables.

$$X = \{x_{it}\} \tag{1}$$

i	index of the investment variables. $i = 1, \dots, I$
t	index of the time periods. $t = 1, \dots, T$

2. Attributes of each investment decision (equipment costs, labor costs, building or land costs).

$$C = \{c_i\} \tag{2}$$

3. Attributes of the investment plan.

$$F = \{f_k\} \tag{3}$$

k	index of the investment plan attributes. $k = 1, \dots, K$

4. Specifications for the production line derived from investment decisions (characteristics of the new machine in speed, throughput, availability, percentage of error). They are a linear function of X.

$$E = \{e_m\} \tag{4}$$

m	index of the specifications. $m = 1, \dots, M$

5. Future demands of each type of product over the planning horizon in each time period. They are stochastic.

$$D^{\omega'} = \{d_t^{\omega'}\}, \quad \omega' \in \Omega' \tag{5}$$

6. Random factors affecting production line operations (equipment failure, human behavior).

$$R^{\omega} = \{r^{\omega}\}, \quad \omega \in \Omega \tag{6}$$

7. Production line decision variables.

$$Y = \{y_{jt}\} \tag{7}$$

j index of the production line decision variables. $j = 1, \ldots, J$

8. Attributes of the production line operation (utilization of equipment and personnel, time in system, time to achieve the orders).

$$V = \{v_{lt}\} \tag{8}$$

l index of the production line attributes. $l = 1, \ldots, L$

The *global objective function* is:

$$\min \; g(F, V) \tag{9}$$

and it can be: monoattribute, multiattribute weighted to form a single attribute, multiattribute not weighted treated akin to goal programming. See in [6, 10] different techniques used for multicriteria optimization.

Usually, the global objective function will be total cost (investment plus operating costs). If only investment cost were considered, then $K = 1$, and if they were linear with respect to the investment variables

$$F = \sum_{t=1}^{T} \sum_{i=1}^{I} c_i x_{it} \tag{10}$$

Operating costs ought to be obtained from production line attributes. The computation of the production line attributes V from their inputs E, D, and R (some of them are stochastic) and from their decision variables Y is done by simulation, necessary when no other analytical or optimization techniques can be applied to represent the complexity involved in the production line. The randomness included in R (for each scenario of demand D) is considered by repeating the simulation until a confidence interval small enough is reached. The minimum value of the attributes V are obtained by a simulation optimization process (see section 5).

We are interested in mean values of the attributes over the different stochastic factors affecting the manufacturing process.

$$V^{\omega'\omega} = f(E, D^{\omega'}, R^{\omega}, Y) \tag{11}$$

$$\bar{V}^{\omega'} = E_\omega V^{\omega'\omega} \tag{12}$$

$$\hat{V} = E_{\omega'} V^{\omega'} \tag{13}$$

E_ω expected value in the Ω probability space.

$E_{\omega'}$ expected value in the Ω' probability space.

Restrictions can affect *investment variables* (e.g., maximum investment, labor constraints) and will be deterministic. Also, there can be constraints associated with production line operations. Firstly, those affecting *operating decisions* will be deterministic and explicit (they can be considered prior the simulation and define a feasible experimental region). And secondly, those affecting *operational attributes* (e.g., maximum percentage of late orders, unacceptance of certain throughput of the production line, certain standard variation in an attribute) will be probabilistic and implicit (they require the simulation to be done).

From other point of view, the constraints can be *monoperiod* or *multiperiod*, according to the number of periods affected. Naturally, the later are much more involving because they make interdependent all the periods and have to be included at the capacity planning level.

3 Investment Optimization

The solution method has to deal with the relation between investment and operational decisions and involves *stochastic optimization*. How can we 'organize' the global attributes computation to achieve the optimal investment plan?

Few references have been found on this topic. Possible methods are:

1. *Nested decomposition*

   The difficulty of this method is the computation of the dual variables of the operation subproblem with respect to the investment decisions. See an application of Benders decomposition in [9].

