OPERATION STRATEGIES FOR A SHOE BATCH MANUFACTURING SYSTEM

by

PAULO FERNANDO FLEURY DA SILVA E SOUZA

A Doctoral Thesis

Submitted in partial fulfillment of the requirements
for the award of

The Degree of Doctor of Philosophy of the
Loughborough University of Technology
September 1976

Supervisor: Dr. E. A. Roberts, M.A. Cambridge, PH.D. Nottingham
Department of Engineering Production

Acknowledgements

I would like to thank the following people for making the execution and completion of this work possible.

Dr. E.A. Roberts, my supervisor, for his help and guidance throughout the work.

Professor R.J. Sury, Head of the Department of Engineering Production, for allowing me the use of the facilities of his Department.

Professor T. M. Husband for his encouragement and guidance at the initial phase of the work.

All staff and fellow colleagues in the Department of Engineering Production for their friendship.

All staff of the Loughborough University Computer Centre, and in particular, Dr. B. Negus, for their cooperation and help.

The Brazilian Government, which through the Federal University of Rio de Janeiro (UFRJ) and the 'Coordenação de Aperfeiçoamento do Pessoal de Nível Superior'(CAPES) gave me the financial support.

Miss N. Shimizu for the invaluable encouragement and help in the final preparation of this report.

SYNOPSIS

A computer simulation model has been developed of a manufacturing system which produces a multi-style, multi-size product range, and which utilizes multiple stations machines. Manufacturing systems of this type can be found in the shoe and textile industries.

The model has been used to examine such a system in relation to its major variables, and under different operating and control rules.

The utilization of a factorial design has allowed the identification of the main effects and interactions between the major variables, which provided information about the mechanisms governing the system's behaviour.

Heuristic priority scheduling rules have been specially developed to fit the characteristics of such a system. Those rules have been tested against other priority rules which are known to perform well in more traditional batch manufacturing systems. A simple priority rule developed for this class of systems (FIFOMB) was shown to perform best in relation to the other rules.

Different strategies for capacity manipulation have been studied, both in terms of variable inputs, such as inventory, additional shifts, and overtime, and also in terms of capital input such as the acquisition of extra tools and machinery. Trade-off curves have been constructed, which allow operational decisions and comparisons between strategies,

to be made in a three dimensional decision grid, which have capacity costs per unit produced, delivery performance, and length of delivery promises as its parameters. Results indicated that the relative performance of different strategies depend on the values of the parameters chosen in the three dimensional decision grid.

CONTENTS

CONTENTS (CONT.) Page

CONTENTS (CONT.) Page

CONTENTS (CONT.) Page

CONTENTS (CONT.)

CONTENTS (CONT.) Page

CONTENTS (CONT.)

CONTENTS (CONT.)

CHAPTER 1

INTRODUCTION

1.1 - Initial considerations

The planning and control of manufacturing systems in general and of batch production in particular, have attracted the attention of research workers over the years. The advent of computers and the development of computer simulation techniques in the past 15 to 20 years, have created new scope for investigation which makes possible the study of the more complex problems of batch production systems, which could not be dealt with by analytical methods.

A large number of problem areas in batch production have been studied through the use of computer simulation models, chief amongst them are the problems of job scheduling, inventory control and design of production lines. In order to allow the results of the studies to be transferable to as wide a range of situations as possible, many research workers have used abstract models, which are designed to represent a general class of batch production systems. Unfortunately some batch production systems have characteristics so different from the ones assumed in those abstract models, that conclusions obtained from them cannot be easily transferred. One such class of systems can be found, for example, in sectors of the shoe and textile industries.

In this study a class of production systems producing in batches and having unique characteristics, is investigated through the use of a computer simulation model. Three of the major characteristics which makes this class of production system unique among the general class of batch production systems are:

i) The pattern of demand, which is characterized by the fact that orders arriving at the system require the production and quick delivery of a multi-size product range. The quantities required for individual product sizes being different from each other.

ii) The manufacturing units composed of multiple station machines which are able to manufacture different products at a time, and which require setting up.

iii) The tooling requirement which is characterized by the fact that every product requires a special tool (mould). This, combined with the fact that the system manufactures a multi-style, multi-size product range, makes availability of tools a major variable in the system.

In order to conduct a thorough investigation on the problems of planning and control of this class of production system, a fairly powerful simulation model was built, based initially on information obtained from an industrial company.

One of the major short term decision (control) 'tools' in a batch shop environment is the use of priority scheduling rules for choosing the sequence in which jobs should be processed in the machines. Therefore one of the first objectives of this investigation was to develop and test priority scheduling rules specially designed to fit the characteristics of this class of system. To this end the investigation with the model was divided into two phases. The first phase, which was exploratory in nature, had the objectives of generating information about the system's characteristics, with the aim of devising the priority rules and other control procedures, and of identifying the major parameters of this class of systems.

The second phase of investigation was designed with the preliminary results in mind, and consisted of three major sets of experiments, involving the testing of the priority scheduling rules, the study of the sensitivity of the system to changes in the values of some of its more relevant parameters, and the study of different strategies for capacity manipulation.

The variables whose effect on the system were analysed included the load factor, the number of tools (moulds), the mean value of setup times, the average size of orders and the ratio between the number of product styles and the number of machines.

The study of priority scheduling rules consisted of comparing the performance of the rules developed in the first phase of experiments with well known priority rules, which have been shown to perform well in traditional batch manufacturing (or job shop) systems, over a range of system configurations generated by changing the values of the major parameters of the system.

The study of capacity manipulation consisted of comparing different strategies obtained by the variation in some of the model's capacity parameters, i.e. the number of moulds, the number of working hours per week (both overtime and extra shift) and the use of strategic inventory. To this end a series of trade-off curves relating capacity costs per unit to delivery performance were built for the different strategies. Apart from allowing a comparison between strategies, such curves were also intended to demonstrate how to provide management with information which allows decision-making in a three dimensional grid which has costs, level of delivery performance and length of delivery promises as the decision parameters.

Leading up to the presentation and discussion of results obtained, a series of preliminary points are discussed, including a literature survey (paragraph 1.2), the presentation of the system's characteristics (chapter 2), the description of the simulation model (chapter 3), a discussion of the experimental design (chapter 4), and a discussion of statistical and tactical problems in simulation experiments, including the results of a pilot study designed to choose an efficient sampling procedure (chapter 5)

The presentation and discussion of results are presented in chapters 6, 7 and 8, followed by conclusions and recommendations in chapter 9.

1.2 Literature Survey

The intention of this section is to examine available literature on the area covered by this study.
As a first step towards this objective, an attempt will be made to classify the study in relation to the spectrum of problem areas to which it is related.

This study is concerned with an investigation of a particular type of manufacturing system, producing in batches, in which a computer simulation model is used to study the effect of various system parameters and operation control variables on the performance of the system. In the course of investigation, such aspects as queueing priorities, plant capacity manipulation and inventory control are considered. Since this is a manufacturing system which produces in batches, comparison with traditional batch manufacturing system is desirable.

There seems to exist a lack of uniformity between various authors when classifying production systems. As Edwards (1974) points out: "The terms used in production textbooks such as process, job, batch, flow, mass, groups and the like, often seem to have different meanings for different authors, moreover, words themselves are often used in an imprecise manner".

A close look at some of the classifications used by different authors will tend to confirm this. Coales (1971), for example, divides production systems into three classes: Batch, Continuous and Quasi-batch; Lockyer (1974), defines three main types and calls them job, batch and flow production; Moore and Jablonski (1969) divide them into two major groups which they call Job lot and Mass production; Starr (1972) confirms the difficulty of the classification problem by pointing to the highly specialized nature of each industry and even of companies within industry. He divides production systems into three different groups called flow shop, job shop and project. Finally Buffa (1972) uses two major groups: Continuous and Intermittent, and subdivides the Intermittent group into three subgroups called closed job shop, open job shop and large scale one time project.

The situation gets even more confusing when the definitions given by different authors are compared.

Considering this wide variety of terms and definitions, it is important to make sure what is meant by a batch manufacturing system. As Hollier and Corlett (1966) pointed out: "Although the term 'Batch Production' may at first sight appear too familiar to need further definition, any discussion of batch production planning and control methods requires that the activity it describes shall be clearly distinguished from one-off or jobbing work on the one hand and flow production on the other. Since batches may vary in size from two or three components to many hundreds, or even thousands, this distinction may not always be obvious". They then define batch production as:"The processing of discrete groups of a particular component or product through a series of operations at a production rate exceeding the average demand, in anticipation of repeat orders which will justify the provision of special production equipment and, possibly the holding of finished stocks".

If a comparison is made between the system being analysed in this study and batch manufacturing systems it is possible to say that there are major similarities and some differences. The major difference concerns the flow of work. While a traditional batch manufacturing system is usually supposed to produce a component through a series of operations, the system considered here produces components through a single operation. The major consequence of this difference will be in relation to the complexity of work flow and the problem of work in progress inventory, which is a serious problem in traditional batch production systems, and is almost non-existent in the system under study.

However major similarities can be found, such as the existence of a number of different components in various proportions, uncertainty in customer's demand, wide variation in batch sizes, the necessity of meeting delivery dates, etc.

1.2.1 - Organizational aspects of batch manufacturing systems

Much attention has been given by research workers to the various aspects of organization and control of batch manufacturing systems. This is justified by its importance in relation to other methods of production and the complexity of the problems it generates. Hollier and Bhattacharyya (1974) expressed this point well:"Batch manufacturing systems are among the oldest, the most common and the most complex methods of manufacture".

Research on batch production systems covers a wide range of problems that spans from the long-term problem of organization of production, to the very short-term, day-to-day problem of loading jobs to individual machines.

Among all the areas studied over the years, the problems of job schedule, organization of production and inventory control seems to be the ones which have received the most attention from research workers.

Organization of production is the way in which production facilities and jobs are organized in order to manufacture the desired products.

Traditionally, batch manufacturing systems have been organized by process layout, and a large amount of research has been devoted to developing techniques which could help to improve its effectiveness. Optimization models have been developed in order to optimize some of the parameters of performance, the most common of all being the transportation cost. El-Rayah and Hollier (1969), present a good survey on the subject.

More recently, a large effort has been devoted to developing and implementing the concept of Group Technology which is said to eliminate most of the difficult problems of traditional batch manufacture.

The evidence of such interest in group technology can be seen by the large amount of research published in the last few years, dealing with many aspects, such as techniques for implementation, economic factors and human aspects. Gombinski (1967), Opitz and Wiendahl (1971), Burbidge

(1971), Carrie (1973), Gallagher et al. (1971), Knight (1974), Edwards (1974) and Fazakerley (1974 and 1976), are good examples of recent research on different aspects of group technology.

Although regarded as being able to minimize most of the more serious problems of traditional batch manufacture, group technology has had a limited application. The great majority of batch manufacturing systems still use the process layout type of organization.

The major problem affecting production control in traditional batch manufacturing systems seems to be the complexity of its work flow, which is reflected in long throughput times, high work in progress, and poor utilization of man and equipment in productive activities.

Hollier and Corlett (1966) have analysed the actual work flow of samples of batches in the shoe and machine tool industries with the objective of identifying causes of delay and their effect on production control. They have found, for example, that batch size had no detectable effect on throughput time, that total load on the factory influences throughput time and amount of work in progress, and that control of work flow can have a large influence on the volume of output and its performance in terms of delivery time.

One of the most effective ways of controlling the flow of work and consequently the quality of delivery performance is by making use of scheduling procedures. Particularly in queueing situations such as the ones found in batch production systems, the use of priority dispatching rules has enabled improvements in delivery performance.

1.2.2 - Scheduling in batch manufacturing systems

A considerable proportion of all research dealing with planning and control of batch manufacturing systems has been dedicated to the problem of shop scheduling.

Two major distinctive lines have been followed: one theoretical, the other experimental.

The theoretical approach assumes a static situation in which n jobs or batches are to be processed in m available machines.
The possible routings can vary from the pure 'job shop' situation, which assumes that all the jobs are to be processed in all the machines, following any desired sequence, to the pure 'flow shop' situation which assumes that all jobs are to be processed in all machines following the same predetermined sequence.

Many analytical methods have been developed, which are able to optimize some single parameter objective function, like the total makespan, in the theoretical 'job shop' and in particular the pure 'flow shop' problems.
Typical examples of such analytical techniques are mathematical programming, both mixed and integer formulations (Bowman, 1959; Wagner, 1959; Manne, 1960), branch and bound algorithms (Lomnicki, 1965; Eastman, 1964; Brown and Lomnicki, 1966; Iggnall and Schrage, 1965), and heuristic algorithms (Johnson, 1954; Campbell et al. 1970; Palmer, 1965).

The main problem with these theoretical approaches is the gap between theory and practice created by the very restrictive assumptions made by most of these models. Most important among them are the assumption that the shop is a static entity in which all the jobs are known prior to scheduling, that the order of processing of the job is the same on each machine, that machines do not need setting up, and that splitting of jobs is not allowed. As King (1976) pointed out, "fundamental difficulties in solving the real practical problems led to so much simplification that, in some cases, has reduced the problem to a shadow of reality".

The experimental approach sees the problem in quite a different way. Instead of considering a static situation with n jobs and m machines, it considers the shop as a network of interrelated queues of jobs awaiting service by the available machines, in which jobs arrive at a certain rate, and join the respective queue, in accordance with predetermined routing. With such a viewpoint, the problem reduces to that of determining what form the priority rule should take, in order to minimize certain desired criteria.

To check the efficiency of different rules, experimental studies are usually conducted in which different priority rules are used and their efficiency compared. To this end many computer simulation experiments have been conducted in which the performances of different priority rules have been compared using different shop characteristics.

The works of Conway et al. (1960, 1962, 1965) are among the earliest in the area, and have undoubtedly influenced many other research workers. In his experiments he made use of a hypothetical shop, which is now considered a classical job shop model. The model assumes that job interarrivals and service times follow exponential distributions, job routings are completely random, machines can produce only one job at a time and never stop for setup or breakdown, and there is no batch splitting or transportation time between machines.

Many other authors (Eilon and Coterill, 1968; Aggarwal et al. 1973; Oral and Malouin, 1973; Elvers, 1974; Day and Hottenstein, 1975), have used this basic job shop model. Many priority rules have been tested and compared in relation to different criteria. They are usually simple heuristic rules based on such variables as order of arrival (FIFO, FCFS, LIFO), characteristics of individual jobs (SPT, LPT), due date (SLACK, SLACK/OPN), etc. The most common criteria used for measuring the performance of the rules have been average throughput time, machine utilization, capacity for meeting due dates, total tardiness, amount of work in process, and some measures of variance in delivery performance. These studies have also analysed the effect of some parameters like the overall load factor, and tightness of due dates on the relative performance of the rules.

The SPT rule, which gives priority to the job with the shortest processing time, has performed well in these investigations, when measured in terms of average throughput time and work in process inventory. It has done reasonably well even when its performance is measured in terms of its ability to meet due dates. Its main disadvantage is that it tends to generate high variance with respect to flow time, meaning that some jobs are delayed for quite a long time. Attempts have been made (Conway and Maxwell, 1962; Eilon and Cotterill, 1968) to modify the SPT rule in order to avoid such large variances, by having a cut-off point such that if a job is too late or has waited too long in the queue, then the priority is suspended for a period of time until the delayed jobs are through. Only then the priority comes back to operation. Results of such attempts have not been conclusive, because of a trade-off which seems to exist between variance and mean flow time. As the variance is reduced by the cut-off procedure the mean flow time tends to increase and so wipe out the advantages of the rule.

Many other studies have been conducted in connection with analysis of priority rules, which have lifted some of the main restrictions of the classic job shop model.

Hollier (1968), for example, considered a shop in which only a limited number of different parts are processed, each having a fixed process routing. The parts are processed in batches of varying sizes and they may require setup, and have a transit time when moving from one machine to the next.

Eilon and Hodgson (1967) considered a shop with two identical machines in which jobs required only one operation in any of the machines, and in which the main objective was to measure the effect of different due date tightness on the effectiveness of the rules.

Others based their investigations in real world job shops. Rochette and Sadowski (1976), conducted their study using a model based on the needle trade industry, in which orders arrive at the begining of the work day and each order follows a predetermined routing through the shop which may include assembly operations. The model also considers workers as a limited resource, which can move between machine centres. They found that the SPT rule performed best in relation to some tardiness based criteria. Hosein and Ross (1975), on the other hand, conducted their investigations using a job shop based in an electroplating environment, which had constant operation times, a fixed number of operations and a constant sequential job routing. They found that an LPT rule, which gives preference to the job with the longest processing time, and is the antithesis of the SPT rule, performed best in relation to the average throughput time, average lateness, and deviation from due date.

Other aspects of the problem which have been investigated are: The effect of scheduling rules on the combined performance of shop and inventory systems for a shop which produces both for inventory and orders (Berry, 1972, and Berry and Rao 1975); the effect of shop size, Labour flexibility and machine limitations (Fryer, 1975); and the effect of setup times (Wilbrecht and Prescott, 1969).

No definite conclusion can be drawn from all those studies. Although they have shown that the use of simple priority rules can greatly improve the performance of batch manufacturing systems, there is no clear winner as to the best rule when the many aspects of efficiency are considered. However, the studies have greatly helped to develop a better understanding of the mechanisms which influence the behaviour of such systems. They have also helped in finding some conclusive answers for more specific situations which unfortunately cannot be generalized.

1.2.3 - Stock control and batch manufacturing systems

The presence of some sort of stock in any batch manufacturing system is almost inevitable, either as raw material, work in process or finished products.

The existence of stocks of finished products depends largely on the policy followed by each individual company which usually have the choice of either producing in advance of future demand with the objective of providing immediate delivery out of stock, or producing only under customer's order and so incurring a delivery delay. The decision will of course be influenced by the market in which each individual company is competing, and can be seen as a strategic decision.

Unlike the stock of finished goods, the existence of work in progress stock is not something which depends only on company policy, but is highly dependent on the manufacturing process itself. Although its general level can be controlled, it is almost impossible to avoid it altogether in a typical batch production system. Similarly with raw material inventory, although its level can be monitored, it is very difficult, if not impossible, to avoid it altogether.

In fact this presence of stock of one kind or another is a phenomenon which occurs in any productive system, whether manufacturing or not.

The almost unavoidable presence of stock in almost all productive activities must be one of the reasons for the immense amount of research dedicated to the problem of stock control.

Lampkin (1967) in a survey through the relevant journals covering a twelve-year period from 1954 to 1965 found a total of 394 papers published on the subject. In the nine years since his survey, the trend has continued and a very considerable number of publications have been added to that initial number.

It is interesting to note that from the point of view of controlling stock, the questions to be answered are quite simple: which, when, and how much new stock to order? The situation gets complicated when one decides to find values for the parameters which will optimize the chosen objective criteria.

The problem of how much to order had its first analytical solution more than 50 years ago with the development of the well known EBQ, or economic batch quantity formula, which optimizes the combined cost of ordering and keeping the stock. The first analytical approach to the question of when to reorder was made about ten years later (Wilson, 1935) and consisted of analysing events between stock being re-ordered and the order being received. It was assumed that a probability distribution for demand during lead time could be estimated, and that through this probability distribution one could choose the correct value <u>s</u> for the reorder point, to give any required degree of certainty that there would not be a stockout during the lead time. This degree of certainty is often called the service level.
The combination of both approaches gave rise to the stock control procedure known as the reorder point method or two-bin system and usually

represented by the symbol (s,Q).

This method, in spite of its known weaknesses and unsuitability for many practical situations is still widely used in many companies.

The major weaknesses of this method are caused by the series of simplified assumptions it makes: demand is considered to occur at a constant rate, lead times are constant and independent for different items, ordering costs are constant and independent of the order size. Apart from that, the 'service level' approach is not a very sensible measure of service. In fact, the probability of a stock shortage between ordering and receipt of the order does not depend only on the reorder level, but is a function of both the reorder level and the reorder batch size. For the same value of the reorder level a bigger value for the reorder quantity Q will result in a smaller probability of stockout. This being so, the 'optimum batch size' formula must underestimate the best value of Q, since in calculating the optimum lot size, one is weighting only the reduction in ordering cost for a higher Q against the increase in stock holding costs. If one considers also that an increase in Q improves service, the best value of Q may be increased considerably.

It was in order to reduce the shortcomings of the (s,Q) procedure that the mathematical theory of inventory theory was advanced.

Most of that literature has dealt with a class of policy called the (s,S) procedure, which tries to find a combined optimum value for s and D, in which s is the reorder point and D = S - s, is the reorder quantity, S being the maximum stock level which is only reached when no demand occurs between reorder and arrival of the order. The complications involved in obtaining the optimum parameters for this procedure can be seen by the increasing sophistication and large number of mathematical models which have been developed over the years.

Although they have contributed a great deal to the understanding of the problem, they have sometimes reached such a level of mathematical sophistication that makes them either too abstract or very complex to operate.

Aggarwal (1974), who presents a good review of current inventory theory and its applications, makes a clear distinction between the theoretical models and the models bridging the gap between theory and practice.

On the theoretical models he makes the following comments: "The large number of research studies and the models available have covered a large number of situations, but they do not in any way exhaust the possibilities of formulating the possible millions of additional models. However, they do indicate that virtually for each group of similar items, there must be a specific inventory policy suiting individual items stocked by a company". He also points out that in order to optimize inventory operations the company needs to determine optimum parameters for each of its items, and considering that most of the system's uncontrollable parameters like demand, costs, and supply change from time to time, it means that these changes must be continually monitored and the optimum parameters recalculated after each change. He then goes on..."Even with the present high speed computers, so wide a use of all the different models by a company seems impractical because it is highly unlikely that the company's analysts will be easily able to programme the computers for deciding for each item, when to switch from one type of an inventory model to another type of inventory model. Further, to incorporate all the optimizing models in a single inventory system can be an extremely costly proposition". He concludes by saying that the costs which may be incurred by trying to maintain an optimum inventory control will most likely offset the savings resulting from the extra effort made.

In order to close the gap between the theory and the practical problems encountered in most manufacturing systems, a number of studies have been conducted in the last few years, in which an experimental approach, usually a computer simulation model, have been used.

Berry (1972), for example, examined the interdependence between the problems of priority scheduling and inventory control. The main objective of the experiment was to test the gains on the combined performance of a shop and its associated inventory system, achieved by the use of inventory data in the decision process of some priority dispatching rules. The inventory system was controlled by an (s,Q) policy, where s was the reorder point and Q the fixed order size, and the performance of three different priority rules using the inventory information were compared with two other commonly used rules, SPT and FIFO. The results indicated that the scheduling rules incorporating inventory information improved the performance as regards the total inventory related costs and percentage machine utilization. But it was also found that the performance obtained by the FIFO rule, which does not use any inventory information, was very close to the others which do use such information. He points out the practical importance of such a result. Because of the simplicity of application of the FIFO rule, one should carefully evaluate the cost of processing inventory information for use in making scheduling decisions against the benefits which it could bring.

Oral et al. (1974) used a simulation model based on a large concern manufacturing power transmission equipment. The objective of the study was to find optimal values of the parameters for an inventory policy of the (Q,s,R) type where Q is the constant size of replenishment orders, s is the reorder level, and R the upper limit indicator for backlogged demand. The model also incorporated a routine which takes into consideration the "impatience" of the customers in relation to delivery delay. The measures of performance were the amount of demand satisfied, number of customers fully satisfied, total cost and discounted cost. A search procedure was used to find optimal values for S and Q. It was found that the optimal solution with respect to cost criteria is different from the optimal solution with respect to service level criteria.

Eilon and Elmaleh (1968) is a further example of the use of an experimental approach to study inventory control systems. The paper is concerned with a dynamic inventory situation in which demands are subject to wide fluctuations, seasonal pattern and trends. The objective is to compare five alternative inventory control policies in relation to their performance, which is obtained by the use of a computer simulation model incorporating a forecasting rule that takes account of seasonal and trend factors. The results show that of five policies tested, a procedure called (T,s,S) was the most promising. The (T,s,S) policy relies on a cyclical review every T periods, when the stock level is replenished to an upper level S, but if in between reviews the stock declines below a lower level s, it triggers a reorder quantity equal to S - s. Among the other control methods tested were the well known (s,Q) and (S,T) policies, which are still used in many companies.

By comparing the theoretical and experimental approaches it can be seen that a very large amount of research has been done into the mathematical theory of inventory control and a large number of cases have been examined. Nonetheless their use in practical situations seems to be restricted and this might be explained both by the difficulty in determining and maintaining the optimal parameters of control, and also by the restrictive assumptions on which the models are based, in order that a mathematical solution can be obtained.

The experimental approach on the other hand, has the advantage of being able to lift most of the restrictive assumptions made by the mathematical models. However it has some disadvantages, chief amongst which is the fact that the simulation model usually needed to perform the experimentations can be expensive in modelling and operating.

Finally, it should be noted that most studies on inventory theory look at stocks just as a component of production systems which needs to be controlled at an optimum level. In reality stock can and should be looked at not as an isolated component, but as an alternative way of manipulating capacity, in the same way that plant, work force and amount of working hours are manipulated.

1.2.4 - Capacity planning in batch manufacturing systems

The problem of capacity planning is basically a strategic decision as opposed to tactical or short-term decisions such as the daily activities of loading machines and controlling inventory.

The level of detail involved in capacity planning can vary from highly aggregate decisions which bypass the details of individual products and the detailed scheduling of facilities and men, to very detailed decisions involving the load imposed on a single machine or group of machines.

At the high level of aggregation, usually called 'aggregate planning' or 'aggregate scheduling' one is usually interested in finding an economic balance between the general level of work force, inventory and working hours, with the objective of meeting a forecasted demand, usually having a fluctuating or seasonal component.

Various models have been developed for determining the parameters of an optimum plan.

There is, for example, a linear programming application (Hanssmann and Hess, 1960), a transportation algorithm (Bowman, 1956), and a linear decision rule (Holt, 1955) to name just a few. Most models in this area consider a system which produces inventoriable items, and has a fixed amount of plant investment, and where a cost trade-off is made between the amount of finished goods inventory and employment level.

At the other end of the scale, there is the problem of planning the capacity needed for a single operation in a machine or a group of machines. In this case the problem is usually dealt with as a queueing or waiting line situation.
The most common objective in such cases is to study the behaviour of the queue in relation to such aspects as its size, the delays caused to its members and the level of machine utilization. With this information an economic decision can be made, which could result in the provision of extra or more efficient equipment in the case of bottlenecks, or an increased load on the facility in the case of under utilization.

Queueing problems, if they are not too complex, can be mathematically handled. But if several factors work together to produce effects, then mathematical solutions become quite complex and very difficult to handle. On these occasions computer simulation models can be used. This is certainly the case in complex batch manufacturing systems in which the queues are interrelated and a mathematical model of the complete system would be impractical. Various studies of capacity planning using computer simulation models have been reported in the literature.

Cantellow et al. (1973), for example, present a case study of a real machine shop, in which a computer simulation model was used to analyse the effect on the performance of the shop of increases in both demand and nominal available production capacity, in terms of machinery and number of working hours.

The problems facing the shop were high investment in work in progress, throughput time consistently longer than planned and a future increase in demand. Performance was measured in terms of costs, made up of three components, viz.

i) interest cost of machine investment

ii) wages

iii) interest cost of work in progress inventory

By adding new machinery and men, and by increasing the number of shifts the resultant total costs could be measured, and bottlenecks eliminated, until the total demand could be fully satisfied.

Dolton and Moody (1975), describe a simulation model which represents an aero engine's repair and overhaul cycle, and the way in which the model has been embedded in the planning and control of an engineering workshop.The model consists of a workshop, the stock of both unserviceable and serviceable engines and parts, and the pool of engines actually in service. In the case of a stockout of serviceable engines, the capacity of the shop can be increased by the utilization of overtime, in order to shorten the normal repair time. The model is simulated on a regular basis such that plans can be made in advance to match capacity to requirements, and to evaluate the costs of different alternatives.

Further examples of the experimental approach to capacity planning can be found in Lipton (1969) and Aley and Zimmer (1974).

These studies of capacity planning show that important relationships exist between different components of production systems, and how they can be manipulated.

These also show that capacity can be varied in a number of ways, and that a smaller load on the system can be a better economic proposition than a higher load, as long as the spare capacity is chosen in the appropriate way.

1.3 - Summary

In this chapter an introductory explanation of the problem under study was made, and a survey was presented on the available literature on the areas related to this study.

A discussion of the typical problems of a traditional batch manufacturing system was followed by detailed analysis of three of the most important problems of planning and control in batch manufacturing system, viz. shop scheduling, inventory control and capacity planning.

The relationship between the work reported in the literature and experiments carried out in this investigation are drawn out in subsequent chapters.

CHAPTER 2

DESCRIPTION OF THE PRODUCTION SYSTEM

2.1 - Introduction

Although the intention of this study is to analyse a class of production systems which can be found in the shoe industry and might be relevant to the textile industry as well, the original information which led to this research was obtained from one particular company within the shoe industry which manufactures shoe components.

In order to define the area of investigation, the original production system, as it was found in the particular company from which the information was first gathered, will be described in detail. This will be followed by a formal characterization of the class of production systems, and the identification of the relevant variables.

2.2- The actual production system

For the purpose of description the production system will be divided into three major components: the products, the production process and the demand or orders input.

2.2.1 - The products

The products manufactured by the company are injection moulded shoe insoles, produced in different shapes and sizes. Each particular shape is called a style and is made up of a range consisting of up to 13 different sizes. This range of product sizes results in the need for a range of sizes of injection moulds. Because of product design and a policy of standardization, it is possible for a mould of a certain size to be used in the manufacture of insoles of both its nominal size and half size above it. The design of the product is relatively simple, consisting of a half split fibreboard in which plastic is inserted between the split parts through an injection moulding process (for technical specifications see Johnson, 1974).

Figure 2.1 is a schematic representation of an insole, which shows how plastic is inserted into the half split fibreboard. A major characteristic, as shown in view BB, is the fact that plastic injection takes place only in the back part of the insole. This means that although this part must be standardized to fit the mould, the front part is independent of the injection moulding process and consequently of the mould design. This characteristic of the product design allows a mould of a certain size to be used in the manufacture of insoles one half-size bigger than that size. It also allows a greater flexibility in the design of the front part of the insole, meaning that a mould of a certain style can be used in the manufacture of insoles of the same class of style. but with different front part shapes. By using a system of standardization for the backpart of the insole, the company is able to limit the number of mould styles without corresponding limitations in the insole style (front part shape). This is of great importance as injection moulds represent one of the major investment costs in the production process.

2.2.2 - The production process

a) Process description

The production process consists of the injection moulding of plastic into a previously cut fibreboard, the cutting operation being a minor one in relation to the major operation of injection moulding. This is true both in terms of production time, which is about 1/100 of the moulding operation, and in the simplicity of the operation, which is reflected in a relatively minor investment cost. As a consequence there is a decoupling between the two stages, caused by a buffer stock, which in practical terms eliminates interference between the two operations.

The injection moulding operation is executed in multiple stations machines, which require a single operator for each machine. A schematic representation of one of these machines, together with the operator's cycle is given in figure 2.2.

The whole injection moulding operation can be best understood by looking at the machine and operator cycles individually.

The machines have twelve stations laid out on a circular turntable which moves around its axis in such a way that, after a complete rotation, each station has been in the position to receive an injection shot. The machine cycle is made up of two phases: The injection phase, in which a screw ram injects plastic into the mould, and the movement phase, in which the table moves around its axis to position the next station in front of the screw ram for the injection cycle. The duration of the machine cycle can be adjusted to fit the requirements of production.

The operator cycle has three phases, represented in figure 2.2 by dotted lines. Phases 1 and 2 are productive phases, and correspond respectively to the operations of unloading and loading the machine with the fibreboards. Phase 3 is a non-productive phase and represents the operation of changing moulds and setting up the machine.

Phases 1 and 2 of the operator's cycle are coordinated with the machine cycle, so that the operator should be able to unload and load a station during the interval of time taken by the screw ram to execute an injection cycle. In practice, because of variability in the operator's cycle, the operations are never perfectly coordinated, resulting in either the operator or the machine waiting for each other. The start of the table's movement is dependent on the operator finishing his operations, so that the machine will always wait, if the operations of unloading and loading are longer than the injection cycle. On the other hand, if the operator finishes before the injection cycle is over, he then has to wait for the machine.

At this point, attention should be paid to the effect of another important cycle time on the process, which has not yet been mentioned. This cycle time is represented by a full rotation of the table which contains the stations, and corresponds to the interval of time which elapses from the loading of a station with a split fibreboard by the operator, to the unloading of a finished product from that same station. This cycle, which also represents the manufacturing time of an individual product, will be called 'process cycle time' as opposed to the 'production cycle time' given by the machine/operator cycles.

It is important to note that while the 'process cycle time' represents the minimum time to complete the manufacture of a component on the machine, the 'production cycle time' represents the production rate of the machine.

It should be noted that any mould can be fitted to any station, so that at any point in time the machine could be producing up to twelve different products.

b) Moulds

A major feature of the production process is the injection moulds. In order to produce a product style in its full size range, a range of moulds must be provided.

The different sizes in a style correspond to the range of sizes of shoes manufactured by companies and includes full and half sizes. As described in 2.2.1, the design of the products is such that a mould of a certain size can produce insoles of both its nominal size and half a size above it. There is however an exception for the extreme lower sizes. In these cases a mould can manufacture its nominal size, a half-size below nominal size, and a half-size above nominal size. The mould and product characteristics, together with the fact that moulds are designed to produce both right and left foot simultaneously, means that six moulds are sufficient, in terms of technological requirements, to produce a full range of thirteen sizes in any one style.

2.2.3 - <u>The demand</u>

In relation to demand, the policy of the company is to produce only against customer's order. This is partially justified by the fact that it produces components for manufacturers belonging to a fashion industry, viz. ladies and men shoes. However the effect of fashion on the components is partially minimized by a policy of standardization, and the product design, so that insoles with different front part shapes can, in many cases, be manufactured by the same set of moulds.

A major characteristic of demand is that when a customer orders an insole of a particular style, he usually requires the full size range, with varying quantities for each size. In terms of production this means the manufacture and joint delivery of up to 13 different batches of components.

Previous studies (McKay,1929) show that the distribution of foot sizes for adult men and women follow a normal distribution, and so it would be expected that the statistics for demand according to shoe sizes, should also fit a normal curve. However when actual data from sales are plotted and compared with data from actual foot sizes distribution (figure 3.3), it shows a drop in demand for half sizes, when compared with full sizes. These results agree with a similar comparison with mens shoes made by McKay (1929).

Although the general level of demand for different sizes follows the above reported distribution, the total quantities ordered in each order vary considerably, resulting in a large variation in batch sizes for production. In spite of such variations, there is a need for a speedy delivery which is usually set at 3 weeks, from the posting of an order to the final dclivery.

2.3 - Formalization of the problem and identification of variables

In order to conduct a systematic analysis of the problem, and to identify the relevant variables affecting the system a graphical model of a production system with its information and material flow is going to be used.

Figure 2.4 is a simplified model of a production system, where blocks represent subsystems, single lines represent information flow, and double lines represent material flow. Six blocks or subsystems are represented in the diagram of figure 2.4. Four are internal components, while the other two are external and represent the inputs to the system in terms of material and information flow.

In the model the flow of information generates and controls the flow of material. The whole process starts in the subsystem representing the pool of customers, which from time to time places an order on the production system. This order is received by the subsystem 'control', which after checking the information on the state of subsystems 'raw material stock', 'machine shop' and 'finished goods', takes a decision which will generate a flow of material starting somewhere in the system and finishing with a delivery of finished goods.

During the whole process a continuous flow of information is processed and decisions made by the control system. These decisions can vary from loading machines to planning overtime or extra shifts.

By using this model it is possible to identify variables and define the characteristics of the class of production systems under study. This will be done by analysing each of the component subsystem in turn.

2.3.1 - Customers and the order input

The type of order input, together with the machine shop, are the major characteristics of the system under study. It is assumed that only a limited number of different product styles are ordered during any interval of time. Each style S_i is made up of a range of sizes, and a component of size j and style i is represented by Z_{ij}

Each order which arrives at the system can be specified by two variables: the style, S_i, required and the individual demands, q_j, for each of the sizes, Z_{ij}, where $q_j \geq 0$ and $Q = \Sigma\ q_j$ is the total amount demanded. It is assumed that the values q_j in an order are not completely independent because of a relationship which exists within j. This relationship can be better expressed by the proportions $p(j) = q_j/Q$, where $0 \leq p(j) \leq 1$ and $\Sigma\ p(j) = 1$. The values of p(j), when plotted against j gives a histogram similar to the distribution described in 2.2.3 and dependent on the particular class of style.

Both the total quantity demanded and the interarrival times follow probability distributions, which are also dependent on the particular class of style.

Delivery delay is a major feature of the system and customers rely on a short and precise delivery date which, within certain limits, is fixed independently of the quantity demanded. A delivery can only be made after all the batches Z_{ij} in an order have been completed, i.e., customers require a full range of sizes to be delivered together.

2.3.2 - Machine shop subsystem

The machine shop is in the heart of the production system and is represented by machines, men and tools.

The machines are multiple stations, and able to produce different components at the same time. Each is manned by a single operator. The cycle times of machines and operators are two variables in the sub-system, which need to be identified. Machine cycle time is deterministic, but operator cycle time is variable.

All components are manufactured in batches of varying sizes, through the same process, by the multiple stations machines. Each one of the different components requires a specific mould for manufacture, and any mould can be fitted in any station on the machines.

In relation to moulds, two variables need to be specified: the list of moulds available in the system, and the specifications which will determine which components each mould is able to manufacture.

Each time there is a change of a mould, the machine has to be stopped, and consequently all products being produced in the other stations are delayed by a period corresponding to the time spent in executing the mould change. Setup times for mould changes are random variables which follow a probability distribution.

2.3.3- Raw material stock and supply system

Stocks of raw material are maintained in order to create a buffer between suppliers and production. If they are well controlled, the chances of a stockout should be small.

In this study it is assumed that raw material stocks are such that stockout never occurs, and raw materials are always readily available for production.

2.3.4 - Finished products

Finished products are the physical output of production systems. A company is usually free to choose, as a matter of policy, whether to start manufacture of a product, only after a firm order is placed by a customer, or alternatively it can produce in advance of customers orders, with the expectation of future demand. In the latter case a stock of finished products will be created, and this will help to shorten delivery delays.

In the system under study it is assumed that the usual policy is to start production only after the receipt of a firm order. However the effect of a change in this policy is also considered, as an alternative to improve delivery performance. In this case stock will be maintained for individual components, and this will imply the need of a policy to control the stocks. A policy can be established by the determination of three parameters: which components to stock, when to order a new batch for stock, and how large a batch should be.

2.3.5 - Control system

The control system can be considered the brain of the production system. It is responsible for the application of rules which guide the day-to-day operation of the production system. In general terms it is responsible for providing answers to the questions of what to produce, how to produce and when to produce.

This is done through the use of such tools as priority rules, the selection of batch sizes, the utilization of overtime and extra shift, the control of stocks and the selection of moulds.

One of the objectives of this study is to determine effective policies for the control system in order to obtain a high efficiency of the production system under study. Some questions which this work aims to answer are: what sort of priority rules should be used; whether to split large production orders into smaller batches; which and how many moulds to have available; whether to use overtime and/or extra shifts; whether to hold stock, and if so, which is the most efficient control policy.

2.4 - Summary

In this chapter the problem under study was presented, and its characteristics discussed.

A description of a particular production unit, from which real information had been gathered, was used in order to identify a class of production systems.

Finally an abstract model of a production system was used in order to formalize the problem and identify the variables which characterize the class of production systems under study.

FIGURE 2.1

INJECTION MOULDED INSOLE

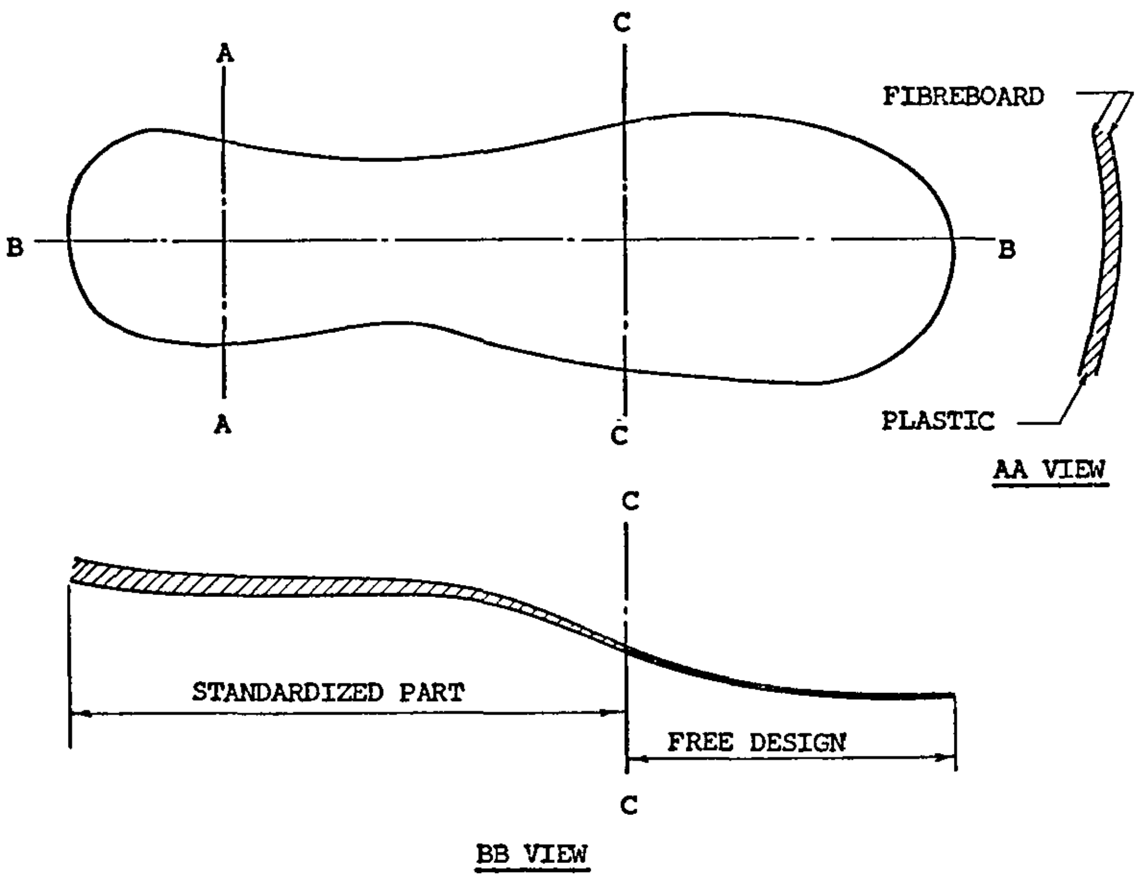


FIGURE 2.2

THE PRODUCTION UNIT

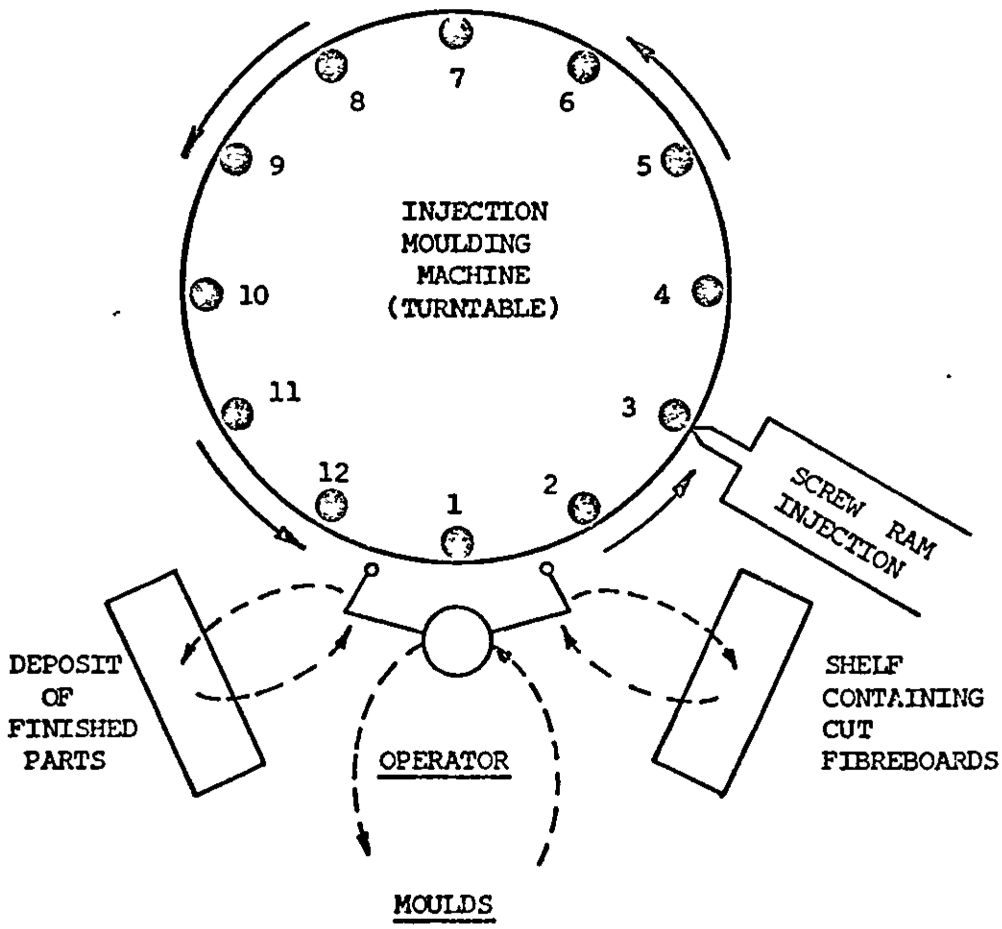


FIGURE 2.3

COMPARISON BETWEEN STATISTICAL DATA FROM ACTUAL ADULT WOMEN FOOT SIZES AND ACTUAL DEMAND FOR INSOLE SIZES

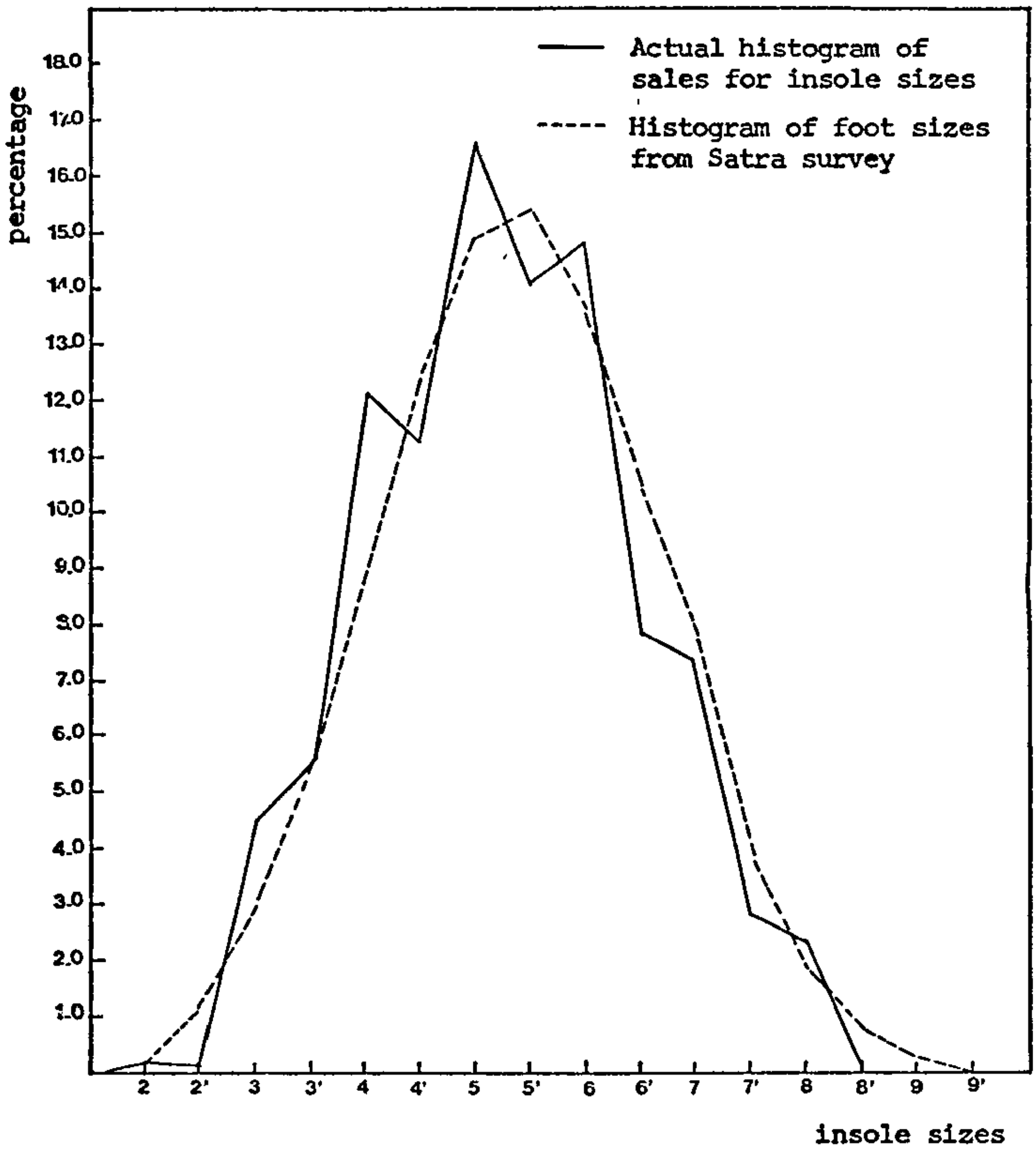
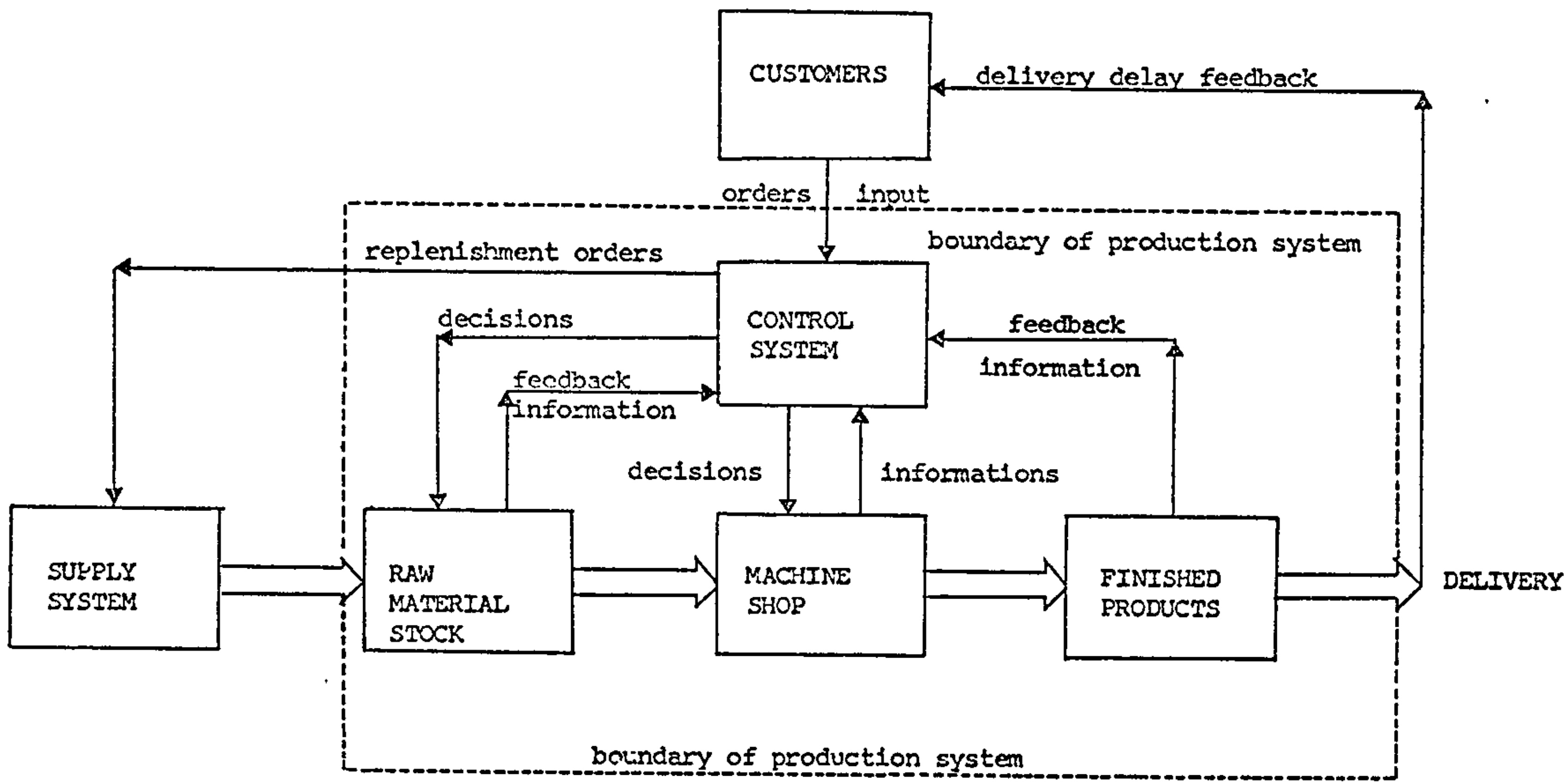


FIGURE 2.4

GRAPHYCAL REPRESENTATION OF A PRODUCTION SYSTEM



CHAPTER 3

MODEL DESCRIPTION

3.1 - Introduction

From the description given in the previous chapter it is evident that the system under study involves among other things a queueing problem with strong stochastic components. It is known that queueing problems,if they are not too complex, can be mathematically modelled, and their behaviour analysed through the model. However, in cases where several factors work together to influence the queueing process, the use of a mathematical model can become infeasible, due to the complexities involved in modelling. This is certainly the case for the system under study. The characteristics of the arrival process, and the complexities of the production activity, involving multiple stations machines, special tools, and possibly stocks, makes mathematical modelling an infeasible proposition. It is therefore necessary in this investigation, to use a computer simulation model.

