Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

DATA REQUIREMENTS FOR ANALYSIS OF MANUFACTURING SYSTEMS USING COMPUTER SIMULATION

Steve Hatami

CPC Center of Expertise – Simulation
General Motors of Canada Limited
1908 Colonel Sam Drive
Oshawa, Ontario, Canada

ABSTRACT

Simulation is often said to be more of an art than a science. One area where this is clearly illustrated is in determining the data requirements for a simulation project. Data and related concerns are all issues for which there are few standards or clear-cut rules. Although a great deal of literature is available on the analysis of input and output data (e.g., to determine the underlying statistical distributions, designed experiments, etc.), few practical guidelines and standards exist in the area of data requirements for conducting computer simulation studies.

Despite the subjectivity in the area of data requirements for simulation models, some relevant issues and modelling approaches are discussed in this document that may be of assistance in determining the information needed for simulation of manufacturing systems. The approach adopted in this work is based on the mapping of the objectives of simulation studies in manufacturing settings over their process lifecycle.

1. INTRODUCTION

Setting proper objectives is perhaps one of the most important steps in a simulation project. The objectives must reflect the need for understanding a system for the purpose of making decisions regarding its attributes with accuracy and efficiency. Furthermore, simulation analysis is most effective when it is applied during all phases of design; from the initial design concept selections to the continuous improvement efforts during the ramp up and on-going production phases. Hence, one must establish time-dynamic objectives that are representative of the various stages of a system's lifecycle.

Analyzing a system in order to evaluate various alternatives and to verify the design parameters may be sufficient for the initial phases of a system's design. The need for further analysis often persists and simulation studies are needed as part of efforts in improving a system's performance and productivity, once it has been implemented. As the system is implemented and experience is gained from it, fine tuning of the system's performance may be required through what-if analysis on the simulation model. At this stage in the process lifecycle, the model may be required to be updated with the newly observed data (i.e., machine reliability data and their underlying statistical distributions) to better resemble the real system (moving from simulation to near emulation). Only then can the simulation model be used as a tool for identifying areas for improvement, relative to the overall system performance.

The level of accuracy of a particular simulation model relies on the level of detail that is included in it. The level of detail required for a model in turn determines data requirements. In practice, one must weigh the marginal value of the additional accuracy and its potential impact on the decisions versus the additional effort and cost to collect the data.

In this document, the approach that is used to determine the data requirements is based on mapping simulation modeling over the process/system lifecycle. This is achieved by establishing objectives for simulation studies for: the planning stages (phase 1), approval of the detailed plans (phases 2), and during production and continuous improvement stages (phase 3) of a manufacturing program. Using this approach, a recommended level of model detail and data requirements for different stages are provided.

2. DATA REQUIREMENTS FOR ANALYSIS DURING PLANNING (PHASE 1)

During the early stages of the introduction of a new system, various manufacturing concepts and processes are examined. A substantial level of detail is not available or finalized yet to be accurately modeled. Furthermore, during the planning phase of a new project, simulation models must be time responsive with fairly short turn-around times so that they will be most effective in analyzing the potentially large number of alternatives. The objective of the simulation project, therefore, should be to *provide insight by evaluating alternatives to complete manufacturing and assembly plans for concept approval.*

On the basis of the above discussion, simulation models at this level of the process lifecycle should be made simple by making assumptions about some of the unknowns, emphasizing only those factors which are suspected to have the greatest impact on the system's performance. In many manufacturing setups, the key parameters that best explain a system's behaviour (both in terms of inherent random variations and dynamic interactions) are machine reliability, speed, and buffer capacities. As a result, the data requirements should be reflective of the focus of the model at this early stage and in terms of the key system variables. For every alternative being considered, the following information may be needed:

2.1 Process Flow Diagram

The purpose of the process flow diagram is to illustrate the scope of the project as well as the flow of the parts through the system. This graphical description summarizes the entire system that is to be modelled through the use of different (but consistent) symbols for operations, storage spaces, and material movements. Although the process flow diagram can be detailed to an extensive level, it is best to maintain simplicity and avoid excessive descriptions on this diagram.

2.2 Storage Space Capacities

These are the proposed buffer sizes between successive machines. It should be pointed out here that, one of the reasons for conducting a simulation study is to conduct what-if analysis on those parameters of the system that best capture the essence of a system. During planning stages of a manufacturing system, a great deal of effort is often focused on determining acceptable buffer sizes. This implies evaluation of various inventory/manufacturing concepts (e.g., just in time) in terms of cost of inventory versus throughput and efficiency. This evaluation is accomplished using throughput statistics as well as other measures of a systems performance such as the average time and average number of jobs in the system. Hence, proposed buffer sizes must be viewed as a starting point and models should incorporate flexibility in this area for many what-if analysis.