   The only way to obtain derivatives for the production line we are dealing with is by infinitesimal finite differences. They are meaningful in the case of buffer capacities (if the buffer capacities are big enough to be thought as continuous) but meaningless in the case of machines and operators.

   Besides, one has to demonstrate or assume the convexity of the production problem.

2. *Combinatorial-enumeration*

   The number of possible solutions and their combinations have to be small enough to be calculated, see [14] as an example. Small examples allow this type of solution.

3. *Guided heuristic search*

   Which are the criteria that could be used to guide the search of solutions?

A totally different approach to the investment optimization process known as *financial simulation* has been found in [23].

4 Production Scheduling Optimization

A basic algorithm is needed to obtain a master production scheduling plan. This algorithm should capture the effects in the line derived from the investment decisions.

The combinatorial nature of the problem directs the effort to find good feasible schedules, rather than find optimality. Heuristics approaches are the most promising, see for example [2].

Recent surveys on production scheduling and lot-sizing are found in [26, 4].

5 Operations Optimization

How can we get an optimal production plan with a simulation model? We have a combinatorial, discrete, multivariate stochastic optimization called *simulation optimization*. Running the simulation model we can obtain conclusions about:

- what parameters are most sensitive, where and how do they affect the performance of the line?
- what are the changes in the line performance measures with respect to the parameters?
- which are the possible actions to improve the performance?

From a knowledge of the line (by the local expertise) and using of a simulation model the parameters that are decision variables in the production line optimization process are identified.

The simulation model requires an optimization process for the current activities of the plant. A change produced in an activity of the production line (replacement of a machine, introduction of another person in certain inspection position) will require the adjustment in the operations decisions that otherwise would mask the effects of the changes proposed.

Two main difficulties arise associated with simulation optimization. Firstly, the objective function is not analytical, then classical nonlinear programming techniques are not applicable in strictu sensu. And secondly, the objective function (attributes) is stochastic and the criteria used to improve the solution have to be aware of that, specially those regarding the evaluation and convergence properties of the algorithms.

5.1 Statistical comparison of attributes and constraints

Attributes are uncertain due to stochastic factors affecting production line performance and constraints can also be stochastic. The simulation is replicated until the confidence intervals for the attributes of interest are small enough (see [18, 24]). Mean and variance of the attributes are used to draw inferences about the performance of a system operation compared to another. Statistical procedures or decision criteria are applied to ensure improvement of a solution with respect to another or to test the violation of the constraints in the optimization process, see references [3, 18, 22].

In [18] a modified two-sample-t confidence interval is shown for comparing two independent operation points with unequal and unknown variances. After obtaining mean and variance for each one the estimated degrees of freedom $\hat{f}$ are computed

$$\hat{f} = \frac{\left[S_A^2(n_A)/n_A + S_B^2(n_B)/n_B\right]^2}{\left[S_A^2(n_A)/n_A\right]^2/(n_A-1) + \left[S_B^2(n_B)/n_B\right]^2/(n_B-1)} \tag{14}$$

$\bar{X}_A(n_A)$ sample mean at point A with n_A observations.

$S_A^2(n_A)$ sample variance at point A with n_A observations.

and the equation

$$\bar{X}_A(n_A) - \bar{X}_B(n_B) \pm t_{\hat{f},1-\alpha/2}\sqrt{\frac{S_A^2(n_A)}{n_A} + \frac{S_B^2(n_B)}{n_B}} \tag{15}$$

is used as the $100(1-\alpha)\%$ confidence interval of the difference. If the interval does not include the 0, we can say that A differs from B with a $100(1-\alpha)\%$ confidence level and that one is superior to the other depending of the interval position. A similar approach is found in [22].

In [3] an heuristic technique is found. When the upper limit of a confidence interval of an attribute is less than the lower limit of another, then a conclusive decision is taken: the first solution is better than the second if we are minimizing. If this conclusion can not be reached the simulation replication is extended to decrease the confidence interval width for the second solution until a maximum number of replications.

Analogously, the constraints should be satisfied with a certain confidence level. That means that the upper (lower) limit of the confidence interval of the left hand side should be less or equal (greater or equal) than the right hand side in less or equal (greater or equal) constraints.

