Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

USING INPUT PROCESS INDICATORS FOR DYNAMIC DECISION MAKING

Michael Freimer

School of Operations Research
and Industrial Engineering
Cornell University
Ithaca, NY 14853, U.S.A.

Lee Schruben

Department of Industrial Engineering
and Operations Research
University of California at Berkeley
4135 Etcheverry Hall
Berkeley, CA 94720-1777, U.S.A.

ABSTRACT

In a continually changing environment, a simulation study that integrates the activities of data collection, model analysis, and decision making has some distinct advantages. In this paper we look at simulation project dynamics from a high-level and examine some ways for integrating these activities. From this perspective, the processes used to drive a simulation model are forecasts of environmental changes, and the parameters for models of these processes are viewed as leading indicators. A simple decision-making scenario having some of the characteristics of semiconductor manufacturing is used to illustrate the ideas.

1 INTRODUCTION

From a high-level perspective, a simulation study consists of data collection, model analysis, and decision making. Each of these is generally viewed as a separate activity, the interactions of which are usually not considered. The goal of this paper is to provide a framework for integrating these activities, with the aim of producing better decisions. The particular example we develop is motivated by the problem of tool set selection in semiconductor manufacturing, but the ideas generalize to any situation where simulations are used to make decisions in a rapidly changing environment.

The following simple decision-making scenario is used for illustration. Consider the situation where we must make a choice between two alternative system configurations sometime in the future. The system is currently configured as *A*, and by some future date, T_d , (which, for simplicity, we will assume is fixed) we must decide whether to switch the configuration to *B* or to maintain configuration *A*. We would like to choose the configuration that will maximize the expected performance of the system (e.g. profits) over a finite time horizon [T_d, T_h], conditioned on the history of the system up to time T_d. Figure 1 shows a timeline of this decision process. T_0 represents the current time.

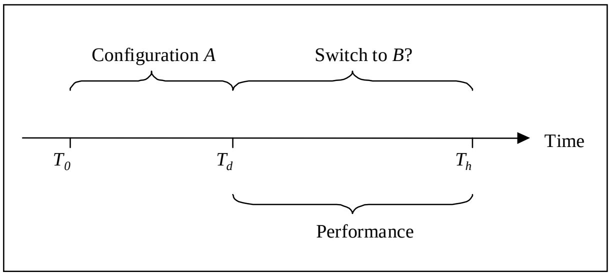

Figure 1: Timeline of the decision process.

We will make the decision between configurations *A* and *B* based on the results of a simulation study. The study (data collection, simulation execution, and output analysis) typically requires large amounts of time, on the order of weeks or months. The current state of practice tends to separate data collection from the other tasks. In other words, during the first stage of the study, say from T_0 to T_{sim}, data is collected or a data base is updated and queried to parameterize the model. (See Figure 2.) Simulation execution and output analysis are performed from T_{sim} to T_d, at which point the decision is made. This means that the decision is based only on observations of the real world up to time T_{sim}, which may be some time before T_d.

Under the presumption that the quality of the decision increases with the amount of real-world data, there is a tendency to delay simulation execution and output analysis (and hence the decision) as long as possible. We believe there is value in having some information about the decision earlier than T_d. For example, if early conditions look favorable, we may wish to expedite the decision. If conditions are less favorable, we may wish to delay the decision and search for other alternatives. In particular, we would like to take into account conditions right up to time T_d.

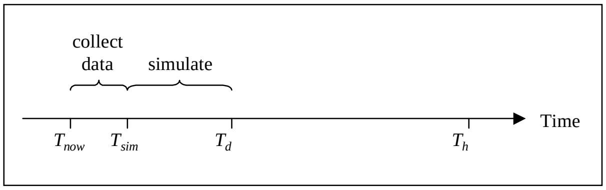

Figure 2: Revised timeline of the decision process.

A sequential approach to analysis is cumbersome if we desire up-to-date information about the decision as conditions change over the interval T_0 to T_d. We are required to repeatedly parameterize and execute the simulation model. An alternative is to make multiple runs of the simulation model at the beginning of the study (i.e. at time T_0) using a range of parameter values. We calculate various "leading indicators" (e.g. estimated demand rate) over the course of each sample path, up to simulated time T_d. Then at any time T_i between T_0 and T_d, we can construct a regression model to predict performance over the interval $[T_d, T_h]$ based on the values of the leading indicators over the interval $[T_0, T_i]$. By calculating the actual values of the leading indicators from the real-world data, we can predict the expected performance, conditional on the environment up to time T_i. The idea is similar to response surface methodologies for predicting simulation results, but for input processes rather than decision variables.