In order to improve the time responsiveness of simulation models at all phases of design (and particularly at the early stages), a machine could be defined as more than one machine, as long as there are zero or very small buffers between them. This implies that lines that are synchronous in nature (e.g., machining transfer lines and press lines) could be treated as one machine due to the fact that as one of them fails, the operation of all of them is halted within their respective cycle times. If a number of machines or stations are treated as one machine, it is important to know the number of stations/machines grouped together as a number of parts may be stored in those stations or machines and thereby change the effective output buffer capacities of proceeding machines.

It must be stated that for more detailed analysis (i.e., for analysis during phases 2 and 3) fitting downtime data to statistical distributions is necessary. Stratification of data by machine/stations may be required if multimodal distributions are observed.

2.3 Machine/Operator Speeds

Speed is simply the rate at which the processes are expected to operate at. At this level, cycle times can be assumed to be constant from cycle to cycle unless the processes have been sufficiently defined so as to be able to determine whether they will be automatic or manual operations. For machines with automatic cycles (or those whose manual portion of the cycle is paced by automatic processes), a constant cycle speed can be assumed. For operations involving manual processing, cycle times may vary considerably from cycle to cycle. This information is extremely important to incorporate into the model as the variability in time to complete a task can have a dramatic impact on the results.

In the absence of observed data, triangular distributions can be assumed to provide a rough representation of variable processing times [Law and Kelton 1982]. The data needed to reflect the variability in the cycle times is as follows:

i- Minimum time for completion of the task.
ii- Maximum time for completion of the task.
iii- Most likely time for completion of the task.

2.4 Information on the Processes

The purpose for this information is to ensure that the analyst has a good understanding of the operations and can, therefore, make appropriate inquiries and make reasonable assumptions about the system.

2.5 Data on Machine Reliability

During the initial phases of the design when no experience with the real system is available, assumptions must be made regarding the statistical distributions that best characterize the reliability of the manufacturing equipment. Exponential distribution can be assumed for modeling time to failures and time to repairs [Banks and Carson 1984; Thesen and Travis 1989]. Furthermore, the choice of this particular distribution is particularly attractive as it requires estimation of only one parameter, the mean.

2.5.1 Time To Failure/Machine Uptime

Machine breakdown information can be obtained in a number of ways. It is important to ensure that the simulation model generates downtimes similar to the manner in which they have been collected. For example appropriate measures must be taken in the model to ensure downtimes are scheduled when a part is being processed and the machine is not idle.

2.5.1.1 Mean Jobs or Mean Time Between Failures

Machine reliability can be represented by average (or mean) number of jobs between successive equipment failures (MJBF). The average number of jobs between failures provides a very good representation of a machine's reliability as it excludes those times during which the machine was interlocked (i.e., starved or blocked) or was being repaired. Moreover, if a machine's speed is changed during the data collection period, it would not effect the accuracy of this measure since it is independent of the cycle time. Multiplication of the machine cycle time by the MJBF would result in Mean Time Between Failures (MTBF).

MTBF = MJBF * Cycle Time

2.5.1.2 Percent Uptime

Machine reliability may be expressed by expected (or quoted) percent uptime. This is the most common form of expressing machine reliability when prior experience or data on that particular machine or technology is not available. When expressing uptime, stating the basis of the calculation is important. The percent uptime based on total time has a different meaning than that based on total processing time. The former includes blocked and starved times, and hence, is not a very good measure of reliability. The latter, however, excludes those times during which the machine was not operating due to part shortages or blocking and as a result it is a more accurate representation of a machine's reliability. Mean time to failures can be calculated by knowing the percent uptime and the average time to repair according to the following formula:

MTBF = {100*MTTR/(100-%uptime)} - MTTR

Note that the above relationship excludes the time to repair from the calculation of the mean time to failures.

2.5.2 Mean Time to Repair

This is the average time it takes to repair maintenance related faults. One modeling approach would be to define MTTR such that it would include the times during which parts are processed on an intermittent basis as part of the machine diagnosis. For example, a machine is taken off the automatic cycle mode and is put on the manual cycle, parts are processed and checked for quality related or other characteristics one part at a time. The time spent in checking the part characteristics must be treated as the machine diagnosis time and hence included as part of the repair time.