5.2 Simulation optimization techniques

Mathematically the simulation optimization problem can be expressed as:

$$\begin{aligned} \min \quad & \bar{V}^{\omega'} \\ \text{subject to:} \quad g_p(Y) &= G, \quad p = 1, \ldots, P \end{aligned} \tag{16}$$

References [12, 17] present surveys of the methods used in simulation optimization. Particular applications to manufacturing systems are presented in references [3, 6, 8, 20, 22, 30] although with simple examples. Moreover, some include capacity planning decisions in the simulation optimization process.

Techniques used to find local or global optima for discrete-event simulations, according to reference [17], can be classified in:

1. *pattern search methods*

   These methods work with *nonconvex, nonconcave* objective functions, do *not* need *continuity* in decision variables and attributes and only use specific function values, no derivatives. They can be applied to constrained as well as unconstrained problems.

   In general, these methods make rapid early progress toward the optimum but iterate close to the final solution. The search procedures are very simple and could be easily implemented in a parallel computer.

   - modified Hooke-Jeeves (see [22])

     The method performs two types of search routines cyclically: an exploratory search and a pattern search. The exploratory search is conducted along the individual coordinate directions in the neighborhood of a reference point to find the search path. The pattern search proceeds along the direction defined by the starting and ending points of the exploratory search.

   - simplex procedure, Nelder-Mead simplex, constrained simplex (complex) (see [3, 5, 6])

     Randomly or uniformly spread generate a set of points forming a simplex satisfying the explicit constraints. Perform a simulation for each point and evaluate the attributes. Eliminate the points violating the implicit constraints. Evaluate the objective function. Find the worst point. Obtain the centroid of the remaining. Determine a new point image of the worst with respect to the centroid. Perform the simulation at this new point and iterates until no further improvement is achieved. Through this process, the simplex moves around in the feasible region while vertices get closer to each other until they collapse in the optimum.

     In the complex search special effort is made to prevent the simplex from leaving the feasible region.

2. *path search methods*

   They involve estimating a direction to move from a current variable set to an improved point in the feasible variable set. Only local information is used. Once a direction is found, a distance to move in that direction is determined.

   Typically, these methods assume that decision variables and attributes are *continuous*. Assumption not necessary in pattern search methods.

- response surface methods (see [6, 18])

  They fit locally a polynomial response surface model to a set of simple observations obtained by orthogonal experimental design (2^J factorial, 2^{J-p} fractional factorial or J simplex) in a particular region.

  An estimation of the linear response surface is required from observations using a multiple linear regression model for the objective function. The coefficients of the regression correspond to the gradients of the objective function. A line search is performed along the optimal direction in Phase I. This procedure is repeated until the linear model stops providing a sufficiently good fit. Phase II is then implemented, where a quadratic polynomial model is fit to the response. Techniques borrowed from nonlinear programming are adapted and used in these methods.

- stochastic quasi-gradient, stochastic approximation (see [1, 13])

  Use a sequence of estimates of the solution that converges almost surely to the solution. The sequence is based on the gradient of the stochastic performance measure.

- perturbation analysis (infinitesimal, finite, smoothed) (see [16, 19])

  It is a technique for the computation of the gradient of performance measures using only one simulation of the system.

  A parameter change can generate perturbations in the timing of events in the simulation of a discrete event dynamic system. Perturbations can be propagated to other events. Since all performance measures depend on the timing of events on the simulation, they will induce perturbations in the sample performance measure. This approach has only been applied to limited simple production lines. A lot of effort is being allocated on extending its application field. At present, it can not be applied to production lines with several types of products and a continuous production.

  Applications to manufacturing systems are found in [9, 30].

- likelihood ratio, weak derivation (see [16, 19])

3. *random search methods*

   Designed for global minimization rather than local. Use a random approach to select the set of variables with the hope of obtaining an improving and, eventually, optimal solution. They are slow to converge and inconsistent and seem the last resort. A comparison between this method and simplex procedure is found in [3].