A difficulty is deciding how to select the parameter values to use for the simulation runs at time T_0. One might sample from an appropriate posterior distribution function, given any available information. Another difficulty is choosing appropriate leading indicators. When there is no dependency in the input process, the indicators can be functions of the current system state and the parameter values used to generate the sample path. (The real-world indicators would use estimates of the parameters.) When there is dependency, the indicators can also be functions of the history of the system up to the current time.

We hope to make these ideas more concrete with the example described in the next section. We describe how the standard approach and the leading indicators methodology can be applied to this example and compare the quality of decisions made with each methodology at various time points between T_0 and T_d.

2 EXAMPLE: TOOL SELECTION IN A SEMICONDUCTOR FAB

We consider a simple example that has some of the characteristics typical of the semiconductor industry. This industry is characterized by highly uncertain demand for products, extremely expensive equipment, and long lead times for ordering equipment. The economic life cycle of a new product may only be a matter of several months and consists of a ramp-up phase, a mature phase, and a ramp-down phase. (See Figure 3.) Production strategies may be build-to-order since inventory rapidly becomes obsolete.

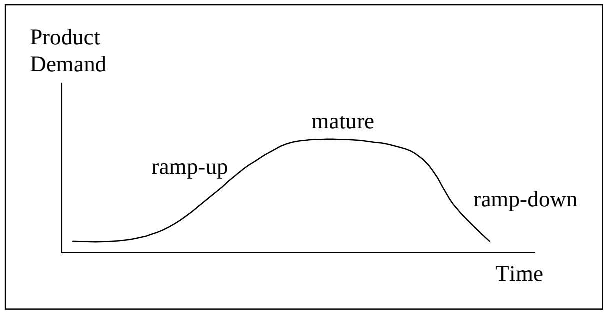

Figure 3: Economic life cycle of a new product.

In this example the manufacturer has started to experience demand for a new product, at the beginning of its ramp-up phase. The factory has some existing capacity for manufacturing the new product, however by time T_d the manufacturer must decide whether to order another tool set for the factory or to make do with the existing equipment. If a new tool set is ordered, it goes online immediately. The current configuration (A) for our model factory consists of five parallel bottleneck tools, each with an exponential processing rate μ. Configuration B is obtained by adding a sixth tool to the system.

The manufacturer is concerned with the factory's ability to satisfy demand during the ramp-up phase of the new product, when the market allows them to ask a high price for the new product. In this analysis T_h is the end of the product's ramp-up phase, the time at which demand enters the mature phase. (Here we consider T_h to be fixed, although it would more realistically be an uncertain quantity.) We would like to minimize a cost function over the period $[T_d, T_h]$, which includes the cost of the tool set (if purchased). Since the manufacturer would like to fill orders quickly, the cost function also penalizes long lead times (the delay between an order's arrival and fulfillment). The cost function is:

$$\begin{aligned} \text{cost} = {} & c_1 \times 1_{[\text{purchase tool set}]} + \\ & c_2 \times (\text{average lead time of orders received}) \end{aligned}$$

For this example we take $c_1 = 0.5$ and $c_2 = 1$.

Orders for the new product arrive according to a nonstationary Poisson process with arrival rate $\lambda(t)$. We model $\lambda(t)$ with a logistic function:

$$\lambda(t) = \gamma\left(\frac{e^{\alpha + t\beta}}{1 + e^{\alpha + t\beta}}\right).$$

In this example the "actual" order arrival process has values -6, 0.1, and 5 for α, β, and γ. (The arrival rate $\lambda(t)$ and the integrated rate function $\Lambda(t)$ are shown in Figures 4 and 5.) We suppose that the manufacturer knows γ (the size of the market) but must estimate α and β for the simulation study. We describe an estimation procedure for these parameters in the next section. The values for T_d and T_h are 80 and 120 respectively.

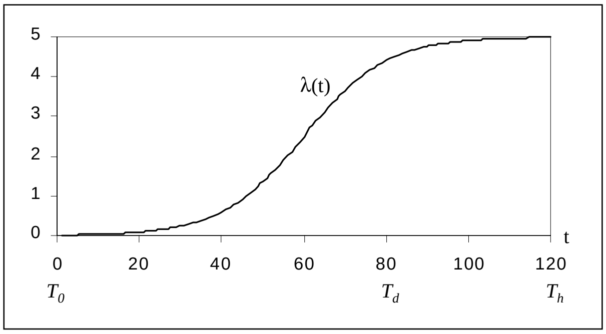

Figure 4: The arrival rate function, $\lambda(t)$.