This chapter describes a computer simulation model, which represents a class of production systems, with the characteristics of the system described in the previous chapter. The model is fairly flexible in respect to the variables influencing the system, which can be easily varied.

In order to program the model for computer runs, a general purpose simulation language (CSL), which is a package provided by ICL, was used. From the point of view of programming, CSL has the advantage of being based on Fortran, which means that apart from offering its own internal simulation routines, it allows the introduction by the user of most of the Fortran facilities. A full listing of the computer program, together with detailed explanations are given in appendix 1.

From a macro point of view the model can be seen as an input-output cycle, in which customer's orders arriving at the system are the input, and delivery of components are the output. A macro block diagram of this input-output cycle is given in figure 3.1.

For the purpose of examining the components of this input-output cycle, the model will be divided into five major blocks: demand or order input; machine shop; inventory subsystem; operation control system; and output variables.

3.2 - Order input

'Order input' represents the demand for the products manufactured by the system. As described in paragraph 2.3, the system manufactures only a limited number of product styles, with each style having a range of product sizes.

The arrival of a customer demand results in the generation of orders which are input to the machine shop, if no stock is being held The whole process for a non-inventory system can be seen as a three stage cycle:

i) Generation of the interarrival time and requirements for next customer demand.
ii) Arrival of a customer order.
iii) Issuing and input to the machine shop of production orders to satisfy customer requirements.

For the execution of this cycle the model makes use of three classes of temporary entities called 'style', 'order' and 'job'.

Entities 'style' are used in the first phase of the cycle, to store the values of interarrival times and requirements of demand. Each class of product style is represented by an entity of the class 'style', such that generation of interarrival times and requirements of demand can be made independently for each 'style'

Each 'style' has associated with it three probability distributions corresponding respectively to interarrival times, total quantity demanded per order, and proportion of the total quantity which is required for each of the product sizes in the range. Each probability distribution is accessed by an independent stream of random numbers, and the model allows for the use of either theoretical distributions or empirical ones provided by the user, in the form of histograms.

The generation of interarrival times and total quantity demanded is executed by sampling from the respective probability distributions, the procedure being repeated each time a new demand arrives at the system. However, the determination of the proportions of the total quantity that are required from each product size in an order, is a more complex procedure and needs detailed description.

As explained in paragraph 2.3 the proportions ordered in each customer demand are not completely independent, but follow a certain pattern which is representative of the general level of demand for the different product sizes.

It is assumed that when customers make a decision to order a certain style, they first decide on the total quantity to be ordered, and then decide on the proportions for each size.

The distribution of proportions in an order can be represented by the variable $p(j)$ where $0 \leq p(j) \leq 1$, $\Sigma p(j) = 1$, for $j = 1, n$ with n representing the number of different sizes in the range. Theoretically there are infinite combinations of $p(j)$ which satisfy the above constraints. In practice, however, it is assumed that only a limited number of combinations are used, following predetermined patterns.

In the model, each style has associated with it a limited number of combinations of p(j), each combination being called a 'pattern', which is input to the model by the user. Each 'pattern' is in turn associated with a probability corresponding to its chance of occurrence, and has a correction factor for the total demand, to take account of the number of different sizes ordered.

The whole process of demand generation can be represented by the following procedure, which is independently repeated for each different product style.

(i) Generate the next arrival time of demand from the corresponding distribution of interarrival times;

(ii) Generate the total quantity Q required for the next demand;

(iii) Select the 'pattern' of proportions required for the product sizes;

(iv) Adjust the value of total quantity Q, corresponding to the 'pattern' selected, and calculate the quantities q_j required from each individual size in the range;

(v) Store this information, (i) to (iv), until the arrival of the demand at the system, when they are transferred to other entities which are created;

(vi) Repeat the whole procedure as from the first step, after each arrival.

The second stage of the 'order input' cycle is the arrival at the system of a customer order which is represented by the event called 'demand'. At each 'demand' the model generates and inputs to the system a temporary entity of the class 'order', which represents the set of requirements made by a customer. Each 'order' is characterized by four attributes representing respectively the style, the total quantity, the number of different sizes ordered, and a serial number which identifies each 'order'. An 'order' stays in the system until all its requirements have been satisfied, with no partial delivery being allowed.

The third and last stage in the cycle is the generation and input to the system of production orders for the required product sizes. The procedure at this stage depends on whether or not the system holds a stock of finished goods. If the answer is no, then immediately after the generation of an entity 'order' the model generates and inputs to the machine shop a series of temporary entities of the class 'job', where each 'job' represents a production order for a batch of components of a particular size, as required in the customer order. If the system holds stock of finished goods, then, before the 'jobs' are generated, a check is made to see whether the order requirements can be satisfied directly from stock. The procedure used in this latter case will be explained when the inventory system is discussed.

Each entity 'job' is characterized by five attributes, namely the 'job' serial number, the serial number of the 'order' to which it belongs, the product size it represents, the quantity of items required, and a code number to identify the tools which can be used in the manufacture of the 'job'.

The whole procedure for demand generation is represented in figure 3.2 by a macro block diagram of the process.

3.3 - Machine shop

The machine shop is represented in the model by machines, moulds and a queue of 'jobs' waiting to be processed.

'Jobs' on arrival join a queue, and wait there until they are selected to be processed on the multiple station machines. The selection of a 'job' depends on the availability of both a free station and a suitable mould, and on the priority scheduling rule used by the model.

3.3.1 - Multiple station machines

Machines are represented by two classes of permanent entities called 'machine' and 'station' respectively. Each 'machine' is linked to a number of 'stations', where each 'station' represents a specific station in the multiple station machines.

a) Entity 'machine'

All 'machines' are identified by six parameters and two loading states, viz.,'idle' and 'busy', which are represented in the models by two distinct queues called 'idle' and 'busy'. It is assumed that a 'machine' can operate both fully loaded (all the 'stations' loaded), or it can operate partially loaded (at least one station unloaded), in which case it will be in the idle state and join the queue 'idle'. In both cases the machine will join the queue 'busy'. The only occasion in which the machine is not in the queue 'busy' is when all stations are unloaded. Apart from the idle and busy states each 'machine' is characterized by the following parameters:

(i) A time cell which keeps the record of time events;

(ii) An identification serial number;

(iii) The number of stations loaded at a certain instant of time;

(iv) A parameter to determine whether the 'machine' is operative or not in an experiment;

(v) Two parameters used to determine the 'process cycle time' as defined in 2.2.2.

The time cell is used to record the times of the events which represent the completion of a job by one of its stations. The identification serial number, which has an unique value for each machine, is used to link a 'machine' to its corresponding 'stations'. The third parameter is used as a record of the number of loaded 'stations' in a machine, at a certain instant of time, as it is assumed that a machine can operate partially loaded. The fourth parameter is to 'switch off' a 'machine' in an experiment. If in the initialization period of a run the value of this parameter is set to 1 the 'machine' will be operative, otherwise it will be 'switched off' and left out of the experiment. The last two parameters are used in connection with the specification of the 'process cycle time', which is the manufacturing time for a single component.

As described in 2.2.2 each machine is manned by a single operator who has the function of unloading and loading, and who has partial control over the machine as far as 'production cycle time' is concerned. That is, each machine has a fixed cycle time which is overruled only when the operator's cycle is longer than the machine cycle.

In order to define the problem of cycle time, one has to consider the following variables:-

- N - number of stations per machine
- NSL - number of loaded stations in a machine in a given instant of time; $NSL \leqslant N$
- OC_i - operator's cycle time
- $\overline{OC}$ - mean value of operator's cycle time
- MC - machine cycle time
- PC_i - 'production cycle time'
- $\overline{PC}$ - average value of 'production cycle time'

Operator's cycle times OC_i have a variability, and their distribution tends to be skewed. The machine cycle time, on the other hand, is a deterministic variable. If the model was to simulate the manufacturing operation, cycle by cycle, the following procedure would be used:

i) generate an operator's cycle time OC_i

ii) if $OC_i \geqslant MC$, make $PC_i = OC_i$, but
if $OC_i < MC$, then $PC_i = MC$

If the machine was manufacturing the same product in all of its stations in a given production run, then the average 'production cycle time' $\overline{PC}$ would be equivalent to the average production rate of the machines, which would be the main parameter of interest. However this is not so. Usually the stations are manufacturing different products, and although $\overline{PC}$ can still be seen as the general production rate, the variable of major interest is now the production rate of each station, because a batch of a certain product is allocated to a single station, and one is interested in determining the processing time for that batch, which on average is equal to ($N * \overline{PC} *$ batch size) where N is the number of stations per machine. For a batch size equal to S the processing time would be determined by simulating N * S 'production cycles', as a station takes N 'production cycles' to manufacture one component. If one considers, for example, that N = 12 and S = 150 are typical values for the system being analysed, this would mean the simulation of 1800 cycles for a typical batch. Since the variable of major interest in the process is the time taken for the completion of a full 'order', which means the completion of the last batch ('job') belonging to that 'order', it seems reasonable to assume that a 'cycle by cycle' simulation of the manufacturing process would result in an extremely high computational effort, which would not be justified by the additional precision that would be obtained.

In order to increase the computational efficiency when running the model, the 'cycle by cycle' simulation approach was substituted by a 'batch manufacturing time' approach, in which mean values instead of individual samples are used. This approach is based on the utilization of the 'process cycle time' which is the time that elapses from the loading of a 'station' by an operator, to its unloading after N 'production cycles'. The model assumes a linear relationship between 'average process cycle time' and the number of stations loaded, which is given by the following expression:

$$\overline{PCT} = N * MC + I * NSL \qquad \ldots \quad 3.1$$

where:

$\overline{PCT}$ - 'average process cycle time'

N - number of stations in a machine

MC - machine cycle time

I - $\overline{PC}$ - MC, is the difference between average 'production cycle time' and 'machine cycle time', and measures the average interference of 'operator cycle time' in the 'machine cycle time'.

NSL - number of stations which are loaded in a machine

The reasoning behind this equation is based on the following assumptions:

i) 'Stations" will always be subject to the machine cycle time, whether or not they are running loaded.

ii) 'Stations' are subject to operator cycle time, only if they are running loaded.

iii) 'Production cycle times' PC_i are given by either operator cycle time OC_i or 'machine cycle time' MC, whichever is bigger.

If NSL 'stations' are running loaded, the 'process cycle time' of each 'station' is given by NSL 'production cycle times' plus (N - NSL) 'machine cycle times'. Rearranging equation (3.1):

$$\overline{PCT} = \overline{PC} * NSL + MC * (N - NSL) \qquad \ldots 3.2$$

The two extreme loading cases would be:

$$NSL = N \therefore \quad \overline{PCT} = \overline{PC} * N$$

$$NSL = 1 \therefore \quad \overline{PCT} = MC * (N - 1) + \overline{PC}$$

Expression 3.1 is used to generate manufacturing times for individual batches, and MC and I are the parameters of each machine which will determine the 'process cycle time' $\overline{PCT}$, used to calculate batch manufacturing times.

b) <u>Entity 'station'</u>

The second class of entities used to represent the multiple stations machines are the 'stations'. Each 'station' is identified by six parameters and two loading states, represented by two classes of queues called 'loaded' and 'unloaded'. For each 'machine' there is a corresponding pair of 'loaded' and 'unloaded' queues, such that when a station is running loaded it joins the 'loaded' queue corresponding to its 'machine', otherwise it joins the corresponding 'unloaded queue'.

The following six parameters are used by each individual station:

i) a serial number used as an identification for each 'station'

ii) the identification number of the 'job' which is being manufactured by the 'station'

iii) the identification number of the mould which is setup in the station

iv) a time cell used to record the completion time of 'jobs'

v,vi) two parameters which are used in connection with the updating of the programmed completion time of 'jobs' allocated('loaded') to the station

The first parameter is a unique identification number for each 'station' and is used together with the 'machine' identification number to link each 'station' to its corresponding 'machine'. The second and third parameters are used in order to identify which 'job' and mould are loaded in the station at a given instant of time, as they change from time to time. The fourth parameter is used to record the value of expected completion times of 'jobs' allocated to the station. The last two parameters are needed because of the variability associated with $\overline{PCT}$, the 'average process cycle time' which, as defined in (3.1), has a linear relationship with NSL, the number of stations loaded in the 'machine'. As NSL in each 'machine' changes from time to time, then $\overline{PCT}$ will also change and, as a consequence, the programmed completion time of 'jobs' allocated in the stations which remain 'loaded', will also be changed.

When a 'job' j is committed to a 'station' its expected completion time is determined by multiplying its batch size S by the 'average process cycle time', $\overline{PCT}$. The first step is the determination of $\overline{PCT}$, as it is a function of the number of 'stations' loaded, NSL. By determining NSL and using equation 3.1, $\overline{PCT}$ is calculated, and the expected completion time t_j can be determined. If t_o represents the time at which production starts on 'job' j, then $\Delta t = t_j - t_o$ represents the expected processing time for 'job' j.

During Δt the number of 'stations' loaded, NSL can be modified by either an 'unloaded' station becoming 'loaded' or vice versa. In both cases, the production rate of the remaining loaded 'stations' would be modified due to the change in the value of $\overline{PCT}$. The consequences of such changes are reflected in the value of the expected completion time, t_j, which have to be revised.

If a change in the value of NSL happens at instant t_i, where $t_o < t_i < t_j$, then t_j and all the other expected completion times on that machine will have to be recalculated. To do this, the following procedure is used:

i) Calculate $\Delta S = (t_i - t_o) * \overline{PCT}_o$, where ΔS measures the amount produced during the interval $(t_i - t_o)$ and $\overline{PCT}_o$ is the value of $\overline{PCT}$ at instant t_o.

ii) Calculate $S_i = S_o - \Delta S$, where S_i is the outstanding production from the original batch S_o.

iii) Calculate the new value $\overline{PCT}_i$, using equation 3.1 with the new value of NSL.

iv) Calculate $t'_j = S_i * \overline{PCT}_i + t_i$, where t'_j is the revised expected completion time for 'job' j

v) Replace the values of t_o, t_j and S_o by t_i, t'_j and S_i respectively

In order to execute this procedure each 'station' needs to record the values t_j, t_o and S_o which are atributes of each 'station'.

3.3.2 - Moulds

The second group of components in the machine shop are the moulds used in the manufacture of different product sizes and styles. They are represented in the model by a class of permanent entities called 'mould' where each 'mould' is characterized by an identification serial number and a list of products it is able to manufacture. A mould is either set up in a 'station' or it is free, in which case it joins a queue called 'free'. It is assumed that a 'mould' only leaves a 'station' when there is another 'mould' to be set up in its place, otherwise it stays in the 'station' even if it is 'unloaded'. Every time a new mould is setup in a station a setup time is generated from a probability distribution and its value is added to the expected completion time of all 'jobs' 'loaded' in the other 'stations' in the 'machine'. The model allows the use of both an empirical distribution of setup times, which is provided by the user in the form of histogram data, or a theoretical one, in which case it is assumed that setup times follow a normal distribution.

3.3.3 - Queues

The third and last component of the machine shop is the 'queue' used by 'jobs' during waiting times, from arrival until start of production. It is assumed that all 'jobs' join the same queue, and that they wait there until they are selected for manufacture in any station of any one of the 'machines', in accordance with the priority scheduling rule in operation.

Because of the characteristics of the arrival process which generates simultaneous arrivals of 'jobs' at the shop, and also for computational necessity, the model actually uses two 'queues' called 'inqueue' and 'atqueue', to handle these simultaneous arrivals.

At the moment of arrival all 'jobs' join 'inqueue', which is an empty queue, such that they can be organized in sequence following the priority rule in operation. If there are one or more 'stations' 'unloaded', the model tries to allocate as many 'jobs' as possible to the stations, and this is done by matching 'jobs' to 'available' moulds, using the priority system in operation. After all possible 'jobs' have been loaded, the remaining ones are taken from 'inqueue' and transferred to 'atqueue' where they wait until a 'station' and a proper 'mould' becomes available. If at the moment of the 'jobs' arrival there is no 'unloaded' 'station', the 'jobs' are transferred straight from 'inqueue' to 'atqueue'.

The activities in the machine shop are generated by the occurrence of two major events, namely the arrival of 'jobs' to the shop, and the completion of 'jobs' by the machines. In order to describe the logic of these activities, two macro block diagrams of events 'job arrival' and 'job completion' are presented in figures 3.3 and 3.4 respectively.

3.4 - Inventory subsystem

Although the production system under study is basically a non inventory system, in the sense that customers expect a delivery delay in completion of their orders, an inventory subsystem is provided by the model, which can be 'switched on' and off between consecutive experiments. The objective of using inventory in this study is not to eliminate or decrease the length of delivery promises, but instead to improve the efficiency in those promises.

The inventory system is defined by four groups of variables, where one group is used to keep the record of the stock levels, and the other three are used to control the level of stock, by determining which products to stock, when to place a replenishment order, and how many to order.

If i represents a particular style and j a product size in that style, then the following variables represent the inventory system:

STOCK (i,j) - A one-zero variable used to indicate whether or not product (i,j) is an inventory item.

QTSTCK (i,j)- A variable used to record the value of stock levels. If STOCK (i,j) is zero, then QTSTCK (i,j) remains always zero.

RPOINT (i,j) A control variable used to record the values of reorder levels for each individual product.

EBQ (i,j) - A control variable used to record the values of reorder batch quantities, for each individual product.

For experiments in which the inventory system is not used, it can be 'switched-off', simply by setting the values of all STOCK (i,j) to zero.

The inventory system can be 'switched-on' for any of the products by setting the value of STOCK (i,j) to one.

When switched on for any of the products (i,j), the inventory sub-system has a strong influence on the general behaviour of the production system, and in particular on the machine shop, which has to produce 'jobs' for both customers and stock replenishment orders. The interference of inventory in the behaviour of the model starts at the moment of generation of 'jobs' after the arrival of a customer 'order'. As described in 3.2, in cases where the inventory is 'switched-off', the arrival of an 'order' causes the immediate generation of 'jobs' and their input to the machine shop. However if the inventory is 'switched on' the procedure is changed, and before a 'job' is generated the model checks the inventory subroutine to determine whether or not the requirements that would be contained in that 'job' can be satisfied from stock. Four alternatives can happen when this check is made.

i) The component required is not kept in stock, meaning that STOCK (i,j) is set to zero.

ii) The component required is kept in stock and there is enough stock to fully satisfy the demand.

iii) The component required is kept in stock but the present stock level can only partially satisfy the demand.

iv) The component required is kept in stock but the level of stock at present is zero.

For each one of these alternatives a different course of action is taken and full description of the procedure is given in the block diagram of figure 3.5

Another aspect of the inventory system worth mentioning is the problem of priority scheduling when there are both 'stock' and 'customer''orders', competing for the same facilities. Any procedure used will inevitably interfere with the whole behaviour of the system as far as delivery performance is concerned, and the design of effective procedures could on its own be the subject of a full investigation. However, as far as this study is concerned, there was not much scope or time for such investigation, and so a decision was made to use a single priority scheme, in which customer 'jobs' are given absolute priority over inventory replenishment 'jobs' such that an 'inventory job' is loaded in a machine only when there is absolutely no customer 'job' available for 'loading'. The detailed procedure is described in the section where scheduling rules are discussed.

3.5 - Control system

The control system is represented in the model by decision variables whose values are selected by the experimenter in order to study their influence on the behaviour of the production system.

Five areas of control are considered in the model:

i) the use of priority scheduling rules

ii) the splitting of jobs into smaller batches

iii) the selection of extra moulds and machines

iv) the use of additional working hours

v) the control of inventory

3.5.1 - Priority scheduling rules

Priority scheduling rules are used in connection with the process of selecting 'jobs' from queues in order to 'load' them into machines. As described in paragraph 3.3 'jobs' can be selected from queues on two occasions: firstly when they arrive at the shop, if there is an 'unloaded' station, and secondly when a previously 'loaded' station completes a 'job' and becomes 'unloaded'. In both cases, when the number of jobs in the queue is bigger than the number of stations available and/or the number of suitable moulds, a priority rule is used in order to decide which 'job' should receive 'loading' priority.

As discussed in 2.2, a considerable amount of research has been directed towards the problem of shop scheduling, in which both theoretical and experimental approaches have been used. Of particular interest to this study are the experimental investigations carried out with the objective of comparing the performance of different heuristic priority rules. A number of relatively simple rules have been proposed and analysed for a variety of situations, and results have shown that they are able to influence the delivery performance of production systems.

In order to analyse the effect of heuristic priority rules on the performance of the system under study, this model was provided with facilities for using any of eight priority rules. Three of them, namely SPT, SLACK and FIFO are well known rules, and the other five, viz. FIFOB, FIFOM, FIFOMB, SPTM and SLACKM are modifications of the original three.

Before describing the procedures followed by each of the eight priority rules, it is worth mentioning the particular characteristics of this production system which make scheduling procedures slightly different from most of the models described in the literature.

Firstly there is the arrival process. In most models 'jobs' arrive individually at the machines as independent entities, while in this model they arrive in groups. 'Jobs' in each group are related to each other by the fact that they are part of the same 'customer order' and must be delivered together.

Secondly, there is the problem of mould (tool) requirements and setup times. Most models which considered the problem of setup times have assumed that their expected values, as well as their actual values, depend on either the 'job' or the particular machine in which it is being 'loaded'. It is also assumed that tools are always available. In this model the availability of suitable moulds (tools) is as important as the availability of machines (stations). They are both restrictions that must be satisfied before a 'job' can be 'loaded'. Furthermore, the expected value of setup times is constant and does not depend on the 'job' being 'loaded'. Every time a change of mould takes place an actual setup time is generated from the same probability distri - bution, irrespective of the 'job' being 'loaded'.

Finally there is the problem of the method for setting due dates. Most studies have assumed that lead times used to fix due dates are variable and a function of the amount of work required by a 'job'. In this study lead times are assumed to be constant and independent (within certain limits) of the amount of work required by the 'jobs'.

These differences between models lead to a series of modifications in the procedure for selecting 'jobs' and loading them into the machines. These changes are due to the following factors:

i) In models where tools are not a restriction, the order of selection of a 'job' from a queue depends only on the availability of a free machine and the position of the job in that queue. This means that if a 'job' is the 'first' in 'queue' at the time a machine (station) becomes free, it will be immediately selected and 'loaded'. In this model, however, because of mould restriction, the fact that a 'job' is the 'first' in the queue does not necessarily mean that it will be the first to be selected and 'loaded', when a machine (station) becomes free. Unless a suitable mould is also 'free', the 'job' can lose its priority and be overtaken by another 'job', for which a 'free' mould can be found.

ii) The characteristics of the arrival process on this model, together with the mould restriction, means that the activities which follow the arrival of a 'job' are different from a 'traditional' model. In both cases the two options for a 'job' at its arrival are either to join a waiting queue or be 'loaded' in a machine (station). However, the circumstances in which either of these options happens differ for the two models. In a 'traditional' model, due to the fact that 'jobs' arrive independently and do not require special tools, a 'job' will always be loaded at the moment of its arrival, if it finds a free machine (station). This happens because there is no other 'job' to compete with it for the facility, and so there is no need to consider priorities. In this model however, due to the fact that 'jobs' arrive simultaneously and require specific moulds, a job will not necessarily be 'loaded' at the moment of its arrival, even if it finds a free station. Before being selected for 'loading' a 'job' has to compete with other 'jobs' for both a station and a mould, and the selection is decided by using the priority rules in operation.

A final point worth mentioning, before the description of the priority rules, concerns the moulds. As described in 3.3, there is always a mould set up in every station, whether the station is 'loaded' or not.

It is assumed that when a station finishes a 'job', it retains its mould until there is a need for a change. This means that at arrival a 'job' could find a station free with a suitable mould already set up, and so no setting up would be needed.

The problems raised above point to the need for considering two stages of decision in the scheduling process:

i) how to sequence the 'jobs' in the queue

ii) how to select 'jobs' and moulds to 'load' a station.

The first stage of decision is a straightforward procedure. After the criterion for queue priority is chosen, jobs are arranged in sequence in accordance with that criterion.

The second stage considers the problem of finding suitable moulds for the 'jobs'. The problem arises because there are three states in which a mould can be found.

i) it can be set up in a 'loaded' station, in which case it is said to be 'unavailable'

ii) it can be set up in an 'unloaded' station in which case it is said to be 'available'

iii) it can be out of station, in which case it is said to be 'free'.

The states of the moulds are not considered in the first stage of the scheduling procedure which is only concerned with sequencing the 'jobs' in the queue. Only in the second stage is this aspect considered. Because of this, a 'job' which gets the highest priority in the queue is not necessarily the first one to be selected and 'loaded'. In the second stage, because of mould considerations, priority for 'loading' could be given to a job which is behind in the queue.

In the section which follows all eight priority rules are defined, by describing their procedure in each one of the two decision stages:

a) SPT

i) organizes the queue by giving highest priority for the job with the minimum imminent processing time. In case of a tie it chooses the job with smallest generation serial number.

ii) selects the first 'job' in queue for which a mould can be found either in the 'available' or 'free' state.

b) SPTM -(is a modification of the SPT)

i) organizes the queue in exactly the same way as in the SPT rule.

ii) selects the first 'job' in queue for which an 'available' mould can be found. If no such mould can be found, selects the first 'job' in queue for which a 'free' mould can be found.

c) SLACK

i) organizes the queue by giving priority to the job with the minimum slack time for the due date. In case of a tie it chooses the 'job' with minimum arrival serial number.

ii) selects the first job in queue for which a mould can be found either on the 'available' or 'free' state.

d) SLACKM - (modification of SLACK)

i) organizes the queue in exactly the same way as the SLACK rule

ii) selects the first 'job' in queue for which an 'available' mould can be found. If no such mould can be found, selects the first 'job' in queue for which a 'free' mould can be found.

e) <u>FIFO</u>

i) organizes the queue by giving priority to the first 'job' to arrive at 'queue'. In case of a tie chooses the 'job' with smallest generation serial number

ii) selects the first 'job' in queue for which a mould can be found either in the 'available' or 'free' state.

f) <u>FIFOM</u> (modification of FIFO)

i) organizes the queue in exactly the same way as FIFO

ii) selects the first 'job' in queue for which an 'available' mould can be found. If no such mould is found, selects the first 'job' in queue for which a 'free' mould can be found.

g) <u>FIFOB</u> (modification of FIFO)

i) organizes the queue by giving priority to the first 'job' to arrive at 'queue'. In case of a tie chooses the 'job' with the <u>largest</u> imminent processing time

ii) selects the 'job' in exactly the same way as FIFO

h) FIFOMB (modification of FIFOB)

i) organizes the queue in exactly the same way as FIFOB

ii) selects the first 'job' in queue for which an available mould can be found. If no such mould is found, selects the first 'job' in queue for which a 'free' mould can be found.

Attention should be paid in the FIFO class rules to the fact that jobs belonging to the same 'order' arrive simultaneously at the 'queue'. For the FIFO and FIFOM rules this means that the ties decisions are dependent on the sequence in which 'jobs' in the same 'order' are generated. In this model it is assumed that generation of 'jobs' in an 'order' are made in sequence, starting with the smallest product size and finishing with the largest one. For the cases of FIFOB and FIFOMB rules the tie decision means that production of an 'order' should always start from the 'job' with the largest batch size. This is justified by the fact that no partial delivery is allowed, and so an 'order' tends to be delayed by its largest 'job'. By giving preference to this particular 'job' it would tend to minimize the 'order' waiting time.

Another important aspect which should be noted refers to the fact that all the eight rules described above are designed for a non-inventory system. For the cases in which the inventory system is 'switched on', the procedure is as follows:

i) separate the customers 'jobs' from the inventory replenishment 'jobs'

ii) apply the priority 'loading' rule in operation to the 'customers' 'jobs'. If no such 'job' can be 'loaded', apply the priority rule to the 'inventory replenishment' 'jobs'.

3.5.2 - Splitting of 'jobs' into smaller batches

Quantities demanded in each customer order vary considerably from order to order. They also vary for the different product sizes belonging to the same order, meaning that some of the 'jobs' require much longer manufacturing times than others. This tends to create an imbalance in the completion times of 'jobs' belonging to the same order . One possible way of controlling this is to split large 'jobs' into smaller batches, such that production of two or more batches of the same original 'job' could be running simultaneously in different stations. For this to be possible there must exist at least one replicate 'mould' of the same style and a suitable size of the 'job' to be split.

This model is supplied with a facility which allows the user to decide which 'jobs' should be split into smaller batches. This is done through the use of control variables, whose values are input as data, such that they can be changed between experiments. The first of these variables is an array whose elements correspond to individual product sizes. If the value of an element is zero, 'jobs' for the corresponding product size are in no circumstances considered for splitting, whereas if the value is 1 'jobs',whose batch sizes are bigger than a certain limit, are split into smaller batches.

The second control variable (MAXLOT), is used to set the limit mentioned above, such that all 'jobs' whose batch sizes are bigger than MAXLOT, and whose corresponding array element is one, are splitinto smaller batches. The number of batches into which a job is split, is obtained by dividing the 'job' batch size by MAXLOT and approximating the result to the next integer number. This means that batch sizes for 'split batches' are limited to values between MAXLOT/2 and MAXLOT.

Splitting of 'jobs' occurs during the process of 'order' arrival, and before 'jobs' are sent to the queue in the machine shop. The whole procedure is shown as a block diagram in figure 3.6.

3.5.3 - Selection of extra moulds and machines

In order to study the effect of extra moulds and machines on the performance of the system, the model allows the user to 'switch' machines and moulds on and off between experiments.

As described in paragraph 3.3.1 a machine can be easily 'switched-on' or off by setting one of its parameters [MACHINE. I (1)] to one or zero respectively. Considering that all machines are equivalent, in terms of production capability, the decision is limited to choosing the number of machines in operation.

The decision about moulds is more complex because they are not equivalent to each other. Apart from deciding on the number of moulds, one has also to decide which moulds to select. As described in 2.3.2, each product requires a special mould, so that for each product style, there must exist a minimum number of moulds in order to satisfy the technological requirements of production. However this minimum technological requirement may not be enough to satisfy the capacity requirements of production.

Because moulds play such a vital role in the production process, it is important to analyse the effect of extra moulds on the performance of the system. This is made possible by the use of control variables which determine the number and specifications of the moulds in operation. After fixing the total number of moulds for a given experiment, the user can specify each one of them, using one of the mould's parameters described in 3.2.2, i.e. the list of products that a mould is able to manufacture.

The right selection of moulds is an important decision and one of the objectives of this study was to devise a procedure which would allow rational decisions to be taken in this area. This procedure is based on the use of information from some of the output variables, and carried out externally to the model. For this reason, discussions about the procedure will be left until the section in which the output variables are discussed (paragraph 3.6.1).

3.5.4 - Use of additional working hours

In order to study the effect of overtime and an extra shift on the performance of the system, the model is provided with a facility for modifying the number of working hours per day. This is done by the use of a control variable (VCONV), which is input as data, and specifies in minutes, the actual amount of working time per day. The procedure is based on the utilization of a correction factor obtained by the ratio between VCONV and the number of minutes in a normal working day, which is applied to all time-based variables such as interarrival time and delivery delay.

To be able to make comparisons with industrial data, the concept of real time was used, such that a normal working day is equivalent to 540 units (minutes), and a week is made up of five days. By varying the value of VCONV, for instance, from 540 to 1080, one would have a correction factor of 2, which would be applied to all time-based variables, such that the actual amount of working time per day would be 1080 minutes, which is equivalent to two shifts of 540 minutes each.

3.5.5 - The control of inventory

Control of inventory is effected by 3 sets of variables, whose values are input to the model as data such that they can be easily changed in between experiments.

The first set, as described in 3.3, is composed of the variables STOCK (i,j), which are used to specify whether or not a product of size i and style j will be kept in stock. If STOCK (i,j) is set to one, the item is a stock item, otherwise its value should be set to zero. When the values of all STOCK (i,j) are set to zero, the inventory subsystem is completely 'switched off' from the model. The last two variables, RPOINT (i,j) and EBQ (i,j), are used to control the level of stock [QTSTCK (i,j)], by establishing the value of the reorder point, and the reorder batch quantity respectively. The values of both RPOINT(i,j) and EBQ (i,j) are chosen externally by the user between experiments.

The use of the reorder point method for controlling inventory means that every time a demand causes the inventory level QTSTCK (i,j) to drop below the reorder point RPOINT (i,j), an inventory replenishment order of batch size equal to EBQ (i,j) is issued to the machine shop.

Issue is executed by generating an 'inventory job' representing product (i,j). No other 'inventory job' will be generated for product (i,j) until the original 'job' has been completed, so that at no time will there be more than one 'inventory job' for a product (i,j). If after the receipt of a replenishment batch the level of QTSTCK (i,j) is still below

RPOINT (i,j), a new 'inventory job' of batch size equal to EBQ (i,j) is generated and input to the shop.

3.6 - The output variables

The output variables are the response of the system to the various inputs placed on it. They can be divided into two groups: one is made up of variables which measure the internal behaviour of the system, and the other of variables which measure the performance of the system in terms of delivery performance.

3.6.1 - Measures of internal behaviour

In order to have a picture of the internal behaviour of the system, the model outputs a series of diagnosis variables. Some of these variables were particularly useful in the initial stages of the study when the model was being validated and decision rules were being devised. Below is a list of these variables, which is followed by the explanations of how they are calculated:

a) Average number of 'jobs' waiting in the queue
b) Standard deviation of number of 'jobs' waiting in the queue
c) Average 'process cycle time'
d) Standard deviation of 'process cycle time'
e) 'Average load factor' on the system(actual)
f) Total demand
g) Machine idle time due to setup (percentage)
h) Machine idle time due to lack of work (percentage)
i) Mean waiting time in queue
j) Standard deviation of waiting time in queue
k) Mould's idle times (percentage)
l) Mean processing time of 'jobs'
m) Total production delivered
n) Average level of stock

a,b) The average and standard deviation of the number of jobs in queue is obtained by taking samples of the number of 'jobs' waiting in queue every time the state of the queue is modified.

c,d) The average and standard deviation of 'process cycle time' is obtained by taking samples every time a 'job' is completed in a station. Each sample is obtained by dividing the 'job' processing time by the 'job' batch size.

e) 'Average load factor' is obtained by the ratio between production requirement and production capacity. Production requirement is calculated by multiplying total demand by the average 'process cycle time', and production capacity is obtained by multiplying the total simulation time by the number of available stations.

f) 'Total demand' is obtained by adding the total quantities required in each customer's order.

g) 'Machine idle time due to set up' is calculated as a percentage, obtained by dividing the total time spent in setting up, by the result of the product of the total simulation time and the number of machines.

h) 'Machine idle time due to lack of work' is calculated as a percentage, obtained by the ratio between stations idle time, and total production capacity. Stations idle time is obtained by adding together the times of all the 'unloaded' periods of the stations and total production capacity is obtained by multiplying the total simulation time by the total number of stations.

i,j) Average and standard deviation of waiting time in queue is calculated separately for each individual product. They are obtained by measuring the waiting time of all 'jobs' for each product, and calculating their mean and standard deviation.

k) Mould's idle time is calculated for each single mould, and is expressed as a percentage obtained by the ratio between total time that each mould stays idle, and the total simulation time.

l) Average processing time of 'jobs' is calculated individually for each product size, and is obtained by taking samples of processing time each time a 'job' is completed in a station.

m) Total production delivered is calculated by adding together the batch sizes of all orders delivered to customers.

n) Average level of stock is calculated separately for each class of style. Their values are obtained by taking samples of the level of stock and the time between variations, each time there is a variation in the level of stock. The average value is obtained by weighting each stock level sample by its corresponding interval of time, and dividing at the end by the total simulation time.

If T = total simulation time

dt_i = interval between instants t_i and t_{i-1}

Q_i = stock level between instants t_i and t_{i-1}

then, Average stock level $= \frac{1}{T} \sum Q_i . dt_i$

A few comments should be made on the uses to which some of the measures of internal behaviour were put. Chief among them is the use of the variables, mean and standard deviation of waiting time in the queue, moulds' idle times, and average processing times of jobs, in the execution of a procedure for selecting extra moulds. In paragraph 3.5.3 the importance and characteristics of the selection problems have already been discussed, and now the selection procedure is described.

The selection procedure is based on the fact that an 'order' is held back from delivery until its last 'job' (product size requirement) is completed. If the distributions of throughput times for the different product sizes could be determined, it would be possible to calculate the probabilities that each will have a throughput time bigger than the promised delivery time.

These probabilities could then be compared and the product sizes classified in accordance with their probabilities of delaying the delivery of orders. Based on this classification, moulds could then be selected. One problem however is that it might be not very easy to identify the distributions of throughput times for the product sizes, and a solution would be to build histograms for each product size. However in order to build these histograms, a very large amount of information would have to be stored. One way of avoiding this is to use parameters of two distributions which make up the throughput time, viz. waiting time in queue and processing time. The use of these two distributions provides more information than the throughput time by separating the waiting time, which can be reduced by providing extra moulds, from the processing time, which cannot be reduced by extra moulds (for the same batch size).

The selection procedure starts by analysing the outputs from an initial experiment in which the machine shop is provided with the minimum possible number of moulds, which are necessary to satisfy the technological requirements of production, as described in 3.5.3. For this first experiment, there is no problem in choosing between moulds, because the selection is bound by the technological requirements. However from this point on a decision about extra moulds should take into consideration the contribution that each mould would make towards improving the delivery performance of the system. This can be done by looking at the lateness probabilities of the product sizes, together with their distribution of waiting time and the percentage utilization of each mould. Preference should be for duplicating moulds which have a high utilization level and whose corresponding product sizes have a long waiting time and a high probability of lateness. After each selection a new experiment, including the additional mould, could be executed, and the new results analysed.

3.6.2 - Measures of performance

In order to measure the performance of the system, such that different configurations and operation rules can be evaluated, the model outputs the following variables:

a) Average delivery delay of orders

b) Standard deviation of delivery delay of orders

c) Average delivery delay of 'production'

d) Percentage of late orders

e) Tardiness index of orders

f) Percentage of 'production' delivered late

g) Tardiness index of 'production'

a,b) The average and standard deviation of delivery delay of orders are calculated from the samples obtained by measuring, for each order which is completed, the time elapsed from the arrival of the order at the system to its delivery. The delivery corresponds to the completion of the last 'job' belonging to that order.

c) Average delivery delay of 'production' is a weighted measure of the delivery delay of orders. For each order which is delivered, the model measures the time it spent on the system, and weights this measure by the total quantity delivered with that order. At the end of the simulation the weighted measures are averaged.

d) Percentage of late orders is calculated by the ratio between the number of orders delivered after the due date, and the total number of delivered orders.

e) Tardiness index of orders is a measure of lateness dispersion. It is the summation of the products of the proportions of orders late and the number of days late.

If d = promised delivery delay (lead time) (days)

i = actual delivery delay (days)

$p(i)$ = proportion of orders with delivery delay equal to i days, then

$$\text{Tardiness index of orders} = \sum_{i=d+1}^{\infty} (i-d) * p(i)$$

In practical terms it is assumed that 45 days is the limit of lateness.

F) Percentage of 'production' delivered late is a weighted measure of the percentage of orders delivered late. It is calculated by the ratio between the number of items delivered after their due date, and the total amount of delivered items.

g) Tardiness index of 'production' is a weighted measure of lateness dispersion. It is calculated in the same way as the tardiness index of orders, with each order weighted by the total quantity delivered.

Apart from these measures of delivery performance which are output direct from the model, other measures of performance involving costs are also used, but are calculated outside the model. For this reason they are not described here and will be discussed in a later chapter together with the experiments in which they are used.

The measures of performance, percentage of late orders; tardiness index of orders; percentage of 'production' delivered late; and tardiness index of 'production', are calculated as a function of a delivery promise (lead time) which is always the same for all orders. The value of these promises however, can be varied from, say, eight days for all orders to, say, ten days for all orders, and in fact the model calculates the above measures of performance for seven different values of the delivery promises (lead times).

3.7 - Summary

In this chapter the characteristics of the simulation model used in this study are presented and discussed.

For the sake of explanation, the model is divided into five major components, viz. order input, machine shop, inventory subsystem, operation control systems and output variables. The characteristics of each of these components are explained in detail, and a series of block diagrams are used to help the explanation.

Details of the computer program used to implement the model, and which was written in CSL, are given in appendix 1.