   - simulated annealing algorithm (see [8, 20, 28])

     The algorithm begins randomly choosing a point in the variables space and evaluating the attributes. Subsequently, an adjacent feasible point is randomly selected and evaluated. The second point becomes the current point with certain probability depending on the values of the attributes and on the temperature of the process. Temperature is then decreased and this iteration is repeated. At high temperatures, the probability of accepting an adjacent point is large. As the temperature decreases, the probability of accepting downhill moves decreases.

The tuning process of this method is the election of the annealing schedule (initial and final temperature and its evolution).

- genetic algorithms (see [28])

4. *integral methods*

   Designed specifically for global optimization rather than local. Not successfully implemented in stochastic environment. An interesting combination of the last two is referenced in [7].

6 Model Implementation

The model presented so far is very complex because it includes investment decisions optimization, production scheduling, simulation optimization, interface with an external simulation package and uncertainties at several levels. All these characteristics require that algorithm design and model development have to be carefully implemented.

The model has been developed to run in a PC type computer using FORTRAN version 5.1 of Microsoft Corp., see [21], and SIMAN IV release 1.0 of Systems Modeling Corp., see [24, 29].

The algorithm is as follows:

1. Determination of the series of demands along the whole planning horizon (1-3 years) divided into periodical goals (monthly).

2. Production scheduling of periodical demands into shop floor orders.

3. Simulation of the production for the first period.

   Here, the interfaces between the capacity planning model and the simulation model are presented, see figure 1:

   - SIMAN has been used as simulation language to represent the production line due to its special manufacturing characteristics. SIMAN requires two types of input files: *model* and *experiment*. The model file (whose name is `SPL.MOD`) is a functional description of the system's components and their interactions. Therefore, the model file is completely production line dependent. Capacity planning for another production line will require rebuilding the model file.

     Besides, as a restriction of the capacity planning model all the possible investment alternatives have to be included into the model file as if they were present. Then, no modification to the model file is needed during the capacity planning model execution.

     The experiment file (whose name is `SPL.EXP`) defines the experimental conditions under which the simulation model is exercised to generate specific output

Figure 1: Interfaces among decision levels. Dashed lines represent where the stochasticity is present.

data. Three other files are included into this one: `PARAMT.DAT`, `ARRIVL.DAT` and `REPLIC.DAT`. The first contains the parameters associated with the random variables used in the model file. The second includes the arrival scheme of entities entering into the production line at specified times, result obtained from the production scheduling module. The third contains the number of replications for each simulation run.

As a consequence, one only needs to change data in the experiment file to make simulation runs under varying conditions. Any modification of the experiment file requires the recompilation and relinking of the production SIMAN input files. This task can be done within the capacity planning model.

- The interface between the operations optimization module and the simulation model uses two files: `SBPINP.DAT` and `SBPOUT.DAT`. The first contains the current values of the operational decision variables needed by the simulation model and the second the production line attributes for the current simulation passed to the simulation optimization algorithm.
- The current investment decisions are passed to the simulation model into the `INSTAL.DAT` file.

The simulation is defined as terminating with the event "production scheduling plan for this period has been completed" defining the end. In this case the replications can be assumed to be statistically independent, by properly controlling the random number streams.

4. Replication of the previous simulation to minimize the number of simulations required to achieve a confidence interval for the operating variable cost small enough.

   This task can be accomplish in parallel. In that case it is necessary that not only a sequence of random numbers generated on each processor is of good statistical properties but also sequences generated on different processors are uncorrelated. Algorithms to generate sequences of random numbers with long periods on parallel processors are found on [11, 25].