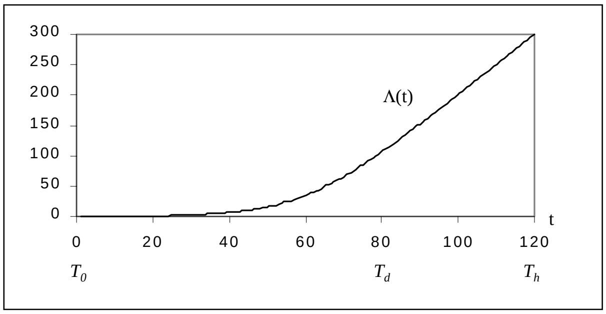

Figure 5: The integrated rate function, $\Lambda(t)$.

To understand how the decision is affected by the values of α and β, we simulated the system at a variety of parameter settings. Each replication had 5 servers up to time T_d. Using CRN we then split the replication, either keeping 5 servers (the no-purchase decision) or adding one (the purchase decision). Table 1 shows some of the results. A "Y" means the purchase decision had the lower mean cost, and an "N" means the no-purchase decision had the lower mean cost. The actual parameter values (-6, 0.1) correspond to a purchase decision.

Table 1: Purchase decisions at different values of α and β, with $\gamma = 5$.

α,β	0.1	0.2	0.3
-10		Y	
-9	N		Y
-8		Y	
-7	N		Y
-6		Y	
-5	Y		Y

3 FORECASTING THE ORDER ARRIVAL PROCESS

At various times T_i during the simulation study we must estimate the parameters α and β from the actual order arrival process. Suppose that there were n order arrivals during the interval $[T_0, T_i]$. Let $t_0 = T_0$, and let $t_1,\ldots,t_n$ be the times of the order arrivals between T_0 and T_i. The arrival rate has the form

$$\lambda(t) = \gamma\left(\frac{\exp\{\alpha + t\beta\}}{1+\exp\{\alpha + t\beta\}}\right),$$

where α and β are values that we wish to estimate. The integrated rate function is

$$\Lambda(t) = \int_0^t \lambda(a)da = \frac{\gamma}{\beta}\ln\left(\frac{1+\exp\{\alpha + t\beta\}}{1+\exp\{\alpha\}}\right).$$

Define random variables X_i as:

$$\begin{aligned} X_i &= \Lambda(t_i) - \Lambda(t_{i-1}) \qquad i = 1,\ldots,n \\ &= \frac{\gamma}{\beta}\ln\left(\frac{1+\exp\{\alpha + t_i\beta\}}{1+\exp\{\alpha + t_{i-1}\beta\}}\right) \end{aligned}$$

The X_i's are exponentially distributed with mean 1. We obtain moment estimators $\hat{\alpha}$ and $\hat{\beta}$ by solving the following equations for α and β:

$$1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\gamma}{\beta} ln\left(\frac{1+exp\{\alpha + t_i\beta\}}{1+\exp\{\alpha + t_{i-1}\beta\}}\right)\right],$$

and

$$2 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\gamma}{\beta} ln\left(\frac{1+exp\{\alpha + t_i\beta\}}{1+\exp\{\alpha + t_{i-1}\beta\}}\right)\right)^2.$$

The solution may be approximated using, for example, the Gauss-Newton method.

Other approaches for estimating the parameters of Poisson processes are described by Cox and Lewis (1966). For a recent paper on the subject, refer to Kuhl, Damerdji, and Wilson (1977).

4 APPLYING THE LEADING INDICATORS METHODOLOGY

To investigate the performance of the leading indicators methodology, we simulated its use on the example described in the previous sections. For each replication we generated an "actual" arrival sample path using the values (-6, 0.1) for (α,β). At timepoints $T_1 = 40$, $T_2 = 60$, and $T_3 =$

80 both the leading indicators method and the standard approach were used to make a decision. We first describe the implementation of the two methods, then their relative performance at the three timepoints.

The first step of the leading indicators methodology is to generate sample paths for the regression model. For each study we made 100 simulation runs, drawing values for α and β from U(-10,0) and U(0,1) distributions, respectively. Regression models were then constructed for each timepoint T_i, i = 1,2,3 to predict the difference in performance between the two system configurations over the interval $[T_d, T_h]$. The independent variables (leading-indicators) were α, β, and the system state (number of free servers and number in queue) at time T_i. Finally, at times T_i, i = 1,2,3 the regression models were used to make the purchase decisions. Estimators $\hat{\alpha}$ and $\hat{\beta}$ (based on the actual data up to time T_i) were used for α and β.