FIGURE 3.1

MACRO BLOCK DIAGRAM OF MODEL

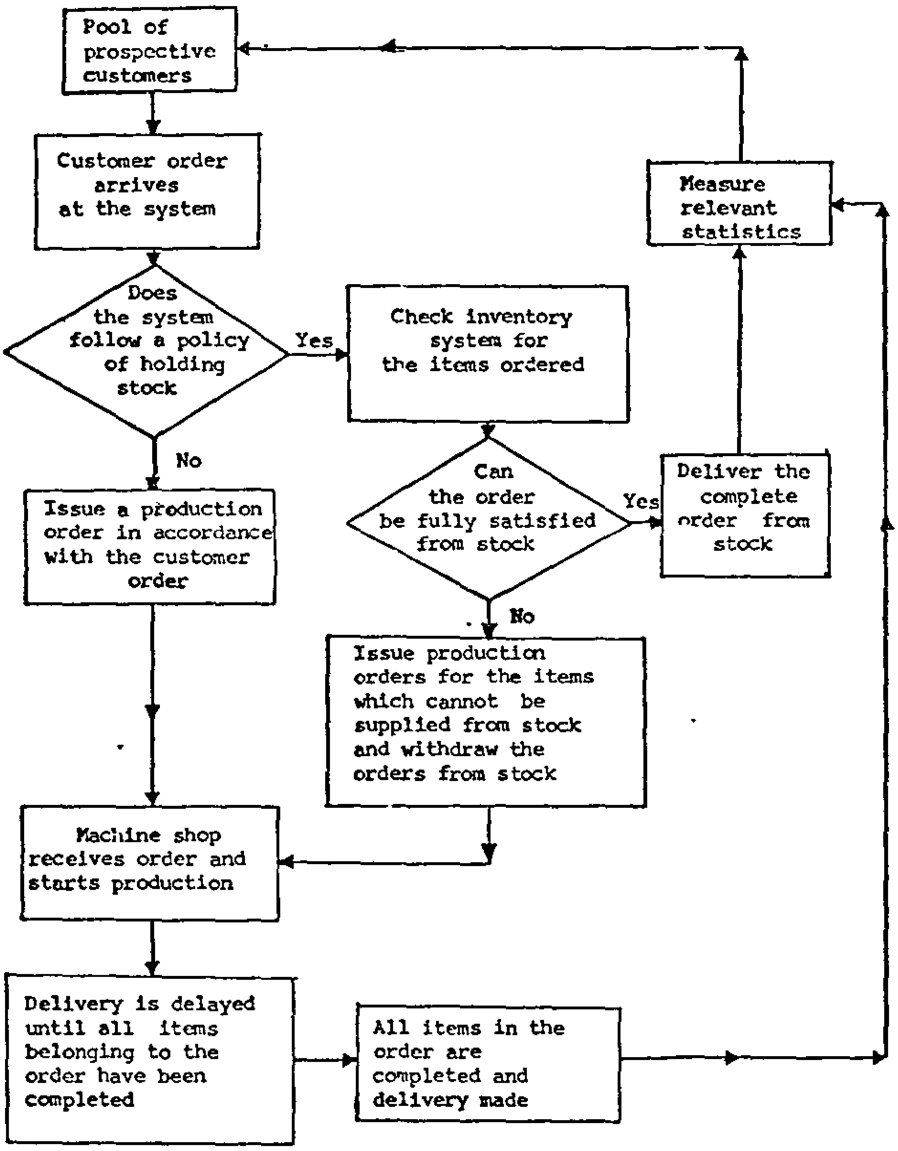


FIGURE 3.2

MACRO BLOCK DIAGRAM OF ORDER INPUT

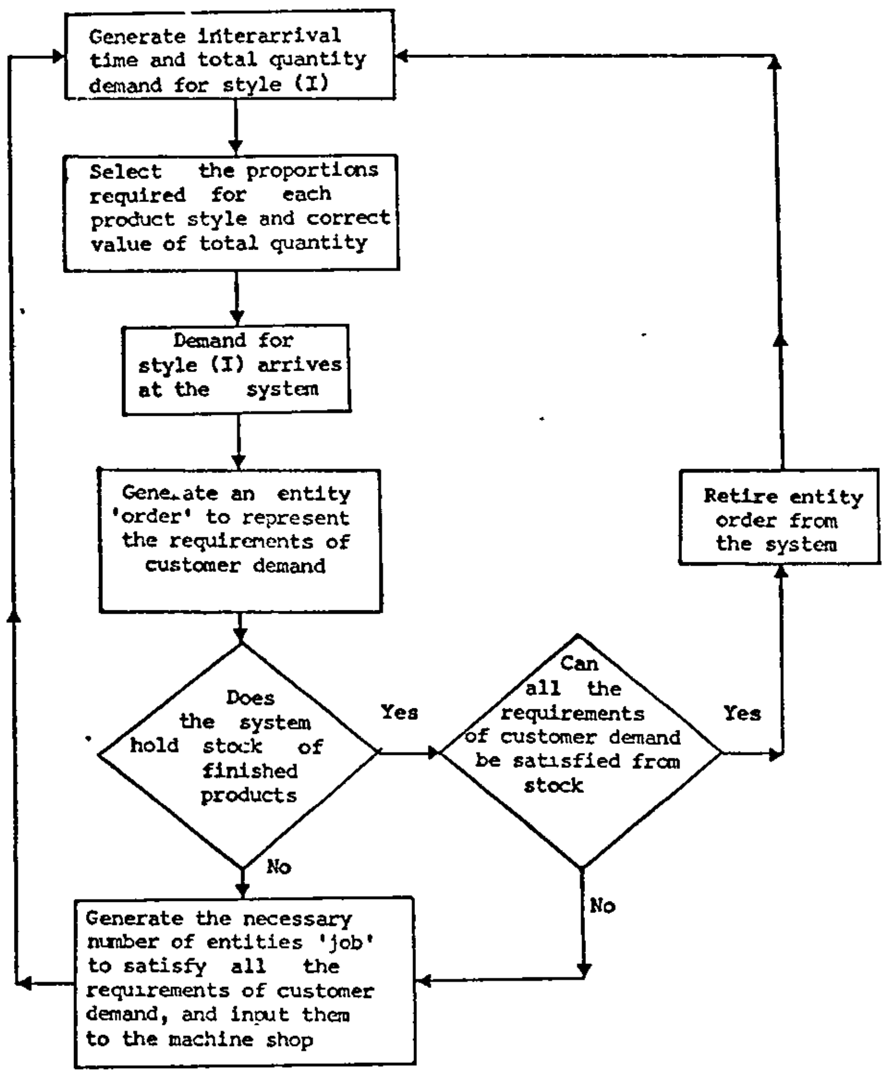


FIGURE 3.3

MACRO BLOCK DIAGRAM OF ARRIVAL OF 'JOBS' AT MACHINE SHOP

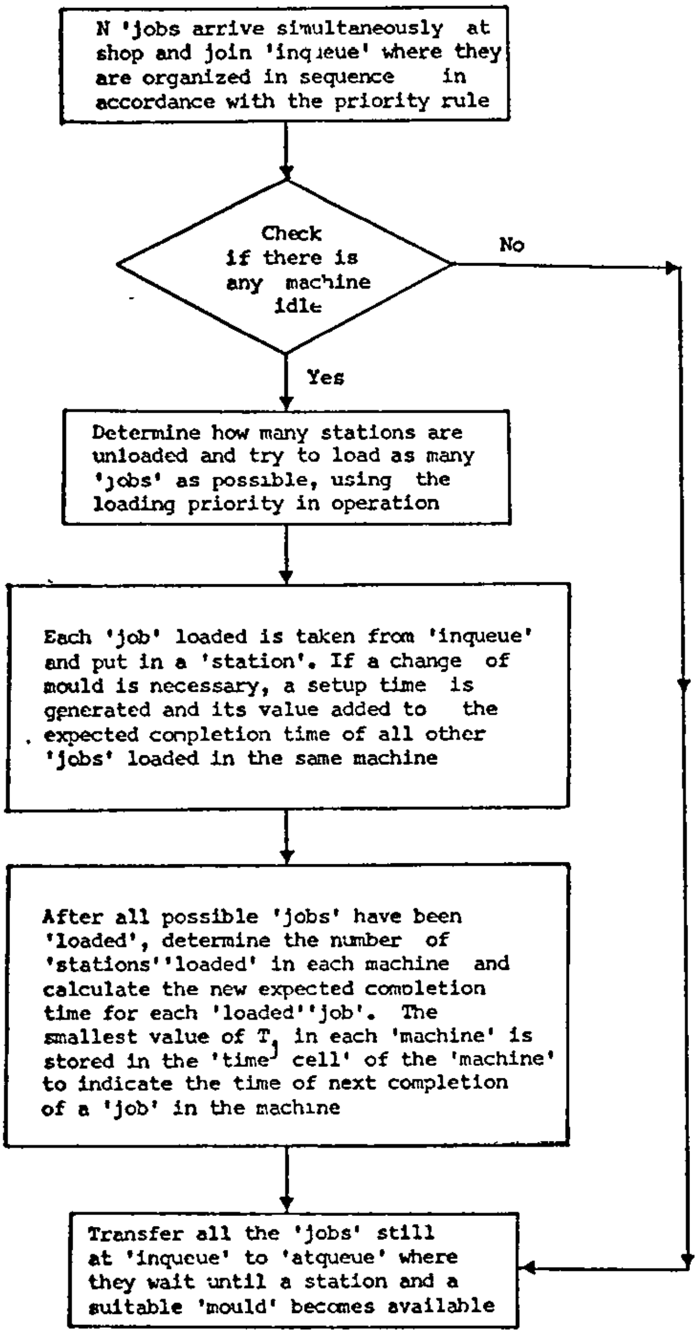


FIGURE 3.4

MACRO BLOCK DIAGRAM OF ACTIVITIES GENERATED BY THE EVENT COMPLETION OF 'JOB'

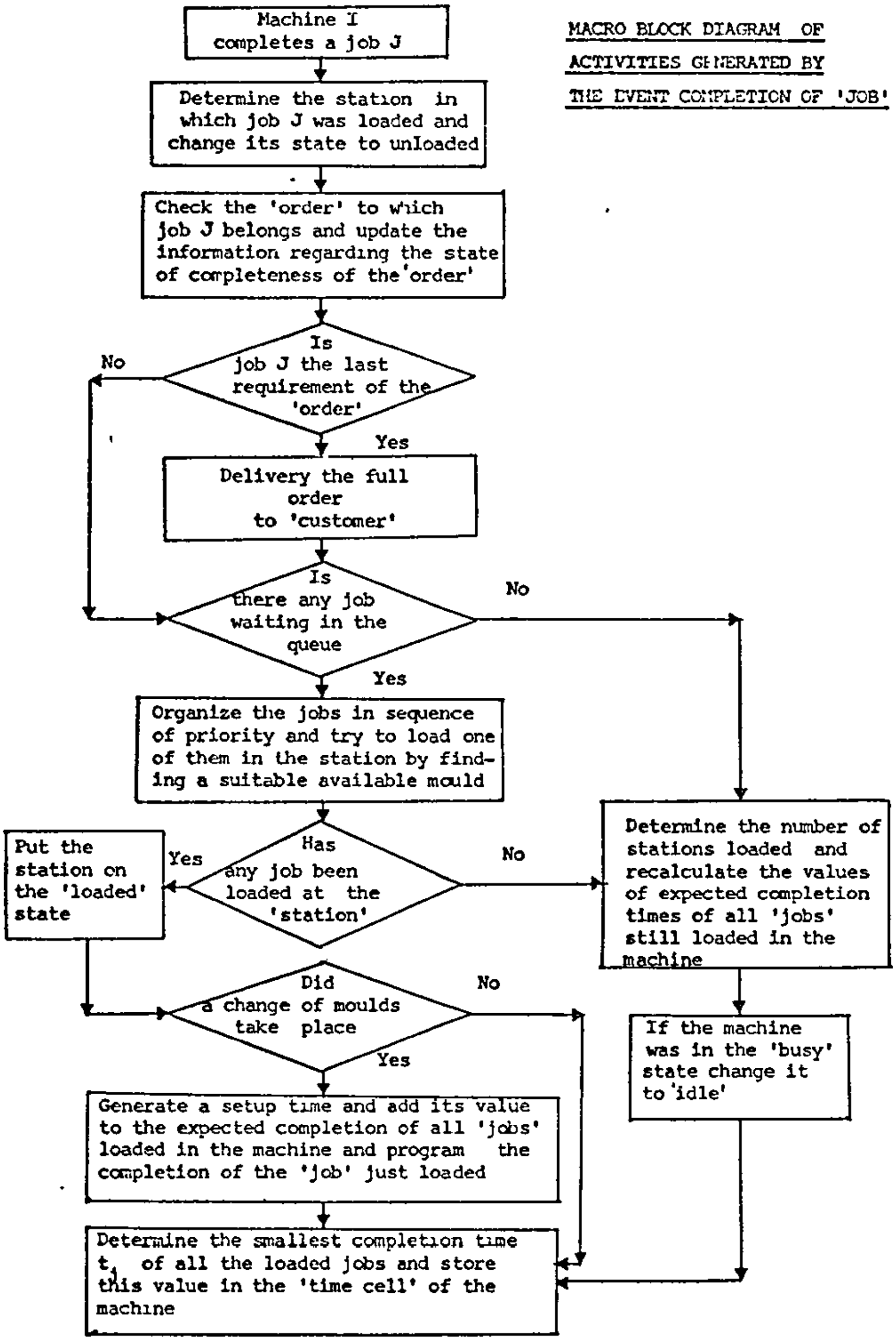


FIGURE 3.5

MACRO BLOCK DIAGRAM OF PROCEDURE FOR GENERATION OF 'JOBS', WHEN THE INVENTORY SUBSYSTEM IS 'SWITCHED ON'

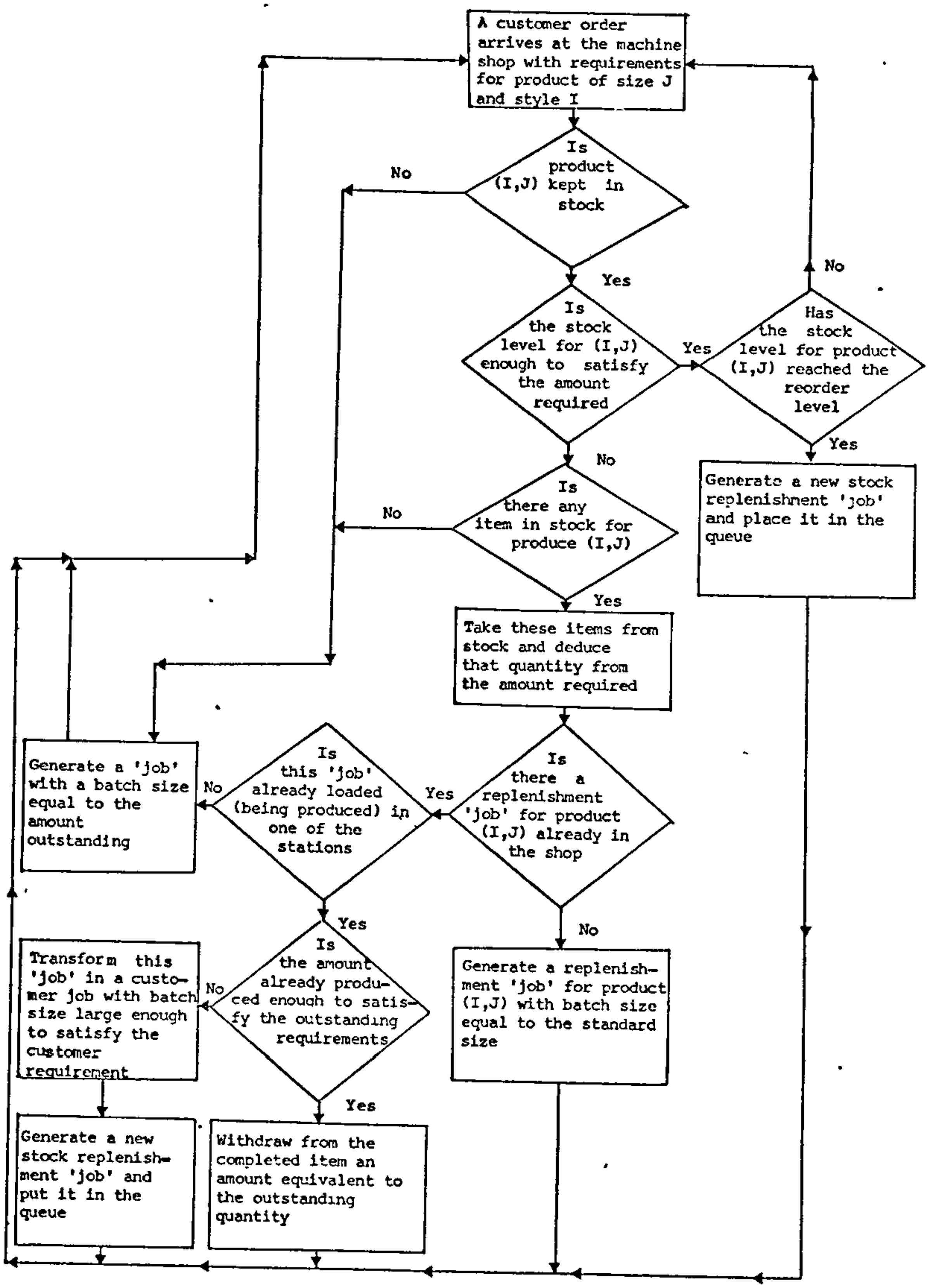
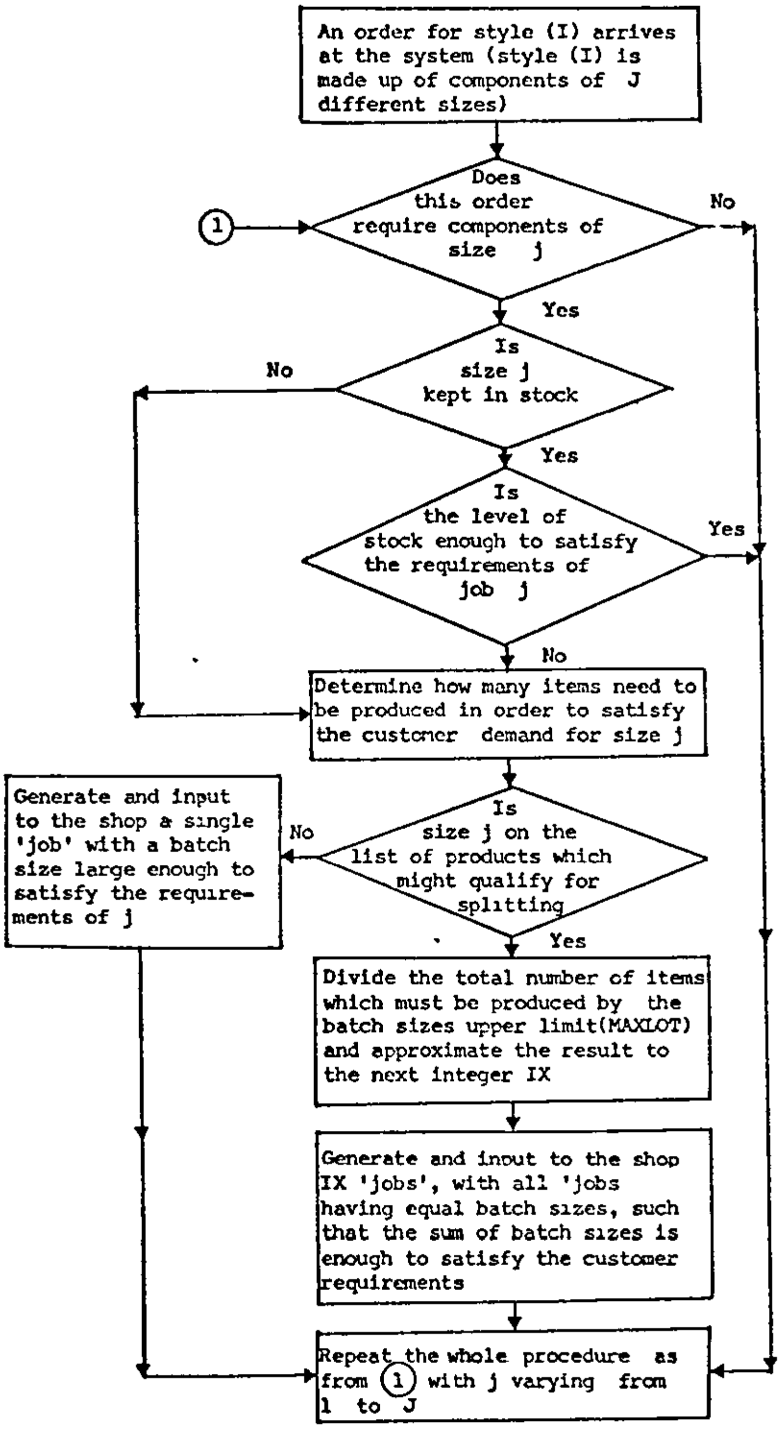


FIGURE 3.6

MACRO BLOCK DIAGRAM OF 'JOB' SPLITTING PROCEDURE



CHAPTER 4

PRELIMINARY INVESTIGATION AND EXPERIMENTAL DESIGNS

4.1 - Introduction

The investigations carried out with the model can be divided into two phases. The first phase (in which empirical information obtained in a particular company was used), was exploratory in nature. The main objectives of this phase were to validate the model and identify the major characteristics of the system, in order to determine typical values for the parameters, and work out possible control rules which might be appropriate to the characteristics of the system under study.

The second phase of the investigation consisted of a more formal set of experiments, which were designed with the objective of generating information which could lead to a more general set of conclusions about this class of production systems. To this end three different sets of experiments were conducted, in which individual experimental designs were organized. The first set of experiments was related to the study of priority scheduling rules. The second set was related to the study of the effects of some of the system's parameters on the system's behaviour. The third set consisted of the study of operation strategies for capacity manipulation.

The two phases of the investigations complemented each other, in the sense that the choice of parameters and the experimental designs of the second phase were largely based on the information obtained from the first phase of the investigation. This chapter is concerned with the description of the results obtained from the first phase of the investigation and with the discussion of the experimental designs of the second phase of the investigation.

4.2 - Preliminary investigation

To validate the model and gain an insight into the behaviour of the system, the model was initially run with information obtained from an industrial company which was intended to reproduce as close as possible the conditions of that production unit. At this time, the actual production unit was in its early stages of development and only a relatively small amount of information was available, concerning the distributions of demand.

During development of the model more comprehensive data was collected which, although not used in this initial phase, formed the basis of the major experimental programme. At this point it should be noted that the initial information was found to be sufficiently representative, since the conclusions from these initial experiments were confirmed during the major series of tests. Some of the parameters used in this preliminary investigation are:

a) Products: There were three classes of styles, where two were made up of eleven sizes and one made up of thirteen sizes. The mean values of interarrival times were 11.6 days, 6.07 days and 4.10 days respectively, while the average size of orders were equal to 2404, 1971 and 1611 items respectively. The three distributions of interarrival times plus the three distributions of order size and the distributions of proportions are given in appendix 2.

b) Machines: there was one twelve station machine. The value of average 'process cycle time' is given by $\overline{PCT} = (3.0 + 0.1\ (NSL))$ minutes [(1)].

c) Moulds: there were 26 moulds covering the three classes of styles. A list of moulds is given in appendix 2.

d) Setup time: the mean value of setup time was 8 minutes. The distribution of setup times is given in appendix 2.

e) Working hours: normal working hours were 5 shifts of 9 hours (45 hours) per week.

f) Due date: due dates were established in accordance with a fixed lead time which meant that any order which spent more than eight days in the machine shop was considered late.

g) Stocks: production was initiated only after the receipt of a customer order, i.e. no stock of finished parts was held.

(1) For description of 'process cycle time' see paragraph 3.3.1

Because no formal priority rule was in operation in the company, it was decided to use the FIFO priority rule[(2)] in the initial runs. It was also decided not to give much consideration to tactical and statistical problems in these preliminary runs, and a single long run, equivalent to a period of three years was used, in which statistics obtained from the first eight weeks of simulation were discarded, to allow for stabilization of the model. A full study of tactical and statistical problems was however conducted for the second (and major) phase of experiments, as reported in chapter 5.

4.2.1 - Validation-runs and analysis of internal behaviour

With the model fed with the above data and before any long runs were made, a series of short runs were executed, in which the program was instructed to output information at the occurrence of every event, such that the logic of the model could be checked and compared with empirical information. This procedure was repeated each time a new modification was introduced to the model (program).

After careful checking that the logic was correct an initial long run was conducted in which the general pattern of demand and the delivery

(2) For description of operation procedure for FIFO rule see paragraph 3.5.1

performance were compared with available information from the company. This initial run also provided useful information about the internal behaviour of the system.

Information output by the model included the histograms of mean waiting time in queue, and mean processing time of 'jobs'; the level of mould utilization; the percentage of time spent setting up the machine; and level of delivery performance obtained by the system.

The histograms of mean processing time of 'jobs' and mean waiting time in queue are shown respectively in figures 4.1 and 4.2. Each figure shows three histograms, where each histogram represents the range of shoe sizes belonging to a particular shoe style. As should be expected, all three histograms of figure 4.1 have bell shaped formats, similar to the distributions of demand (see paragraph 2.2.3), for the shoe sizes in a style. On the other hand, the histograms of figure 4.2 (mean waiting time in queue) differ markedly from each other. The histogram for the product sizes of style one, is similar to what would be expected from the use of the FIFO priority rule. As discussed in paragraph 3.5.1, the FIFO rule gives priority to the 'jobs' having the smaller arrival serial number, which means that 'jobs' belonging to the same 'order' are given priority in accordance with the product size they represent. The smaller product sizes get preference over the larger product sizes, as this is the sequence in which they are generated. This priority scheme means that 'jobs' representing the larger product sizes would tend to wait longer in queue, than 'jobs' representing the smaller product sizes. This effect is confirmed by the first histogram which represents product sizes of style one. However the other two histograms (for styles two and three) do not show the same effect. Although there is still a tendency for longer waiting times for the larger sizes, the two histograms do not follow the same smooth pattern as histogram one does. What characterizes those two histograms is the fact that a few product sizes in each style (4-1/2 and 6-1/2 for style two; 4-1/2 and 7 for style three) have distinguishably longer waiting time in queue than the other product sizes. The reasons behind this effect will be shown later to be related to the restricted number of moulds for those particular sizes.

Another histogram of interest as far as delivery delay is concerned, is the histogram of average throughput time of 'jobs' representing the different product sizes in a style. This histogram can be obtained by adding the histogram of mean processing time of 'jobs' to the histogram of mean waiting time in queue.

In figure 4.3 the histograms of average throughput times for the different product sizes, for each of the three styles are presented. They show that for each style, there are few product sizes (5, 5-1/2 and 6 for style one, 4, 4-1/2, 6 and 6-1/2 for style two; and 4-1/2 and 7 for style three) which have markedly higher throughput times than the other product sizes in the corresponding range. The implication is that the product sizes with the higher throughput times have a large share of the delivery delay of orders for their style.

In order to detect possible reasons for the long waiting times suffered by some of the product sizes, an analysis was made of the level of mould utilization. It is known from paragraph 2.2.1 that a mould of a certain size can be used in the manufacture of more than one product size. Take, for example, product size 5 of style one. In accordance with the list of moulds held by the company (appendix 2), there are two moulds of size 5 and two moulds of size 5-1/2. Considering that a mould of size 5 can manufacture product sizes 5 and 5-1/2, and that a mould of size 5-1/2 can manufacture product size 5-1/2 and 6, it means that there are four moulds available for the manufacture of product size 5-1/2, where each mould is shared with another product size. When relating mould idle capacity to individual product sizes, this fact must be taken into consideration. As described in 3.6.1-e, the model outputs the percentage of idle time for each individual mould held by the system. By analysing these outputs it is possible to calculate an index to represent the percentage of mould idle capacity available for each product size. This index will be called 'index of idle capacity'. Take again the case of product size 5. The outputs from the initial run show the following percentages of idle time for the four moulds related to it.

MOULD SERIAL NUMBER	MOULD SIZE	PRODUCT SIZES RELATED TO MOULD	PERCENTAGE OF IDLE TIME
5	5	5; 5-1/2	68.85
6	5	5; 5-1/2	88.84
7	5-1/2	5-1/2; 6	65.52
8	5-1/2	5-1/2; 6	65.28

When calculating the 'index of idle capacity' allocated to each product size, it is assumed that the percentage of idle time of each mould is equally shared by all the product sizes it is able to manufacture. In this way, the 'index of idle capacity' for product size 5 will be equal to 143.37% (68.85/2 + 88.84/2 + 65.52/2 + 65.28/2). This index can be similarly calculated for each product size, and then compared with the values for mean waiting times in queue for individual product sizes.

In figure 4.4 three histograms are presented, which show the 'index of idle capacity' allocated to each product size, for each of the three styles. It can be seen that, in general, there is a fair amount of idle capacity in terms of moulds, but some of the product sizes (3-1/2, 7-1/2 and 8 for style one; 4, 4-1/2, 6 and 6-1/2 for style two; 4-1/2, 5, 6-1/2 and 7 for style three) have a much smaller share of the idle capacity than some of the others. The average level of mould utilization was equal to thirty two per cent, i.e., there was sixty eight percent of mould idle capacity, most of which allocated to style 1.

In figure 4.5 the histograms of 'index of idle capacity' per product size (figure 4.4) is superimposed on the histogram of mean waiting time in queue (fig. 4.2). An analysis of figure 4.5 indicates a relationship between mould idle capacity per product size and average waiting time in queue. This relationship is clearer for product sizes 4-1/2 and 6-1/2 of style two and 4-1/2 and 7, of style three. In those cases, although the amount of idle capacity is still nearly twenty percent, the fact that two product sizes had to share a single mould (see list of moulds in appendix 2) meant that one of the product sizes had always to wait for the completion of the other before it could be 'loaded' in a station.

Other interesting information obtained from the initial run refers to the 'percentage of machine idle time due to setup', which was relatively small (three percent), and the 'machine idle time due to lack of work', which was relatively high (thirty percent). It is also interesting to note that although there seemed to be plenty of spare plant capacity, the delivery performance was relatively poor: twenty percent of the orders were late, with a tardiness index of orders equal to 0.73, and thirty five percent of production was late, with a tardiness index of 1.17.

4.2.2 - Development and initial tests of priority scheduling rules

In view of the information obtained from the initial run, it was decided that more experiments should be made in order to test the possible effects of different operation control procedures, on the behaviour of the system.

One possible modification of the control procedure is the use of priority scheduling rules better suited to the characteristics of this production system. To this end three modified versions of the FIFO rule (FIFOB, FIFOM, and FIFOMB) were devised and compared against two well known priority rules (SPT and SLACK) and modified versions of them (SPTM and SLACKM). The description of each of these priority rules have already been given in paragraph 3.5.1.

Before discussing the results obtained by using these priority rules, a few comments should be made on the reasons behind the modifications introduced to the FIFO rules, which were later partially extended to the SPT and SLACK rules. The idea behind the FIFOM rule, and by extension, the SPTM and SLACKM, was to reduce the amount of time lost with setting up (changing moulds), by giving an extra priority to 'jobs' which could be 'loaded' in a machine without the need of changing moulds. Although the results from the initial runs suggested that only a small percentage of time was spent with setting up, it was thought that the reduction in the number of mould changes could bring some improvement on the performance of the system. The idea behind the FIFOB rule was to give priority to 'jobs' with larger batch sizes, over their companion 'jobs' in the same 'order', which have smaller 'batch sizes'. This procedure would tend to reduce the waiting time in queue for the 'jobs' belonging to the high demand (large batches) product sizes, with a possible reduction of their average throughput time. To check the veracity of this assumption an initial run was made with the FIFOB rule, in which the histograms of mean waiting time in queue were analysed. Fig. 4.6 presents the three histograms of average waiting time in queue, each corresponding to a different product style. A comparison between figure 4.6 and figure 4.2, which presents similar histograms for the FIFO rule, shows that the use of the FIFOB rule has caused a desirable modification in the shape of the histogram for style one. The new histogram is now U shaped, with the high demand (high processing times) product sizes having a smaller mean waiting time in queue, and the low demand (low processing time) product sizes having larger waiting time in queue. This U shaped waiting time histogram, combined with a bell shaped processing time histogram, would tend to create a more uniform histogram of average throughput times. For style two and style three, however, the U shaped effect did not occur, and their histograms have maintained their original shape. This might be explained by the small number of moulds available for style numbers two and three, which most probably has hindered any changes in the 'loading' sequence which was brought about by the FIFOB rule in the case of style number one, which had more moulds available.

Finally the FIFOMB priority rule was designed in order to combine the characteristics of both FIFOB and FIFOM priority rules.

The results of delivery performance [3] and percentage of machine idle time due to setup, obtained by the use of the eight priority rules are presented in table 4.1. From the results it appears that introduction of modifications in the FIFO rule can bring some improvements in the system performance. All the three modified versions, FIFOB, FIFOM, and FIFOMB, have produced slightly better results than the FIFO rule in all five measures of performance, with FIFOMB producing the best results among them. The differences however are relatively small and

(3) For definition of measure of performance see paragraph 3.6.2

might not be statistically significant.When FIFOMB is compared with the other rules (SPT, SPTM, SLACK and SLACKM) it performs quite well. Apart from the result of percentage of late orders, in which it comes third to SPTM and SPT, the FIFOMB rules comes first in all the other measures of performance used. It is interesting to note that although SPT and SPTM did well in relation to percentage of late orders, they did particularly badly in terms of lateness dispersion (tardiness index of order and tardiness index of production), which indicates their tendency to delay certain 'jobs' for a very long time.

Another observation which came out of this series of experiments was that although the use of priority scheduling rules could have some positive effect on the performance of the system, the general level of performance was still poor.

4.2.3 - The effects of 'job splitting'; extra machines; and new demand pattern

To further explore possible ways of improving the delivery performance and to get more information about the nature of the system, another series of three experiments were conducted:

i) The first experiment consisted of using the control procedure which would allow the splitting of larger 'jobs' into smaller batches, such that it would be possible (in cases where there was more than one suitable mould available), to manufacture more than one batch of the same original 'job' simultaneously in two different stations. By doing so, one would hope to reduce the average throughput time of 'jobs' (batches) belonging to the high demand (high processing time) product sizes. As explained in 3.5.2, two variables are used in order to control the splitting of jobs: the first variable, SSTYLA (I,J), is used to specify which products should have their jobs considered for splitting, and the second variable (MAXLOT) is used to determine how large a 'job' should be before it is split into smaller batches. It is evident that in cases where there is only one mould suitable for the manufacture of a certain product size, the 'jobs' belonging to that product size should not be split, because the split batches would have to wait for each other, as no partial delivery is allowed. The decision taken was therefore to split all 'jobs' for which there were at least two suitable moulds, and which have a batch size bigger than four hundred and fifty components (MAXLOT = 450). The choice of four hundred and fifty for MAXLOT was based on the results obtained from the previous experiments, which had shown that the longer mean waiting times in queue varied between four and five days. Therefore if the batch sizes for 'jobs' are limited to four hundred and fifty items (between three and four days production), there should be a better chance of delivering orders inside the eight days promise.

ii) The second experiment consisted of modifying the demand pattern, though maintaining the same level of demand. The idea was to look at the influence of the total amount required per 'order' on the delivery performance of the system. This can be done by modifying both the distribution of interarrival times (by halving its mean value) and the distribution of total quantity per order (by also halving its mean value). This modification in the demand pattern means that twice the number of orders would be arriving at the system but the average size of orders would be half that of the order size in the original experiments.

iii) The objective of the third experiment was to observe the effect on the shop delivery performance, caused by a large increase in machine capacity, brought about by the addition of an extra machine to the machine shop. This would mean doubling the overall production capacity and halving the load factor.

Each experiment was completely independent from one another. In all three experiments the FIFOMB rule was used as the scheduling rule. The results obtained are shown graphically in figure 4.7 and numerically in table 4.2. Figure 4.7 presents the distribution of lateness for the three new experiments, compared with the same distribution for the previous experiment using the FIFOMB rule. Table 4.2 shows the numerical results of the 'average delivery delay of production', 'percentage of production delivered late', and 'tardiness index of production', for the three new experiments and for the original FIFOMB experiment.

The results of table 4.2 and figure 4.7 suggest the following:

i) The splitting of jobs in the present shop configuration (1 machine, 26 moulds) has almost no effect on the delivery performance. The percentage of 'production delivered late' came down from 31.11 to 29.94 percent and the tardiness index has not changed. This is probably due to the limited number of moulds for style numbers two and three.

ii) The inclusion of an additional machine had a large impact on delivery performance. The percentage of production delivered late came down from 31.11 to 1.80 percent and the tardiness index changed from 1.05 to 0.04. This was due to a large reduction in the average waiting time in queue, which came down from 1.96 days to 0.79 days, a drop of more than 50 percent, and also to the reduction in the average value of 'process cycle time', which came down from 4.47 minutes to 3.50 minutes, a change of -21.70 percent. This reduction in the 'process cycle time'(4) can be understood by the fact that its value is a function of the number of stations 'loaded' and the amount of setup time. As the load factor went down, the average number of stations 'loaded' also went down.

iii) The new demand pattern also had a large effect on the delivery performance of the system, showing that for the same level of production there could be large improvements in

(4) For definition of 'process cycle time' see paragraph 3.3

delivery performance, if the pattern of demand was favourable. The 'percentage of production delivered late' came down from 31.11 to 4.65 percent and the tardiness index was reduced from 1.05 to 0.13. The main reason for this improvement is that smaller orders mean smaller batch sizes for the 'jobs' and consequently smaller 'processing time' on the machine. Although there is an increase in the average waiting time in queue, and the amount of time spent in setting up, they are not large enough to offset the decrease in the 'jobs' processing time.

4.2.4 - The influence of moulds and of MAXLOT (Job splitting parameter) on the behaviour of the shop

In order to complete the preliminary investigation two additional series of experiments were devised in order to throw some light on two questions.

The first question relates to the possible influence that the value of MAXLOT could have on the effectiveness of the splitting procedure. The second question relates to the influence exerted by the moulds on the delivery performance of the system.

The first series of experiments had the objective of getting information on the influence that the value of MAXLOT and the 'mould restriction'(5) had on the delivery performance of the system. This was done by making a small change in the model which eliminated the 'mould restriction',

(5) The expression 'mould restriction' is used to express the fact that 'jobs' have to find a suitable mould before they can be 'loaded' at a station.

such that 'jobs' would always find a suitable mould available. This is theoretically equivalent to having an infinite number of moulds. With such modifications implemented it was possible to isolate the effect caused by MAXLOT values from the effect caused by the 'mould restriction'. This series consisted of 10 experiments in which the value of MAXLOT was changed in steps, from one hundred and fifty to one thousand and then to infinity (no splitting at all).

Results of these experiments are presented in figure 4.8 and table 4.3. Figure 4.8 shows the variation on the 'percentage of production delivered late' caused by variation in the value of MAXLOT. In table 4.3 the results of 'percentage of production delivered late'; 'tardiness index of production'; 'average number of 'jobs' waiting in queue'; and 'percentage of idle time due to setup', are presented for each of the ten experiments. From the analysis of the information, the following observations can be made:

i) 'Mould restriction' seems to have a large influence on the delivery performance of the system. The 'percentage of production delivered late', which was equal to 31.11, in the case where the 'mould restriction' was in operation (table 4.2), came down to 13.30 percent when the 'mould restriction' was lifted, a reduction of two thirds. The same effect was observed for the tardiness index of orders (down from 1.05 to 0.27).

ii) The splitting of jobs seems to have a positive influence when there is no 'mould restriction'. There was a drop from

13.30 percent to 6.30 percent in the percentage of 'production' delivered late, when 'jobs' were split with a MAXLOT value of 350.

iii) An analysis of figure 4.8 and table 4.3 seems to indicate that the curve of 'percentage of production delivered late' has a point of minimum, which occurs when the value of MAXLOT is between 250 and 350. The results also show that the 'percentage of production delivered late' is not very sensitive for values of MAXLOT above 350, but is quite sensitive for values below 250. For example, for a value of MAXLOT equal to 450, the percentage of 'production delivered late', changed from 6.30 to 6.70, a difference of only 0.40 percent, which most probably is not statistically significant. On the other hand, when the value of MAXLOT was put at 120, there was a sharp increase in the percentage of production delivered late, which moved from 6.30 percent to 8.50 percent, a change of 2.20 percent, which is still not a large difference in absolute terms, but which is considerable in relative terms. The reason for this sudden increase can be explained by the sharp increase in the average number of 'jobs' waiting in queue, which went up from 13.1 to 32.1, when the value of MAXLOT changed from 250 to 120. This increase in the number of 'jobs', resulted in a large increase in the percentage of time spent with setting up the machine, which went up from 3.71 percent to 6.11 percent, a relative increase of nearly sixty percent.

The second series of experiments in this last phase had the intention of following up from the results obtained in the above series, by going back to the more realistic 'mould restriction' situation. The idea was to get information about the effect that variation in the number of moulds might have on the delivery performance of the system. This was done by having a series of seventeen runs in which the number of moulds was increased in steps of one, from eighteen (which is the minimum technological requirement as described in paragraph 3.5.3) to thirty two, and then from thirty two to forty five in a single step. The selection of moulds for all seventeen experiments was made through the use of the procedure described in paragraph 3.6.1. The FIFOMB priority rule was used throughout the experiments, and apart from the last experiment in the series (forty five moulds), all the other experiments were executed without 'jobs' being split into smaller batches.

Results obtained from these experiments are presented in table 4.4, which contains the results of 'average delivery delay of production'; 'percentage of production delivered late'; and 'tardiness index of production'. A graphical presentation of the results is also given in figures 4.9 and 4.10 which show respectively the 'percentage of production delivered late', and the 'tardiness index of production', as a function of the number of moulds. The following observations can be made from the results:

i) There is a clear relationship between the number of moulds and delivery performance of the system. By looking at figures 4.9 and 4.10 it can be seen that the curves obtained have a shape similar to a negatively exponential curve, which means that increases in the number of moulds result in a large decrease in lateness and tardiness, when the number of moulds is small (between 18 and 26). However, when the number of moulds gets larger (above 27), increases in the number of moulds result in very small decreases in lateness and tardiness. For example, when the number of moulds was increased from eighteen to twenty seven (a net increase of nine moulds) the 'percentage of production delivered late' went down from 64.3 to 20.9 percent, a relative drop of 67.50 percent, and an absolute drop of 43.4 percent. However when the number of moulds was increased from twenty seven to forty five (a net increase of eighteen) the value of the 'percentage of production delivered late' went down from 20.9 to 18.0 percent, a relative drop of only 13.9 percent, and an absolute drop of only 2.9 percent. The same kind of observations are true for the tardiness index and average delivery delay of production.

ii) In experiment 17 with the number of moulds equal to forty five, 'jobs' were split into smaller batches with a value of MAXLOT equal to 450. When the results of experiment 17 are compared with the results of experiment 16 (which also had forty five moulds, but no splitting of 'jobs') it can be seen that splitting of jobs caused a drop in the 'percentage of production delivered late' from 18.0 to 12.7 percent, an absolute drop of 5.3 percent and a relative drop of nearly 30.0 percent.

iii) When the results of experiment 9 (which had twenty six moulds, selected by the procedure described in paragraph 3.6.1), are compared with the results obtained from a previous experiment (table 4.1 - FIFOMB) in which the selection of moulds was the one used by the company from which data was obtained (also twenty six moulds), it is possible to sce that using the twenty six moulds selected by the described procedure caused a drop in 'percentage of production delivered late' from 31.11 to 25.9, an absolute drop of 5.2 percent and a relative drop of nearly 17.0 percent.

iv) Finally it should be said that forty five is a very large number of moulds for this situation. An indication of this is the fact that some of the moulds were hardly used, with a level of idleness above 99 percent. This was true even for the case when 'jobs' were being split, and also when different selections of moulds were tried. The average mould's utilization factor in this situation was equal to only 18.0 percent.

It should be said that apart from helping a better understanding of the system, the results of these preliminary experiments were also used as a first guideline to the particular company from which the initial data was gathered.

4.3 - Experimental designs

The information obtained from the preliminary runs suggested three areas of investigation to be followed in order to get a better understanding of this class of production systems and to suggest possible ways of efficiently running the system. The first area consists of the analysis and comparison of the scheduling rules described in paragraph 3.5.1 and which were initially tested in the preliminary runs. The second area is concerned with the analysis of the effects caused to the system by variations in some of its parameters like, for example, the mean value of setup times, the load factor, etc. The third area refers to the study of strategies for capacity manipulation, where options like increased number of working hours per week, extra moulds, and the use of finished product inventory are compared in terms of costs and benefits (represented by better delivery performance).

4.3.1 - Experimental design for the study of priority scheduling rules

The preliminary tests with the priority scheduling rules, reported in 4.2.2, have produced some results which, although useful, were by no means conclusive. This is because of two main reasons. Firstly, the differences between some of the results were too small in order to allow any firm conclusion to be made without the backing of a statistical test. Secondly, and more important, the priority scheduling rules were tested for a single system configuration, and as a consequence, any conclusion which might have been reached, would be restricted to that single situation.

Ideally one would like to test the priority scheduling rules under as many different system configurations as possible. For example, one would like to see how the rules behave for different values of setup times, different load factors, different number of moulds, etc. As far as the model is concerned, it would be possible to analyse the priority rules for an almost limitless number of system configurations, simply by modifying the parameters of the variables provided in the model and described in chapter three. However, because of time and cost considerations one has to limit the number of experiments to a manageable size. As Bonini (1963) points out, when the number of changes that can be made is quite large, one must select some for study and ignore others. When referring to his particular study he says: "Since there is no concreteness or "boundedness" about the universe of all possible changes, we shall have to take a judgment sample, that is, we shall use our own judgment in deciding which changes in the firm to study".

Judgement was also used in this study in order to select the variables whose values were to be changed. After considering such aspects as the amount of time and experimental effort which would be required, the information available from industrial data, and the usefulness of the conclusions which might be obtained, it was decided that the priority scheduling rules should be tested for a number of system configurations, which would be obtained by varying six of the system's parameters:

i) the load factor on the system

ii) the mean value of the distribution of setup times

iii) the number of moulds

iv) the mean value of the distribution of total quantity demanded per order (order size)

v) the value of 'MAXLOT' for splitting jobs

vi) the ratio between the number of product styles and the number of machines (stations)

Because the second area of investigation was concerned with measuring the effects on the system of variations in the same set of variables it was at first thought that it might be worthwhile to combine the first and second series of experiments in a single experimental design. In order to analyse such a possibility it was necessary to decide how many different values (levels) each variable (factor) should have, and the actual values for each variable (factor). Here again judgment must be used. In relation to the priority scheduling rules the number of levels is limited to a maximum of eight (the eight priority rules described in 3.5.1). Also there is no problem in setting the particular level of the appropriate variable (factor), as priority scheduling is a qualitative variable. On the other hand, for the cases of the other six variables, there is no limitation on the number of possible levels as they are quantitative variables. However if one is mainly interested in analysing the effect of changes in the variables, it is possible to limit the number of levels to two. For example, it would be possible to measure the effect of load factor by comparing the results obtained from the model when the system was subjected to a low load factor, as opposed to the results obtained when the system was subjected to a high load factor.

In view of the objectives of this part of the study it was decided that the use of two levels for each of the six variables (factors) would be sufficient to generate the necessary information and give generality to the results.

If the two areas of investigation were to be combined in a single experimental design, a possible solution would be to have a factorial design, in which priority scheduling rules and the six other variables would be the factors, and their values, the levels of the factors (Davies, 1967 (1)). A full factorial experiment for this case would require a total of 512 experiments ($2^6 * 8$), which would be too large a number as far as time and computer resources availabilityare concerned. For this reason consideration was given to another experimental design which could economize on the number of experiments, such that the study would be kept within manageable size, and still generate sufficient information.

In order to plan a more economical experimental design, two questions had to be answered:

i) is it necessary to test all the eight priority scheduling rules?

ii) is it necessary to compare the priority scheduling rules for all the possible system configurations which will be generated from a full factorial design?

To answer these questions, the objectives of the experiments with the scheduling rules should be considered. In the main, these objectives are to compare the efficiency of the priority scheduling rules, specially developed for the characteristics of this class of production system, against rules which have shown to perform well in traditional 'job shop' or batch manufacturing systems. It is also the intention that these comparisons be made over a range of situations created by variations of the value of some of the system variables. This can be obtained by taking a 'sample' of the total system configurations which would be generated, by a complete factorial design. It was also thought that not all the eight priority scheduling rules needed to be tested. The FIFO and FIFOM rules could be left out of the experiments as FIFOB and FIFOMB are no more than modifications of them, and as far as the evidence of paragraph 4.2.2 shows, performed better, and are based on more logical principles (as far as this system is concerned).