   Iteratively compute mean and variance of the operating variable cost using the following expressions:

$$\bar{X}(n) = \frac{1}{n}\left[(n-1)\bar{X}(n-1) + X_n\right] \tag{17}$$

$$S^2(n) = \frac{1}{n-1}\left[(n-2)S^2(n-1) + \frac{n}{n-1}(\bar{X}(n) - X_n)^2\right] \tag{18}$$

   X_n value of the last observation (replication) n.
   $\bar{X}(n)$ sample mean with n observations. $\bar{X}(1) = X_1$
   $S^2(n)$ sample variance with n observations. $S^2(1) = 0$

Simulation replication will be stopped when the half-length of the confidence interval is less or equal to τ times the mean, being τ a prescribed tolerance:

$$t_{n-1,1-\alpha/2}\frac{S(n)}{\sqrt{n}} \leq \tau\bar{X}(n) \tag{19}$$

or rearranging

$$S(n) \leq \frac{\tau\sqrt{n}}{t_{n-1,1-\alpha/2}}\bar{X}(n) \tag{20}$$

$t_{n-1,1-\alpha/2}$ Student-t with $n-1$ degrees of freedom and $100(1-\alpha)\%$ confidence level.

In [18] a minimum value of replications $n \geq 10$ and a maximum value of tolerance $\tau \leq 0.15$ are recommended.

This sequential procedure has not been implemented due to SIMAN limitations and a fixed number of simulation replications is done instead.

5. Simulation optimization algorithm to assess the decision variables giving the optimal values of the attributes. That means repetition of points 3 and 4 many times.

   The Hooke-Jeeves algorithm has been used for the simulation optimization with the following steps:

   (a) Choose an arbitrary initial point $Y_1 = \{y_j\}$ and prescribed step lengths Δy_j in each of the coordinate directions $u_j, \quad j = 1 \dots J$. Set $n = 1$.
   (b) Compute $V_n = f(E, D, R, Y_1)$. Set $j = 1, Y_{n,0} = Y_n$ and start the exploratory move as stated in step c.
   (c) The variable y_j is perturbed about the current temporary base point $Y_{n,j-1}$ to obtain the new temporary base point as

$$Y_{n,j} = \begin{cases} Y_{n,j-1} + \Delta y_j u_j & \text{if} \quad f^+ = f(Y_{n,j-1} + \Delta y_j u_j) < f(Y_{n,j-1}) \\ Y_{n,j-1} - \Delta y_j u_j & \text{if} \quad f^- = f(Y_{n,j-1} - \Delta y_j u_j) < f(Y_{n,j-1}) < f^+ \\ Y_{n,j-1} & \text{if} \quad f(Y_{n,j-1}) < \min(f^+, f^-) \end{cases}$$

   This process of finding the new temporary base point is continued for $j = 1, \dots, J$ until y_j is perturbed to find $Y_{n,j}$.
   (d) Obtain the new point as $Y_{n+1} = Y_{n,J}$ and go to step e.
   (e) With the help of the base points Y_n and Y_{n+1}, establish a pattern direction S as

$$S = Y_{n+1} - Y_n \tag{21}$$

   and find a point $Y_{n+1,0}$ as

$$Y_{n+1,0} = Y_{n+1} + \lambda S \tag{22}$$

where λ is the step length which is taken 1 for simplicity.

(f) Set $n = n + 1$, $V_n = f(E, D, R, Y_{n,0})$, $j = 1$ and repeat step c.

(g) The process finishes when no further improvement can be achieved.

The buffer sizes are the decision variables and are assumed to be discrete.

The operating objective function to be minimized represents the variable costs including cost to run the workstations (energy, O&M) and land cost and has been been computed from the following production line attributes:

- Equipment utilization (including the time that the workstation was blocked) V_1
- Buffers size Y
- Makespan (time to achieve the prescribed orders) V_2

$$\min(V_1 c_1 + Y c_Y) V_2 \tag{23}$$

The equipment utilization and the buffers size are directly proportional to the O&M and surface cost per unit of time, respectively.

Information about the derivatives of the objective function with respect to the investment variables is drawn by increasing each investment variable by one and running again the simulation optimization algorithm. That perturbation can cause non feasibility in the constraints, if any.

6. Repetition of tasks 3, 4 and 5 the next period of the planning horizon. Machines are supposed to be idle and the production line empty at the beginning of each period.

7. Repetition of steps 2 to 6 for another series of demands.

7 Simple Production Line

In order to test the model previously presented a simple production line that schematically represents the characteristics of the electronic board production line at the Hewlett-Packard plant in Puerto Rico, described in [27], is used. Basically, it is a serial production line with different paths for the different classes of memory boards. The demand is fixed for each period and known in the beginning.