For the standard approach, estimators $\hat{\alpha}$ and $\hat{\beta}$ were calculated at timepoints T_i, i = 1,2,3 based on the actual data available up to time T_i. We then used these values to generate 100 simulations of the system over the interval $[T_i, T_h]$ and chose the configuration with the lowest average cost.

In order to determine the correct decision for each replication, conditional on the "actual" sample path up to time T_i, we also simulated each system configuration over the interval $[T_i, T_h]$ using the actual parameter values α = -6 and β = 0.1. Table 2 shows the frequency of correct decision for each method at timepoints T_1, T_2, and T_3. The table also shows the average squared difference between the predicted performance over the interval $[T_i, T_h]$ and the expected performance given the correct parameter values α = -6 and β = 0.1. As expected, the quality of decisions made with the standard approach improves with the amount of data. The leading indicators approach works reasonably well early in the time horizon, meaning that it may be a useful tool for providing early information about the decision.

Table 2: Relative performance of methodologies.

	Standard Approach (Leading Indicators)		
Early Decision Time	T_1 = 40	T_2 = 60	T_3 = 80
% correct	33 (81)	81 (76)	90 (81)
MSE	0.33(0.07)	0.05(0.04)	0.03(0.04)

5 COMMENTS AND TOPICS FOR FURTHER RESEARCH

The leading indicators method describe here seems promising as a tool for providing early information about decisions. Further investigation is required concerning the choice of leading indicators and the method's performance on problems with dependent arrival processes. Another area of interest is the method's relationship to the Bayesian approach to input distribution selection developed by Chick (1999). Rather than point estimates for α and β, one would use an appropriate posterior distribution based on data collected up to the time of the decision. For example, suppose we start with prior $\pi(\alpha, \beta)$. Since the X_i's are iid exponential and

$$X_i = \frac{\gamma}{\beta} ln\left(\frac{1+exp\{\alpha+t_i\beta\}}{1+\exp\{\alpha+t_{i-1}\beta\}}\right),$$

we have

$$f(t_1,\ldots,t_n \mid \alpha,\beta) = \prod_{i=1}^{n}\left(\frac{\gamma\exp\{\alpha+\beta t_i\}}{1+\exp\{\alpha+\beta t_i\}}\left(\frac{1+\exp\{\alpha+\beta t_i\}}{1+\exp\{\alpha+\beta t_{i-1}\}}\right)^{\frac{-\gamma}{\beta}}\right).$$

The posterior distribution is then:

$$h(\alpha,\beta \mid t_1,\ldots,t_n) = \frac{f(t_1,\ldots,t_n \mid \alpha,\beta)\pi(\alpha,\beta)}{\iint f(t_1,\ldots,t_n \mid \alpha,\beta)\pi(\alpha,\beta)d\alpha\, d\beta}$$

The leading indicators would be appropriate functions of this distribution.

ACKNOWLEDGEMENTS

The research reported here was partially supported by a joint src (fj-490) and nsf (dmi-9713549) research project in semiconductor operations modeling.

REFERENCES

Chick, S. E. 1999. Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty. Submitted for publication.

Cox, D.R. and P.A.W.Lewis. 1966. *The Statistical Analysis of Series of Events.* London: Methuen & Co.

Kuhl, M.E., H. Damerdji, and J.R. Wilson. 1997. Estimating and Simulating Poisson Processes with Trends or Asymmetric Cyclic Effects. In *Proceedings of the 1997 Winter Simulation Conference*, ed., Sigrun Andradottir, Kevin Healy, David Withers, and Barry Nelson, 287-295. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

AUTHOR BIOGRAPHIES

MICHAEL FREIMER is a Ph.D. candidate at Cornell University's School of Operations Research and Industrial Engineering. He received his undergraduate degree in mathematics from Harvard. Prior to attending graduate school he worked for an operations research consulting firm, Applied Decision Analysis, Inc. in Menlo Park, CA. His research interests are in simulation modeling and revenue management.

LEE SCHRUBEN is a Professor in the Department of Industrial Engineering and Operations at the University of California at Berkeley and is currently on a leave of absence from Cornell University where he is the Andrew S. Schultz Jr. Professor of Industrial Engineering. His research interests are in statistical design and analysis of simulation experiments and in graphical simulation modeling methods. His simulation application experiences and interests include semiconductor manufacturing, dairy and food science, health care, banking, and the hospitality industry.