After all the above considerations it is now possible to devise the experimental design which will be used in the study of the priority scheduling rules. There are six priority scheduling rules to be tested (FIFOB, FIFOMB, SPT, SPTM, SLACK and SLACKM), under a range of system configurations which will be generated by varying six of the variables which might have an influence on the system. The variations are based on two levels for each variable. Below is a summary list of the changes (levels) that will be made in each of the six variables:

i) <u>Load factor: a low average load factor (65%) against a higher average load factor (85%)</u>. It should be pointed out that these values of load factor are nominal average values. The actual values will change depending on the system configuration being simulated. As described in paragraph 3.6.1, the actual average load factor is a direct function of the 'process cycle time'. The 'process cycle time' on the other hand, depends on the amount of time spent in setting up, which in turn depends on the mean value of the distribution of setup times and the number of changes of moulds. Considering that the mean of the distribution of setup times is one of the variables which is changed, and that the number of changes of moulds depends on the relation between the number of moulds and the number of stations (which also changes), it is expected that the actual load factor will vary considerably around the nominal values of 65% and 85%. Because of this it was decided that 85% should be the upper average value, in order to avoid the possibility of the model 'blowing up' with a load factor too close or above 100%, which could happen for some of the tighter system configurations. The choice of 65% as the lower average level was made because of two main reasons. Firstly, to allow a large enough difference between the two values, such that any effect could be easily detected, and secondly, in order to keep its value close to 70% which was the actual value at which the company referred to previously was working. Finally it should be said that the variation in the nominal average load factor is obtained by changing the mean of the distribution of interarrival times. Eight different mean values have to be used in order to take account of the number of machines, which (as will be seen later) vary between one and two, and the average size of orders which will also vary.

ii) <u>Setup time: A lower value for the mean setup time (8 min.), against a higher value (16 min.)</u>. The choice of 8 minutes and 16 minutes was made in order to keep within the bounds of this class of production system. The difference between the values should be large enough in order to allow any possible effect to be detected, but not too large, such that the variations in load factor would become too wide (see relationship between setup times and load factor above).

iii) Number of moulds: a medium number of moulds (27 moulds) against a large number of moulds (42 moulds). The decision on the number of moulds was based on the results obtained from the preliminary investigation (paragraph 4.2.4) which showed that the addition of extra moulds after 27, had little influence on system performance for that particular system configuration. The idea was to check whether this lack of effect would hold true for a wider range of situations. Considering that at 45, some of the moulds were not used at all, a decision was made to have 27 and 42 moulds.

iv) Mean value of the total quantity demanded per order: a lower (1000 items), against a higher value (1600 items). The choice of these values was based on the analysis of actual data from the industrial company which showed that the average order size varied between 1100 and 1900 for different product styles. (See appendix 3). The choice of values far away from those two, were incompatible with the range of lead times which is expected from this class of production system.

v) Splitting of 'jobs': The two situations tested were:
first - 'jobs' were never split (MAXLOT = ∞)
second - 'jobs' were split (MAXLOT = 450)
The idea was to check whether 'splitting of jobs' had any

effect at all on the system's performance. The choice of 450 for MAXLOT was based on the results obtained from the preliminary experiments reported in paragraphs 4.2.3 and 4.2.4.

vi) The ratio between number of styles and number of machines: a lower ratio (3:2), against higher ratio (3:1). The choice of these two ratios was based on the actual situation found on the industrial example and reported in paragraph 4.2. It should be noted that the variation in the ratio is obtained by maintaining the number of styles constant and varying the number of machines. For this reason, and for simplicity of expression, from now on, reference will be made to both the ratio and the number of machines (two or one).

In order to make further references and manipulations easier, the variables and their values will be tabulated and associated with letters and abbreviated names, as shown on the table below.

		Values of the variables' parameters	
		0	1
Variables to be changed	Symbol	Standard values	Alternative values
A - nominal load factor	a	65%	85%
B - Setup time	b	8 min.	16 min.
C - Number of moulds	c	42	27
D - Size of orders	d	1000	1600
E - Splitting of job	e	450	∞
F - Number of machines	f	2	1

It can be seen on the table above that the values of the variables have been divided into two groups: the first group is called standard values and is denoted by a zero, while the second group is called alternative value and is denoted by 1. The intention of doing that is to facilitate reference such that a system configuration in which a = 85, b = 16, c = 27, d = 1600, e = ∞ and f = 1, can be denoted as abcdef, while the system configuration a = 65, b = 8, c = 42, d = 1000, e = 450 and f = 2 will be denoted (I). This means that when a variable is at its standard value, it is associated with zero and the letter which represents the variable does not appear on the denomination of the system configuration. However when the value of a variable is set at its alternative value, the variable is associated with one and its symbol (letter) appears on the denomination of the system configuration. In this way if one starts from the standard configuration (I), and then change the value of a from 65 to 85, the new system configuration will be denominated a. (For further reference to this method of referring to variable changes, see Davies, 1967 (2)). It should be noted that the parameters of the variables were divided in such way that (I) and abcdef would represent respectively the most 'loose' system configuration (low load factor; low setup time; large number of moulds; small size for the orders; more favourable ratio style/machine),and the most 'tight' system configuration (high load factor; high setup time; smaller number of moulds; large size for the orders; less favourable ratio style/machine).

Using the notations introduced above it is now possible to propose an experimental plan for testing the priority scheduling rules. The idea is to test them for a set of system configurations which will include the two extreme cases, (I) and abcdef, and some 'intermediate' cases obtained by the joint variations of 'some' of the other variables. In other words, the rules will be tested over a set of system configurations, which represent a sample of the total number of system configurations, which would be obtained if a full factorial design were used.

In order to limit the amount of experimental work, the number of system configurations will in principle be limited to six. This will allow the testing of the two extreme cases and four intermediate cases (samples), obtained by varying three parameters at a time. This design will not allow the systematic study of the influence of the variables on the scheduling rules, but will allow the comparison of their performance for a range of situations, which is representative of the situations which would be otained by a complete factorial design. However in the second part of the investigation, the influences of those variables on the system behaviour will be fully analysed.

The six initial system configurations chosen as a 'sample' of the universe of sixty four (2^6) possible system configurations are:

i) (I)

ii) abc

iii) def

iv) bdf

v) ace

vi) abcdef

The experimental design will consist of testing each one of the six priority scheduling rules (FIFOB, FIFOMB, SPT, SPTM, SLACK, SLACKM) for each of the six system configurations above, giving a total of thirty six experiments; At the end of these thirty six experiments results can be analysed, and a decision can be made on whether or not more experiments should be carried out.

4.3.1.1 - Discussions on the choice of parameters for the variables

Before describing the experimental designs of the other two phases of experiments, there is a need to discuss all the other variables of the model whose values have still to be determined and justified. In accordance with the descriptions of chapter three, there are a total of eighteen different variables, some having more than one value. In order to determine each of these values, the variables will be listed in accordance with the subsystem they belong to in the model, as described in chapter 3.

a) Variables belonging to the demand or order input subsystem

i) the number of product styles was fixed at three, as it was in the industrial example. It should be considered that three product styles already represent a total of thirty nine product sizes and at least eighteen moulds. To increase the number of styles would result in a large increase in the program core size. However by varying the number of machines it is possible to assess the effect which might result from a different number of product styles, because then the ratio between product style and number of machines (stations) is modified.

ii) the number of product sizes in each style was fixed at thirteen as this is a typical number of sizes for ladies and men's shoes.

iii) the distributions of quantities demanded for the different product sizes in each style and which are represented by a distribution of proportions, were fixed in accordance with the values of the histograms of figure 2.3, which was discussed in paragraph 2.2.3. For details of the distribution see appendix 3.

iv) the distribution of total quantity demanded per order was one of the variables whose parameters would be varied. However as a probability distribution it has a qualitative parameter (the type of distribution) and quantitative parameters (mean and variance). In order to determine those parameters, the Kolmogoroff-Smirnoff goodness-of-fit test was applied to the industrial data referred in paragraph 4.2 (the details of the goodness-of-fit test are shown in appendix 3). The test showed that the distributions of total quantity demanded per order fits an exponential distribution. It was therefore decided that an exponential distribution should be used for the generation of total quantity per order. The quantitative parameter of the distribution (the mean) is one of the parameters varied in the model, assuming values of 1000 and 1600. The distributions however have their tail cut at 5,000 in order to take account of the way in which due date is fixed. As described in 2.3.1 and 3.6.2, due dates are based on a fixed delivery delay promise which is independent (within certain limits) of the total quantity demanded. It is assumed that this limit is 5,000, and that any order which might be larger than that would in fact be processed as two independent orders, with different due dates, as this seems to be the practice of the company from which data was obtained.

v) the distribution of interarrival times of orders is one of the variables whose parameters are modified. In order to determine its qualitative parameter (type of distribution) a Kolmogoroff-Smirnoff goodness-of-fit test was performed on the industrial data, and it was found that the data also fits an exponential distribution. It was therefore decided to use exponential distributions to generate the interarrival times of orders. Its quantitative parameter (the mean) is fixed for each experiment in order to generate the desired load factor on the system. Details of the goodness-of-fit test are shown in appendix 3.

b) <u>Variables belonging to the machine shop subsystem</u>

i) the number of stations per machine was fixed at twelve, in accordance with the industrial data. It should be noted that this variable might have a considerable influence on the behaviour of the system. For this reason, considerations were given to the possibility of including this variable in the list of variables whose values were going to be changed during the study. However, the modification of the value of this variable would require a series of economical and technological considerations, for which data were not available. For example, if one were to compare one twelve-station machine with two six-stations machines (to maintain the same production capacity), the following points would have to be taken into consideration: what would be the difference in capital and operation costs between one twelve-station machine and two six-station machines? Would two six-station machines require two operators as compared with one operator for a single twelve-station machine? What would be the technological consequences in terms of 'process cycle time' of having a machine with six stations? Would the 'curing' time interfere with the possible reduction in 'process cycle time'? Unfortunately this information was not available, and the amount of guesswork would have to be so great as to make any comparisons very doubtful. It might however be an interesting area of investigation, in order to help the machine manufacturers to decide on the design of these multiple station machines.

ii) the 'process cycle time' was left constant with the same values as the industrial example, which was described in paragraph 4.2. It should however be noted that there is a relationship between the mean value of the total quantity demanded per order, the delivery delay promises, and the 'process cycle time'. Considering that both, the mean value of the total quantity demanded per order, and the delivery delay promises are varied, this relationship can be assessed.

iii) the product sizes that each mould can manufacture were maintained constant, in accordance with the principle that each mould of a certain size can manufacture a component of both its normal size and half-size above the nominal size. For further details see paragraph 2.2.1. The main consequences of changing this variable would be economical, meaning that more or less moulds would be needed in order to satisfy the minimum technological requirements as described in 3.5.3.

iv) the distribution of setup times was assumed to follow a normal distribution, with the tail cut at three standard deviations. The parameters of the distributions (mean and standard deviation) varied during the runs, but the ratio between the standard deviation and mean was kept constant and equal to 0.20.

c) <u>Variables belonging to the inventory system</u>

The inventory system was 'switched off' in all the experiments concerned with this phase and the second phase of experiments. It was 'switched on' only in the last phase of experiment related to the study of capacity manipulation. Its control parameters, viz. list of products to be manufactured for stock, the reorder point for each stock item, and the reorder batch quantity, are discussed later, when the last series of experiments are discussed.

d) <u>Variables belonging to the operation control subsystem</u>

The priority scheduling rules; the values of MAXLOT; the number of machines; and the number of moulds, are all modified in this series of experiments and their parameters have already been discussed. The only point which should be stressed is that the selection of moulds is made by the use of the procedure described in paragraph 3.6.1. The other variable of the operation control subsystem is the number of working hours per week which, in this experiment, was maintained constant and equal to forty five hours per week. Its value is however modified for the third series of experiments, and discussions will be conducted when those experiments are analysed.

4.3.2 - Experimental design for the study of main effects and interactions of some variables on the behaviour of the system

The main objective of this part of the investigation is to obtain information about the sensitivity of the system to variations in some of its variables. In the last paragraph the six variables to be analysed have been defined, and a notation introduced in order to represent the different system configurations which are generated by making changes in the values of the variables.

The method of experimentation consists of making changes in the values of the variables and then analysing the effects of these changes upon the behaviour of the system. In order to conduct the study in a systematic way, it is necessary to decide upon the proper method of analysis, viz. the experimental design.

It would be a relatively simple matter to make a series of independent alterations in the variables, one at a time, and to note the effect of each of those alterations in turn. Such a procedure however has two drawbacks.

i) Interaction effects would be ignored. It is possible that a change may have an effect upon the system only if some other change is also effected. For example, a high number of moulds may have an effect upon the delivery performance of the system, only if the ratio (number of style/number of machines) is low.

Such interactions might be important, and should not be ignored.

ii) There would be little generality in the results. If single changes were made, the effects of these changes could be said to apply only to situations quite similar to the system configuration in operation when that change was made. Ideally each individual change should be made over a wide variety of system configurations.

As stated at the beginning of this chapter, the objective of this part of the investigation is to have an experimental design which will generate information that will enable the estimation of the main effects and interactions of the various variables. The factorial experimental design is well suited for such an analysis. If a complete factorial design were to be used, it would require a total of 64 (2^6) experiments. It is however possible to economize in the number of experiments by using a fractional factorial design. Such a design enables the estimation of the main effects and low-order interactions, by confounding higher order interactions with lower order interactions. This is made possible by the fact that higher order interactions are generally assumed to be zero. As Bonini,(1963 (2))says: 'Interactions higher than first order are generally assumed to be zero. In addition, their meaning would be difficult to decipher at the present state of knowledge'. He also notes that a detailed examination of references failed to reveal any illustrations in where the higher order interactions were not ignored or assumed to be zero.

If a half-replicate factorial design (Davies, 1967 (3)) is used, only thirty two experiments will be needed, and it will still be possible to measure all the main effects and first order interactions, which would be confounded respectively with fourth and third order interaction. In view of this, it was decided that a half replicate factorial design should be used.

There are two possible designs for a half-replicate design, one being the complement of the other. In table 4.5 these two designs are listed. The first design is obtained by equating the main effect F to the fourth order interaction ABCDE and the second design is obtained by equating F to -ABCDE (Davies, 1967 (3)). If one considers that the third and fourth order interactions are zero, then the two designs are equivalent as far as the estimation of the main effects and first order interactions are concerned. In figure 4.6 the main effects and first order interactions are listed, with their respective aliases (confounding pairs), for the case of the first design, which is the design used in the experiments.

A single priority scheduling rule will be used throughout the thirty two experiments in this series. The choice of the rule to be used will depend upon the results from the previous series of experiments. In relation to the other system's variables, they will be maintained fixed, with their parameters set in accordance with the description given in paragraph 4.3.1.1.

4.3.3 - Experimental design for the study of strategies for capacity manipulation

In paragraph 4.2.4 a series of experiments were discussed in which the relationship between the number of moulds and the delivery performance of the system was analysed. Although the results obtained helped to increase the understanding about the system's behaviour, it fell short of producing more practical information. If one is going to increase the number of moulds, then the operational as well as the economic consequences of the decision should be taken into consideration. The operational consequences can be measured by the effect on the delivery performance, and the economical consequences by the costs incurred when providing the moulds. To increase the number of moulds is however only one way of increasing the system's capacity to meet demand. Capacity could also be increased by working overtime, having extra shifts, maintaining stock, etc.

This series of experiments was therefore designed with the objective of comparing the different strategies for capacity manipulation.

The experimental design for this phase of the investigation consists of making single changes in the system's 'capacity parameters' (number of moulds; number of machines; number of working hours per week; and the other three control parameters related to the inventory control subsystem), and to measure the consequences of those changes on both the delivery performance and the costs incurred by the system. All the remaining variables of the system are maintained constant, in accordance with the parameters set in paragraph 4.3.1.1.

Further discussions about the experimental design and the system's parameters for this phase of the investigation will be conducted in chapter eight, when the results of the experiments are presented and discussed.

4.4 - Summary

This chapter has presented the results of some preliminary experiments and discussed the experimental designs used during the main core of the investigation.

The main core of the investigation was divided into three areas, each one having its own set of experiments for which experimental designs were organized, in accordance with the objectives of the study.

Also discussed in this chapter were the parameters for the other variables of the system whose values were maintained constant throughout the experimentation.

TABLE 4.1

RESULTS OBTAINED BY THE USE OF DIFFERENT PRIORITY SCHEDULING RULES

RULE	PERCENTAGE OF LATE ORDERS	TARDINESS INDEX OF ORDERS	PERCENTAGE OF LATE PRODUCTION	TARDINESS INDEX OF PRODUCTION	IDLE TIME DUE TO SET UP (%)
FIFO	21.02	0.73	35.31	1.17	3.15
FIFOB	19.28	0.71	33.77	1.14	3.11
FIFOM	20.83	0.71	34.38	1.15	2.80
FIFOMB	18.88	0.67	31.11	1.05	2.81
SPT	18.76	1.06	37.38	2.28	3.37
SPTM	15.00	0.79	31.66	1.72	2.82
SLACK	22.04	0.73	33.28	1.09	3.17
SLACKM	22.02	0.80	34.18	1.20	2.88

TABLE 4.2

RESULTS OF DELIVERY PERFORMANCE FOR DIFFERENT SHOP CONFIGURATIONS

EXPERIMENT	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION
ORIGINAL FIFOMB	7.15 days	31.11 %	1.05
SPLITTING OF JOBS	7.01 days	29.94 %	1.05
TWO MACHINES	4.20 days	1.80 %	0.04
NEW DEMAND PATTERN	3.91 days	4.65 %	0.13

TABLE 4.3

RESULTS OF EXPERIMENTS TO TEST THE EFFECT OF MAXLOT VALUE ON SYSTEM PERFORMANCE

VALUE OF MAXLOT	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE NO. OF JOBS IN QUEUE	IDLE TIME DUE TO SETUP %
120	8.50	0.19	32.1	6.11
250	6.30	0.14	13.1	3.71
350	6.30	0.14	9.6	3.17
450	6.70	0.15	8.2	2.93
550	7.40	0.16	7.1	2.72
650	7.90	0.16	6.6	2.66
750	8.10	0.19	6.1	2.63
900	10.40	0.23	5.9	2.61
1000	10.40	0.23	5.9	2.60
∞	13.30	0.27	5.7	2.59

TABLE 4.4

RESULTS OF EXPERIMENTS AIMED AT RELATING NUMBER OF MOULDS TO DELIVERY PERFORMANCE

EXPERIMENT NUMBER	NUMBER OF MOULDS	AVERAGE DELIVERY DELAY OF PRODUCTION	TARDINESS INDEX OF PRODUCTION	PERCENTAGE OF PRODUCTION DELIVERED LATE
1	18	13.34	6.21	64.3
2	19	13.22	6.11	63.5
3	20	10.75	4.21	51.1
4	21	8.65	2.48	42.3
5	22	8.20	2.06	40.4
6	23	7.21	1.19	33.8
7	24	6.84	0.99	30.8
8	25	6.57	0.83	27.6
9	26	6.19	0.71	25.9
10	27	5.93	0.61	20.9
11	28	5.91	0.65	22.5
12	29	5.74	0.52	20.4
13	30	5.73	0.53	20.7
14	31	5.81	0.59	20.6
15	32	5.77	0.56	20.2
16	45	5.56	0.50	18.0
17(*)	45	5.16	0.38	12.7

(*) Experiments 16 and 17 were both made with 45 moulds, the difference being that while in experiment 16 there was no splitting of 'jobs', in experiment 17 all 'jobs' with batch size larger than 450 were split into smaller batches.

TABLE 4.5

TWO POSSIBLE DESIGNS FOR A HALF-REPLICATE FACTORIAL DESIGN

FACTORS						LIST OF EXPERIMENTS FOR	
A	B	C	D	E	F	DESIGN 1	DESIGN 2
-	-	-	-	-	-	(I)	f
+	-	-	-	-	+	af	a
-	+	-	-	-	+	bf	b
+	+	-	-	-	-	ab	abf
-	-	+	-	-	+	cf	c
+	-	+	-	-	-	ac	acf
-	+	+	-	-	-	bc	bcf
+	+	+	-	-	+	abcf	abc
-	-	-	+	-	+	df	d
+	-	-	+	-	-	ad	acf
-	+	-	+	-	-	bd	bdf
+	+	-	+	-	+	abdf	abd
-	-	+	+	-	-	cd	cdf
+	-	+	+	-	+	acdf	acd
-	+	+	+	-	+	bcdf	bcd
+	+	+	+	-	+	abcd	abcdf
-	-	-	-	+	+	ef	e
+	-	-	-	+	-	ae	aef
-	+	-	-	+	-	be	bef
+	+	-	-	+	+	abef	abe
-	-	+	-	+	-	ce	cef
+	-	+	-	+	+	acef	ace
-	+	+	-	+	+	bcef	bce
+	+	+	-	+	-	abce	abcef
-	-	-	+	+	-	de	def
+	-	-	+	+	+	adef	ade
-	+	-	+	+	+	bdef	bde
+	+	-	+	+	-	abde	abdef
-	-	+	+	+	+	cdef	cde
+	-	+	+	+	-	acde	acdef
-	+	+	+	+	-	bcde	bcdef
+	+	+	+	+	+	abcdef	abcde

Note: (-) factor is on its standard value

(+) factor is on its alternative value

TABLE 4.6

LIST OF MAIN EFFECTS AND FIRST ORDER INTERACTIONS AND THEIR CORRESPONDING ALIASES (CONFOUNDINGS)

MAIN EFFECTS	ALIASES
A	BCDEF
B	ACDEF
C	ABDEF
D	ABCEF
E	ABCDF
F	ABCDE

FIRST ORDER INTERACTIONS	ALIASES
AB	CDEF
AC	BDEF
AD	BCEF
AE	BCDF
AF	BCDE
BC	ADEF
BD	ACEF
BE	ACDF
BF	ACDE
CD	ABEF
CE	ABDF
CF	ABDE
DE	ABCF
DF	ABCE
EF	ABCD

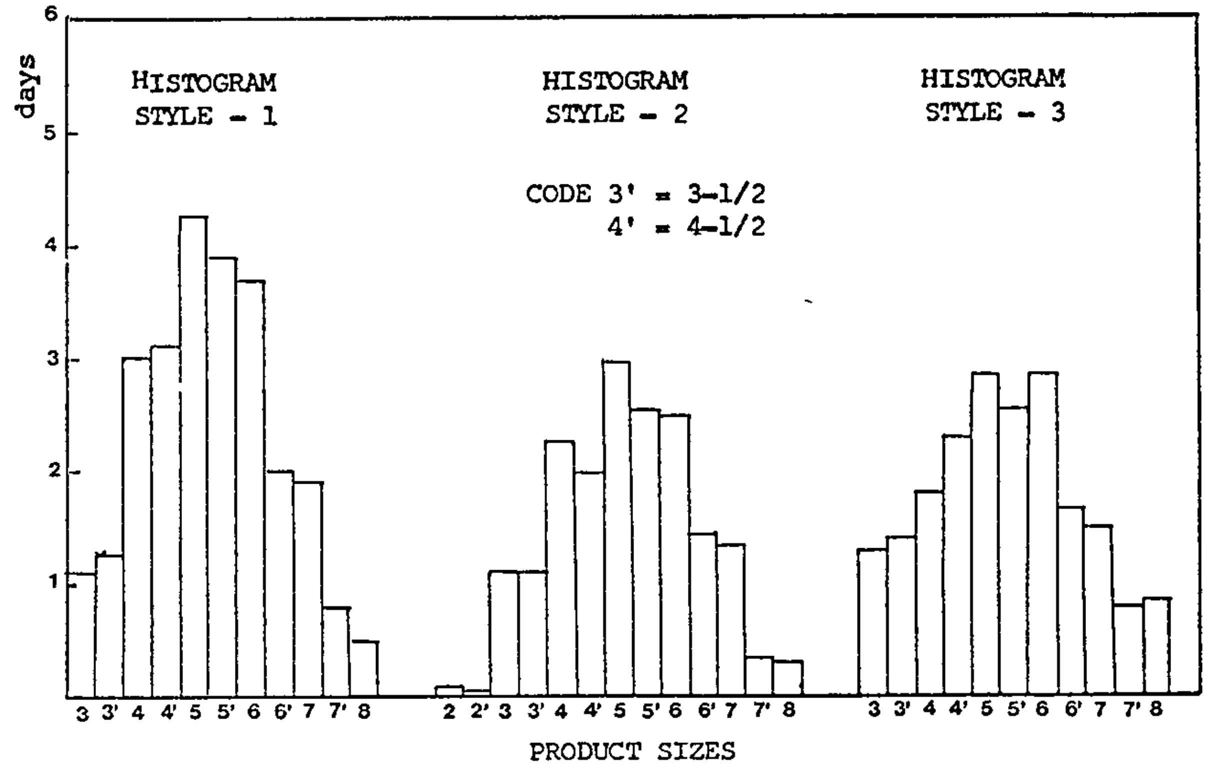


FIGURE 4.1

HISTOGRAMS OF MEAN PROCESSING TIME OF 'JOBS' (PRIORITY FIFO)

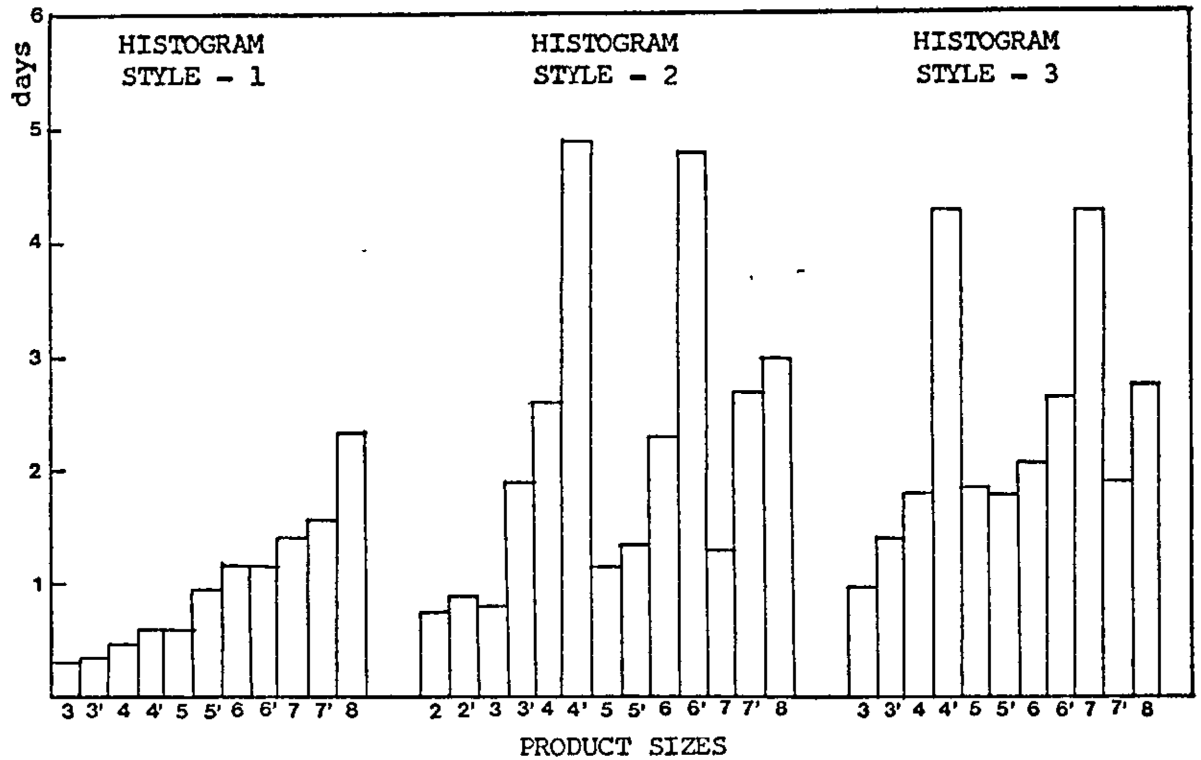


FIGURE 4.2

HISTOGRAMS OF MEAN WAITING TIME IN QUEUE (PRIORITY FIFO)

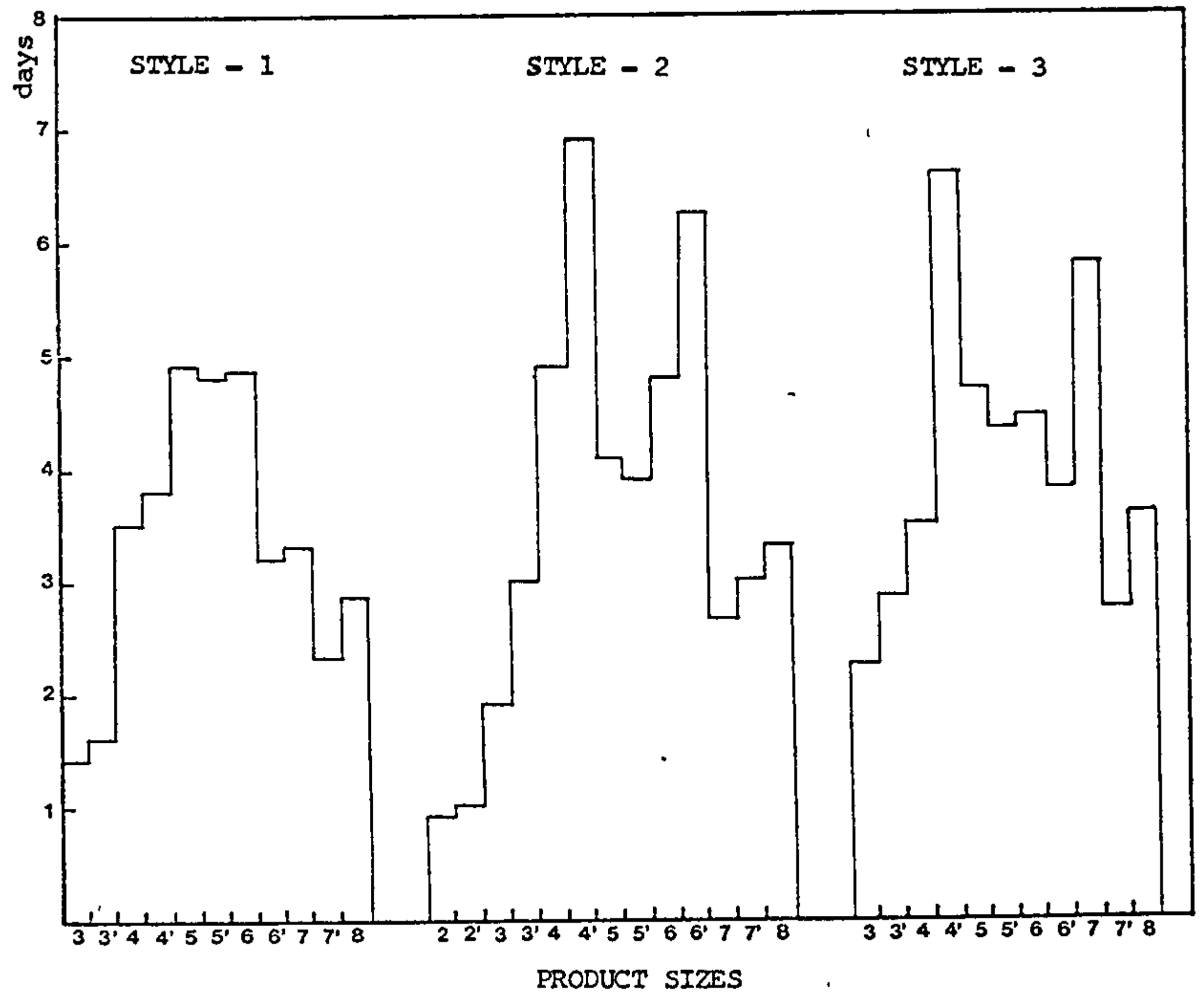


FIGURE 4.3

HISTOGRAMS OF AVERAGE THROUGHPUT TIME - FIFO

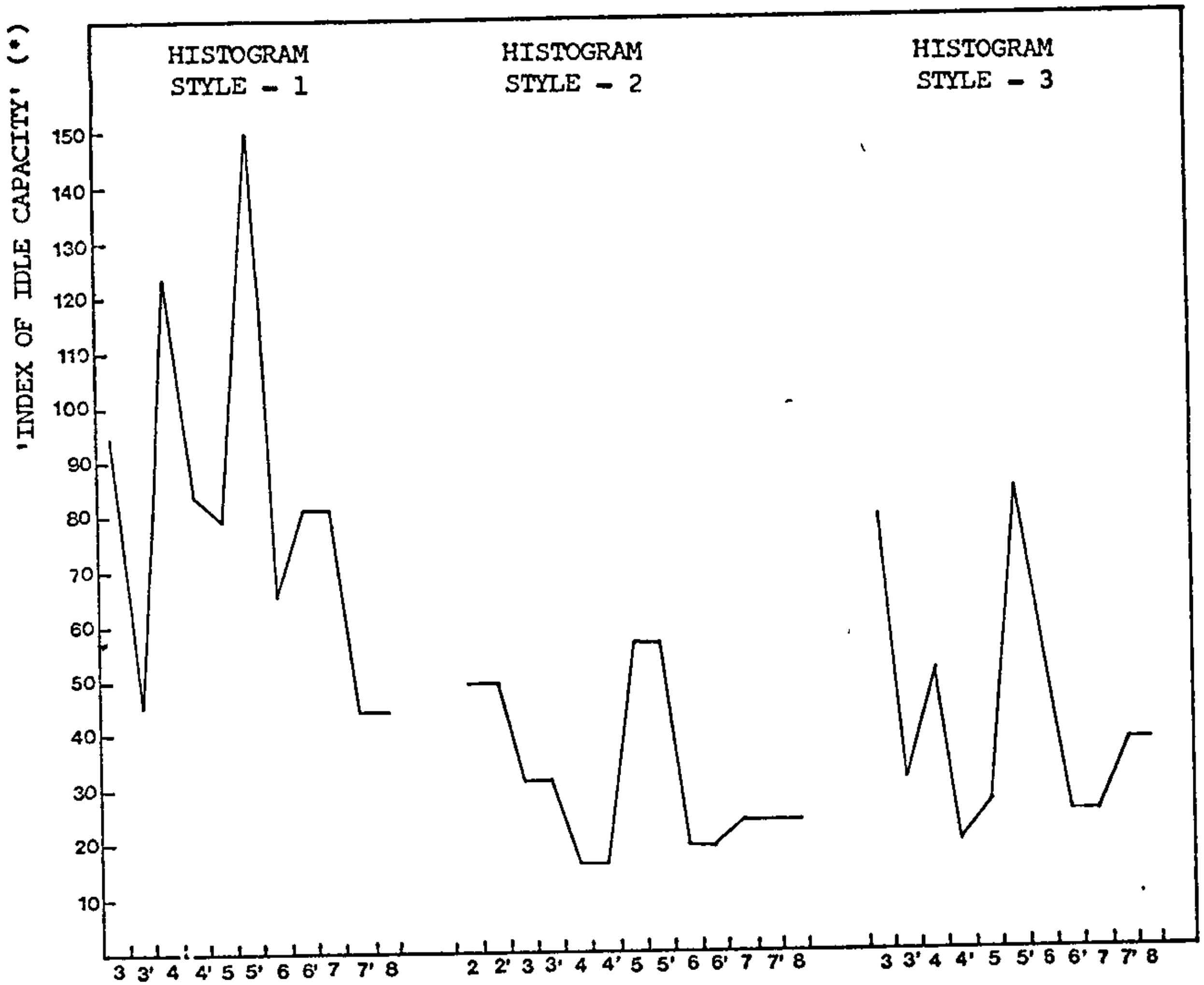


FIGURE 4.4

SHARE OF MOULD'S IDLE CAPACITY ALLOCATED TO INDIVIDUAL PRODUCT SIZES

(PRIORITY FIFO)

(*) For definition of 'INDEX OF IDLE CAPACITY' see paragraph 4.2.1

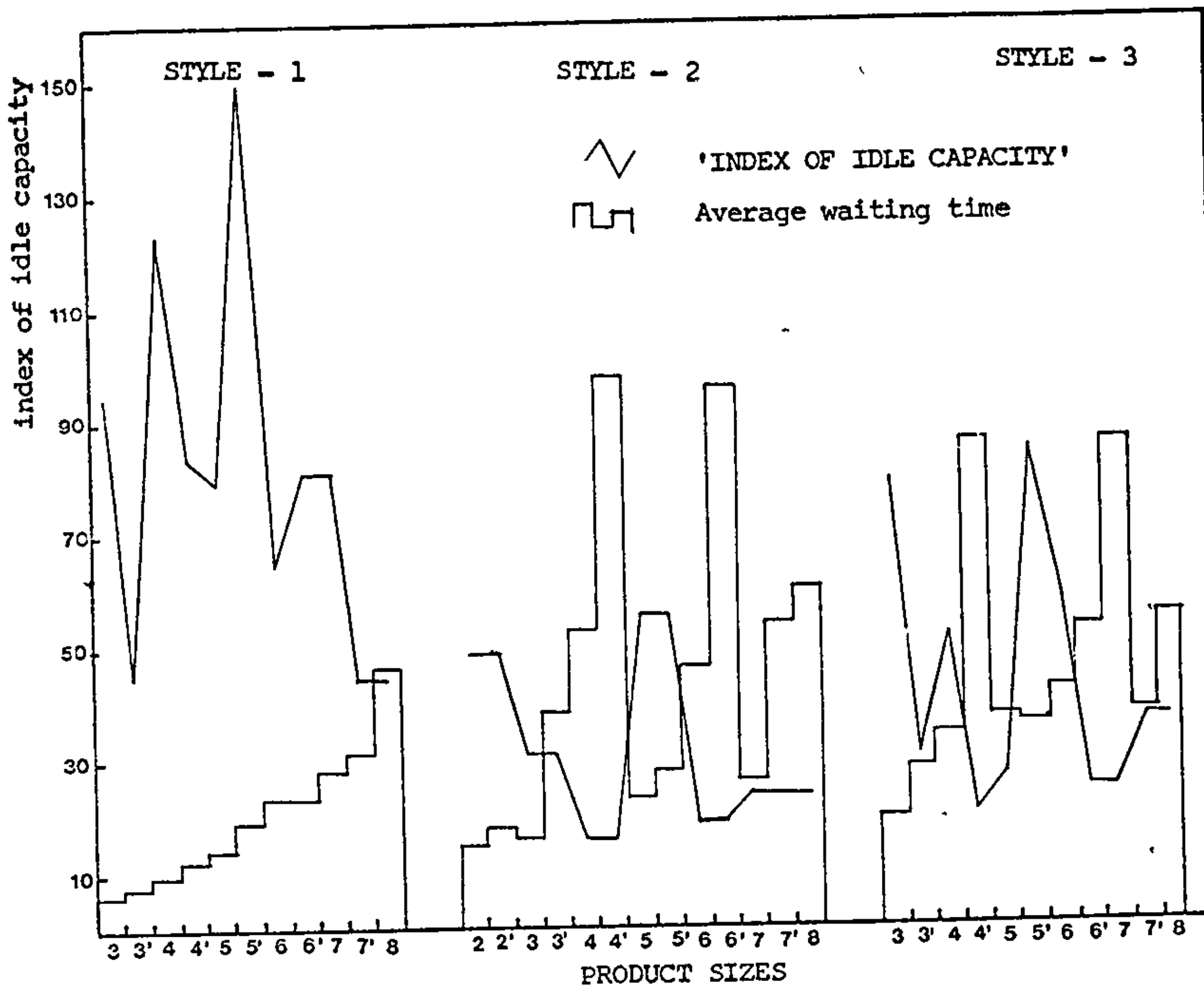


FIGURE 4.5

FIGURE 4.3 SUPERIMPOSED OVER FIGURE 4.4

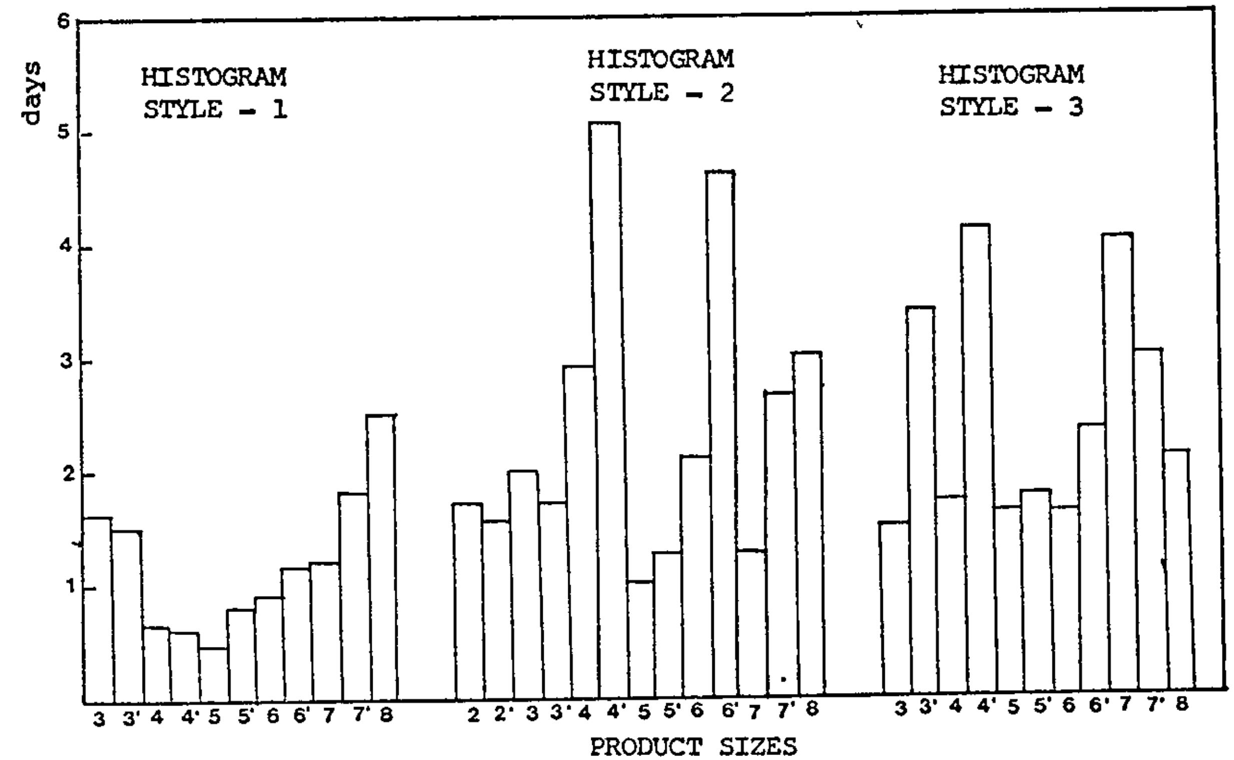


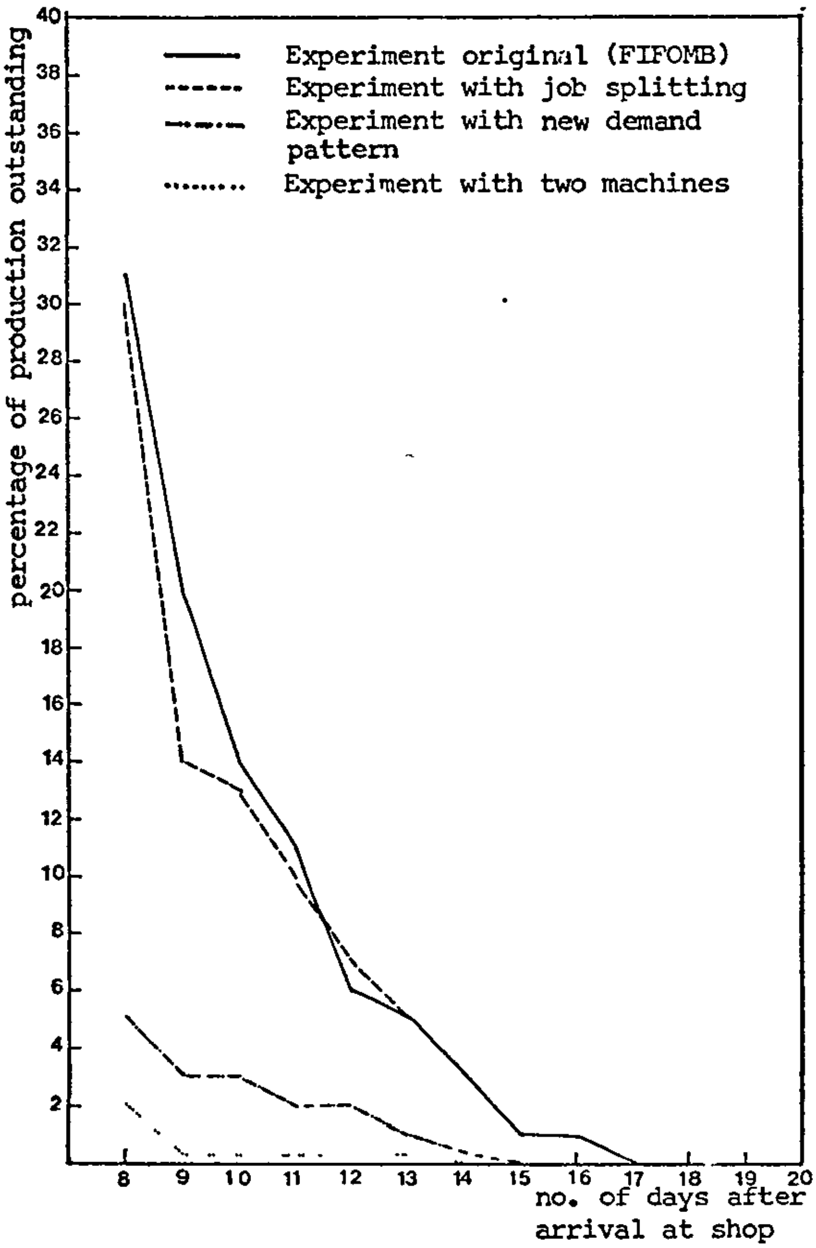
FIGURE 4.6

HISTOGRAMS OF AVERAGE WAITING TIME IN QUEUE FOR DIFFERENT PRODUCT SIZES

(PRIORITY FIFOB)