The layout of this simple production line is presented in figure 2. The line is composed by 6 machines and 6 buffers in front of them. All of them except the very first are limited in size. Two types of parts are flowing through the line with paths 1-2-4-5 and 1-3-4-6, respectively. The processing times for each machine are 10, 30, 20, 60, 40, 50 time units, respectively. No time is involved in transportation between machines.

Figure 2: Simple production line.

8 Remarks

The following remarks have been derived from the optimization of this simple production line with the objective function previously formulated. This kind of conclusions also has to be extracted from a real manufacturing system.

1. Important differences, both in value and sign, have been found among the derivatives obtained as perturbations at the optimal point when a further simulation optimization is allowed or not at this point. Of course, with deterministic parameters the operating costs at the new optimal point always are less or equal that those obtained with no optimization. With probabilistic parameters, due to fail in progress of the algorithm, may not always be the case.

2. Duplicating the machine speed is not equivalent to a duplication in the investment variable. If this could be assumed in general then the derivatives would be computed in a continuous way with different kind of techniques.

3. An increment in an investment variable may not imply an improvement in the objective function. Thus, the objective function is not monotonically decreasing for the investment variables as always happens in generation expansion planning.

4. The derivatives of the operating objective function with respect to the investment variables are about 15 to 20 %. A change in the production scheduling plan from 100 parts of type 2, 100 parts of type 1 to a plan of 1 part of type 2, 1 part of type 1 and so on until 200 parts are completed causes an increment in the objective function of about 5 % (the sign indicates that uniform scheduling is worse in this simple production line with the assumptions made).

5. A nonlinear effect in the objective function due to congestion has been found. A duplication in the demand causes an increment of 234 % in the objective function. Thus, an important reason for investment will be an expected increment in demand.

References

[1] Andradóttir, S. (1991) "A Projected Stochastic Approximation Algorithm" *Proceedings of the 1991 Winter Simulation Conference.* pp 954-957.

[2] Aoki, T., Nakayama, S., Yamamoto, M., Hashimoto, M. and Tanaka, J. (1991) "Combinatorial Scheduler: Simulation & Optimization Algorithm" *Proceedings of the 1991 Winter Simulation Conference.* pp 280-288.

[3] Azadivar, F. and Lee, Y. (1988) "Optimization of Discrete Variable Stochastic Systems by Computer Simulation" *Mathematics and Computers in Simulation.* Vol 30, No 4, pp 331-345.

[4] Bahl, H.C., Ritzman, L.P. Gupta, J.N.D. (1987) "Determining Lot Sizes and Resource Requirements: A Review" *Operations Research.* Vol 35, No 3, pp 329-345. May-June.

[5] Barton, R.R. and Ivey, J.S. (1991) "Modifications of the Nelder-Mead Simplex Method for Stochastic Simulation Response Optimization" *Proceedings of the 1991 Winter Simulation Conference.* pp 945-953.

[6] Biles, W.E. and Swain, J.J. (1979) "Mathematical Programming and the Optimization of Computer Simulations" *Mathematical Programming Study.* Vol 11, pp 189-207.

[7] Boender, C.G.E., Rinnooy, A.H.G. and Timmer, G.T. (1982) "A Stochastic Method for Global Optimization" *Mathematical Programming.* Vol 22, No 2, pp 125-140.

[8] Bulgak, A.A. and Sanders, J.L. (1988) "Integrating a Modified Simulated Annealing Algorithm with the Simulation of a Manufacturing System to Optimize Buffer Sizes in Automatic Assembly Systems" *Proceedings of the 1988 Winter Simulation Conference.* pp 684-690.

[9] Caramanis, M. (1987) "Production System Design: A Discrete Event Dynamic System and Generalized Benders' Decomposition Approach" *Production Research.* Vol 25, No 8, pp 1223-1234.