3. DATA REQUIREMENTS FOR ANALYSIS DURING THE DESIGN FINALIZATION (PHASE 2)

At this level, the engineering and design of all manufacturing and assembly facilities and systems will be completed. During phase 2, as more data becomes available, additional details of the operations and downtime data must be added to the simulation model to improve its accuracy. Accordingly, the objective of the analysis should be to *assist in the finalization of the manufacturing and assembly processes by evaluating the system's performance in more detail.*

At this stage of the process lifecycle, the simulation models could become fairly elaborate and as a result require additional input information. With respect to the processes involved, the following information would be needed in addition to those gathered during phase-1 and the updates made in response to changes in system's parameters:

3.1 Product Mix and Schedules

Product mix and option dependencies must be known if they impact the processing speed or the product process flow. Information on product mix would include percentage of each type, their respective processing times, and process flow diagram for each part type. Schedules must also be known so that the proper algorithm is incorporated into the model.

3.2 Facilities Layout

The facilities layout can assist the analyst in better understanding the system in terms of its physical space, distances, parts flow, etc. Furthermore, the layout can be used in creating an animation of the model. CAD IE layouts can be easily incorporated into some animation programs if they can be supplied in the *.DXF format (Data eXchange Format).

3.3 Operating Philosophies

This refers to those production strategies which may have an impact on the behaviour of the system and should be incorporated in the model. Some examples are: repair rules, mass/tag relief, etc.

3.4 Material Handling Systems

With respect to the material handling, the objectives of the simulation study, the nature of the system, and characteristics of the material handling system determine the data requirements. Many simulation studies may be focused on the processes themselves and based on the belief that an acceptable approximation of the overall system can be achieved by studying the effects of the above mentioned process related factors.

A number of situations exist under which explicit modeling of material handling systems may not be needed or knowing the relative distances and speed, or travel times, may be sufficient. Many forms of material handling systems are sufficiently reliable so that they have little impact on the system's behaviour. These situations are observed in systems that employ more traditional and less sophisticated material handling concepts (e.g., fork lift transporters). In some cases, even if the reliability of the handling equipment cannot be assumed, through a good understanding of the system, the modeler may conclude that few interactions can be attributed to the characteristics of the material handling systems. In such scenarios, one can estimate the effect of the material handling systems on the overall system performance without recourse to a detailed representation.

Very often, however, studying the material handling system itself is of interest. It may be felt that the handling system can have a great impact on the performance measures of interest. This is especially evident in flexible manufacturing systems where more recent and sophisticated technologies (e.g., AGVs) are employed. In such cases, a great deal of information would be needed which is determined by the proposed type(s) of handling systems. The data requirements for effective modeling of such systems deserve special considerations which is outside of the scope of this work.

4. DATA REQUIREMENTS FOR ANALYSIS DURING THE RAMP UP AND ON-GOING PRODUCTION (PHASE 3):

This phase represents the start up and on-going production operations. At this level, all necessary details should have been incorporated into the model and the model validated to assist the plant in making improvements to their operations. Also, a great deal of what-if analysis may be performed at this stage of the design cycle to determine areas of improvement that will have the most global impact. Therefore, the objective of the analysis should be to *validate the model against the system and assist the manufacturing personnel in their continuous improvement efforts.*

In order for the model to be of assistance in the analysis of production systems during start-up, the data representing the initial conditions must be collected from the operations and incorporated into the model. The type of data that is required is exactly the same as collected previously. Changes that have been implemented, however, must be accurately reflected in the model. For example, during the acceleration period machines may run at lower speeds than planned to ensure the quality of products, and repair times may be longer as the learning curve peaks.

One important step in this phase is to validate the model. This involves comparing results of the simulation to those of the actual system under the same operating conditions. If the model was successfully validated, no further data would be required and it can be used for analysis of any future proposals. If, however, the model is found to report different statistics (e.g., throughput, etc.) than those observed from the real system, then further steps must be taken to investigate the sources of inconsistency.

One of the most probable sources of inconsistencies is the validity and accuracy of the data or assumptions about them. This may be more true in the case of data reported through automatic data collection and reporting systems.

In order to perform detailed sensitivity analysis (e.g., concentrating on individual stations) or investigate data related inconsistencies, a substantial amount of data may be required. Ideally, due to the volume of the data, they would be collected by the automatic collection systems and reported in the form of an ASCII file. Two types of analysis and their respective recommended data are discussed below. Data requirements of the first approach can also be used for Pareto analysis and performing sensitivity analysis on sources of lost times (as in studying effects of preventive maintenance efforts).