FIGURE 4.7

DELIVERY PERFORMANCE WITH DIFFERENT OPERATIONAL CONDITIONS



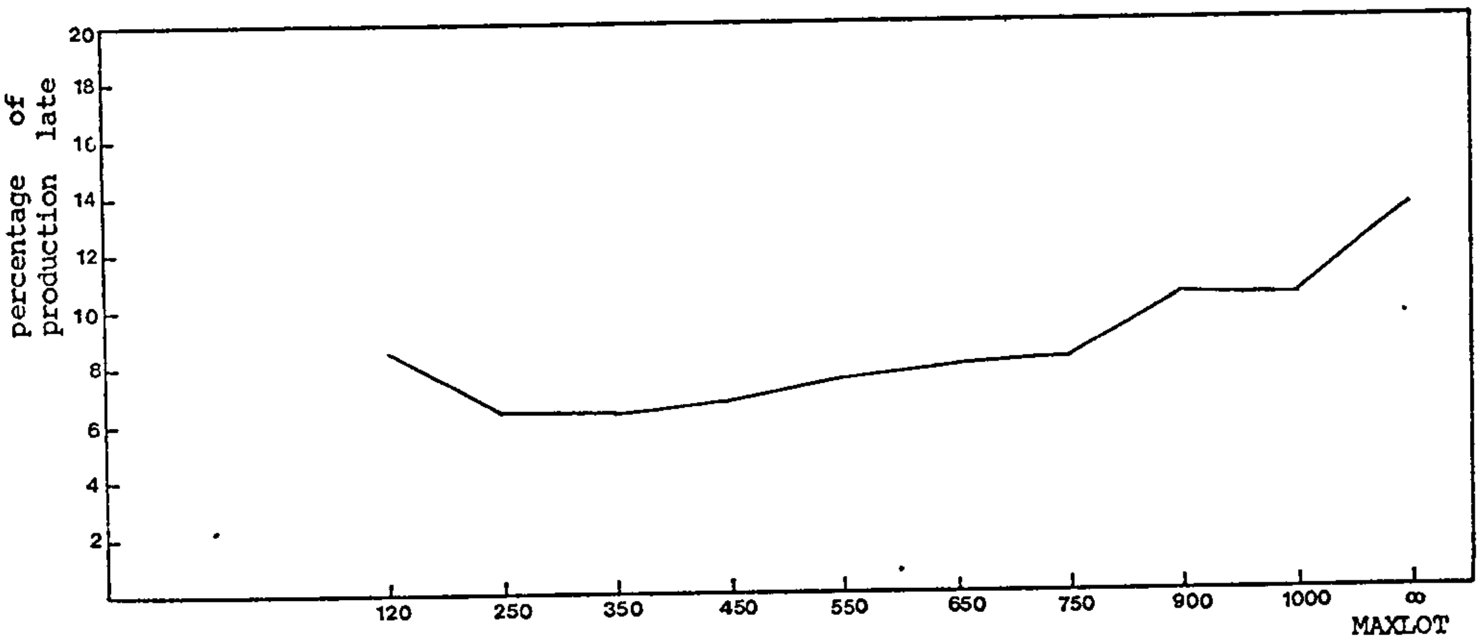
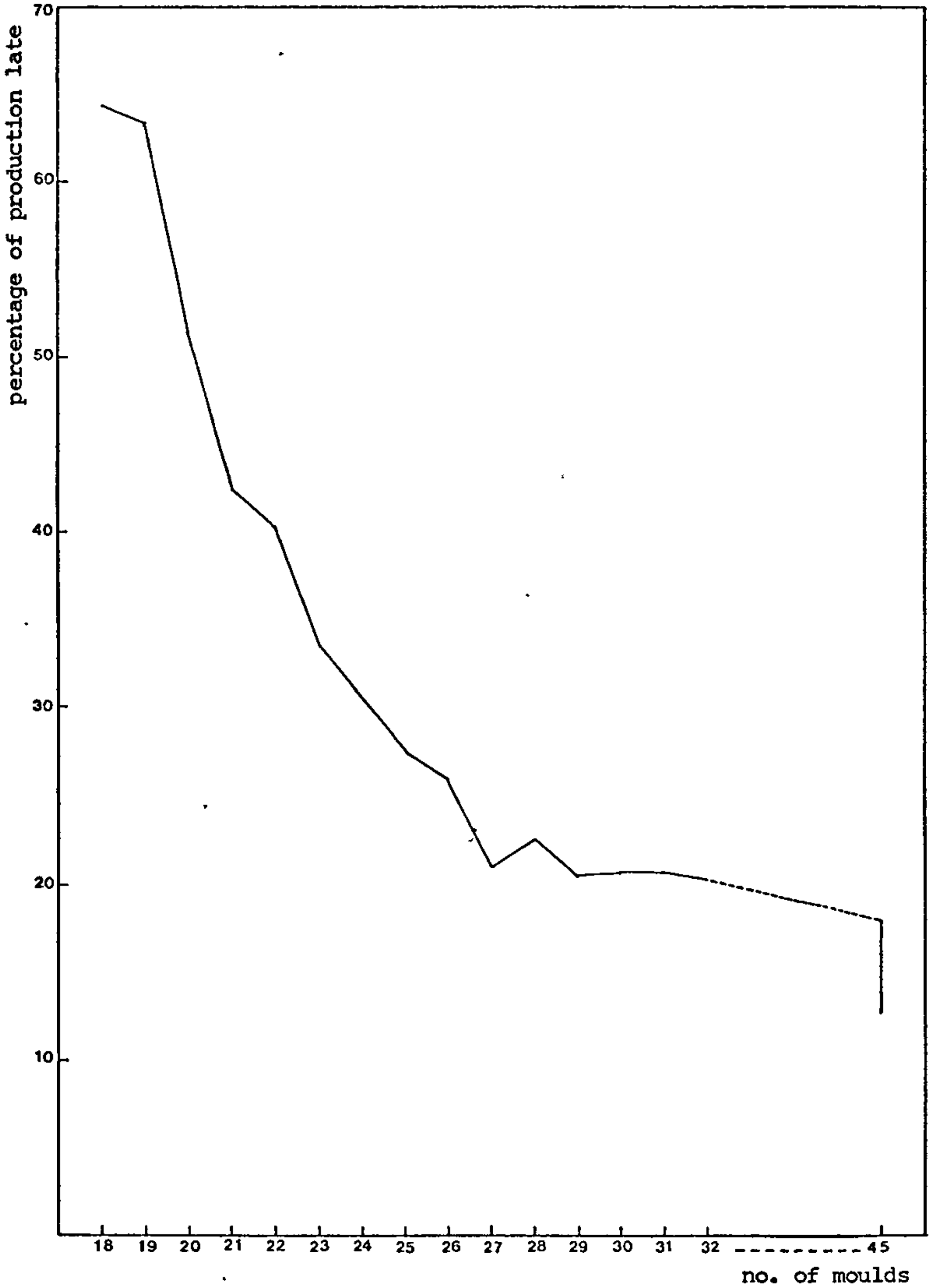


FIGURE 4.8

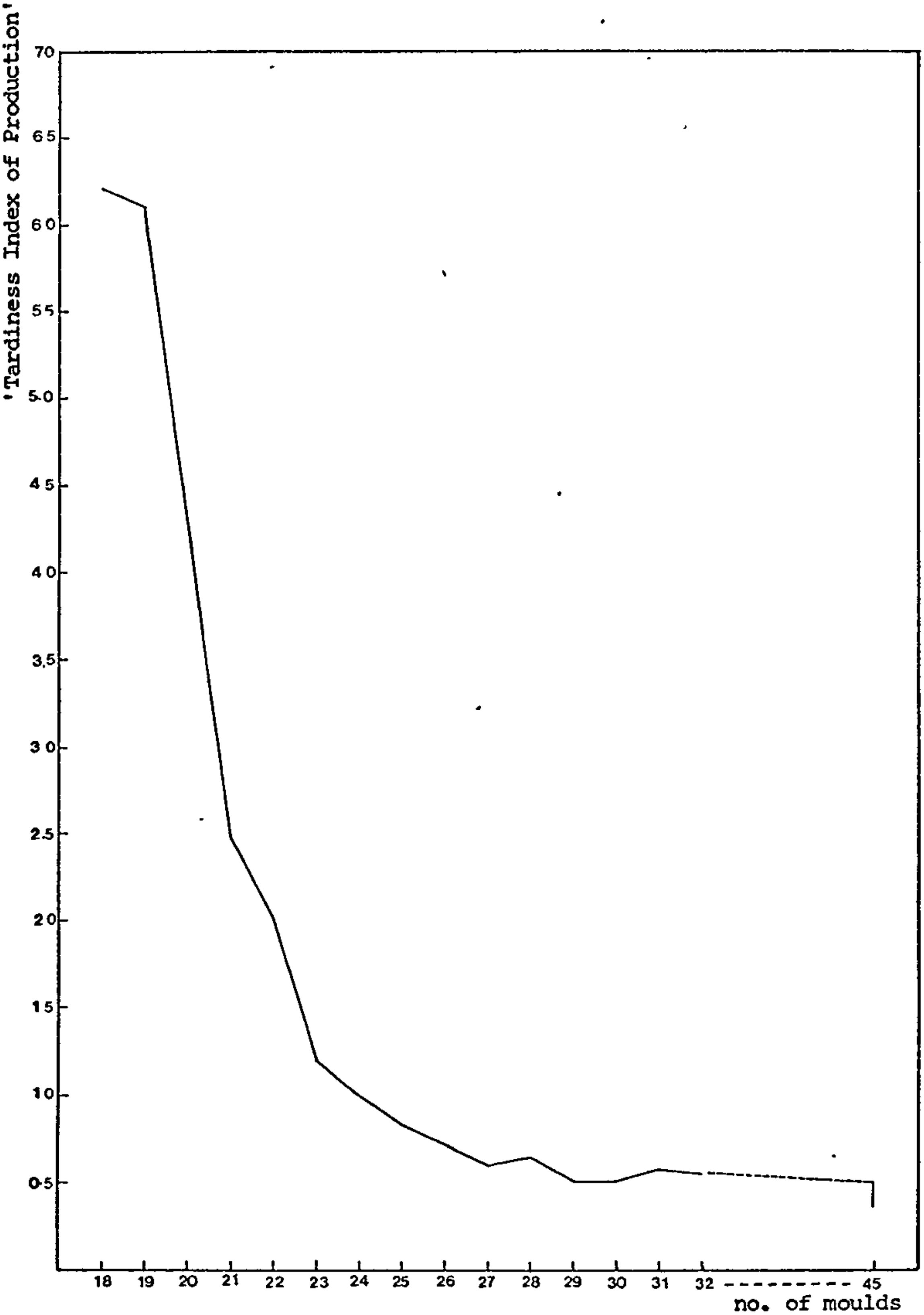
EFFECT OF MAXLOT VALUE ON THE DELIVERY PERFORMANCE

FIGURE 4.9



RELATION BETWEEN NUMBER OF MOULDS AND 'PERCENTAGE OF PRODUCTION DELIVERED LATE'

FIGURE 4.10



RELATION BETWEEN NUMBER OF MOULDS AND 'TARDINESS INDEX OF PRODUCTION'

CHAPTER 5

DISCUSSIONS OF TACTICAL PROBLEMS AND SELECTION OF SAMPLING PROCEDURE

5.1 - Introduction

The objective of this chapter is to discuss the way in which the experiments with the simulation model were carried out.

Conway (1963) defines three main phases in any investigation by simulation, after the problem to be considered has been identified and formulated. They are:

1. Model implementation - description of the model in a suitable computer language;
2. Strategic planning - design of experiments that will yield the desired information;
3. Tactical planning - determination of how each of the test runs specified in the experimental design is to be executed.

Phases 1 and 2 have been described in detail in chapters 3 and 4 respectively; this chapter will concentrate on the description of phase 3 which is concerned with the efficiency of the method used in obtaining the desired information from the simulation model. It is basically an analysis of different ways of obtaining the desired information within a certain statistical precision. If the efficiency of the experimentation was not taken into consideration, then the use of very large sample sizes could overwhelm all of the difficult tactical questions, but this is not a satisfactory or practical answer. A decision was therefore taken to make a series of pilot runs, in order to decide on a tactical plan.

5.2 - Analysis of tactical problems

The necessity of tactical considerations arises because of two problems which are present in most computer simulation experiments. These are the problems of equilibrium and variability in the model with the consequent necessity of determining sample sizes. As Gordon (1969) says, "the introduction of stochastic variables into simulation models causes the variables used to measure the system performance to become random variables, and so the problem of gauging the significance of the results must be considered. The values measured are no more than samples, and they must be used to estimate the parameters of the distributions from which they are drawn".

Each one of the above two problems generates questions which have to be answered before efficient use can be made of the model.

El-Rayah (1973) suggests five questions that should be answered when planning a computer simulation experiment:

"(1) How long do the systems require to settle down to a steady-state (equilibrium) condition?

(2) What starting condition for the simulation runs should be used or whether an 'empty and idle' starting condition is reasonable.

(3) How long a simulation run should be?

(4) How many runs (observations) to have and which method of sampling and replication should be used to obtain the samples?

(5) What, if any, variance reduction techniques to employ?"

Questions 1, 2 and 3 relate to the problem of equilibrium or 'steady-state' situation of the simulation runs, while questions 3, 4 and 5 relate to the problem of variability of the results.

5.2.1 - Problems of stabilization and starting conditions

The problem of stabilization arises from the intermittent nature of simulation models. Each time a new experiment is carried out, an artificial initial condition is used to start the run. This abrupt beginning must be taken into consideration in order to avoid the introduction of any bias which could influence the results of the output variables which the model is intended to measure. A large enough run should be able to overcome this problem because the bias, which would be introduced at the beginning, would have smaller and smaller influence as the size of run increases, until this influence becomes negligible. This of course is not a satisfactory solution because in increasing the run length one is neglecting the efficiency of the experimentation, and moreover, depending on the system being analysed and the starting condition used, it may be that an extremely long run would be needed in order to reduce to acceptable levels the influence of the bias on the results being reported.

If some reasonable starting condition is chosen and some information discarded in the beginning (stabilization period), it should be possible to reduce the size of the run to an acceptable length. Unfortunately it is not very simple either to determine what a good starting condition would be or to decide when measurement should begin. In fact as Conway (1963) says, there is no general objective criteria for determining when measurement should start, the only thing clear is that the problem should be recognized and dealt with. The decision about starting condition and non sampling period becomes even more difficult when the model is being used to compare different system configurations, because then, a starting condition which seems reasonable for one system configuration could well not be so for another.

Conway (1963) suggests 3 alternative solutions to the problem of initial conditions:

"(1) Test each system starting 'empty and idle';

(2) Test each system using a common set of starting conditions that is essentially a compromise between different sets of reasonable starting conditions;

(3) Test each system with its own 'reasonable' starting condition".

Each of the solutions has its own pitfalls. Solution 3 seems to be the most difficult to apply because it would require a previous knowledge of the behaviour of each of the system configurations being tested, and also because there is a danger that the use of different starting conditions could bias the results in favour of a particular system configuration. Solution 2 should be more efficient than solution 1 in terms of reducing the length of the stabilization period, but again it would be very difficult to find a 'compromise reasonable condition' when a large number of system configurations are being compared. In these cases, a modified 'empty and idle' solution could be used. For example, the time scale for the first few exogeneous events can be artificially compressed to accelerate the development of a reasonable 'backlog of work'.

In order to find an answer to these problems a series of pilot runs has to be made, in which the variables being measured are output at short intervals, and the results plotted on a chart to indicate the behaviour of the variables over time. In planning the pilot runs a series of decisions has to be made, concerning initial conditions and system configurations to be tested, together with the variables which it is intended to analyse.

Before these points are considered, it is important to note that the discussions which follow assume that no stock of finished goods is held by the system. This is due to the fact that the inventory subsystem was only incorporated to the model in the later stages of the study, and for the great majority of the experiments it has been 'switched off'. The major consequence of this assumption is that conclusions in respect of initial conditions and stabilization periods might not be valid for the experiments in which the inventory system is 'switched on'. Discussion of this problem will be left for a later stage when experiments with the inventory system are reported.

5.2.1.1 - Choosing initial conditions

In view of the large number of system configurations being studied it was decided to use a modified 'empty and idle' solution for the initial condition, in which the exogeneous event, arrival of orders in the system, was artificially compressed to create a fast backlog of work in the queue. The generation of order arrivals to the system is made through a series of sampling distributions, each using its own stream of random numbers. Basically, each one of the different product styles has two main distributions attached to it, one representing the interarrival time and the other the total size of the order. The method chosen for initialization considers the system in the 'empty and idle' situation, but instead of generating an arrival time for the first orders in each style from the appropriate distribution, each initial order is artificially set to arrive at time zero in the simulation. This still leaves the other random variable, size of order, to be determined. By testing different values for the sizes of the first orders arriving, it is possible to analyse its effect over time on the variables being output.

The values chosen are multiples of the average size of the orders arriving at the system.

Three values, equal to 1, 2 and 4 times the average size of the orders were tested.

5.2.1.2. - Choosing the system configurations to be tested

In view of the large number of system configurations in the study, it would be impractical to test the effect of the initial conditions on all the configurations. Fortunately this is not necessary, as it is possible to choose a few configurations which represent the range of situations covered by the study. Two different system configurations were chosen:

1) $a = 65$, $b = 8$, $c = 42$, $d = 1000$, $e = 450$, $f = 2$, FIFOMB

2) $a = 85$, $b = 16$, $c = 27$, $d = 1600$, $e = \infty$, $f = 1$, FIFOMB

Where: a = load factor on the system (percentage)

b = mean setup time (minutes)

c = total number of moulds available

d = average size of an order

e = maximum size of a batch, before splitting (MAXLOT)

f = number of machines in the system

Analysis of these two configurations will give an indication of the influence of initial conditions on the output variables, and the general behaviour of the system over time.

5.2.1.3 - Choosing the variables to be output

The simulation model was designed to output a large number of variables, some related to the internal behaviour of the system, such as 'process cycle time' and 'average number of jobs waiting in the queue', and others related to the delivery performance of the system, such as 'average delivery delay of orders', 'percentage of late orders' and 'tardiness index of orders'.

Considering the number of variables, it would not be practical to analyse all of them for each one of the situations tested. Fortunately this is not necessary. By carefully selecting some key variables it is possible to check whether the system has reached a point of equilibrium. One such variable in a queueing system is the 'average number of jobs in queue'. Other key variables are those used to measure the delivery performance of the system, because they represent the output of the system. It is expected that if the system is in equilibrium, then all its output variables should also be in equilibrium. However the only way to check that is by comparing their relative behaviour over time. Three of the major output variables in the system are 'average delivery delay of orders','percentage of late orders' and 'tardiness index of orders'.

Figures 5.1, 5.2, 5.3 and 5.4 show the behaviour of the four key variables described above for the case of FIFOMB, a = 85, b = 16, c = 27,

$d = 1600$, $e = \infty$, and $f = 1$, with the size of initial orders equal to the average size of an order (1600 pairs).

The value of each variable is output at intervals corresponding to the completion of 10 full orders for both cumulative and independent statistics.

Figures 5.1 to 5.4 show clearly that the four variables behave in a very similar pattern. Thus it is possible to choose one of them, and to draw conclusions which hopefully will be valid for the other variables as well. The variable 'average delivery delay of orders' was chosen due to its characteristic as a mean value.

5.2.2 - Pilot study - determination of starting conditions and stabilization period

In order to choose the starting conditions and determine the stabilization period, a series of eleven runs were conducted, in which the variable 'average delivery delay of orders' was output at intervals corresponding to the completion of ten full orders, and its behaviour over time plotted for the cases of cumulative and independent statistics.

The first three runs were designed with the objective of analysing the behaviour of the output variable for the three initial conditions already described, for the case of FIFOMB, $a = 85$, $b = 16$, $c = 27$, $d = 1600$, $e = \infty$, $f = 1$.

Figures 5.5 and 5.6 show the results for the cases of cumulative and independent statistics respectively. Looking at the figures it is possible to observe that initial conditions 1 and 2 (average and twice average order size) seem more appropriate than condition 3, because initial outputs are closer to the ones found at the more 'steady' part or the run. It is also possible to observe that the system has a highly random component, which means that any effect of a sudden build-up in the system is quickly dissipated. This can be seen, for example, in the dissipation of the build-up effect caused by the initial conditions. After the completion of twenty full orders the additional effect of starting with condition 2 instead of 1 has been completely dissipated, as shown in fig. 5.5. The same is true for initial condition 3, after thirty full orders have been completed. The conclusion to be drawn from the above is that if initial conditions 1 or 2 are used, and if a period corresponding to the completion of twenty full orders is discarded at the beginning of each run, any initial bias which could have been introduced by the artificial initialization can be considered insignificant for any run of reasonable length.

Runs 4 to 11 were intended as a test of the suitability of initial condition no. 2, for other system configurations. The same procedure of outputting the variable 'average delivery delay of orders' at intervals corresponding to the completion of ten full orders was repeated for the cases of two system configurations;

1) FIFOMB, $a = 65$, $b = 8$, $c = 42$, $d = 1000$, $e = 450$, $f = 2$

2) FIFOMB, $a = 85$, $b = 16$, $c = 27$, $d = 1600$, $e = \infty$, $f = 1$

For each case four different streams of random numbers were used, with the intentinn of analysing the effect of different streams on the output variables. Figs. 5.7 and 5.8 show the cumulative statistics for the output variable for system configurations 2 and] respectively, and for the four different streams. Analysis of figures 5.7 and 5.8 shows that initial condition 2 is quite reasonable for both system configurations, as is illustrated by the relatively small bias introduced at the beginning of the run, which can be made even smaller by eliminating from the statistics the period corresponding to the completion of the first twenty full orders.

Another observation that can be made relates to the strong influence different random streams have on the output variables. In fact it can be noted that even after a relatively large run, when the cumulative statistics from individual streams seem very close to equilibrium, the values of the output variable differ markedly from the different streams.

This observation confirms the results reported by El-Rayah (1973) which found a high statistical significance in the difference of response estimates obtained through different random streams. This fact tends to indicate that in order to obtain precise estimates for the output variables it would be better to have a series of smaller runs with different random streams, than a large run with a single random stream.

In the light of these results, it seems reasonable to suggest that to achieve stabilization and have reasonable starting conditions, condition 2 (modified 'empty and idle' situation with initial orders twice the average size of an order) should be used, and information discarded at the beginning for a period corresponding to the completion of twenty full orders and only then start gathering statistics. The length of the run for the valid period will depend on the number of replications used and the desired level of statistical precision.

5.2.3 - Problems of sample size, replication and variability

The necessity of determining sample size in simulation experiments is due to the introduction into the model of stochastic variables which cause the variables being measured to become random variables, in most cases having a high level of variability. The existence of such uncertainty results in the need for assessing the accuracy of the results being considered. The statistical methods available for estimating population parameters through statistics of random samples are well known, and familiar to anyone concerned with sampling procedures. Considering that the results of simulation experiments are, in essence, samples from stationary populations, it would seem logical, at first, to make use of the familiar statistical models available for traditional sampling experiments.

A traditional sampling technique consists of drawing a variable x_i from a population that has a stationary probability distribution, with a finite mean μ and variance σ^2(Gordon (1969)). If n independent observations of the variable x_i are made, then one would be able to estimate the parameters μ and σ^2 of the real population through the statistics $\bar{X}$ and S^2 which are given by

5.1... $$\bar{X} = \frac{\sum^{n} X_i}{n}$$ and 5.2... $$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)$$

The central limit theorem establishes that the distribution of $\bar{X}$ tends to a normal distribution with mean μ and variance σ^2/n

It happens that the variable, [Guenther (1964) (3)]

$$t_{n-1} = \frac{\sqrt{n}\ (\bar{X}-\mu)}{S} \qquad \ldots 5.3$$

follows a Student t distribution with n - 1 degrees of freedom. Based on this distribution one could then choose a value u such that $f(u) = 1 - \alpha/2$ for $0 \leqslant \alpha/2 \leqslant 1$ and denote u by $u_{\alpha/2}$ The probability that t_{n-1} is greater than $u_{\alpha/2}$ would be $\alpha/2$ Due to the symmetry of the t distribution, it would then be possible to state the probability of t_{n-1} lying between $-u_{\alpha/2}$ and $u_{\alpha/2}$ by

$$\Pr\left[-u_{\alpha/2} \leqslant t_{n-1} \leqslant u_{\alpha/2}\right] = 1 - \alpha \qquad \ldots 5.4$$

From (5.3) and (5.4) it is possible to build a confidence interval for the parameter μ being estimated, which would have attached to it a probability given by

$$\Pr\left\{\bar{X} - \frac{S}{\sqrt{n}} \cdot u_{\alpha/2} \leqslant \mu \leqslant \bar{X} + \frac{S}{\sqrt{n}} \cdot u_{\alpha/2}\right\} = 1 - \alpha$$

where the constant $1 - \alpha$ is the confidence level, and the interval

$$\bar{X} \pm \frac{S}{\sqrt{n}} \cdot u_{\alpha/2}$$

is the confidence interval. It can be seen that the size of the confidence interval and, as a consequence, the precision of the estimate, is an inverse function of the sample size n. So, theoretically, for a set confidence level, it would be possible to reduce the confidence interval to any desired level, by simply increasing the sample size n. However all these procedures are based on the following two assumptions:

a) the observations are drawn from stationary distribution;

b) the observations are statistically independent.

In a simulation experiment the first condition can be obtained by considering the stabilization problem as in paragraph 5.2.1, but the second condition has still to be considered. Unfortunately many simulated systems tend to generate highly autocorrelated variables. This is particularly true for the case of temporary entities in queueing systems, where variables belonging to consecutive elements tend to be highly autocorrelated. This autocorrelation between consecutive elements precludes the use of the classical statistical method of equating the variance of the sample mean to the variance of the sample divided by the sample size. In order to overcome this problem a series of different solutions have been suggested, and they are all related to the method by which samples are obtained.

Different sampling methods have been used in reported simulation experiments, and they differ mainly in relation to what one considers to be a single sample. For example, consider a simple queue problem in which one is interested in measuring the average throughput time of jobs leaving the system. The most obvious way of obtaining the desired statistic would be to measure each single throughput time, x_i, for individual jobs, as they leave the system, and then, by using the collection of all x_i obtained, calculate the desired statistics. If n such jobs were measured the sample size would obviously be equal to n.

By varying the duration of the run from, say, one to two months, one could increase the sample size. Two problems exist with this method: first, the series of values x_i tend to be highly autocorrelated, and so one cannot use the sample variance to estimate the sample mean variance, without the introduction of an extra term which takes account of the autocorrelation. Secondly, if one is interested in measuring some parameters from permanent entities like, for example, the percentage of time the service station stays idle, one would not be able to estimate its precision, because such a statistic would be the ratio of two cumulative variables, namely the total time the station stayed idle divided by the total simulation time.

A second method would be to have a series of n independent runs of, say, one week, instead of a single run of eight weeks. The runs would be made independent by restarting the simulation after the end of each run, and using a different random stream each time. This method would solve the problems discussed above, because then it would be possible to measure the uncertainty associated with all the variables being measured.
Each run would contribute a single value to the calculation of the mean and variance, viz. the mean value of each run, and the use of different random streams would guarantee independence of the samples. The sample size would be the number of runs. The main problem with this method is the fact that each time a new run is started the problems of initial conditions and stabilization period have to be considered, and this could in certain cases result in a large amount of information (stabilization period) being thrown away, which can be seen as sampling inefficiency.

The third method, also called the 'run-subdivision method', consists in having a single long run divided into equal periods, such that the mean value of each period would contribute a single value to the calculation of statistics. With this method, the final condition of a sampling period will be the starting condition of the next one, and as a consequence, subsequent periods would in principle not be independent.
Two solutions could be adopted in order to guarantee independence of consecutive periods. The first would be to make the length of the sub-division periods so long as to render subsequent periods independent from each other. This independence could, and should, always be tested by calculating the autocorrelation coefficient in the time series generated. The second solution would use a decoupling period between consecutive runs, in the same way that the stabilization period is used to render the sampling period independent from the initial conditions.

It is argued, Conway (1963), that the final condition of an individual period should be a more reasonable starting condition for the next period than one set artificially, and because of that, the decoupling period could be smaller than a normal stabilization period, resulting in a more efficient sampling procedure than the one used by the completely independent run method, which always starts from an artificially set initial condition. The main problem with such a method is that the autocorrelation tends to vary depending on the system configuration being analysed. There is therefore the problem of ensuring that autocorrelation is not overlooked in certain cases, which could result in a serious overestimation of the precision of the results.

Apart from the three methods described above, another two have been suggested in the literature concerning computer simulation experiments. They are both more complex than the previous ones, and do not seem well suited for the sort of investigation this project is intended to follow.

The first one, Conway (1964), is based on the idea of eliminating the autocorrelation between observations by performing a linear transformation on the original time series obtained during the experimentation. Traditional analysis is then applied assuming the transformed observations are uncorrelated. However, it has been pointed out, Fishman and Kiviat (1967), that this procedure throws away a considerable amount of information, and that the transformed time series may be inappropriate for comparison purposes.

The second method, Fishman (1967) and Fishman and Kiviat (1967), is based on exploitation of the autocorrelation rather than in its elimination. The approach centres on the utilization of spectral analysis on the generated time series, with the objective of identifying an interval called 'correlation time of the process', which is dependent on the autocorrelation properties of the process. The principle of the 'correlation time' is based on the assumption that if a process is observed for a time interval equal to $\underline{n}$ 'correlation times', then from the point of view of the variance of the sample mean, this time series is equivalent to collecting n/2 independent observations. By using the 'correlation time' together with the total observation interval, one can define the number of 'equivalent independent observations' contained in the autocorrelated time series, with the objective of calculating the sample mean variance.

The main problem with this method is the fact that it needs a preliminary run for each individual system configuration being analysed, with the objective of estimating the population variance σ^2. For studies in which there are a large number of system configurations being investigated, such preliminary runs may become prohibitive. For this reason this method has not been considered in this study. Based on the characteristics of the five methods described above and on the characteristic of this study, it was felt that only two methods, the independent replication method and the run-subdivision method, are suitable for conducting the series of planned experiments in this study. In order to help decide which method to use, a series of pilot runs were made with the objective of determining the most efficient and practical method. Efficiency is understood as the capacity for generating the required information, with the required precision and with the minimum sampling effort. The description of the pilot runs will be given later on in this chapter, but a discussion about variance reduction techniques will be conducted first.

5.2.3.1 - The use of variance reduction techniques

All the discussions in this chapter have been concerned with the efficiency of obtaining the desired information from the simulation model. The main objective is to minimize the sampling effort for a given precision of the generated information. Considering this objective, one cannot fail to analyse the possibility of using some variance reduction techniques in the experimentation with the model.

Variance reduction techniques are statistical methods used with the objective of reducing the variance of the estimated response, through the replacement of the crude 'straight on' sampling procedure by a revised procedure.

Many variance reduction techniques can be found in the literature about sampling procedures, and some of them, like proportional sampling, fixed sequence sampling, importance sampling, use of concommitant information and control variate, have been analysed by Ehrenfeld (1962) to check their efficiency when applied to computer simulation experiments.

Kleijnen (1975) however, points out that the adjustments required by most of these techniques in order to make them applicable to simulation experiments, result in quite complicated procedures and that they have hardly been applied in practice.

Two other techniques however, remain very simple when applied to simulation experiments. They are the use of antithetic variates and common random numbers (or correlated sampling).

The concept of antithetic variates, as developed by Hammersley and Morton (1956), is based on the idea of generating pairs of negatively correlated random variables. The use of antithetic variates in computer simulation experiments has become widely accepted, and the most common procedure is to generate a pair of runs such that the first run is generated in the 'normal' way from the random numbers r_1, r_2... but its companion on the pair is generated 'antithetically' from the complements of these random numbers, i.e., from $(1-r_1)$, $(1-r_2)$, ... By doing so one hopes to create a negative correlation between the responses of the two partner runs. Such correlation decreases the variance of the average output of the two runs, since:

$$\text{var.}\left\{(x_1 + x_2)/2\right\} = \left\{\text{var.}\ (x_1) + \text{var.}\ (x_2) + 2\ \text{cov.}\ (x_1 + x_2)\right\} / 4$$

where x_1 and x_2 are the output of run 1 and 2 respectively [Kleijnen (1975)].

If the two runs were independent, cov(x_1, x_2) would be zero. Whether antithetics indeed create negative correlation in a complicated simulation cannot be proved. However intuition indicates that negative correlation may be expected. Experiments with various simulated systems of moderate complexity show that such correlation is created indeed. [Kleijnen (1975), El-Rayah (1973)] .

Common random numbers can be used when the intention is to simulate and compare two or more systems or system configurations in relation to their response. By using the same sequence of random numbers for all systems being compared, one is comparing them 'under the same circumstances' or, statistically speaking, their responses are supposed to show positive correlation. Such correlation is desired since:

$$\text{var}\ (x - y) = \text{var}\ (x) + \text{var}\ (y) - 2\ \text{cov}\ (x.y)$$

where x and y are the response of systems 1 and 2 respectively. This variance reductinn technique is widely used in practice and Conway (1963) strongly recommended its use by saying:"The use of a common sequence to test all alternatives appears the most important single procedural question in simulation and can very well mean the difference between feasibility and impossibility."

5.3 - Selection of sampling procedure

The previous paragraph has listed a number of sampling methods, together with some variance reduction techniques.

Because of practical considerations only two sampling methods and two variance reduction techniques will be considered. Basically two questions are examined:

a) whether to use completely independent runs, or the run subdivision method of sampling;

b) whether to use any of the variance reduction techniques.

Considering the characteristics of the study, which will compare a series of different operation rules and system configurations, it was decided that the variance reduction technique of common random numbers should be used, independently of the sampling method chosen. The arguments for its use have already been strongly justified in the last paragraph.

In relation to the other variance reduction techniques, the use of antithetic variates seems the most useful, and practical, for utilization in this study. To check its effectiveness a pilot study will have to be made, in which the following four sampling procedures will be compared:

i) to have n samples, from completely independent runs;

ii) to have n samples, using the run subdivision method;

iii) to have n/2 samples, from antithetic pairs of completely independent runs;

iv) to have n/2 samples of antithetic pairs, using the run subdivision method.

Apart from helping to decide upon the choice of the sampling procedure, this pilot study will give information about the relationship between sample size and statistical precision, for this particular problem.

5.3.1 - Preliminary considerations concerning the pilot runs

Before describing the pilot study design and discussing its results, some preliminary considerations have to be made. These considerations relate to sample size, the length of a run, and the problem of statistical independence of the samples. Depending on whether completely independent runs, or the run subdivision method is used, the approach to the problems will differ.

Problems which arise when using completely independent runs are quite simple. The main concern should be about the stabilization period which, if overlooked, could result in bias being introduced to the final results by the initial conditions. This problem has been fully discussed in paragraph 5.2.2 and it was concluded that by using a modified 'empty and idle' starting condition, eliminating the data corresponding to the first twenty full orders (two hundred and sixty jobs), and only then starting to gather valid statistics, it would be reasonable to assume that no significant bias would be transmitted to the output variables.

The decision on the length of each run has to be made in the light of conflicting factors: if the length is too small, the probability of bias (from the initial condition) being transmitted to the final outputs is increased. On the other hand, as the length of runs increases, there is a tendency for a decline in the efficiency of the antithetic variates, because the chance of the simulated elements getting out of sequence increases. There is also a need to consider the relationship between the length and number of runs. For a fixed sample effort (run length * number of runs), an increase in run length results in a decrease in the number of runs.

The main consequence of this is the decline in the number of degrees of freedom available for calculating the precision of the estimate, which increases the chance of a Type II error when using statistical tests.

The main problem with the run subdivision method, apart from the previous ones, is the autocorrelation which tends to exist among subsequent samples.

In order to make sure that autocorrelations are not overlooked, they must be estimated. The only way of doing this is by generating a series of outputs from the model and analysing them as a time series from which the autocorrelation coefficients of lag 1 to K are calculated and plotted as a correlogram.

Kendall (1966) says about serial correlations:

"For series which are not random there will be dependencies of one kind or another between successive terms. One very useful measure of this effect is the product moment correlation between successive observations. Given $\underline{n}$ values u_1, u_2, ... u_n, the so called serial correlations of lag 1 is defined by:

$$r_1 = \frac{\frac{1}{n-1} \sum_{i=1}^{n-1} \left\{ \left(u_i - \frac{1}{n-1} \sum_{i=1}^{n-1} u_i\right) \left(u_{i+1} - \frac{1}{n-1} \sum_{i=1}^{n-1} u_{i+1}\right) \right\}}{\left[\frac{1}{n-1} \sum_{i=1}^{n-1} \left\{ u_i - \frac{1}{n-1} \sum_{i=1}^{n-1} u_i \right\}^2 \frac{1}{n-1} \sum_{i=1}^{n-1} \left\{ u_{i+1} - \frac{1}{n-1} \sum_{i=1}^{n-1} u_{i+1} \right\}^2 \right]^{1/2}} \quad \ldots 5.5$$

Likewise the serial correlation of lag K is the correlation between pairs of K units apart, viz.

$$r_k = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} \left(u_i - \frac{1}{n-k} \sum_{i=1}^{n-k} u_i\right) \left(u_{i+k} - \frac{1}{n-k} \sum_{i=1}^{n-k} u_{i+k}\right)}{\left[\frac{1}{n-k} \sum_{i=1}^{n-k} \left\{ u_i - \frac{1}{n-k} \sum_{i=1}^{n-k} u_i \right\}^2 \frac{1}{n-k} \sum_{i=1}^{n-k} \left\{ u_{i+k} - \frac{1}{n-k} \sum_{i=1}^{n-k} u_{i+k} \right\}^2 \right]^{1/2}} \quad \ldots 5.6$$

In practice (and also for theoretical convenience) it makes for simplicity to modify these definitions to some extent.

Instead of measuring the first (n-k) u's about their mean, we may measure about the mean of the whole set of observations; and similarly for the values at the end. Similarly, instead of taking separate variances in the denominator, we may use the variances of the whole series. Thus, writing $\bar{u}$ for $\sum_{i=1}^{n} u_i/n$ we may put:

$$r_k = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} (u_i - \bar{u})(u_{i+k} - \bar{u})}{\frac{1}{n} \sum_{i=1}^{n} (u_i - \bar{u})^2} \quad \ldots 5.7$$

For series of moderate length the difference between the two formulae is negligible. However for short series, where exactitude is necessary, we should be careful not to use 5.7"

5.3.2 - Analysis of autocorrelations

Because of the necessity of measuring the autocorrelation, a preliminary run was made, with the objective of studying the effect of different period lengths on the autocorrelation coefficients. For a given total run length the number of periods (samples) can be varied, with a consequent variation in the period length. It is logical to expect that smaller periods will generate more correlated values than longer periods. To check that, correlograms corresponding to different period lengths were plotted.

By carefully selecting values for the period lengths, it is possible to obtain all the information from a single run. However it was considered wise to replicate the run using different random streams. Four replications were used and compared. The following lengths and number of periods were chosen for a total run size equivalent to the departure of six hundred full orders from the system:

No. of periods	period length
120	5
60	10
30	20
20	30
15	40

For each of the five period lengths the autocorrelation coefficients were calculated and correlograms plotted. Because the objective at this point was only to have an insight into the behaviour of the process, in relation to the autocorrelations, it was thought that exactitude was not necessary, and so, formula 5.7 instead of 5.6 was used for the calculation of the autocorrelation coefficients.

All correlograms present a wild fluctuation, jumping from negative to positive values and vice versa, very quickly. They also reveal a considerable influence of different random number streams on the model's behaviour. All this indicates a strong random component in the process, confirming the impression obtained in paragraph 5.2.2.

However, it was found that for small period lengths, the autocorrelation of at least lag 1 was always very high, but that it decreased when the period increased, being negligible for period lengths of 30 or more.

In figures 5.9 to 5.13, correlograms corresponding to period lengths of 5, 10, 20, 30 and 40 respectively, are plotted. They correspond to one of the replications, and are a good example of the pattern of the correlograms generated by the process.

In the light of such observations, a decision can now be made in relation to period lengths for the run subdivision method.

It seems reasonable that if a period length large enough is used, the autocorrelations between periods can be discharged as insignificant when assessing the sample mean variance. It is always possible to take them into consideration by using the following expression,

$$\text{var. } (\overline{X}) = \frac{S^2}{n^2} \left[n + 2 \sum_{i=1}^{n-1} (n-k)\, r_k \right] \qquad \ldots 5.8$$

Where S^2 = sample variance

n = sample size

r_k = autocorrelation coefficient of lag k

However it looks as if there is no need to use this expression, at least for the moment, if a reasonably large period length is used.

5.3.3 - Pilot study - Selection of a sampling procedure

Using the results from the previous paragraph it is now possible to design a pilot study with the objective of comparing the efficiency of the four sampling strategies mentioned before.

Sixteen experiments in total were conducted, four for each sampling strategy.

For all cases the same initial conditions and non-sampling periods, as described in 5.2.2, were used. The system configuration used was (FIFOMB, $a = 85$, $b = 16$, $c = 27$, $d = 1600$, $e = \infty$, $f = 1$), the length of each individual period (or run, for the case of independent runs) chosen was equal to 150 departures of orders from the system, and the number of periods (or runs) made equal to 6. The reason for the choice of 150 and 6 respectively for period length and number of samples, is due mainly to computation time and other practical considerations. The choice of 150 for the interval length will also guarantee independence between samples for the case of the run subdivision method, as well as guaranteeing that no significant bias will be transmitted to the final output by the initial conditions, in the case of independent runs.

Experiments 1 to 4 were used to test the efficiency of completely independent runs. The four experiments differ only in respect of the random streams used. Each one consists of six completely independent runs, each run having a non-sampling period equivalent to twenty departures, followed by a sampling period equivalent to one hundred and thirty departures, giving a total of six independent samples. Experiments 5 to 8 were used to assess the efficiency of the run subdivision method. As in the previous series, they differ only in respect of the random streams used. Each experiment consists of a single long run, equivalent to the departure of nine hundred and twenty orders from the system, split in one non-sampling period of twenty departures (stabilization period), plus six sampling periods of one hundred and fifty departures. At the end of each sampling period, the mean value of the outputs for that period is calculated and recorded as a sample x_i, and a new random number stream is selected as this seems to enhance independence between periods and precision in the estimation of the real mean.

Experiments 9 to 12 were designed to test the efficiency of antithetic variates when used in conjunction with completely independent runs.
As in previous cases, the experiments differ from each other in respect to random number streams.
Each experiment consists of six runs, made up of three antithetic pairs of runs. Each pair is obtained by having a 'straight sampling' run followed by an antithetic run.
If the 'straight' sample uses a random stream r_1, r_2, r_3... the antithetic run will use a stream $(1 - r_1)$, $(1 - r_2)$, $(1 - r_3)$, ... Each pair contributes a single value for the calculation of the sample mean variance, this value being the average of the pair.

Experiments 13 to 16 were designed with the objective of testing the efficiency of antithetic variates when used in conjunction with the run subdivision method of sampling.

As in all previous cases, the four experiments differ only in relation to random number streams. All four experiments consist of two long runs, each having a total length equivalent to the departure of four hundred and seventy orders, split in one non-sampling period of twenty departure plus three sampling periods of one hundred and fifty departures.

At the end of each sampling period, the mean value of the outputs for that period are calculated and recorded, and new random number stream selected. After the last sampling period is terminated an antithetic run is started from the original initial condition. This run uses random streams which are the complement of the ones used in the previous runs. If the $\underline{n}^{th}$ sampling period, of the original run used the stream r_1, r_2, ... r_n, the $\underline{n}^{th}$ sampling period of the antithetic run will use the stream $(1 - r_1)$, $(1 - r_2)$, ... $(1 - r_n)$.

At the end of the antithetic run, the outputs of equivalent periods of the two runs are averaged, giving rise to 3 pairs of antithetic samples. If the outputs from the original run are x_1, x_2, and x_3, and the outputs from the antithetic run are y_1, y_2 and y_3, the 3 resulting samples will be given by $(x_1 + y_1)/2$, $(x_2 + y_2)/2$ and $(x_3 + y_3)/2$.

This procedure has a pitfall, concerning the way in which the antithetic pairs are generated.

As the length of run increases, the effectiveness of the antithetic procedure tends to decline, because of the increased chance that the elements being simulated will get out of sequence.

The consequence of elements getting out of sequence is a decline in the negative correlation between the antithetic samples.

A procedure used by El-Rayah (1973) seems more efficient. It consists of making sure that each period from the original run, and its corresponding antithetic companion will have exactly the same initial condition. This is obtained by a process of dumping and later reinstating the initial conditions of each period of the original run. Unfortunately this procedure was not used in this study because of practical considerations. An application package (CSL), with its own 'internal' variables was being used and this made it extremely difficult to dump and reinstate the conditions of the simulated system.

The execution of the sixteen experiments has generated the necessary information for assessing the relative efficiency of the four sampling strategies. For each strategy four independent values of the sample mean, standard error, and relative standard error were calculated.

The efficiency of each method is measured by the average of the relative standard error, over the four independent estimates. The relative standard error is given by

$$RS_{\bar{X}} = S_{\bar{X}} / \bar{X}$$

and measures the relative dispersion of the mean.

The results given in table 5.1, show that the most efficient method of sampling would be to use antithetic pairs of completely independent runs. This method proved to be 1.8 times more efficient than the run subdivision method with antithetics, and more than 3 times as efficient as the other two methods.

The use of antithetics on the other hand proved to be almost twice as efficient as the 'straight' sampling for the run subdivision method and more than 3 times as efficient as the 'straight' sampling for the case of completely independent runs.

The experiments also gave information about the level of precision which can be expected if three pairs of antithetic runs are used.

It was felt that in order to combine the advantages of precise estimate of the mean and narrow confidence intervals, the number of samples should be increased.

Based on these results a sampling procedure was designed which consists of six independent pairs of antithetic runs. All runs start from the same initial condition, as described in 5.2.2, have a non-sampling period (stabilization) equivalent to the completion of twenty full orders, followed by a sampling period equivalent to the completion of another one hundred and thirty orders.

After deciding on the sampling procedure, a final test was made in order to check whether it would give a precise estimate of the mean, without any bias. The simplest way of checking that is to have two sets of experiments, in which the run length is varied, and then to apply a statistical test to check whether there is any significant difference between the two estimates. The result of two sets of experiments, both using the same sampling procedure as described above, were compared. One experiment used run lengths of one hundred and fifty orders and the other run lengths of one hundred orders. The difference in sample mean between the two were compared, through a t test on the paired samples, and no significant difference could be found. For details of the t test see table 5.2.

The results confirm that the estimates that will be obtained by using the chosen sampling procedure will satisfactorily represent the real parameters.

5.4- Summary

This chapter discusses the tactical problems involved in computer simulation experiments, and describes in detail how a decision was reached in choosing the best sampling strategy. Analysis of pilot studies yielded information which helped to decide the initial conditions and stabilization period.

Four sampling strategies have been tested, and it was found that the most efficient method would be to use independent pairs of antithetic runs. It was also shown that the use of variance reduction technique, known as antithetic variates greatly enhances the efficiency of sampling.

Finally a decision was made to utilize the following sampling procedure for the experiments in which the inventory system is not in operation:

(i) Start simulation from an 'empty and idle' initial condition, as described in 5.2.2.

(ii) Consider a stabilization period equivalent to the completion of twenty full orders (two hundred and sixty jobs), and discard the information obtained during this period;

(iii) After the twentieth order has been completed, start sampling, and continue to do so until another one hundred and thirty orders have been completed;

(iv) After the completion of the one hundred and fiftieth order, restart the simulation from the original initial condition, and use an antithetic stream of random numbers. Follow the same procedure as in the first run.

(v) After the antithetic run has been completed, return again to the original initial condition and have a new run similar to the first one, but using a completely fresh set of random numbers;

(vi) Repeat the procedures as from item (iv) until twelve runs have been obtained.

Table 5.1 - Comparison of sampling strategies in terms of confidence in the estimation of the mean.

STRATEGIES		$\bar{X}$	$S_{\bar{x}}$	$S_{\bar{x}} / \bar{X}$
1	Six independent runs	6.82	1.380	0.1962
2	Six samples from continuous run	6.93	1.482	0.2080
3	Three independent pairs of antithetics	6.88	0.420	0.0610
4	Three pairs of antithetics from continuous runs	6.78	0.760	0.1120

Table 5.2 - Student t test on the difference of the sample mean values of runs with different lengths.