[10] Evans, G.W., Stuckman, B. and Mollaghasemi, M. (1991) "Multicriteria Optimization of Simulation Models" *Proceedings of the 1991 Winter Simulation Conference.* pp 894-900.

[11] Fushimi, M. (1989) "Random Number Generation on Parallel Processors" *Proceedings of the 1989 Winter Simulation Conference.* pp 459-461.

[12] Glynn, P.W. (1986) "Optimization of Stochastic Systems" *Proceedings of the 1986 Winter Simulation Conference.* pp 52-58.

[13] Glynn, P.W. (1986) "Stochastic Approximation for Monte Carlo Optimization" *Proceedings of the 1986 Winter Simulation Conference.* pp 356-364.

[14] Gujar, R., Desruelle, P. and Hammer, R. (1990) "Electronic Manufacturing Cell Design Using PC-based Modeling and Simulation Tools" *Proceedings of the 1990 Winter Simulation Conference.* pp 559-563.

[15] Hax, A.C. and Candea, D. (1984) *Production and Inventory Management.* Prentice-Hall. Englewood Cliffs, USA.

[16] Ho, Y. and Cao, X. (1991) *Perturbation Analysis of Discrete Event Dynamic Systems.* Kluwer Academic Publishers. Boston, USA.

[17] Jacobson, S.H. and Schruben, L.W. (1989) "Techniques for Simulation Response Optimization" *Operations Research Letters.* Vol 8, No 1, pp 1-9.

[18] Law, A.M. and Kelton, W.D. (1991) *Simulation Modeling and Analysis. Second Edition* McGraw-Hill. New York, USA.

[19] L'Ecuyer, P. (1991) "An Overview of Derivative Estimation" *Proceedings of the 1991 Winter Simulation Conference.* pp 207-217.

[20] Manz, E.M., Haddock, J., and Mittenthal, J. (1989) "Optimization of an Automated Manufacturing System Simulation Model Using Simulated Annealing" *Proceedings of the 1989 Winter Simulation Conference.* pp 390-395.

[21] Microsoft Corporation (1991) *Microsoft FORTRAN Reference.* Microsoft Corporation. Redmond, USA.

[22] Nandkeolyar, U. and Christy, D.P. (1989) "Using Computer Simulation to Optimize Flexible Manufacturing System Design" *Proceedings of the 1989 Winter Simulation Conference.* pp 396-405.

[23] O'Loughlin, M.J., Driskell, M.K. and Diehl, G. (1990) "Financial Simulation: Combining Cost Information in Systems Analysis" *Proceedings of the 1990 Winter Simulation Conference.* pp 578-581.

[24] Pedgen, C.D., Shannon, R.E. and Sadowski, R.P. (1990) *Introduction to Simulation Using SIMAN* McGraw-Hill. New York, USA.

[25] Reif, J.H. and Tygar, J.D. (1988) "Efficient Parallel Pseudorandom Number Generator" *SIAM Journal on Computing.* Vol 17, No 2, pp 404-411.

[26] Rodammer, F.A. and White, K.P. (1988) "A Recent Survey of Production Scheduling" *IEEE Transactions on Systems, Man, and Cybernetics.* Vol 18, No 6, pp 841-851. November/December.

[27] Rojas, J.C. (1992) *Simulation of a Just In Time Assembly Process for Computer Memory Cards at Hewlett-Packard Puerto Rico.* Eng. Thesis. Stanford University.

[28] Stuckman, B., Evans, G.W. and Mollaghasemi, M. (1991) "Comparison of Global Search Methods for Design Optimization using Simulation" *Proceedings of the 1991 Winter Simulation Conference.* pp 937-944.

[29] Systems Modeling Corporation (1989) *SIMAN IV Reference Guide.* Systems Modeling Corporation. Sewickley, USA.

[30] Suri, R. and Leung, Y.T. (1987) "Single Run Optimization of a SIMAN Model for Closed Loop Flexible Assembly Systems" *Proceedings of the 1987 Winter Simulation Conference.* pp 738-748.

[31] Wantuck, K.A. (1989) *Just-in-Time for America.* The Forum Ltd. Milwaukee, USA.