4.1 Lost Production Time Data Analysis

Using this approach, the integrity of data can be checked by evaluating lost production times and relating them to the cycle times according to mathematical relationships. For this type of analysis, in addition to knowledge of the cycle times, information on lost times and their causes is needed. In many manufacturing settings there are three general sources of lost production times: starved times, blocked times, and production time lost due to maintenance related faults. Lost times on a section of the system for any given period of time can, therefore, be compared to the throughput observed on that machine or line through its cycle time. This form of analysis assumes that starved, blocked, and downtimes are mutually exclusive of one another and hence they are additive. For example, if machine X operates at 60 jobs per hour and for the nth hour experiences 22 minutes of lost production time, we can conclude that it should have produced roughly 38 jobs.

This analysis, due to the linear relationship between its dependent and independent parameters, can be extended over a sufficiently large number of time intervals to build a multiple regression model [Hatami 1989]. The theoretical regression equation is of the following form:

$$Throughput = Machine\ Speed * \{1 - \Sigma(Starvation\ Times) - \Sigma(Blocking\ Times) - \Sigma(Maintenance\ Times)\}$$

Observed values are substituted into the above relationship and correlation coefficients between the dependent and the independent variables are computed. Low correlation coefficients indicate inconsistencies in the data.

Also note that if a particular line is operating in a synchronous manner, then all of the machines in that line can be grouped and treated as one machine for checks on accuracy of the data.

A detailed description of the fault, (e.g., broken tool, bad weld tip, etc.) would be needed if sensitivity analysis is to be performed on individual types of faults (as in studying the effect of preventative maintenance). Obviously, if needed, the information gathered on stations and causes of failure can provide better estimates of mean time between failures and their underlying statistical distributions.

4.2 Throughput Data Analysis

This form of analysis will analyze the throughput, comparing it on a line to line basis. This approach is particularly efficient and useful in environments where limited buffer spaces are allowed. The recommended procedure is to collect throughput statistics on all of the machines or lines for a sufficient number of production hours and during the same time frame. In this manner, average hourly throughputs of various machines/lines can be statistically compared to one another to test for any statistically significant differences.

For example, suppose machine A feeds into machine B and that the calculated average hourly throughputs for machines A and B are 55 and 58, respectively. This implies that, if we had calculated those statistics over the course of 20 hours, then there must have been at least 60 parts in the buffer after machine A and before machine B, prior to collection of the data. If we know that there were not that many parts in the buffer at the start of the data collection period, or the maximum buffer capacity between machines A and B is actually less than 60, then we can conclude that the reported throughput data has flaws.

5. CONCLUSIONS

Analysis of manufacturing systems using computer simulation models requires data which must be accurate and representative of the systems that are being modeled. While collection and compilation of data for simulation may appear to be tedious or time consuming, most or all of the required data are those that would be needed, irrespective of the tool, to perform an accurate and effective analysis of the system. For example, downtime data collection on individual machines must be performed so that proper allowances are included in parts' labour routings, calculation of standard hour, and line balancing.

As was mentioned at the outset of this report, the data requirements of any specific simulation study may differ from the ones suggested here. Data requirements must, however, be based on the objectives of a particular study and in the absence of meaningful data, assumptions must be made. The purpose of this paper has been to provide a set of guidelines and a framework for determining data requirements on the basis of a system's lifecycle.

It should also be emphasized again that the accuracy of the input data is vital to the success of any analysis in general and simulation models in particular. Although the accuracy of simulation models that use reliable data cannot be guaranteed, the inaccuracy of those that rely on inaccurate data can be. Any time spent in proper collection and reporting of the data will be well worth the effort. With the introduction of more automatic data collection systems into many manufacturing environments, the need for testing the validity of the reported data are stressed.

ACKNOWLEDGEMENTS

The author is grateful to Ms. C. Morey, Misters E. Cowley, and J.T. Blake for their valuable input and time to review this document.

REFERENCES

Banks, J. and J.S. Carson (1984), *Discrete-Event System Simulation*, Prentice Hall, Englewood Cliffs, NJ.

Hatami, S. (1989), "Oshawa Truck Plant Body Shop Downtime and Throughput Analysis", Project Report, CPC Center of Expertise-Simulation, Oshawa, Ont.

Law, A.M. and W.D. Kelton (1982), *Simulation Modeling and Analysis*, Mc Graw-Hill, New York, NY.

Thesen, A. and L.E. Travis (1989), "Simulation For Decision Making-An Introduction", In Proceedings of the 1989 Winter Simulation Conference, E.A. MacNair, K.J. Musselman, and P. Heidelberger, Eds. IEEE, Piscataway, NJ, 9-18.