RUN LENGTH		DIFFERENCE
100 ORDERS	150 orders	
RESULTS		
X_1	X_2	$d_i = X_1 - X_2$
7.37	7.12	0.25
6.51	6.66	-0.15
7.15	6.81	0.34
6.05	5.99	0.06
6.43	6.43	0.00
6.92	7.03	-0.11

$$\bar{d} = \frac{\Sigma d_i}{6} = 0.065 \qquad S_d^2 = \frac{\Sigma (d_i - \bar{d})^2}{5} = 0.03818$$

$$t_s = \frac{\bar{d} - \mu_d}{S_d} \sqrt{n} = 0.8142 \qquad t^*_{5,0.90} = 1,476$$

$$t^* > t$$

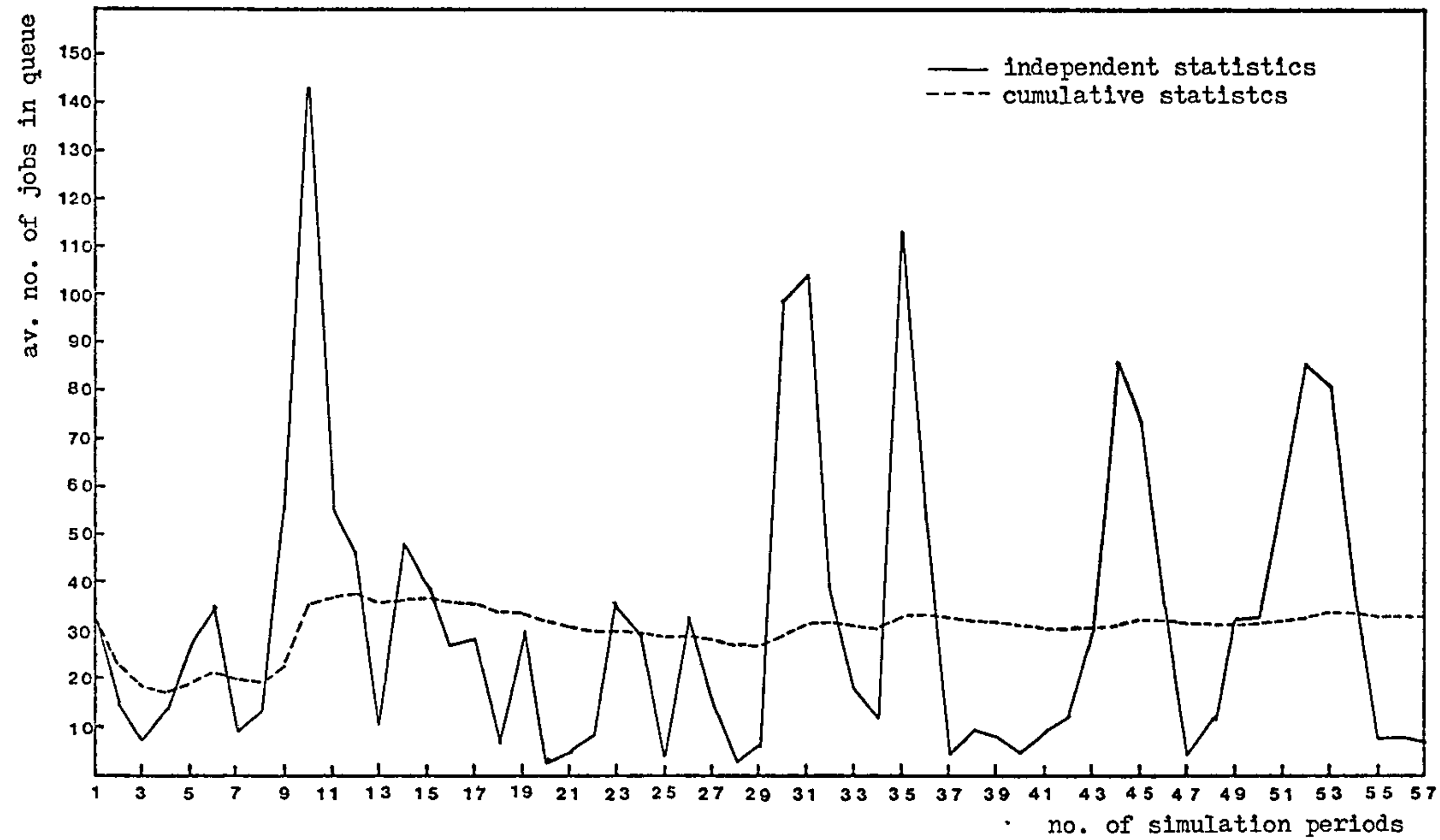


FIGURE 5.1

AVERAGE NUMBER OF JOBS IN QUEUE VS. LENGTH OF RUN
(one simulation period is equivalent to the completion of ten orders)

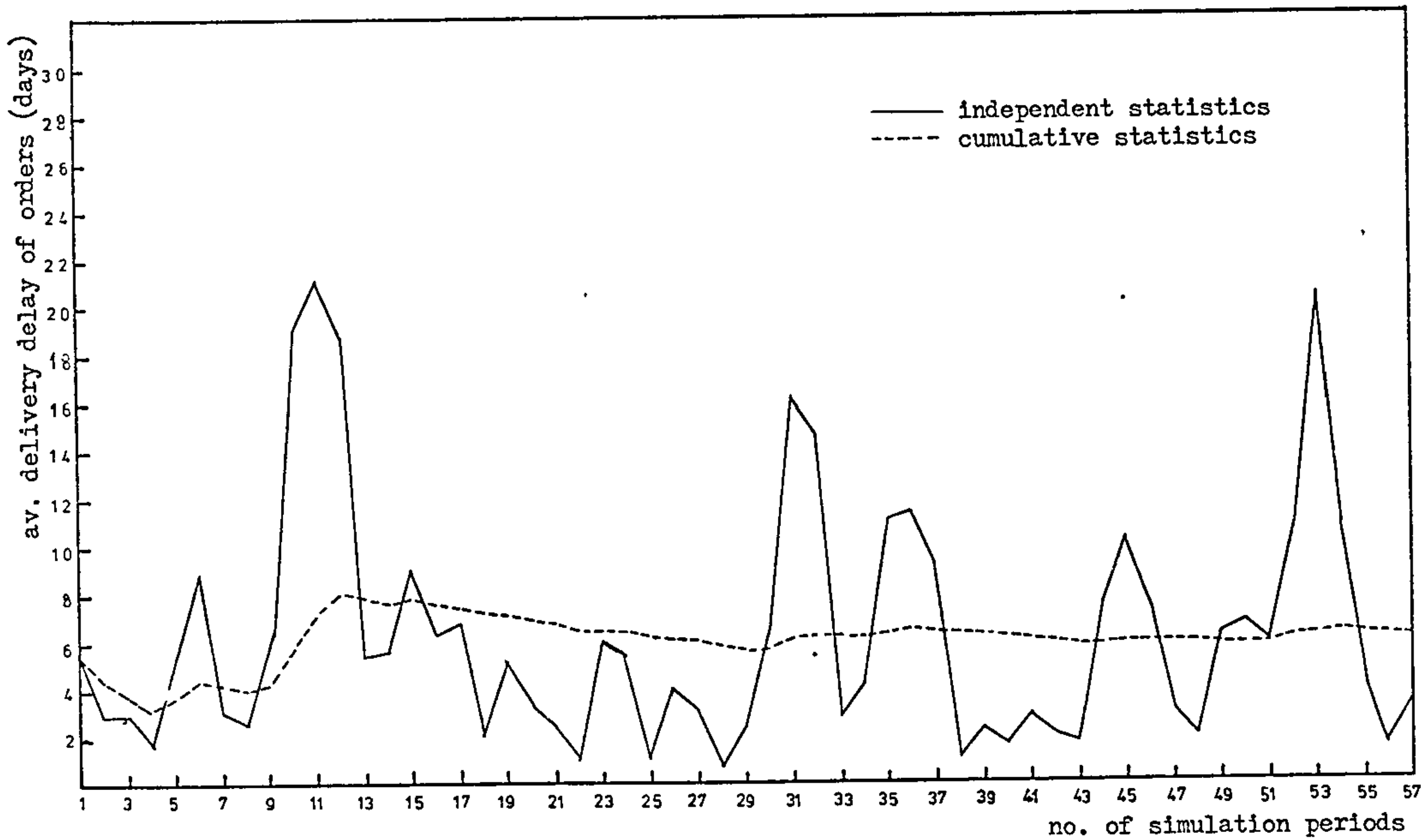


FIGURE 5.2

AVERAGE DELIVERY DELAY OF ORDERS VS. LENGTH OF RUN
(one simulation period is equivalent to the completion of ten orders)

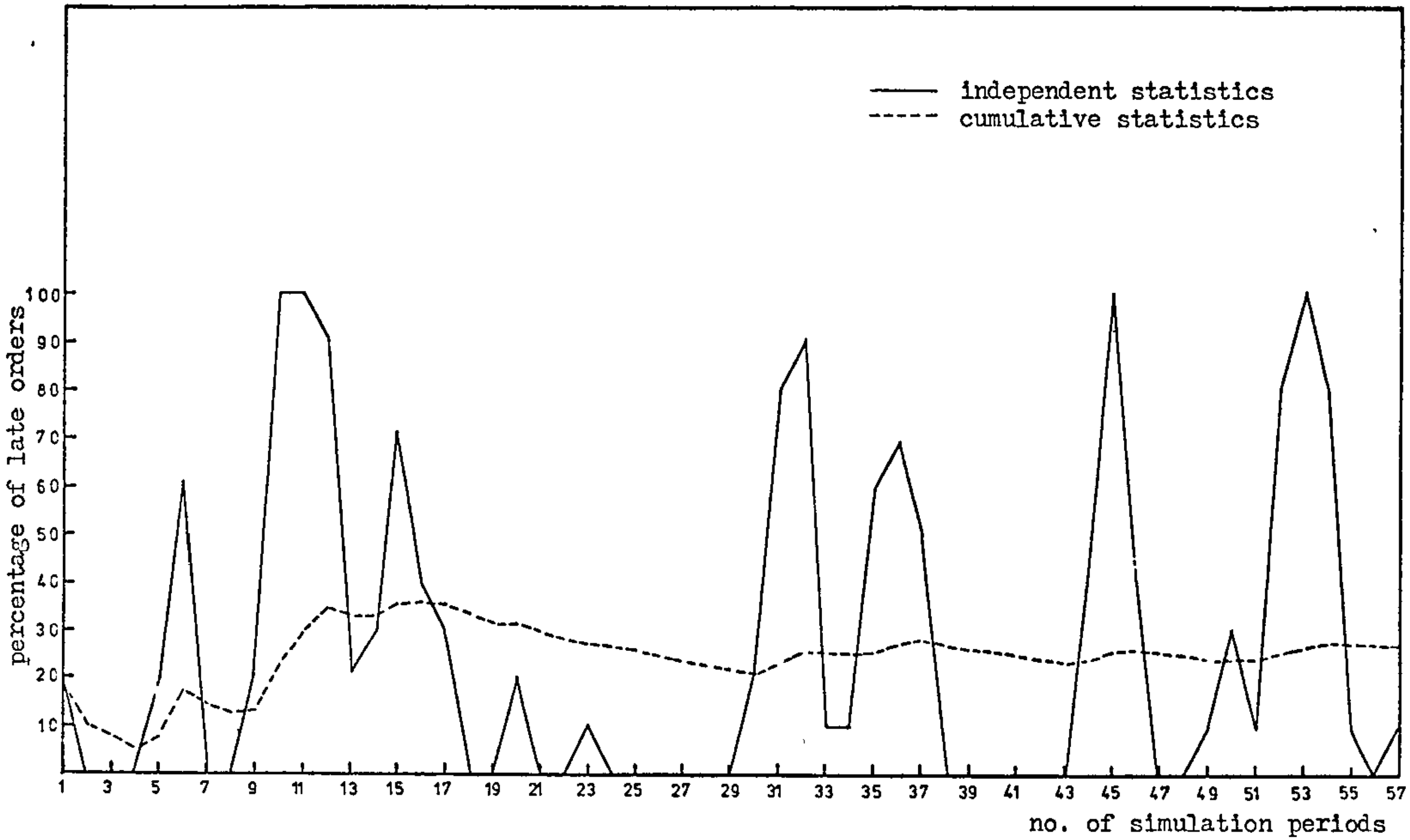


FIGURE 5.3

PERCENTAGE OF LATE ORDERS VS. LENGTH OF RUN
(one simulation period is equivalent to the completion of ten orders)

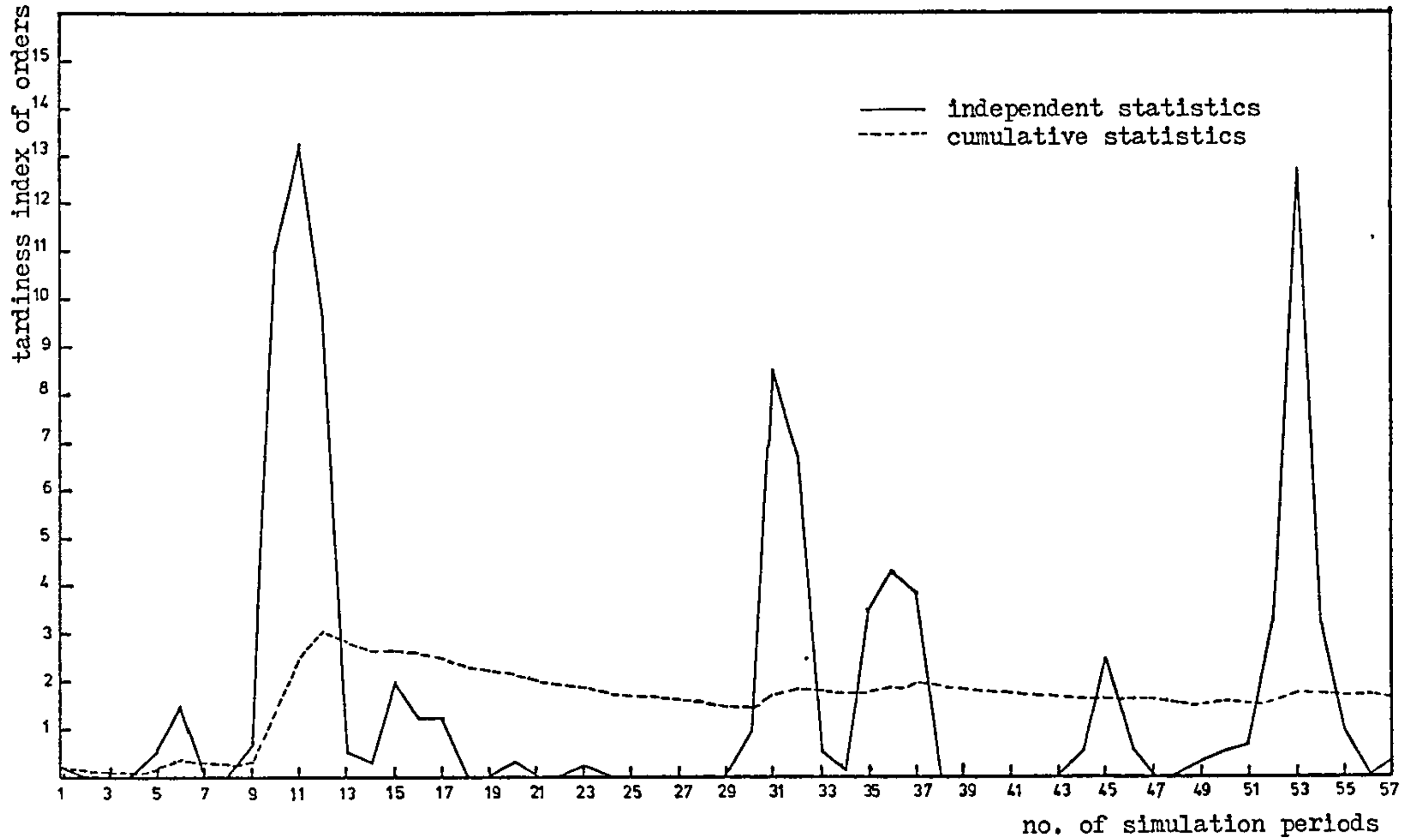


FIGURE 5.4

TARDINESS INDEX OF ORDERS VS. LENGTH OF RUN
(one simulation period is equivalent to the completion of ten orders)

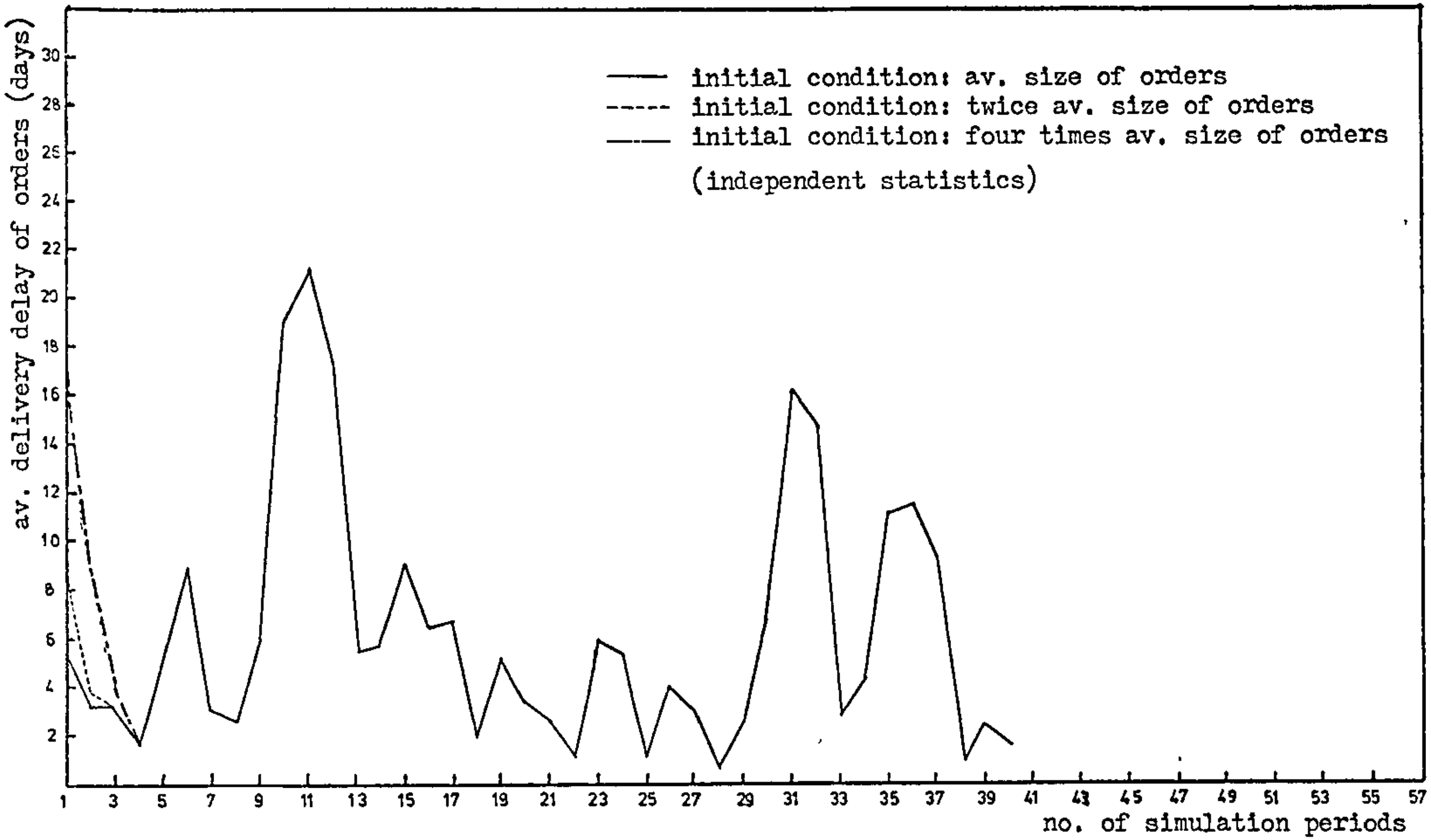


FIGURE 5.5

EFFECT CAUSED BY DIFFERENT INITIAL CONDITIONS ON INDEPENDENT STATISTICS
(one simulation period is equivalent to the completion of ten orders)

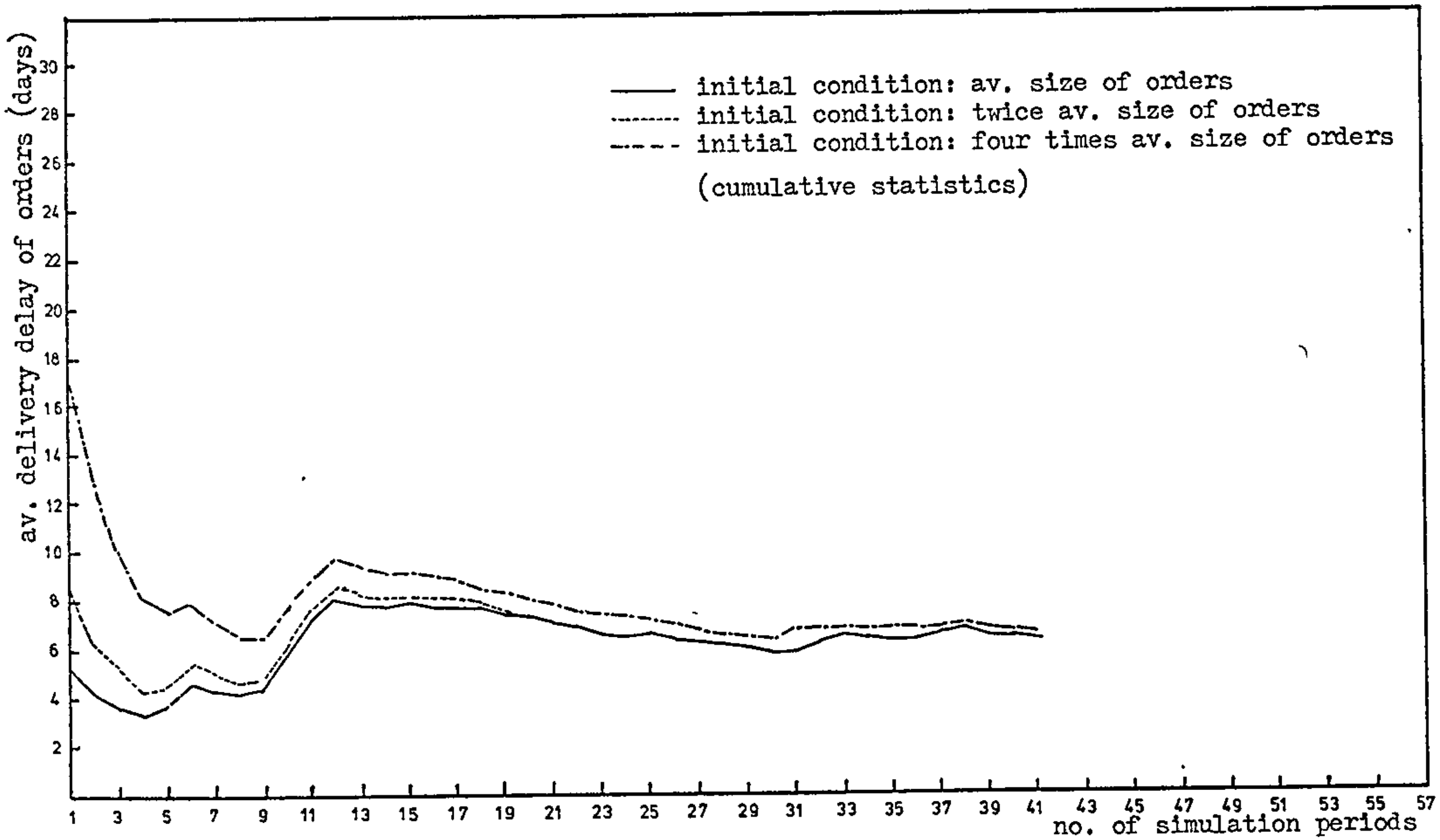


FIGURE 5.6

EFFECT CAUSED BY DIFFERENT INITIAL CONDITION ON THE CUMULATIVE STATISTICS
(one simulation period is equivalent to the completion of ten orders)

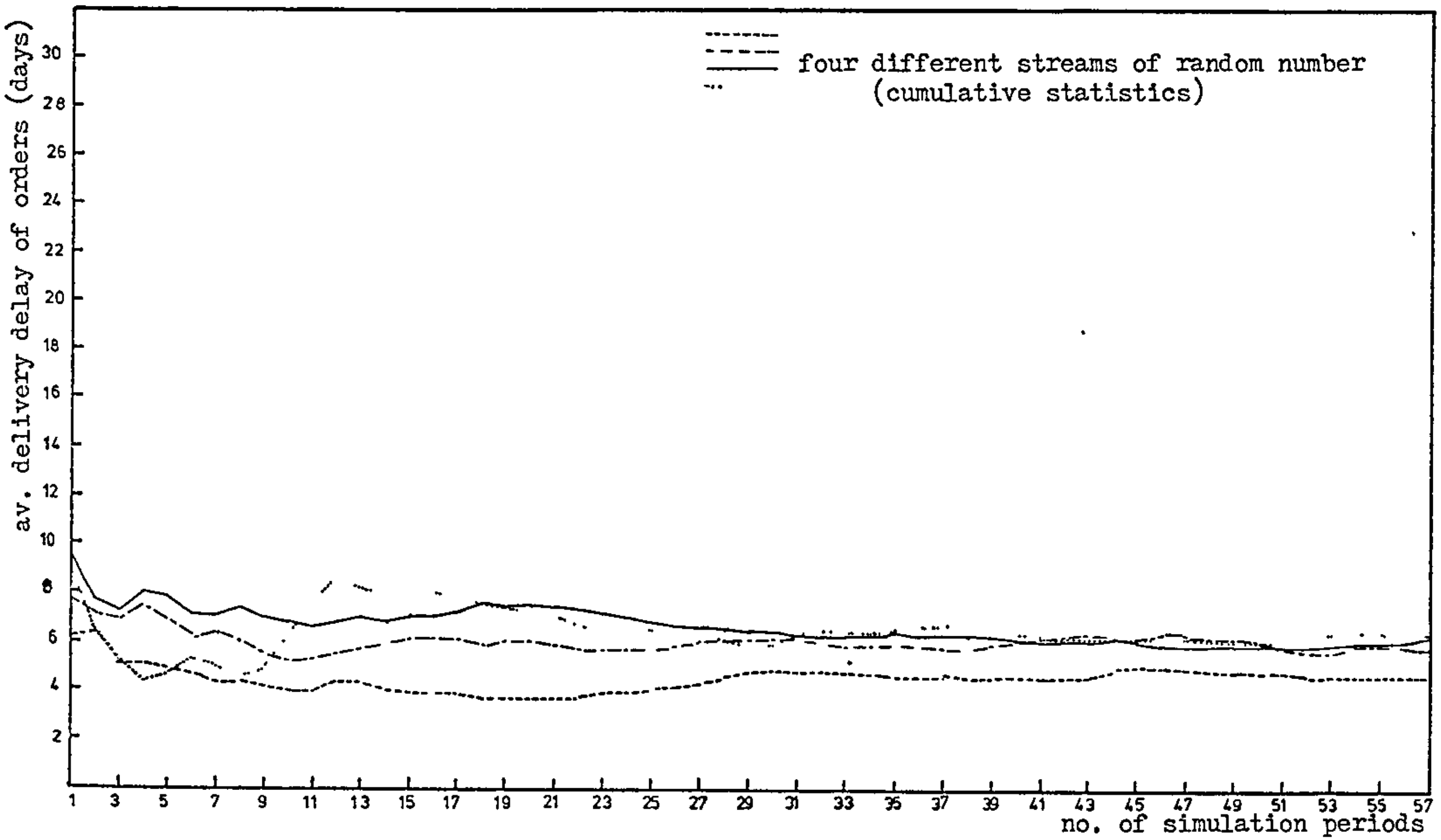


FIGURE 5.7

EFFECT OF DIFFERENT RANDOM STREAMS ON THE CUMULATIVE STATISTICS
FOR SYSTEM CONFIGURATION FIFOMB, a = 85, b = 16, c = 27, d = 1600, e = ∞, f = 1
(one simulation period is equivalent to the completion of ten orders)

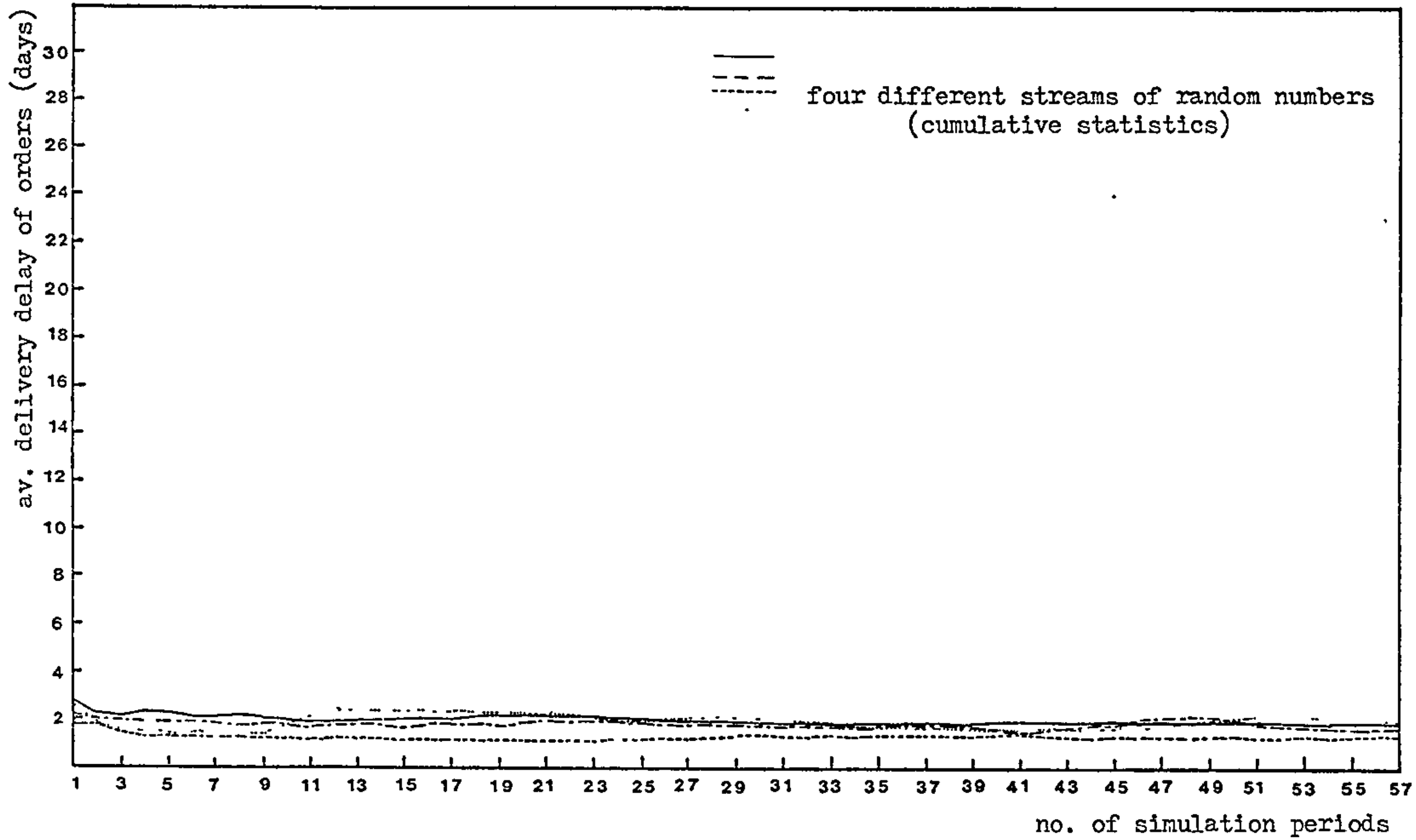


FIGURE 5.8

EFFECT OF DIFFERENT RANDOM STREAMS ON THE CUMULATIVE STATISTICS
FOR SYSTEM CONFIGURATION FIFOMB, a = 65, b = 8, c = 42, d = 1000, e = 450, f = 1
(one simulation period is equivalent to the completion of ten orders)

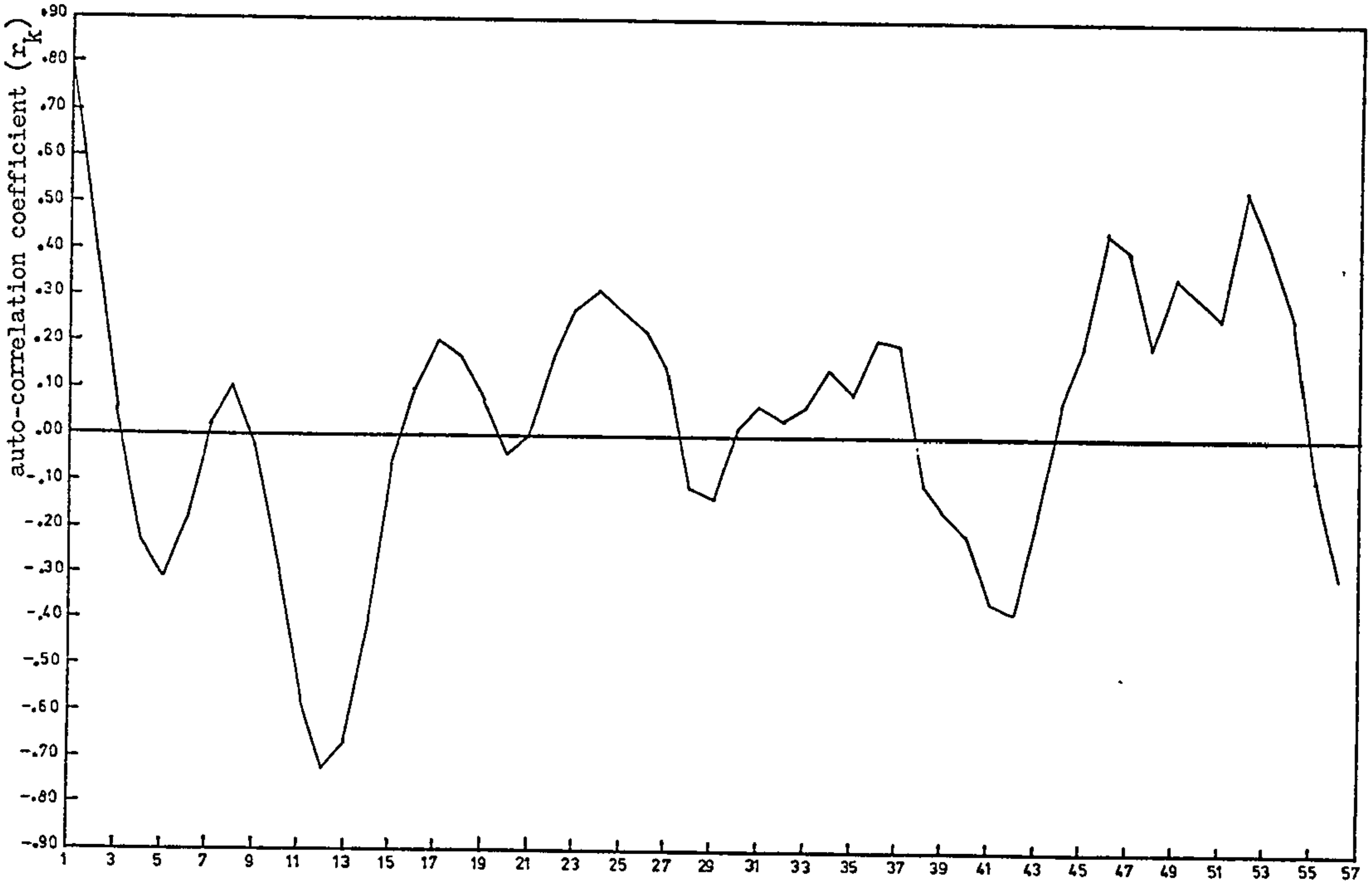


FIGURE 5.9

CORRELOGRAM - PERIOD LENGTH EQUALS FIVE ORDERS

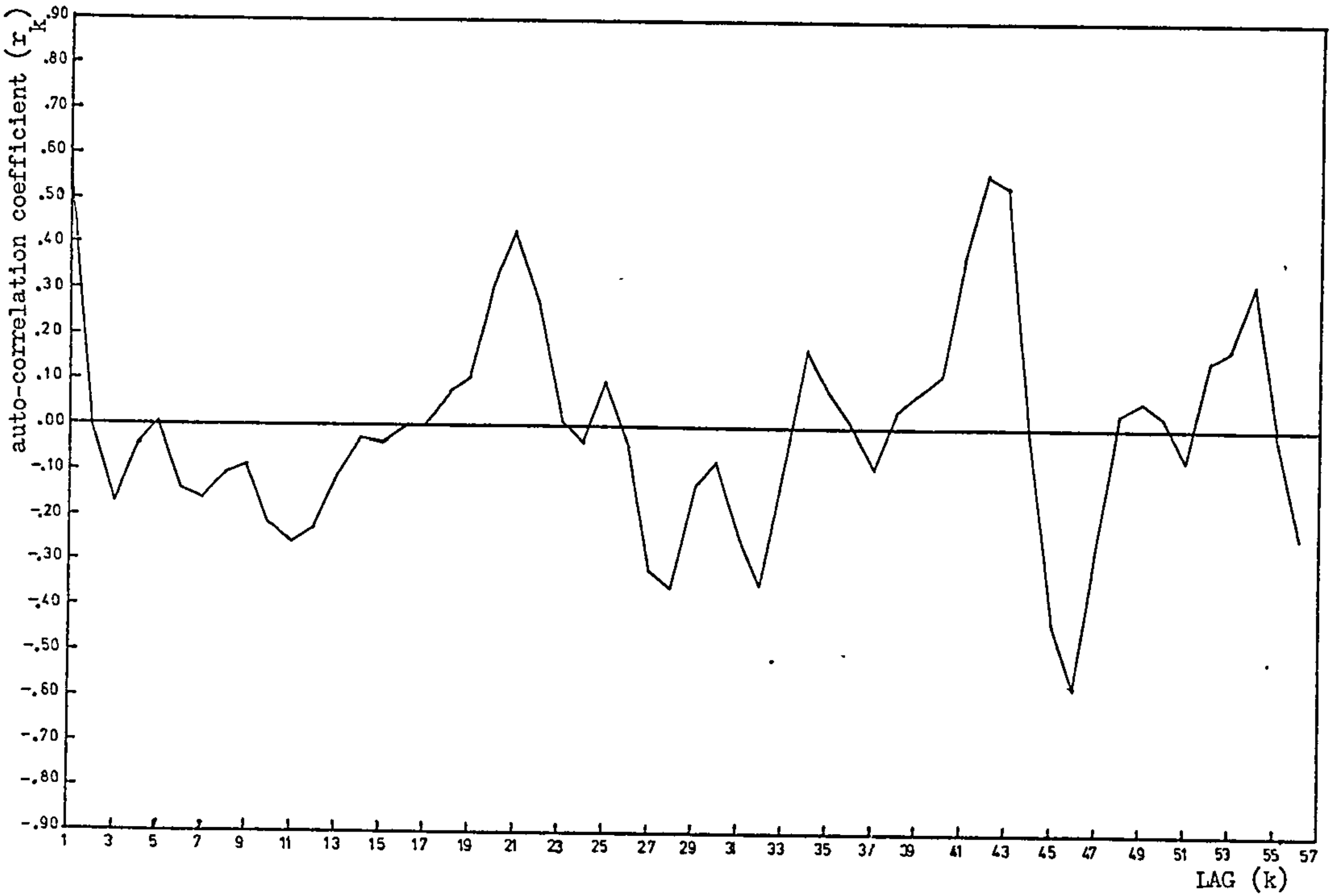


FIGURE 5.10

CORRELOGRAM - PERIOD LENGTH EQUALS TEN ORDERS

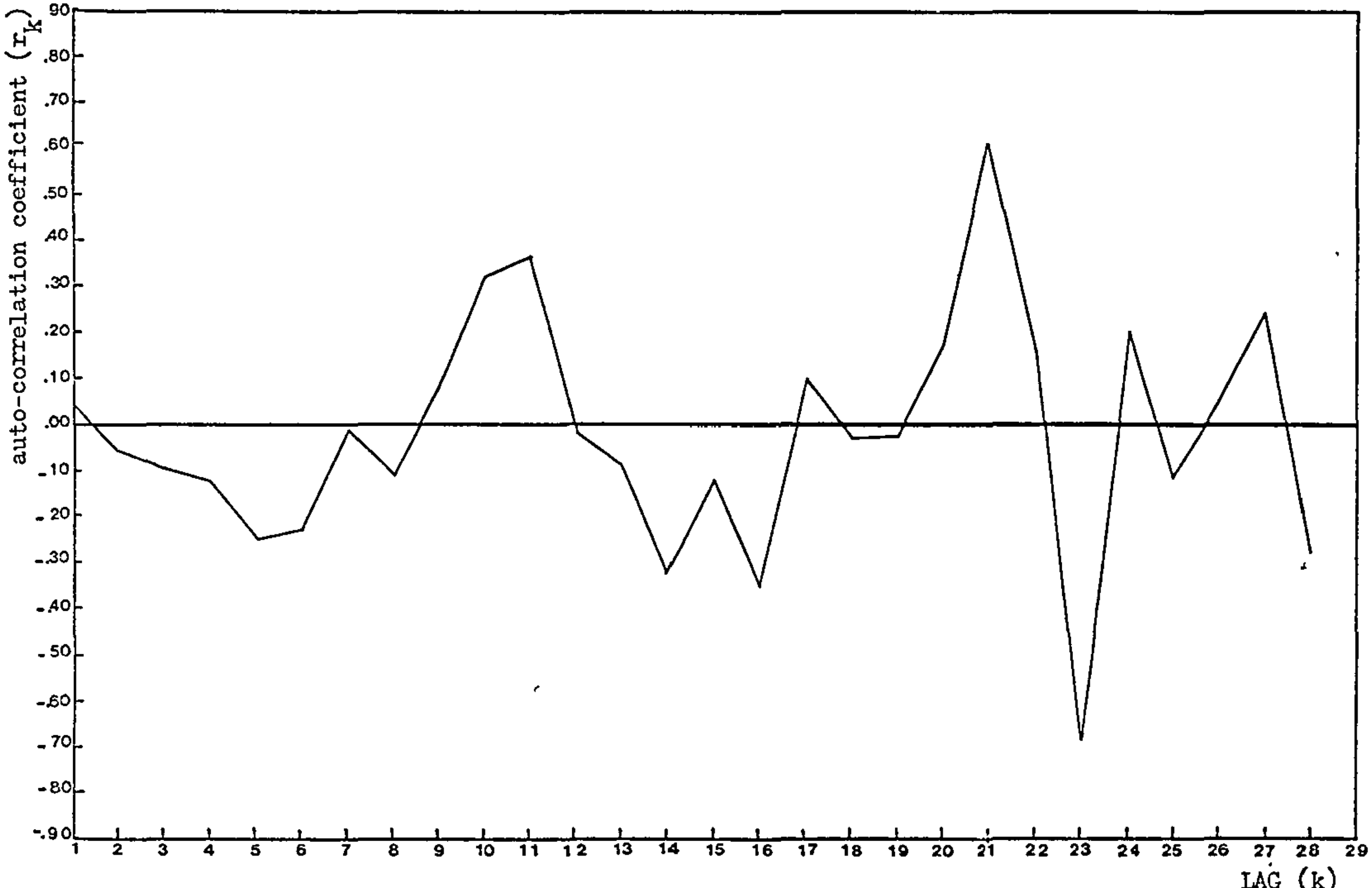


FIGURE 5.11

CORRELOGRAM - PERIOD LENGTH EQUALS TWENTY ORDERS

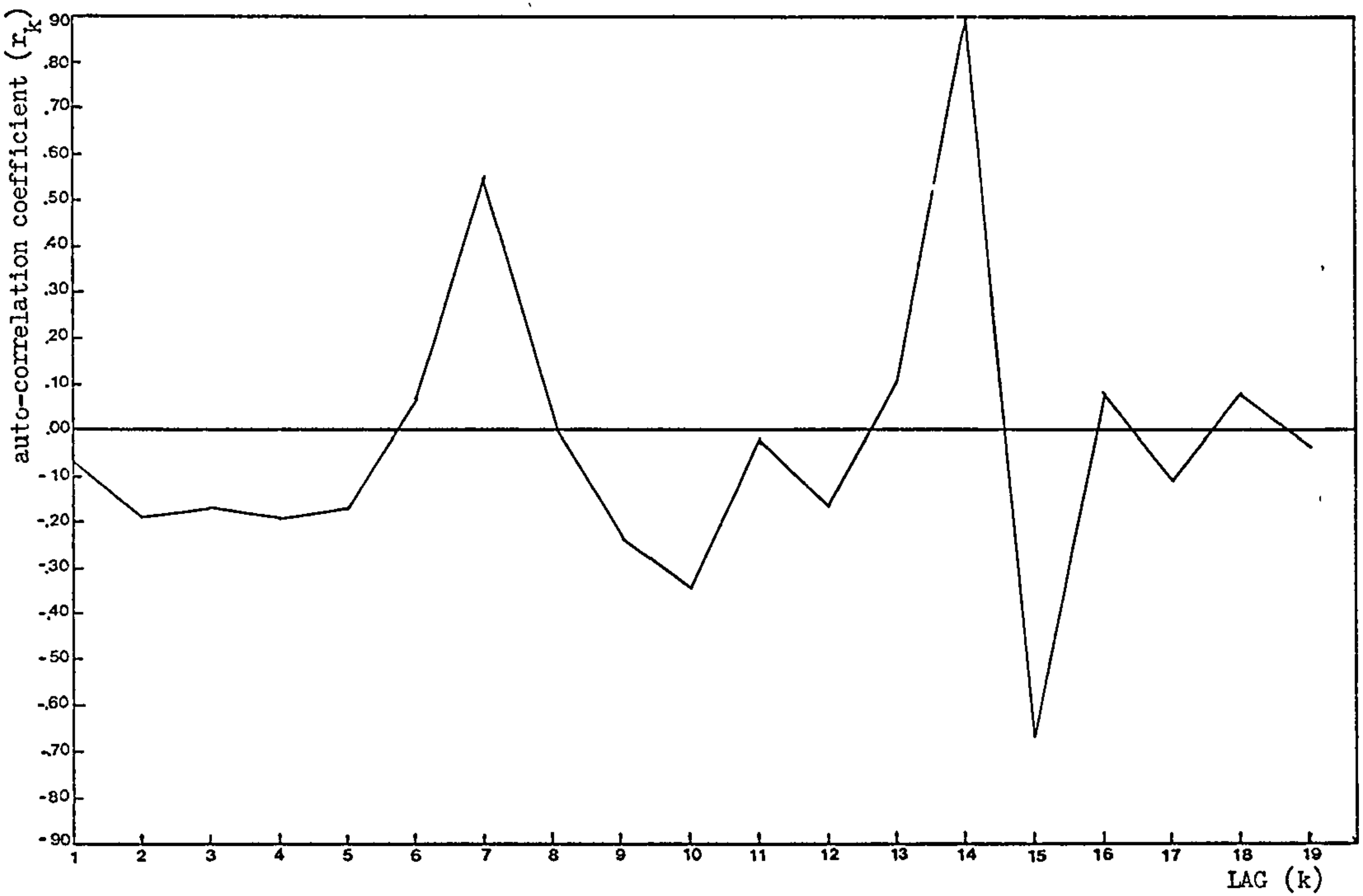


FIGURE 5.12

CORRELOGRAM - PERIOD LENGTH EQUALS THIRTY ORDERS

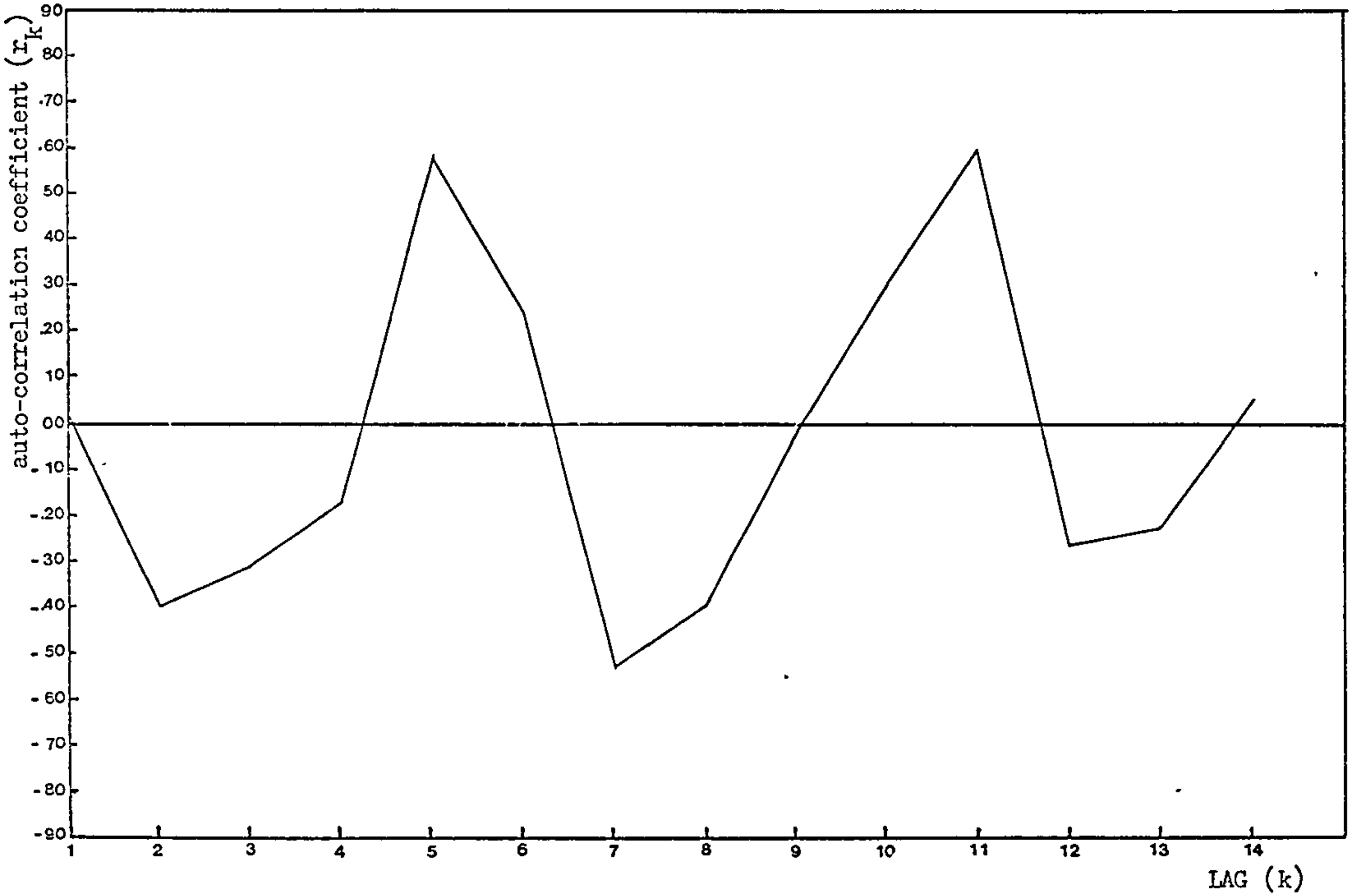


FIGURE 5.13

CORRELOGRAM - PERIOD LENGTH EQUALS FORTY ORDERS

CHAPTER 6

STUDY OF PRIORITY SCHEDULING RULES

6.1 - Introduction

The main objective of this part of the study is to compare the priority rules in relation to their ability to influence the delivery performance of the system.

The experimental design, as described in 4.3.1, consists of testing six priority rules, viz. FIFOB, FIFOMB, SPT, SPTM, SLACK and SLACKM, under six different system configurations. Each of the six rules will be tested for each of the six system configurations, giving a total of thirty six experiments. Each experiment will be conducted in accordance with the sampling procedure described in 5.3.3.

As described in 3.6.2, the model outputs seven variables related to measures of delivery performance. However one of those variables, viz. 'standard deviation of delivery delay of orders' will not be used, as it is thought that 'tardiness index of orders' and 'tardiness index of production' are more significant measures of dispersion than the standard deviation. The remaining six variables can be divided into two groups, where the variables of one group are 'weighted' measures of delivery performance (each order which is delivered is weighted by its total batch size) and the variables of the other group are 'unweighted' measures of delivery performance. In order to facilitate future references to the measure of delivery performance the name of the variables will be shortened in accordance with the table below:

Variable	Short name
Average delivery delay of order	'order delay'
Average delivery delay of 'production'	'production delay'
Percentage of late orders	'orders late'
Tardiness index of orders	'tardiness of orders'
Percentage of 'production' delivered late	'production late'
Tardiness index of 'production'	'tardiness of production'

The variables 'orders late', 'production late', 'tardiness of orders' and 'tardiness of production' are functions of due-date. As described in 3.6.2 the procedure for determining due dates is based on a constant 'lead time', which is fixed at a number D of days. In order to study the influence of the value of D on the scheduling rules, the model calculates the values of 'orders late', 'production late', 'tardiness of orders', and 'tardiness of production' for different values of D, equal to eight, ten, twelve, fourteen, sixteen, eighteen and twenty days respectively.

For the sake of analysis the thirty six experiments are divided into six blocks of six experiments, where each block consists of the results obtained by applying the six priority rules to a specific system configuration.

As far as the delivery performance is concerned, the priority scheduling rules are compared to each other by the use of graphs and by statistical analysis. The graphs are organized by plotting the values of 'orders late', 'tardiness of orders', 'production late' and 'tardiness of production', obtained from each rule, as a function of D. This will give a total of four graphs for each system configuration, each graph having six curves, where each curve represents the delivery performance obtained by a particular priority scheduling rule, as a function of D.

The statistical analysis consists of the application of two statistical tests to each block of results representing a system configuration. First an 'F test' of analysis of variance is applied to the data, with the objective of testing whether or not there is any significant statistical difference between the six priority rules. The hypothesis and alternate are:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$

H_1: at least two means are not equal

Where μ_1, $\mu_2 \ldots\ldots \mu_6$ represent the means of some measure of delivery performance for each of the six priority rules respectively. The statistical tests are applied to all six measures of performance, for the case of D equal to eight days.

The 'F test' is followed by a 'multiple comparison test', with the objective of identifying which differences are significant. A number of 'multiple comparison procedures' are available (Scheffe (1959), Tukey (1953), Dunnet (1955)), which are based on the construction of confidence limits with a specified confidence level. Tukey's and Scheffe's methods are designed to allow the comparisons between all possible pairs of contrasts (differences), while Dunnet's method is designed with the objective of comparing all the means against a given control. For this reason the confidence intervals generated by Dunnet's method are narrower than the intervals generated by Scheffe's and Tukey's methods. In view of the objectives of this study, it was decided that Dunnet's method should be used, such that in each test the lowest mean is used as the 'control'. In this way it would be possible to test the significance of the differences between each priority rule and the priority rule which had generated the lowest (best) value for a measure of delivery performance. For a good reference about the applicability of Scheffe and Tukey's method see Guenther (1964), (1).

Because the sampling method adopted in this study has made use of the variance reduction technique, of 'common random numbers', or 'close replication' (paragraph 5.2.3.1), it is necessary to consider the random numbers as a factor (where the number of levels equals the number of antithetic pairs(six)). This means that instead of a one way classification method, a two way classification (or randomized block) will be used for the statistical analysis.This method has the advantage of separating the variance due to the random streams, from the variance caused by the differences between contrasts (priority rules).(For more detailed description of the two way classification system (randomized blocks), see Davies (1958) and Guenther (1964) (2)).

Apart from the measures of delivery performance the model also outputs all the other measures of internal behaviourwhich were described in paragraph 3.6.1. Some of those variables, i.e., 'average number of jobs waiting in the queue'; 'average process cycle time';'actual average load factor'; 'total demand'; 'total production delivered';'machine idle time due to setup', will also be used in the discussions in order to support the results of delivery performance. Rather than use the actual values of 'total demand' and 'total production delivered' their differences are used as a measure of delivery performance. These values are termed 'remaining content on the shop'.

All the numerical results of the variables used on the discussions which follow are presented in appendix 4.

6.2 - Presentation of results

The results of this series of experiments will be presented individually for each system configuration, both for the graphical and statistical analysis. After the individual presentations which are accompanied by brief comments, the results will be discussed as a whole, in order to take into account the overall performance of each rule over the six system configurations tested.

6.2.1 - Results from system configuration (I)

Table 6.1 presents the results of the statistical analysis carried out on the data for the six measures of delivery performance. The 'F test' indicates a statistically significant difference between the priority rules, for all the six measures of delivery performance.

The results also show that FIFOB produced the lowest value for the variables 'order delay', 'production delay', and 'tardiness of order', and SLACK produced the lowest value for the variables 'tardiness of order'(together with FIFOB), 'orders late', 'tardiness of production' and 'production late', with the SPT and SPTM rules producing the highest values for all six variables. However when the multiple comparison test was applied to the data it indicated few significant differences. Apart from the variable 'order delay', for which all five rules are significantly different from the 'control' (FIFOB), the test failed to find any significant differences between the rules FIFOB, FIFOMB, SLACK and SLACKM. However it indicated that SPT and SPTM were significantly different from the 'control', for all the six measures of performance.

The level of delivery performance with D equal to 8 days was quite good for all of the rules. The lowest values for 'orders late' and 'production late' were respectively 0.76% and 1.11%, while the highest values were 2.24% and 7.09% respectively. This 'good' delivery performance comes as no surprise, as system configuration (I) is the most loose (low load factor; low value for setup times; large number of moulds; small size for the orders; more favourable ratio style/machine) of all configurations. As shown in appendix 4, the actual average load factor was around 66% for FIFOB, FIFOMB, SLACK and SLACKM and close to 69% for SPT and SPTM, which might justify their worse performance, as this higher load factor is a direct function of a higher 'process cycle time' which was around 5.15 minutes for the four better rules and around 5.30 minutes for SPT and SPTM. The lowest value for the 'remaining content' was produced by FIFOMB, and SPTM, with SLACK and FIFOB producing the highest values.

The curves of delivery performance as a function of lead time D, are presented in figures 6.1, 6.2, 6.3 and 6.4 respectively. It should be noted that in order to make the curves distinguishable from each other, the scales for these graphs have been greatly enlarged in relation to graphs for other system configurations. Figures 6.1 to 6.4 show that all the four measures of lateness and tardiness decrease in exponential manner as D increases. They also show that the relative performance between the rules do not seem to be much affected by the value of D. FIFOB, FIFOMB, SLACK and SLACKM are always very close to each other while SPT and SPTM are always worse than the other four. It should however be pointed out that in such a 'loose' situation no big differences between the priority rules should be expected as there is plenty of spare capacity and as a consequence any priority rule should produce reasonable results.

6.2.2 - Results from system configuration abc

Table 6.2 presents the results of the statistical analysis carried out on the data for the six measures of delivery performance. The 'F tests' gave mixed results. For the variables 'order delay', 'tardiness of order', 'orders late' and 'tardiness of production', the test indicated a statistically significant difference between the rules at 0.01 level. For the variables 'production late' the test indicated a statistically significant difference at the 0.05 level , and for 'production delay' the differences between rules were significant only at the 0.10 level. In relation to the individual performance of the different rules the results are also mixed. The SPTM rule produced the lowest value for the unweighted measures, i.e., 'order delay', 'tardiness of order', and 'orders late'. In all three cases (with the exception of SPT which did not show any significant difference) there was a statistically significant difference between SPTM and all the other four rules.

In relation to the weighted measures of delivery performance, FIFOMB produced the lowest value for 'production delay' and 'tardiness of production', while SPTM produced the lowest value for 'production late'. However the multiple comparison test indicated few statistically significant differences in all three cases. In the case of 'production delay' the only rule significantly different from FIFOMB was SPT, while in the case of 'tardiness of production' SPT and SPTM were the only rules significantly different from FIFOMB. Finally, for the case of 'production late', FIFOB, SLACK and SLACKM showed a statistically significant difference from SPTM.

It is interesting to note that there was a tremendous deterioration of the general level of delivery performance when the system changed from configuration (I), to configuration abc (higher load factor, higher setup times, and smaller number of moulds). The lowest value for 'production late' and 'tardiness of production' (for D = 8) which were equal to 1.11% and 0.022 respectively in the case of (I), went up to 34.09% and 2.598 respectively, in the case of abc. The actual 'average load factor', as shown in appendix 4, also went up from around 66% to around 78% for FIFOMB and from around 69% to around 83% for SPT. The lowest values of 'remaining content' were produced by SLACKM and FIFOMB, while SPT and SPTM produced the highest values.

It should be noted that the actual increase in the arrival rate was of the order of 33% and so the actual load factor should in principle be expected to increase by the same rate. Actually it had a relative increase of only 20%. This can be explained by the reduction in the values of 'idle time due to setup' and of 'process cycle time' which in the case of FIFOMB (and similarly for all other rules) went down from 6.15% and 5.18 min.respectively to 4.58% and 4.58 min., when the system changed from configuration (I) to configuration abc. This decrease in total setup times (and consequently of 'process cycle time') can only be explained by the fact that the decrease in the number of moulds (from 42 to 27) had more influence in bringing down the total amount of setup than the increase of the mean value of setup time (from 8 to 16 min.) had on bringing it up. This is understandable when it is considered that for system configuration abc, the system had 27 moulds for 24 stations, while in the (I) configuration the relationship was 42 to 24.

The curves relating delivery performance to lead time D, are shown in figures 6.5, 6.6, 6.7 and 6.8 (it should be noted that the scales for this set of figures are much smaller than the scales used in configuration (I)). In general the figures indicate the same exponential like relationship between D and the measures of delivery performance, for all the six rules. However it should be noted that SPT and SPTM rules tend to lose their advantage over the other four rules as the value of D increases. For example, if SPTM is compared with FIFOMB in terms of 'production late' (figure 6.7), it can be seen that for D equal to 8 days, SPTM produced a value of 34.09% compared with 38.83% for FIFOMB, i.e., an advantage of 4.74% for SPTM. On the other hand when the value of D was 20 days, the advantage was reversed in favour of FIFOMB, which produced a value equal to 5.61%, as compared to 10.30% for SPTM, i.e., an advantage of 4.69% in favour of FIFOMB.

A comparison between FIFOMB, FIFOB, SLACK and SLACKM, indicates that their relative performances are not very much affected by the value of D. FIFOMB tends to produce always the lowest values, among the four, for delivery performance, although the differences are very small and as shown by the case of D = 8 days, not statistically significant.

6.2.3 - <u>Results from system configuration def</u>

Table 6.3 presents the results of the statistical analysis carried out on the data for the six measures of delivery performance. The 'F tests' have indicated that apart from 'orders late', all the other measures of delivery performance show a statistically significant difference at the 0.01 level between the priority rules. The analysis of individual rules show that FIFOB has produced the lowest value for all the six measures of delivery performance. However when the multiple comparison test was applied, it failed to show any statistically significant difference between FIFOB and FIFOMB for all six measures of performance. In relation to the other four priority rules, they all produced results which were significantly different from FIFOB, for the measures 'orders delay', 'production delay' and 'tardiness of orders'. In relation to the other three measures of delivery performance only SPT and SPTM showed some statistically significant differences from FIFOB.

When the general level of performance is compared with the results obtained from configuration (I) it shows that the joint modification of the average size of orders (from 1000 to 1600), of the procedure for job splitting (splitting against no splitting), and of the ratio between number of styles and number of machines (from 3:2 to 3:1) has caused a considerable deterioration on the delivery performance of the system for all rules. The values of 'tardiness of production' and 'production late' for FIFOB, which were equal to 0.035 and 1.97%, went up to 0.318 and 12.01% respectively. However, this deterioration in delivery performance was much smaller than the one caused by changing from (I) to abc as described in 6.2.2. It is interesting to note from appendix 4 that the total time spent carrying out setups were considerably less for the rules which are designed to reduce setup time. FIFOMB, SLACKM, SPTM produced respectively 5.05%, 4.98% and 5.21% of 'percentage of time spent setting up' as compared to 6.30%, 6.29% and 6.51% respectively for FIFOB, SLACK and SPTM.

FIFOB and SLACK produced the lowest values for the 'remaining content', with SPT and SPTM again producing the highest values.

The curves of delivery performance as function of D are presented in figures 6.9, 6.10, 6.11 and 6.12. The curves for all four measures of delivery performance show similar patterns, which indicates that for this system configuration the value of D does not have much influence on the relative performance of the priority rules. The relative rankings of the rules for D = 8 days is maintained when D is varied but with SPT and SPTM producing markedly higher values than the other four rules, which produced very close results all the way through.

6.2.4 - <u>Results from system configuration ace</u>

Table 6.4 presents the results of the statistical analysis carried out on the data for the six measures of delivery performance. The 'F tests' again show mixed results, and indicate the existence of statistically significant differences between the rules for the unweighted measures of performance, viz. 'order delay', 'tardiness of order', and 'orders late', but no significant differences for the weighted measures of performance, with the exception of 'tardiness of production' for which there was significant difference at the 0.05 level. The results obtained from individual rules show that SPTM produced the lowest value for all unweighted measures of delivery performance, and also for 'production late'. In the case of the two other (weighted) measures, viz. 'production delay', and 'tardiness of production', the lowest values were produced by FIFOMB. However when the multiple comparison test was applied it indicated significant differences between SPTM and FIFOB, FIFOMB, SLACK and SLACKM for all unweighted measures, but no significant differences between SPTM and SPT. In relation to the other measures of delivery performance, the only significant difference was for SPT and SPTM in relation to FIFOMB for the 'tardiness of production'.

When the results of this series are compared with the results obtained from system configuration (I) a marked drop in delivery performance is indicated. The lowest value of 'production late' and 'tardiness of production' changed from 1.17% and 0.022 for configuration (I) to 34.81% and 2.398 for this system configuration. If these results are compared with the results obtained for system configuration abc it shows that they are very close to each other. This comes as no surprise as the two configurations are very similar, the only difference being the mean value of setup times (8 min. in ace, and 16 min. in abc) and the fact that 'jobs' are not split in ace. As shown in appendix 4, the difference in setup time is reflected in the values of the 'percentage of time spent with setting up', which came down from 4.65% to 3.63% for the case of SPTM when changing from abc to ace, and similarly for the other priority rules. In relation to the other measures of 'internal behaviour', SLACKM and FIFOMB produced the lowest values for 'process cycle time' and actual 'load factor', with SPT and SPTM producing the highest values. Also

SLACKM and SLACK have produced the lowest values for the 'remaining content', with SPT and SPTM again producing the highest values. Figures 6.13, 6.14, 6.15 and 6.16 for the measures of delivery performance, show a similar picture to the corresponding results of figures 6.5, 6.6, 6.7 and 6.8, which present the results for configuration abc. The SPT and SPTM rules tend to lose their advantage over the other four rules as the value of D increases. The relative results for FIFOB, FIFOMB, SLACK and SLACKM are not much influenced by the value of D, being in all cases very close to each other, with FIFOMB producing the lowest values among the four.

6.2.5 - Results from system configuration bdf

In table 6.5 the results of the statistical analysis carried out on the data for the six measures of delivery performance are presented. For all the six measures, the 'F test' has indicated a statistically significant difference among the priority rules, at the 0.01 level. The analysis of individual priority rules shows that FIFOMB has produced the lowest value for all measures of delivery performance but one, viz. 'production late', for which SLACKM has produced the lowest value. The results of the multiple comparison test indicate that SPT and SPTM are significantly different from the 'control' (SLACKM for 'production late' and FIFOMB for the other measures) for all measures of delivery performance; that SLACK was also significantly different for all measures but one ('tardiness of production'); that FIFOB was significantly different only for 'order delay', 'orders late', and 'production late', and that FIFOMB and SLACKM are significantly different only in relation to 'order delay'.

A comparison between the results obtained for this system configuration and the results for configuration def (which differs from this configuration only in relation to the mean value of setup time, 8 min. vs. 16 min.) and the splitting of the jobs (no splitting vs. splitting), indicates that the general level of delivery performance suffered a small deterioration, which must have been caused by an increased amount of time spent with setting up the machines. This increase is shown by the 'percentage of machine idle time due to setup', which went up from 5.05% to 8.50% for the case of FIFOMB and from 6.51% to 12.00% in the case of SPT. It also went up for all other rules, as shown in appendix 4. It is interesting to note that the difference in the percentage of time spent with setting up, between the rules designed to avoid setup (FIFOMB, SLACKM and SPTM), and the rules which do not try to avoid setup (FIFOB, SLACK and SPT), are relatively large for this system configuration. For example, FIFOB and FIFOMB spent respectively 11.26% and 8.50% of the total time with setting up. The same effect is true for other rules (SPT vs. SPTM and SLACK vs. SLACKM). This effect is naturally reflected in the mean value of 'process cycle time', and consequently on the delivery performance of the rules. In relation to 'remaining content' SPT and SLACK produced the lowest value with SPTM again producing the highest value.

The curves relating delivery performance to D are presented in figures 6.17 to 6.20. Figures 6.17 and 6.19 ('orders late' and 'production late' respectively) show that the differences between FIFOMB vs. FIFOB and SLACKM vs. SLACK, which are considerable for D = 8 days, tend to disappear as the value of D increases. The figures also show that the difference between SPTM (and SPT) and the other four rules tend to become relatively larger as the value of D increases. A comparison between figures 6.17 and 6.18 against 6.19 and 6.20 respectively show that the relative performance of SPT and SPTM against the other four rules is much larger when the measures of delivery performance are 'weighted' (figs. 6.19 and 6.20) by the 'order' batch size.

6.2.6 - Results from system configuration abcdef

Table 6.6 presents the results of the statistical analysis carried out on the data, for the six measures of delivery performance. The results of the 'F test' indicated a statistically significant difference among the rules, for all the six measures of delivery performance, at the 0.01 level. The analysis of individual rules show that FIFOMB has produced the lowest value for all the three 'weighted' measures of delivery performance ('production delay'; 'tardiness of production'; 'production late'),while SPTM produced the lowest value for 'order delay' and 'tardiness of order', and SPT produced the lowest value for 'late orders'. However the results of the multiple comparison test showed no significant differences between FIFOMB and SPTM for 'order delay' and 'tardiness of order', but indicated a significant difference in favour of FIFOMB for the 'production delay' and 'tardiness of production'. The results also indicated significant differences for FIFOB and SLACK for all measures of performance but one, viz. 'tardiness of production' for FIFOB. In relation to SLACKM the multiple comparison test indicated only two significant differences, corresponding' to 'order delay' and 'orders late'. The results for SPT are similar to the results for SPTM, as far as significant differences are concerned.

The general level of delivery performance, obtained from this system configuration was the worst in relation to the other five configurations. This is to be expected as abcdef is the most 'tight' of all configurations. The smallest values for 'order delay' and 'production delay' which in the case of configuration (I) were equal to 1.83 days and 2.94 days respectively, moved up to 6.62 days and 8.43 days respectively, when the system configuration changed to abcdef. It is also interesting to note the considerable differences in 'process cycle time', between the rules which try to avoid setting up (FIFOMB, SLACKM and SPTM) and the other rules (FIFOB, SLACK and SPT). The values of 'process cycle time' for FIFOMB, SLACKM and SPTM were equal to 5.04, 5.03 and 5.26 minutes respectively compared to 5.55, 5.55 and 5.75 minutes for FIFOB, SLACK and SPT. These differences which are a direct consequence of the amount of time spent with setting up for each rule, is reflected in the value of average load factor for each rule, which vary from a low of 80.92% for FIFOMB, to a high of 91.46% for SPT. In relation to the 'remaining content', SLACK and FIFOB produced the lowest values, while SPT and SPTM again produced the highest values.

Figures 6.21, 6.22, 6.23 and 6.24 show the results of delivery performances as a function of D for all the six rules. It should be noted in figures 6.21, 6.22 and 6.23, how the performance of the SPT and SPTM rules tends to deteriorate relative to the other four rules, as the value of D increases. For example, in figure 6.21, if the results of 'orders late' for SPTM is compared with the corresponding results for FIFOMB, it is possible to see that for D = 8 days, SPTM had 23.8%

of orders late compared with 31.98% for FIFOMB, an absolute advantage of 8.14% in favour of SPTM. However for D = 20 days, the results were reversed with FIFOMB producing only 3.84% of orders late as compared to 6.02% for SPTM, an absolute advantage of 2.16% in favour of FIFOMB. The same effect happens for the weighted measures of performance, but it is much more accentuated. The results for 'production late' (figure 6.23) are a good example of this. For D = 8 days the results for FIFOMB and SPTM were respectively 41.91% and 44.55% with the multiple comparison test failing to find any statistically significant difference between them. However for D = 20 days, the results were respectively 4.04% and 14.23%, a considerable difference both in relative and absolute terms.

A comparison between FIFOMB, FIFOB, SLACK and SLACKM show that their relative performance tend not to be much influenced by the value of D, but the absolute differences tend to decrease with the increase in the value of D.

6.3 - Discussion of results

The results obtained for the different system configurations are mixed as far as the performance of individual rules are concerned. The relative performance of the rules is affected by the system configuration; by the way in which delivery performance is measured; and by the value of D (lead time used to fix due dates). However a detailed analysis of the data show some clear points. The first point to be noted is the consequences of incorporating in the priority rules the procedure for reducing the amount of time spent in setting up. Comparisons between FIFOB vs. FIFOMB, SLACK vs. SLACKM, SPT vs. SPTM indicate that in general the introduction of the procedure for reducing the amount of setup time is beneficial to the performance of the priority rules. Take, for example, the results of tables 6.1 to 6.6, for FIFOMB and FIFOB. From there, it is possible to see that for the 36 results of delivery performance (six system configuration X six measures of delivery performance), FIFOMB produced lower values than FIFOB on 25 occasions, while FIFOB produced lower values than FIFOMB on 11 occasions. All the 11 occasions however happened for system configurations (I) and def, which because of their 'slackness' produced very small absolute differences between rules. For example, for system configuration (I), the differences in favour of FIFOB are 0.11 days for 'order delay'; 0.06 days for 'production delay'; 0.008 for 'tardiness of order'; 0.013% for 'orders late'; 0.07 for 'tardiness of production'; and -0.16% for 'production late'. The application of a 't' test on these results (table 6.7) failed to find any significant differences at the 0.01 level. On the other hand, the differences in favour of FIFOMB are much more significant. For example, for system configuration abcdef, the differences in favour of FIFOMB are 1.05 days for 'order delay'; 0.98 days for 'production delay'; 0.59 for 'tardiness of order'; 6.72% for 'orders late'; 0.65 for 'tardiness of order' and 5.52% for 'production late'. A 't' test applied to those differences indicated that they are all significant at the 0.01 level (see table 6.8). Identical comparisons made between SLACK and SLACKM and SPT and SPTM gave similar results. SLACKM produced 25 better results than SLACK and SPTM produced 31 better results than SPT. It should be noted from figures 6.1 to 6.24 that the effect of D on the above differences is mainly in reducing their absolute values, but basically maintaining the pattern obtained for D = 8 days. In relation to the measures of internal behaviour FIFOB, SLACK and SPT consistently produced the highest values for the 'idle time due to setup' and 'process cycle time', the only exception was for configuration (I), where all rules produced similar results. As far as the 'remaining content' is concerned, there is no clear difference between the two groups of priority rules.

A second point to be observed relates to the relative behaviour of FIFOMB, SLACKM and SPTM (similarly FIFOB, SLACK and SPT). A close examination of figures 6.1 to 6.24 show that the behaviour of FIFOMB and SLACKM are very similar for all configurations, while the behaviour of SPTM is quite distinct from the other two. The relative performances of FIFOMB and SLACKM do not appear to be influenced by the value of D (there is no substantial 'crossing' between them), but the performance of SPTM in relation to FIFOMB and SLACKM is clearly influenced by the value of D. This effect can be seen most clearly in figures 6.7, 6.15, 6.17, 6.21, 6.22 and 6.23, where the initial advantages of SPTM over the other rules are reversed as the value of D increases. This characteristic of SPT rule is in accordance with the results obtained from previous studies of priority rules in more traditional job shop and batch manufacturing systems (Conway and Maxwell, 1962; Eilon and Coterill, 1968; Oral and Malouin, 1973), which indicated that SPT rules tend to generate distributions of throughput times with high variance and skewness.

Another important point to be discussed refers to the relative efficiency of the priority rules in terms of performance. Because of the observations made before which indicated the advantage of FIFOMB over FIFOB, SLACKM over SLACK, and SPTM over SPT, the comparisons will concentrate on the relative performance of FIFOMB, SLACKM and SPTM. From the results of individual system configurations it was shown that depending on the measure of performance, the system configuration and the value of D, the relative performance of SPTM in relation to FIFOMB and SLACKM would vary markedly. In general it can be said that SPTM tends to perform better for the unweighted measures of performances 'order delay' and 'orders late', and perform particularly badly in terms of the 'weighted' measures of delivery performance 'production delay' and 'tardiness of production'. A good example of this can be seen in the cases of system configurations abc, ace and abcdef. On these three occasions, with a single exception(abcdef;'order late'), SPTM produced the lowest values among all six rules for 'order delay','tardiness of order' and 'order late' in the case of D = 8 days. However when the 'weighted' measures, 'production delay' and 'tardiness of production' are considered, it can be seen that SPTM has produced the second highest values in all cases with one exception, viz. 'production delay, abc. The performances of

FIFOMB and SLACKM are in general very similar, with FIFOMB producing lower values than SLACKM. If their results (in tables 6.1 to 6.6) are compared it can be seen that FIFOMB produced lower results than SLACKM on 33 occasions out of 36. However the differences between them are in general very small and as far as the multiple comparison tests are concerned, very few are significant. It should be observed from figures 6.1 to 6.24 that although the absolute differences between FIFOMB and SLACKM tend to decrease when D increases, there is in general no reversion of the relative performance of the two rules.

In order to have a more clear picture of the relative performance of the rules two further analysis were made on the delivery performance data. The first analysis consisted of comparing the rules in relation to their average performance over the six system configurations. In table 6.9 the average results of each rule for each of the six measures of delivery performance, for D = 8 and 20 days, are presented and each rule is ranked in accordance with the results obtained. The results confirm the observations made before.
For D equal 8 days, SPTM produced the lowest average result for the unweighted measures of performance 'order delay', 'tardiness of order' and 'orders late', while FIFOMB produced the lowest values for the 'weighted' measures of delivery performance 'production delay', 'tardiness of production' and 'production late'. In all cases SLACKM came behind FIFOMB, although very close in most cases. For D equal 20 days, the advantage of SPTM was reversed in favour of FIFOMB, which produced the lowest value for all measures of delivery performance. It should be noted that although the absolute differences between rules for D equal 20 days are very small, it helps to accentuate the point about

the SPTM (and SPT) losing its advantages over FIFOMB and SLACKM as the value of D increases, which again confirms the tendency of SPT rules to generate highly skewed distributions of throughput times.

The second analysis consisted of computing from the data of tables 6.1 to 6.6, the number of statistically significant differences that each rule had from the lowest value (control on the multiple comparison test). In order to make the analysis more explicit the total number of significant differences for each rule was broken down into two groups, corresponding respectively to the 'weighted' measures of delivery performance, and the 'unweighted' measures of delivery performance. The results of this analysis were plotted in figure 6.25, and they show that for the 'weighted' measures of delivery performance there was not a single significant difference for FIFOMB, while SPTM had 13 out of a maximum of 18 (three measures of performance vs. six system configurations), and SLACKM had only one significant difference. For the unweighted measures SPTM had 7, FIFOMB 8 and SLACKM 12 significant differences out of a maximum of 18. Figure 6.25 also shows the total number of significant differences for the total of the six measures, which is the result of the addition of the previous two histograms.

To complete the discussion about the performance of the priority rules, a few comments should be made about the overall effects of the individual priority rules on the variables of internal behaviour. To help the discussions, the results for the variables 'average number of jobs waiting in the queue'; 'average process cycle time';'average load factor'; 'machine idle time due to setup' and 'remaining content' were averaged over the six system configurations and the final results are presented in table 6.10. The first point to be observed is that there is a direct relationship between the values of 'machine idle time due to setup' and the values of the variables 'process cycle time' and 'actual load factor'. This relationship is due to the way in which those variables are defined and calculated (see paragraph 3.6.1). This relationship is confirmed by the results of table 6.10, which show that the relative rankings of all the priority rules for those three variables are the same. For example, it can be seen that SLACKM has produced the lowest values among all rules for all the three variables, while SPT produced always the highest value for the same variables. The same pattern is maintained for all the other four priority rules. It is therefore possible to concentrate the discussions on one of the above three variables, viz. 'idle time due to setup', as the observations which will be made are also valid for 'load factor' and 'process cycle time'.

The results in table 6.10 show that SLACKM, FIFOMB and SPTM, in this order have produced the lowest values for 'idle time due to setup', while SPT, SLACK and FIFOB have produced the highest values. It should also be noted from the data of appendix 4, that the ranking of individual rules in respect to the above variable, is not much influenced by the system configuration. FIFOMB and SLACKM consistently produced the two lowest values throughout the six configurations, while SPT has always produced the highest values. A second variable of internal behaviour is the 'remaining content'. From the results of table 6.10 it can be seen that SLACKM and SLACK have on average produced the lowest value for it, with FIFOMB and FIFOB producing slightly higher values than the former two, but with SPT and SPTM producing results which are considerably higher. A close look at the data in appendix 4 indicates that the relative performance of the priority rules in respect to this variable is influenced by the system configuration.

The last variable of internal behaviour to be considered is the 'average number of jobs waiting in queue'. Results from appendix 4 show very clearly that the relative ranking of the rules in relation to this variable is not influenced by the system configuration. In all six configurations tested SPTM produced the lowest values, followed closely by SPT. After them followed FIFOMB, SLACKM, FIFOB and SLACK in this order. This capacity of SPT rules for reducing the number of jobs waiting in the queue confirms results of previous research on priority rules. However while in traditional job shop or batch manufacturing systems, the number of jobs in queue can be related to the amount of work in process, this is not so in this system which has a single queue. Therefore the fact that SPTM is able to reduce the 'queue size' is not much advantage in practical terms.

In view of all reported results and analysis it is possible to conclude that FIFOMB seems to be the most appropriate of all the six priority rules, as far as this class of production system is concerned. This conclusion is even more strong if one considers that the weighted measures of delivery performance are more relevant than the unweighted measures, as they take into consideration not only the number of orders delivered, but also their intrinsic value, which is represented by their batch size. Finally it should be pointed out that from the practical point of view FIFOMB has the advantage of being much easier to operate than both SLACKM and SPTM. This can be a significant aspect in the case of real production systems, particularly those which do not have a sophisticated production control department. The FIFOMB was therefore selected for the next two series of investigations.

A final comment should be made about the relationship between the results obtained in this study and the results obtained in other studies of priority rules, and which have been reported in the literature. The first point to be stressed is that comparisons with results from other studies will be difficult because of the large dissimilarities between the characteristics of this class of production system and the characteristics of the models used by other authors in analysing priority rules. Among the most striking dissimilarities are the facts that most authors analysed priority rules for production systems in which machines can only process one job at a time, jobs never have to wait for tools, and job arrival and delivery are independent of other jobs. Most of the models (Eilon and Hodgson, 1967, is one exception) also considered that jobs require multiple operations which are executed on different machines. Another dissimilarity between models relates to the way in which due dates are fixed. Most of the authors who considered due dates in their models used a method for fixing due dates which is based on the work content or the number of operations required for each job, while in this study due date is based on a fixed 'lead time', which (within certain limits) is independent of the characteristics of the jobs.

In spite of this large number of differences, some comments can still be made. Firstly, when the results obtained in this study for SPT (SPTM) rules are compared with the results reported by other authors, it is possible to note some similarities, as far as the 'unweighted' measures of delivery performance are concerned. For example, Eilon and Coterill (1968) have found that for a classical hypothetical job shop, without setup times, and due dates fixed as a function of the jobs' work content, the SPT rule (which they call SI) performs well in reducing the mean throughput time of jobs, but falls behind with respect to variance of throughput times. It also performs well in terms of average delay but not so well in terms of variance of missed due dates. They also found that the performance of priority rules becomes increasingly pronounced as the load ratio increases. Eilon and Hodgson (1967) also reported the results for the SPT rule in a single operation system, consisting of two identical machines, which did not require setting up. They concluded that the SPT rule appeared to minimize among other variables, the mean throughput times and the average queue length, but as expected, it had a clear effect in delaying jobs with long estimated processing times. Other authors like Elvers (1973), Oral and Malouin (1973), Wilbrecht and Prescott (1969), Hollier (1968), have also analysed the SPT or SPT based rules for typical batch or job shop systems, with different characteristics, and they all found that in general the SPT based rules tend to perform well in respect to mean individual delay and number of jobs late, but tend to perform badly in terms of tardiness based criteria. It should be pointed out that none of the above authors seem to have used any 'weighted' measure of delivery performance which take account of the order's batch size. In this study it was shown that the advantages of SPTM (and SPT) rules in reducing the average delay of order (mean throughput times) can be reversed in favour of other rules (FIFOMB, SLACKM) if individual jobs (orders) are weighted by their batch sizes.

One conclusion of this study which does not agree with other studies (one exception is Wilbrecht and Prescott (1969)) is the clear advantage of the rules which are designed to avoid setup time, over the rules which do not try to avoid setup times. The disagreement is understandable when one considers the characteristic of the machines used in this study with their multiple stations, which means that each setup executed in one station affect all the other jobs on the other stations, giving a greater dimension to the importance of setup times.

Another point of disagreement with other studies relates to the results obtained by SPT and SPTM in relation to the variable 'remaining content'. Hollier (1968), for example, used a measure called 'remaining work content', and found that SPT based rules tend to minimize this variable. In this study it was found that SPT and SPTM did particularly badly in relation to the variable 'remaining content'. The reason for the differences in results might be due to the way in which 'remaining content' is calculated in this study as compared with the other study. While in that study 'remaining work content' is calculated as the amount of work left in the shop for processing, in this study, 'remaining content' means the sum of the work still to be processed plus the work ('jobs') already completed but unable to be delivered because of the state of incompleteness of the 'order' of which each 'job' is a part.

Finally, it is stressed again that it is difficult to make meaningful comparisons because of the great dissimilarities between the systems being compared.

6.4 - Summary

In this chapter six priority rules, viz. FIFOB, FIFOMB, SLACK, SLACKM, SPT and SPTM were compared primarily in relation to their ability to improve the delivery performance and also in relation to their effect over some of the system's internal variables.

The measures of delivery performance used included average delivery delay, percentage of orders delivered late, and tardiness indexes. The values of those three variables were calculated for both weighted, and unweighted criteria, where the weight was given by the batch size of each order. The measures of tardiness and percentage late which are dependent on the due date, were calculated for different values of due dates, which were obtained by varying the value of the lead time used to determine the due dates.

An experimental design was used such that each rule was tested over six system configurations, obtained by the joint variation of six parameters of the system. These system configurations represent a sample of the total 64 system configurations which would be obtained by a full factorial design.

The results of the 36 experiments indicated among other things the advantages of the priority rules designed to avoid setup times (FIFOMB, SLACKM and SPTM) over the rules which do not avoid setup times (FIFOB, SLACK, SPT). The results also indicated that SPTM rule tends to produce better results for the unweighted measures of delivery performance, 'average delivery delay of orders', and 'percentage of orders delivered late', and for tight due-dates. However it tends to lose its advantages over the other rules (mainly FIFOMB and SLACKM) when the due date is more 'loose'. In the case of the weighted measures of delivery performance, SPTM (and SPT) tends to perform particularly badly, while FIFOMB tends to perform well.

The variables of internal behaviour used included the 'percentage of time spent with setting up', the 'actual load factor', the 'process cycle time', the 'remaining content , and the 'average queue size'. In general it was found that the rules FIFOMB, SLACKM and SPTM tend to produce lower results for the 'percentage of time spent in setting up', the 'process cycle time', and the 'actual load factor', than the other three rules (FIFOB, SLACK, SPT). SPTM and SPT tend to produce lower values for the queue size, but higher values for the 'remaining content', than the other four rules.

Finally a decision was made to select the FIFOMB rule as the most appropriate for this class of production system.

TABLE 6.1
SYSTEM CONFIGURATION (I)

MEAN VALUES - D = 8 DAYS

PRIORITY RULE	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
FIFOB	(1) 1.827	(1) 2.942	(1) 0.013	0.833	0.035	1.978
FIFOMB	* 1.940	2.998	0.022	0.962	0.042	1.815
SLACK	** 1.998	2.990	(1) 0.013	(1) 0.768	(1) 0.022	(1) 1.117
SLACKM	** 2.028	3.002	0.023	1.025	0.032	1.402
SPT	** 1.952	* 3.575	** 0.072	** 2.178	** 0.252	** 7.098
SPTM	* 1.988	** 3.518	** 0.082	** 2.243	** 0.228	** 5.822

RESULTS OF 'F TEST'

FS	** 5.00	** 29.16	** 4.81	** 5.83	** 5.92	** 12.61
F	3.85	3.85	3.85	3.85	3.85	3.85

MULTIPLE COMPARISON TEST - (CRITICAL VALUE FOR DIFFERENCES)

$D^{*}_{0.05}$	0.11	0.18	0.05	0.95	0.15	2.42
$D^{**}_{0.01}$	0.14	0.24	0.06	1.25	0.19	3.18

Convention: (1) - Smallest value for the measure of performance (control)
* - Significant at 0.05 level
** - Significant at 0.01 level
FS - 'F ratio' calculated from data
F - Critical value for 'F ratio'
$D^{*}_{0.05}$ - Critical value for differences between each contrast (priority rules) and the 'control' at 0.05 level
$D^{**}_{0.01}$ - Critical value for difference between each contrast (priority rules) and the 'control' at 0.01 level

TABLE 6.2
SYSTEM CONFIGURATION abc

MEAN VALUES – D = 8 DAYS

PRIORITY RULE	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
FIFOB	** 7.307	8.535	** 2.193	** 33.65	2.772	* 40.00
FIFOMB	** 7.028	(1) 8.287	** 2.025	** 31.98	(1) 2.598	38.83
SLACK	** 7.857	8.810	** 2.552	** 37.95	2.930	** 42.42
SLACKM	** 7.523	8.505	** 2.335	** 35.83	2.715	* 40.46
SPT	5.802	** 9.292	1.392	21.21	** 3.713	37.24
SPTM	(1) 5.407	8.740	(1) 1.245	(1) 18.59	* 3.385	(1) 34.09
RESULTS OF 'F TEST'						
FS	** 30.05	2.34	** 8.76	** 37.80	** 4.77	* 3.72
F	3.85	2.09	3.85	3.85	3.85	2.60
MULTIPLE COMPARISON TEST – (CRITICAL VALUE FOR DIFFERENCES)						
$D^{*}_{0.05}$	0.60	0.75	0.59	4.84	0.67	4.86
$D^{**}_{0.01}$	0.79	0.99	0.78	5.74	0.88	6.44

Convention: (1) – Smallest value for the measure of performance (control)
* – Significant at 0.05 level
** – Significant at 0.01 level
FS – 'F ratio' calculated from data
F – Critical value for 'F ratio'
$D^{*}_{0.05}$ – Critical value for differences between each contrast (priority rules) and the 'control' at 0.05 level
$D^{**}_{0.01}$ – Critical value for difference between each contrast (priority rules) and the 'control' at 0.01 level

TABLE 6.3

SYSTEM CONFIGURATION def

MEAN VALUES - D = 8 DAYS

PRIORITY RULE	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
FIFOB	(1) 3.272	(1) 4.898	(1) 0.148	(1) 5.90	(1) 0.318	(1) 12.01
FIFOMB	3.298	4.953	0.192	6.73	0.390	12.65
SLACK	** 3.628	* 5.068	* 0.232	8.07	0.348	12.30
SLACKM	* 3.557	* 5.072	** 0.272	8.14	0.430	12.89
SPT	* 3.572	** 6.407	** 0.527	* 9.48	** 1.477	** 22.72
SPTM	* 3.582	** 5.798	** 0.385	8.46	** 0.633	** 19.04
RESULTS OF 'F TEST'						
FS	** 4.01	** 128.14	** 24.64	2.14	** 53.62	** 10.29
F	3.85	3.85	3.85	2.09	3.85	3.85
MULTIPLE COMPARISON TEST - (CRITICAL VALUE FOR DIFFERENCES)						
$D^{*}_{0.05}$	0.26	0.18	0.09	2.91	0.21	4.67
$D^{**}_{0.01}$	0.34	0.24	0.12	3.86	0.28	6.18

Convention: (1) - Smallest value for the measure of performance (control)

* - Significant at 0.05 level

** - Significant at 0.01 level

FS - 'F ratio' calculated from data

F - Critical value for 'F ratio'

$D^{*}_{0.05}$ - Critical value for differences between each contrast (priority rules) and the 'control' at 0.05 level

$D^{**}_{0.01}$ - Critical value for difference between each contrast (priority rules) and the 'control' at 0.01 level

TABLE 6.4
SYSTEM CONFIGURATION ace

MEAN VALUES - D = 8 DAYS

PRIORITY RULE	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
FIFOB	** 6.812	8.053	** 1.887	** 30.89	2.432	36.52
FIFOMB	** 6.708	(1) 7.970	* 1.853	** 30.32	(1) 2.398	36.33
SLACK	** 7.273	8.158	** 2.195	** 33.72	2.500	37.09
SLACKM	** 7.182	8.102	** 2.163	** 33.65	2.468	37.09
SPT	5.465	8.462	1.167	20.44	* 2.973	36.09
SPTM	(1) 5.238	8.248	(1) 1.127	(1) 19.03	* 2.947	(1) 34.81

RESULTS OF 'F TEST'

FS	** 29.78	1.03	** 7.73	** 44.92	* 2.85	0.57
F	3.85	2.09	3.85	3.85	2.60	2.09

MULTIPLE COMPARISON TEST - (CRITICAL VALUE FOR DIFFERENCES)

$D^{*}_{0.05}$	0.53	0.57	0.57	3.26	0.52	3.71
$D^{**}_{0.01}$	0.71	0.73	0.75	4.31	0.69	4.91

Convention: (1) - Smallest value for the measure of performance (control)
* - Significant at 0.05 level
** - Significant at 0.01 level
FS - 'F ratio' calculated from data
F - Critical value for 'F ratio'
$D^{*}_{0.05}$ - Critical value for differences between each contrast (priority rules) and the 'control' at 0.05 level
$D^{**}_{0.01}$ - Critical value for difference between each contrast (priority rules) and the 'control' at 0.01 level

TABLE 6.5

SYSTEM CONFIGURATION bdf

MEAN VALUES - D = 8 DAYS

PRIORITY RULE	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
FIFOB	** 4.085	5.570	0.323	** 11.22	0.562	** 18.22
FIFOMB	(1) 3.723	(1) 5.253	(1) 0.262	(1) 8.01	(1) 0.490	14.24
SLACK	** 4.503	** 5.843	* 0.438	** 13.91	0.633	** 19.47
SLACKM	* 3.935	5.353	0.308	8.91	0.513	(1) 14.01
SPT	** 4.875	** 8.712	** 1.182	** 15.64	** 3.275	** 34.10
SPTM	* 3.900	** 6.422	** 0.563	* 10.35	** 1.298	** 21.90

RESULTS OF 'F TEST'

FS	** 64.69	** 163.36	** 41.77	** 18.07	** 113.15	** 50.71
F	3.85	3.85	3.85	3.85	3.85	3.85

MULTIPLE COMPARISON TEST - (CRITICAL VALUE FOR DIFFERENCES)

$D^{*}_{0.05}$	0.18	0.34	0.18	2.30	0.34	3.46
$D^{**}_{0.01}$	0.24	0.45	0.23	3.04	0.45	4.57

Convention: (1) - Smallest value for the measure of performance (control)

* - Significant at 0.05 level

** - Significant at 0.01 level

FS - 'F ratio' calculated from data

F - Critical value for 'F ratio'

$D^{*}_{0.05}$ - Critical value for differences between each contrast (priority rules) and the 'control' at 0.05 level

$D^{**}_{0.01}$ - Critical value for difference between each contrast (priority rules) and the 'control' at 0.01 level

TABLE 6.6
SYSTEM CONFIGURATION abcdef

MEAN VALUES - D = 8 DAYS

PRIORITY RULE	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
FIFOB	** 7.795	* 9.412	* 2.345	** 38.65	3.072	** 47.44
FIFOMB	6.748	(1) 8.433	1.753	** 31.98	(1) 2.418	(1) 41.91
SLACK	** 8.997	** 10.250	** 3.222	** 47.56	** 3.712	** 54.23
SLACKM	* 7.402	8.805	2.175	** 36.54	2.662	44.65
SPT	7.118	** 13.203	1.820	(1) 23.33	** 7.005	** 48.44
SPTM	(1) 6.620	** 10.875	(1) 1.745	23.84	** 4.792	44.55

RESULTS OF 'F TEST'

FS	** 23.50	** 35.78	** 12.43	** 43.58	** 37.40	** 15.10
F	3.85	3.85	3.85	3.85	3.85	3.85

MULTIPLE COMPARISON TEST - (CRITICAL VALUE FOR DIFFERENCES)

$D^{*}_{0.05}$	0.60	0.97	0.54	4.68	0.94	3.67
$D^{**}_{0.01}$	0.79	1.28	0.71	6.19	1.24	4.85

Convention: (1) - Smallest value for the measure of performance (control)
* - Significant at 0.05 level
** - Significant at 0.01 level
FS - 'F ratio' calculated from data
F - Critical value for 'F ratio'
$D^{*}_{0.05}$ - Critical value for differences between each contrast (priority rules) and the 'control' at 0.05 level
$D^{**}_{0.01}$ - Critical value for difference between each contrast (priority rules) and the 'control' at 0.01 level

TABLE 6.7

SYSTEM CONFIGURATION (I)

't' TEST ON THE DIFFERENCES BETWEEN FIFOMB AND FIFOB

VALUES OF DIFFERENCES (SIX SAMPLES)

	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
	0.03	-0.06	0.00	0.00	0.00	0.00
	0.16	0.11	0.00	0.00	0.00	0.00
	0.03	-0.01	0.00	0.00	0.00	0.02
	0.14	0.10	0.00	0.38	0.01	0.76
	0.07	0.01	0.00	0.00	0.00	0.00
	0.25	0.19	0.05	0.39	0.03	- 1.76
$\bar{X}$	0.110	0.057	0.008	0.128	0.07	- 0.163
S	0.086	0.092	0.020	0.199	0.012	0.839
Limit	0.118	0.126	0.027	0.237	0.016	1.151

* Significant at 0.05 level

** Significant at 0.01 level - $t_{0.01,5} = 3.365$

TABLE 6.8

SYSTEM CONFIGURATION abcdef

't' TEST ON THE DIFFERENCES BETWEEN FIFOB AND FIFOMB

VALUES OF DIFFERENCES (SIX SAMPLES)

	ORDER DELAY	PRODUCTION DELAY	TARDINESS OF ORDER	ORDERS LATE	TARDINESS OF PRODUCTION	PRODUCTION LATE
	2.22	2.03	1.35	15.38	1.51	12.27
	1.26	1.02	0.97	7.69	0.84	7.30
	0.70	0.83	0.30	5.30	0.51	4.06
	0.74	0.67	0.30	5.77	0.29	6.06
	0.85	0.75	0.48	4.23	0.51	2.61
	0.51	0.57	0.16	1.92	0.26	0.84
$\overline{X}$	**1.05	**0.98	**0.59	** 6.72	**0.65	** 5.22
S	0.627	0.54	0.37	4.65	0.47	4.04
Limit	0.86	0.74	0.51	6.39	0.65	5.50

* Significant at 0.05 level

** Significant at 0.01 level - $t_{0.01,5} = 3.365$

TABLE 6.9

RESULTS OF THE AVERAGE DELIVERY PERFORMANCE OF PRIORITY RULES OVER THE SIX SYSTEM CONFIGURATIONS

D = 8 Days

PRIORITY RULE	ORDER DELAY (DAYS)	PRODUCTION DELAY (DAYS)	TARDINESS OF ORDER	ORDERS LATE (%)	TARDINESS OF PRODUCTION	PRODUCTION LATE (%)
FIFOB	(4) 5.183	(3) 6.568	(4) 1.152	(4) 20.19	(3) 1.532	(3) 26.03
FIFOMB	(3) 4.908	(1) 6.316	(2) 1.018	(3) 18.33	(1) 1.389	(1) 24.29
SLACK	(6) 5.709	(4) 6.853	(6) 1.442	(6) 23.65	(4) 1.691	(5) 27.72
SLACKM	(5) 5.271	(2) 6.473	(5) 1.213	(5) 20.68	(2) 1.470	(2) 25.08
SPT	(2) 4.797	(6) 8.275	(3) 1.026	(2) 15.38	(6) 3.116	(6) 30.95
SPTM	(1) 4.456	(5) 7.267	(1) 0.828	(1) 13.74	(5) 2.267	(4) 26.70

D = 20 Days

PRIORITY RULE	ORDER DELAY (DAYS)	PRODUCTION DELAY (DAYS)	TARDINESS OF ORDER	ORDERS LATE (%)	TARDINESS OF PRODUCTION	PRODUCTION LATE (%)
FIFOB	SAME AS FOR 8 DAYS	SAME AS FOR 8 DAYS	(2) 0.112	(2) 1.95	(3) 0.169	(3) 2.85
FIFOMB			(1) 0.101	(1) 1.76	(1) 0.146	(1) 2.49
SLACK			(5) 0.151	(5) 2.76	(4) 0.189	(4) 3.33
SLACKM			(4) 0.122	(3) 2.06	(2) 0.153	(2) 2.54
SPT			(6) 0.178	(6) 3.16	(6) 1.019	(6) 8.27
SPTM			(3) 0.115	(4) 2.26	(5) 0.571	(5) 5.92

TABLE 6.10

RESULTS OF AVERAGE VALUES OF VARIABLE OF INTERNAL BEHAVIOUR FOR SIX SYSTEM CONFIGURATIONS

MEAN VALUES

PRIORITY RULE	NO. OF JOBS IN QUEUE	PROCESS CYCLE TIME	ACTUAL LOAD FACT.	TOTAL SETUP %	REMAINING CONTENT
FIFOB	45.42	5.33	76.17 %	7.18	11,037
FIFOMB	41.36	5.10	72.96	5.71	11,076
SLACK	48.35	5.34	76.15	7.27	10,838
SLACKM	43.35	5.08	72.71	5.66	10,831
SPT	35.07	5.50	78.86	7.75	12,163
SPTM	29.72	5.27	75.67	5.87	12,129

FIGURE 6.1

SYSTEM CONFIGURATION (I)

PERCENTAGE OF LATE ORDERS

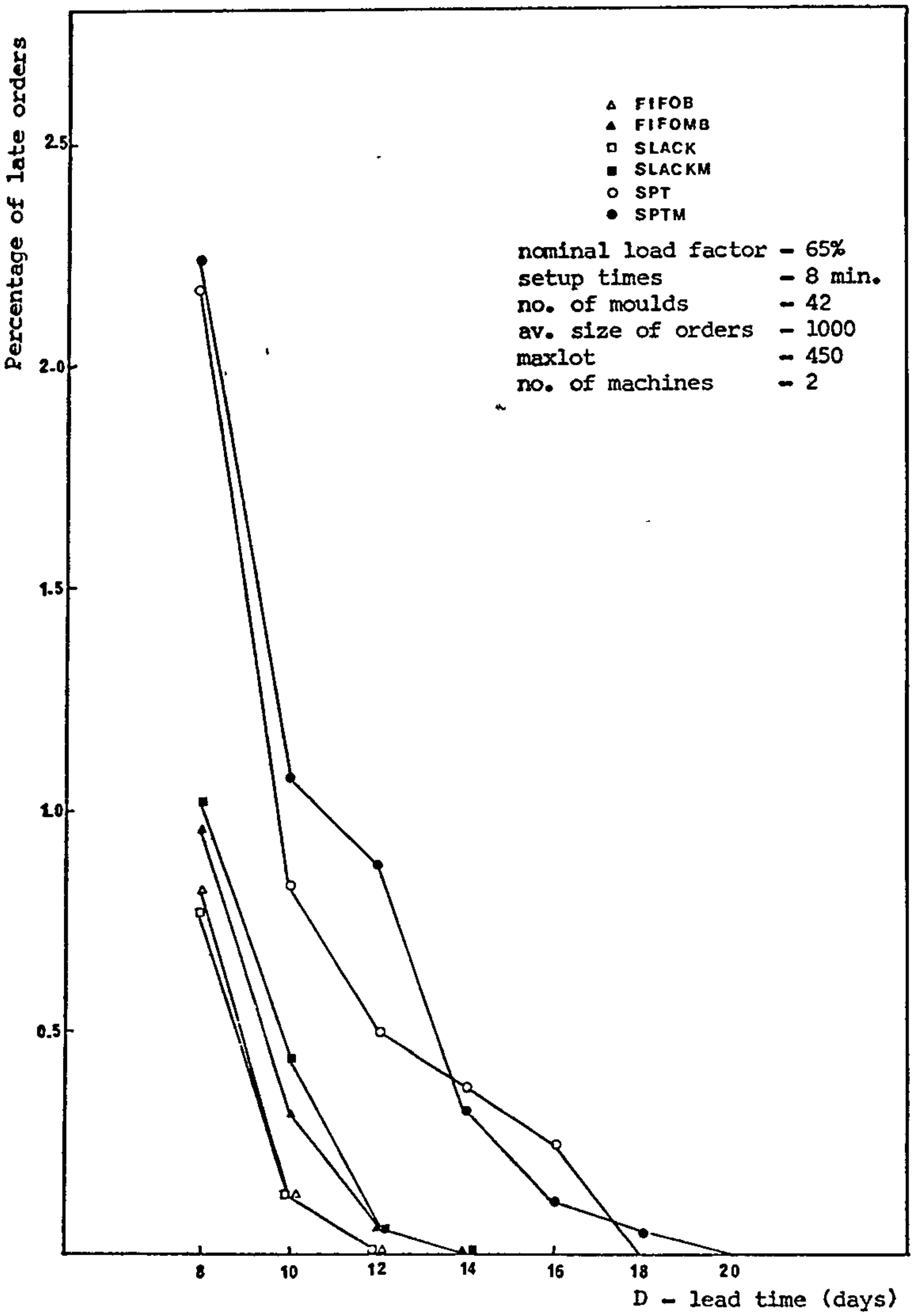


FIGURE 6.2

SYSTEM CONFIGURATION (I)

TARDINESS INDEX OF ORDERS

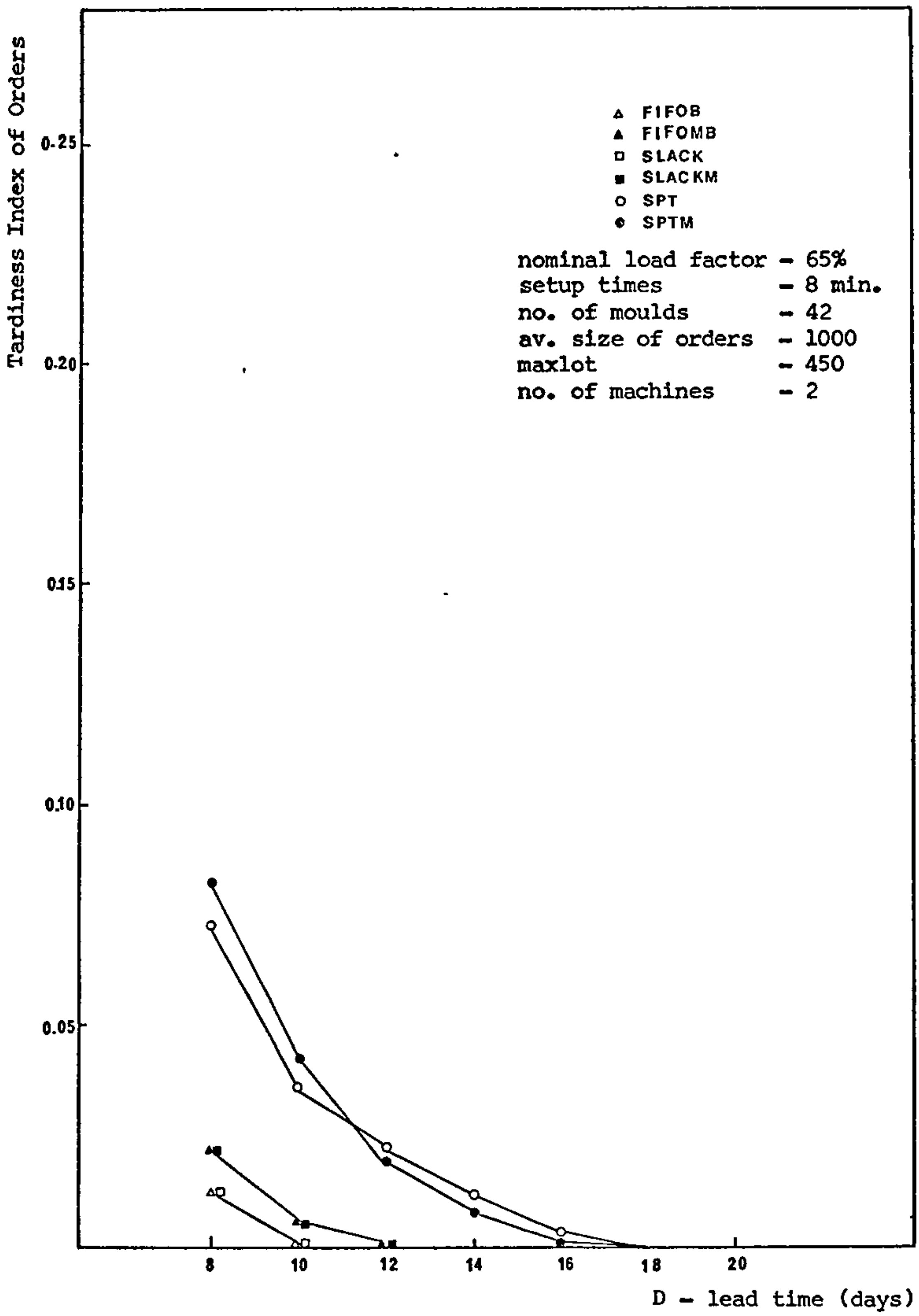


FIGURE 6.3

SYSTEM CONFIGURATION (I)

PERCENTAGE OF PRODUCTION DELIVERED LATE

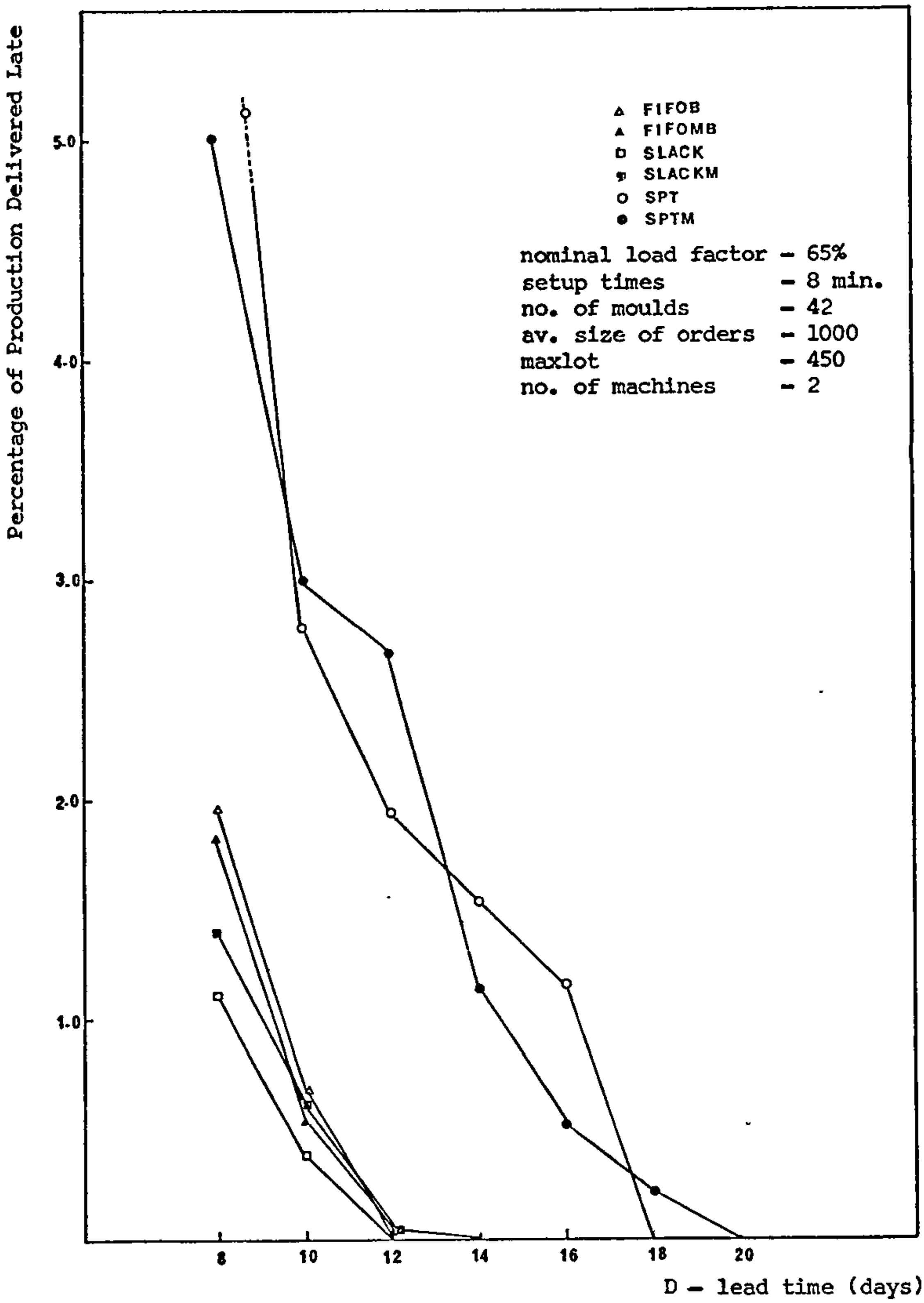


FIGURE 6.4

SYSTEM CONFIGURATION (I)

TARDINESS INDEX OF PRODUCTION

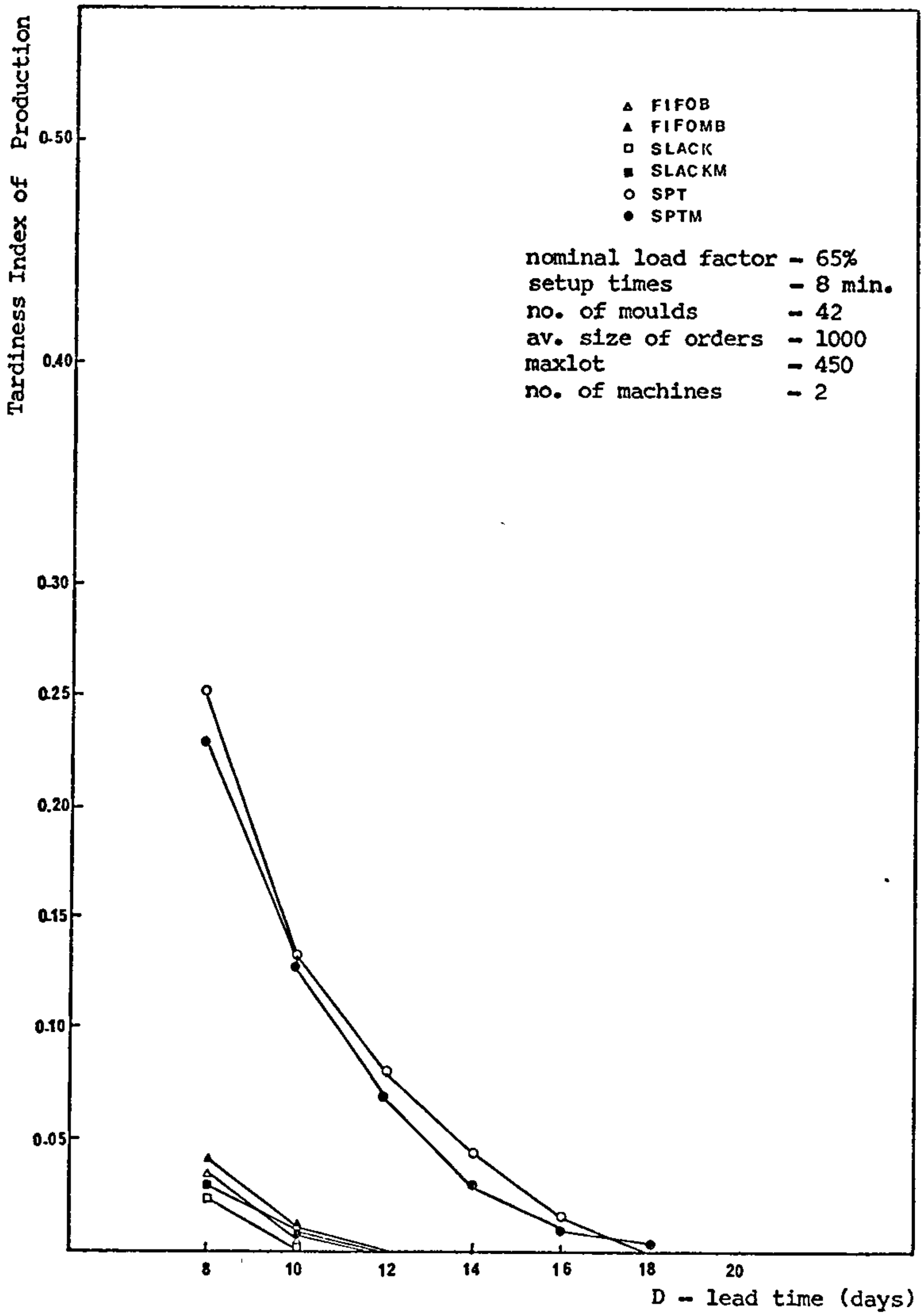


FIGURE 6.5

SYSTEM CONFIGURATION abc

PERCENTAGE OF LATE ORDERS

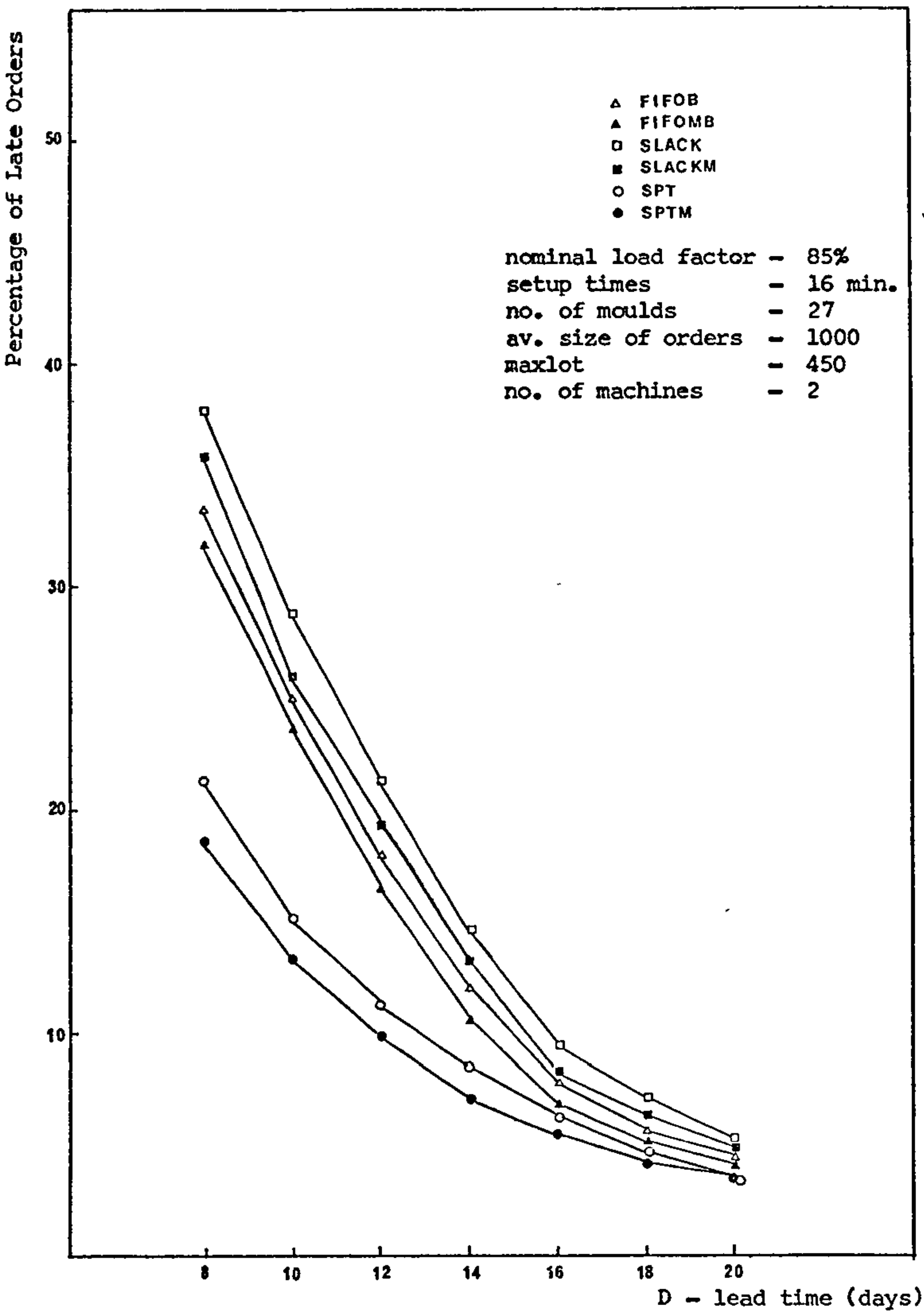


FIGURE 6.6

SYSTEM CONFIGURATION abc

TARDINESS INDEX OF ORDERS

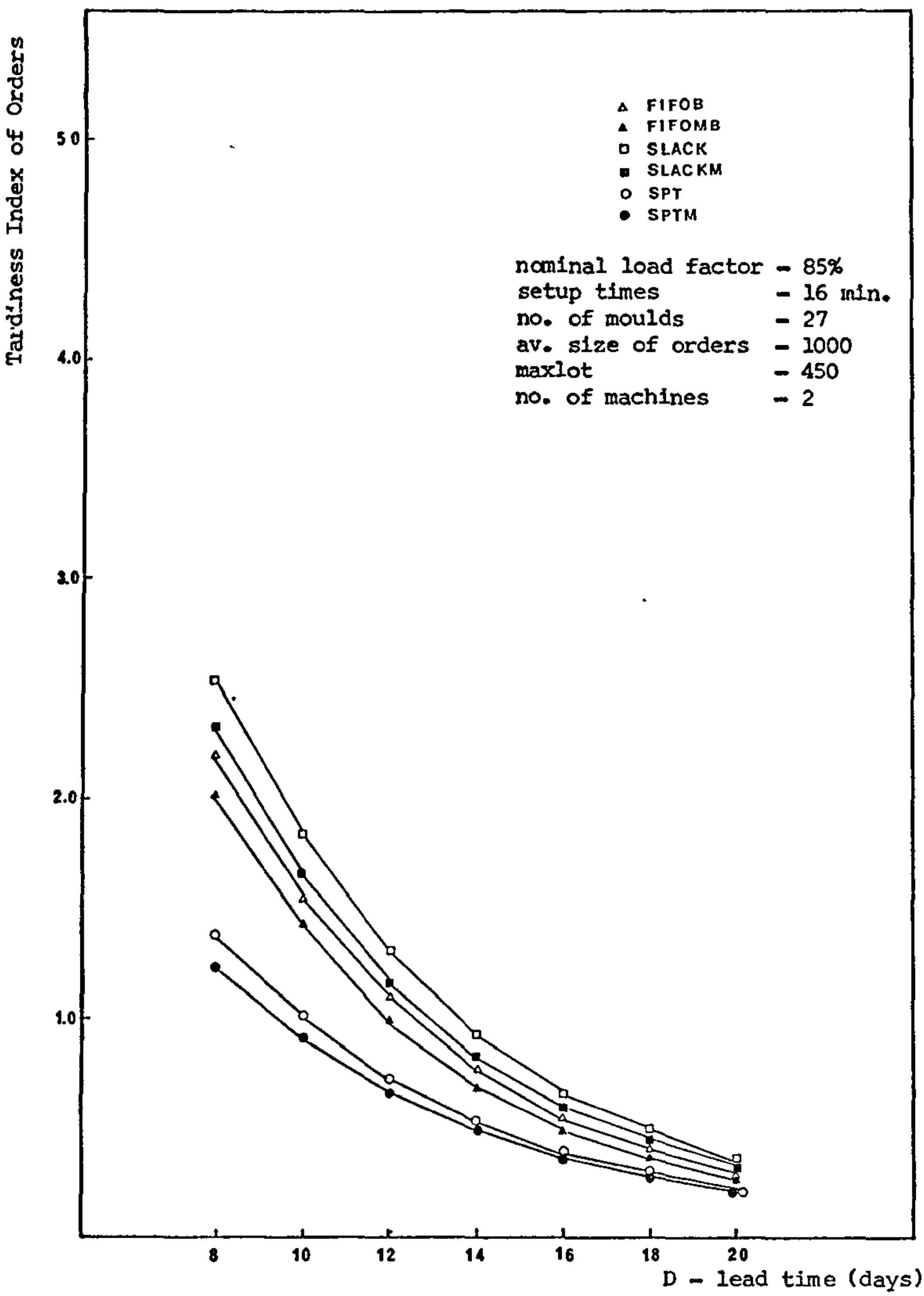


FIGURE 6.7

SYSTEM CONFIGURATION abc

PERCENTAGE OF PRODUCTION DELIVERED LATE

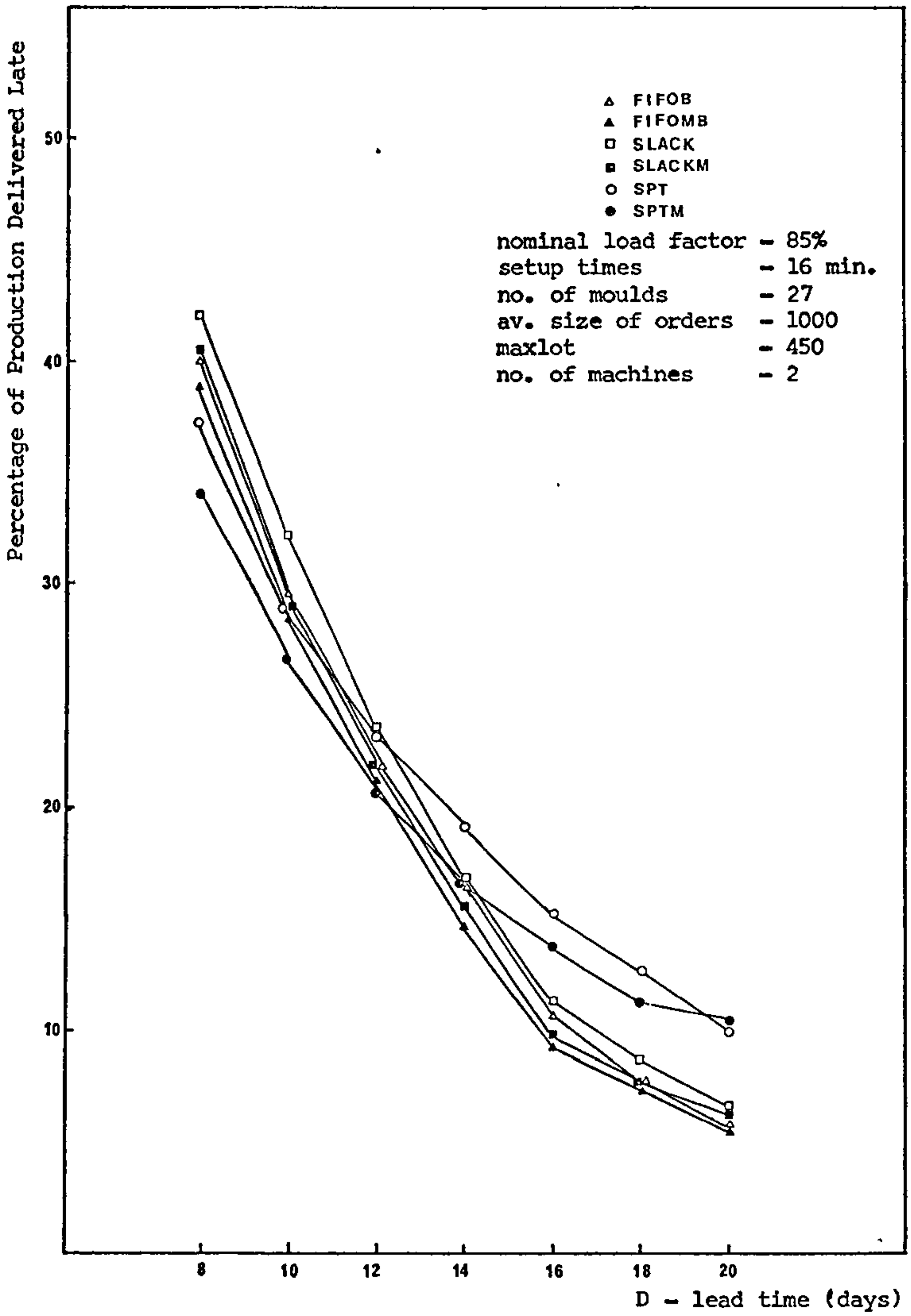


FIGURE 6.8

SYSTEM CONFIGURATION abc

TARDINESS INDEX OF PRODUCTION

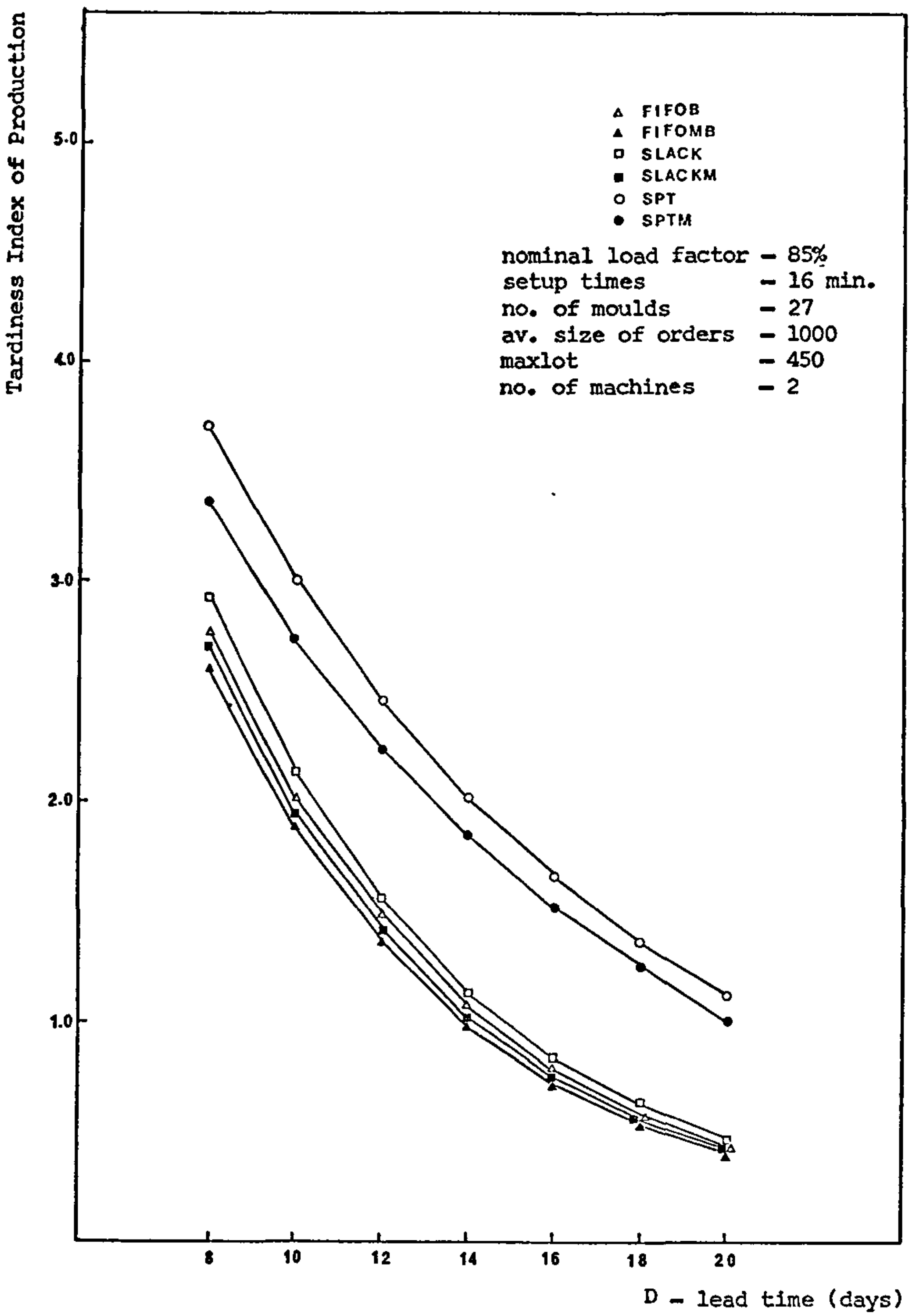


FIGURE 6.9

SYSTEM CONFIGURATION def

PERCENTAGE OF LATE ORDERS

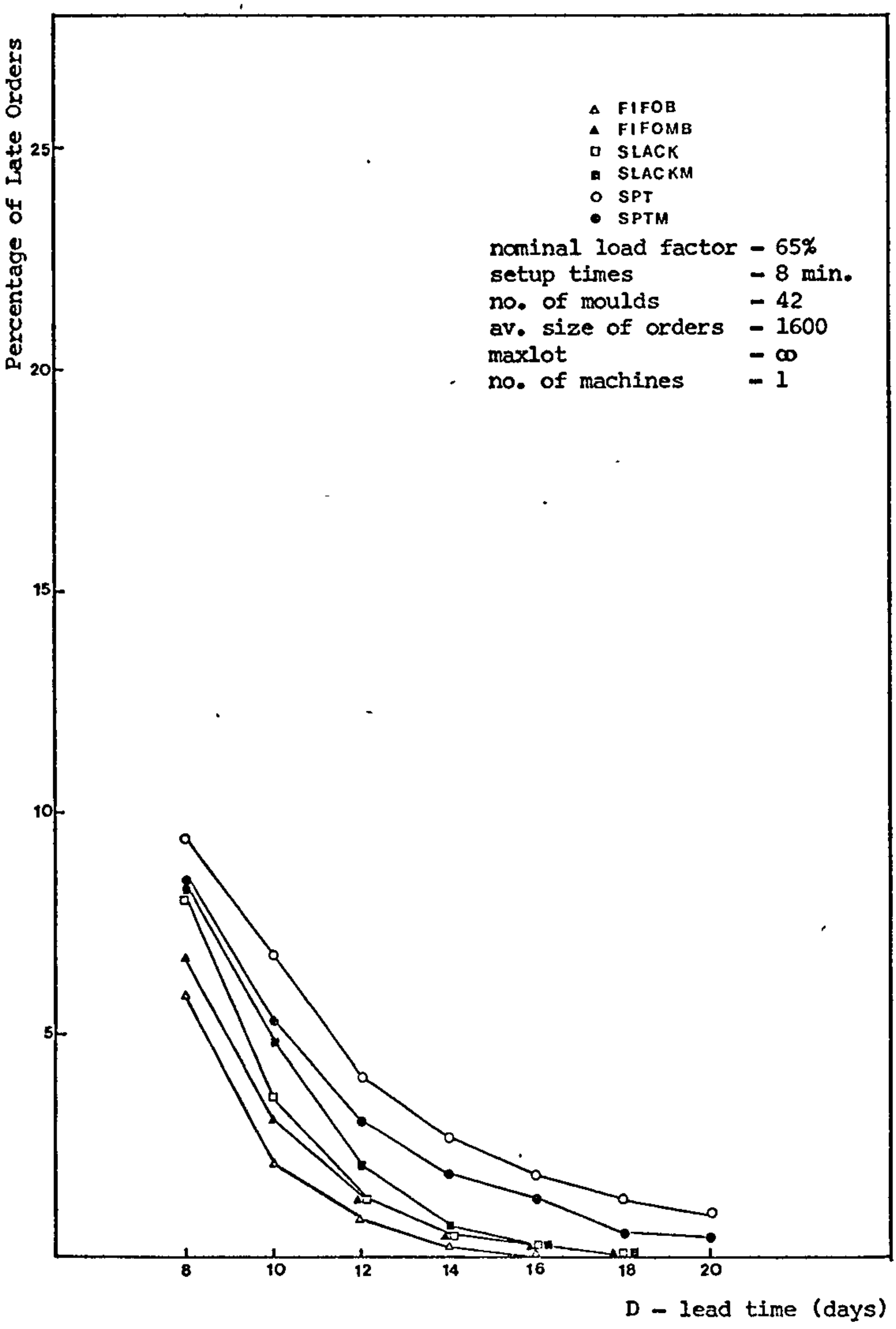


FIGURE 6.10

SYSTEM CONFIGURATION def

TARDINESS INDEX OF ORDERS

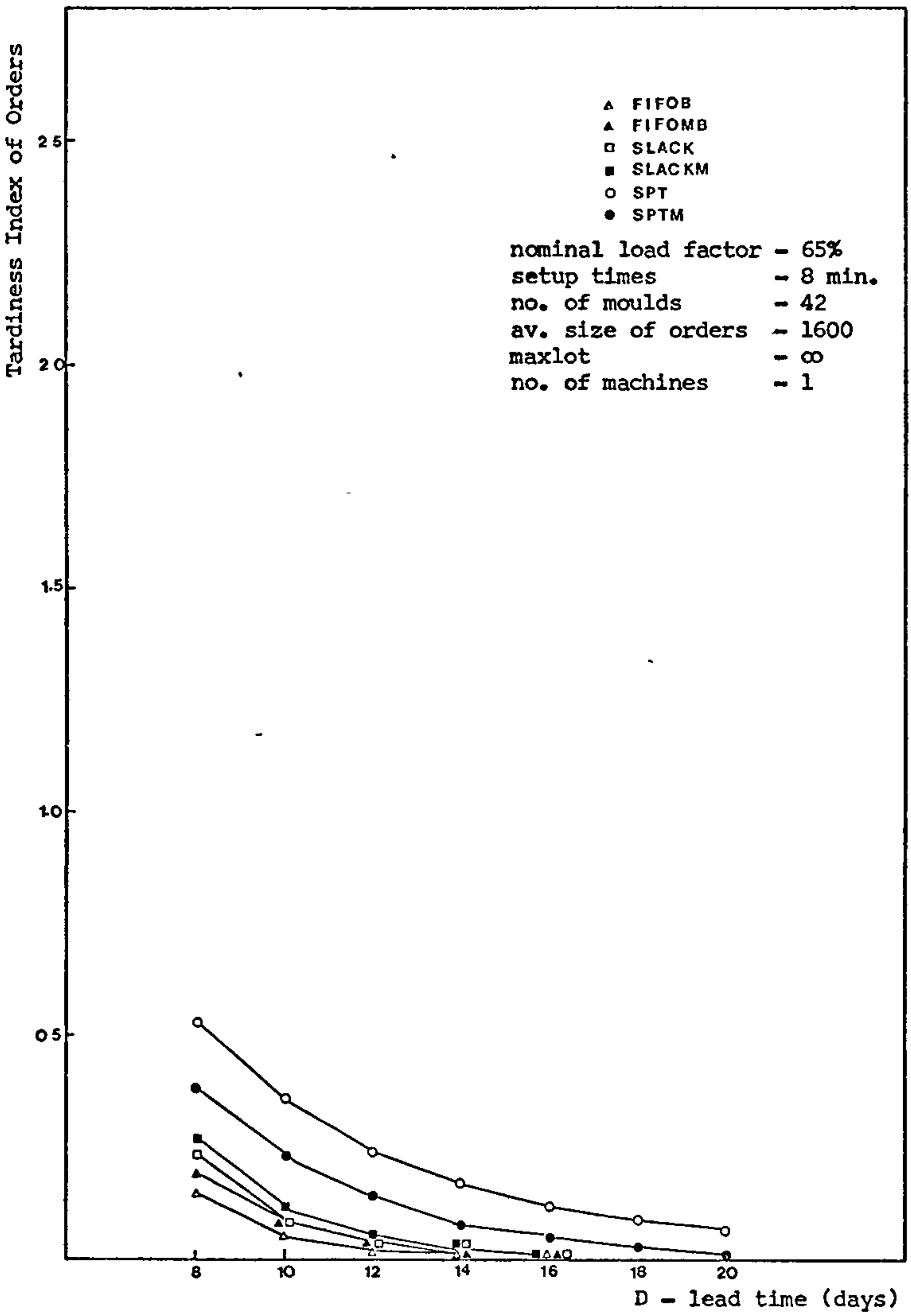


FIGURE 6.11

SYSTEM CONFIGURATION def

PERCENTAGE OF PRODUCTION DELIVERED LATE

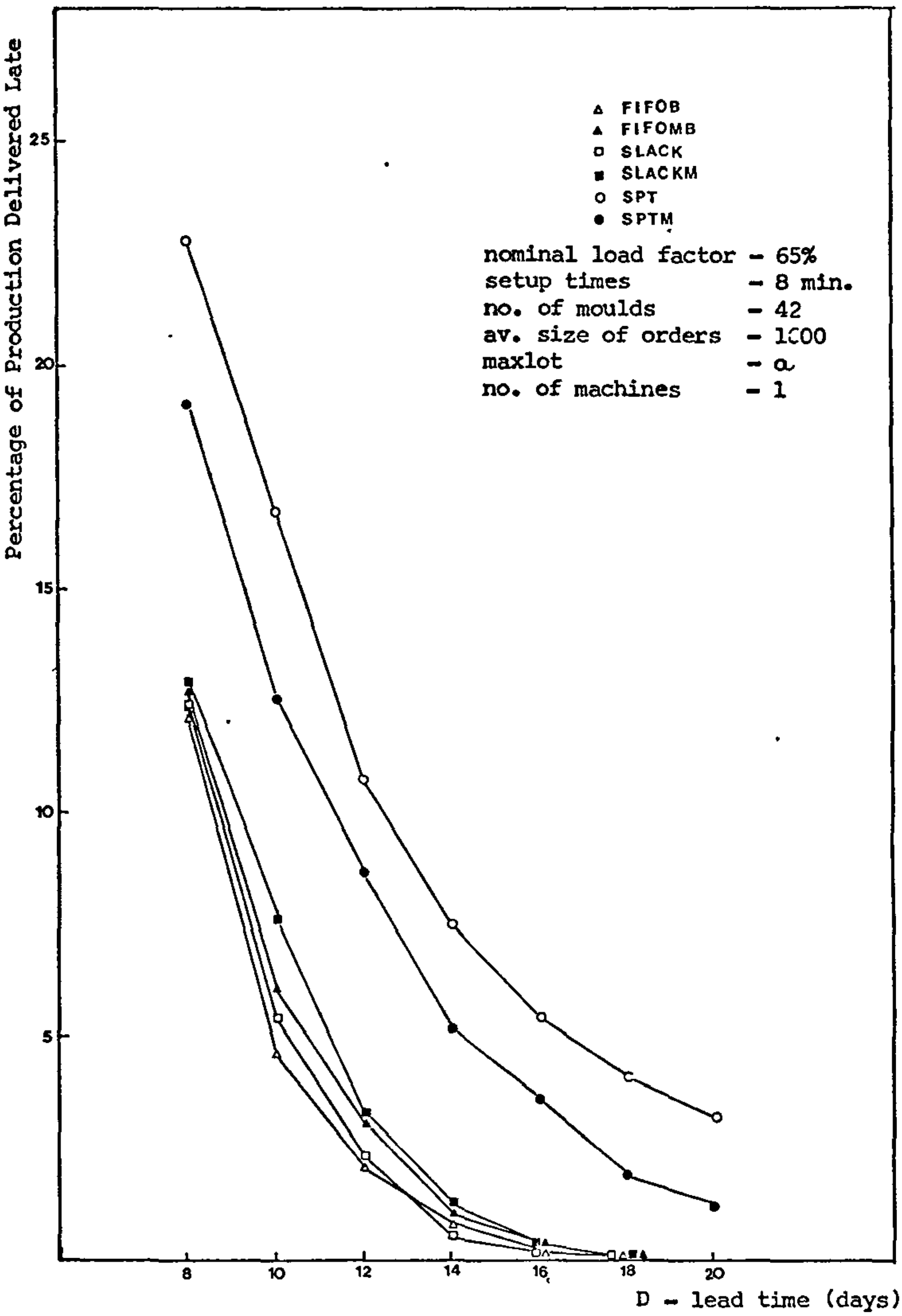


FIGURE 6.12

SYSTEM CONFIGURATION def

TARDINESS INDEX OF PRODUCTION

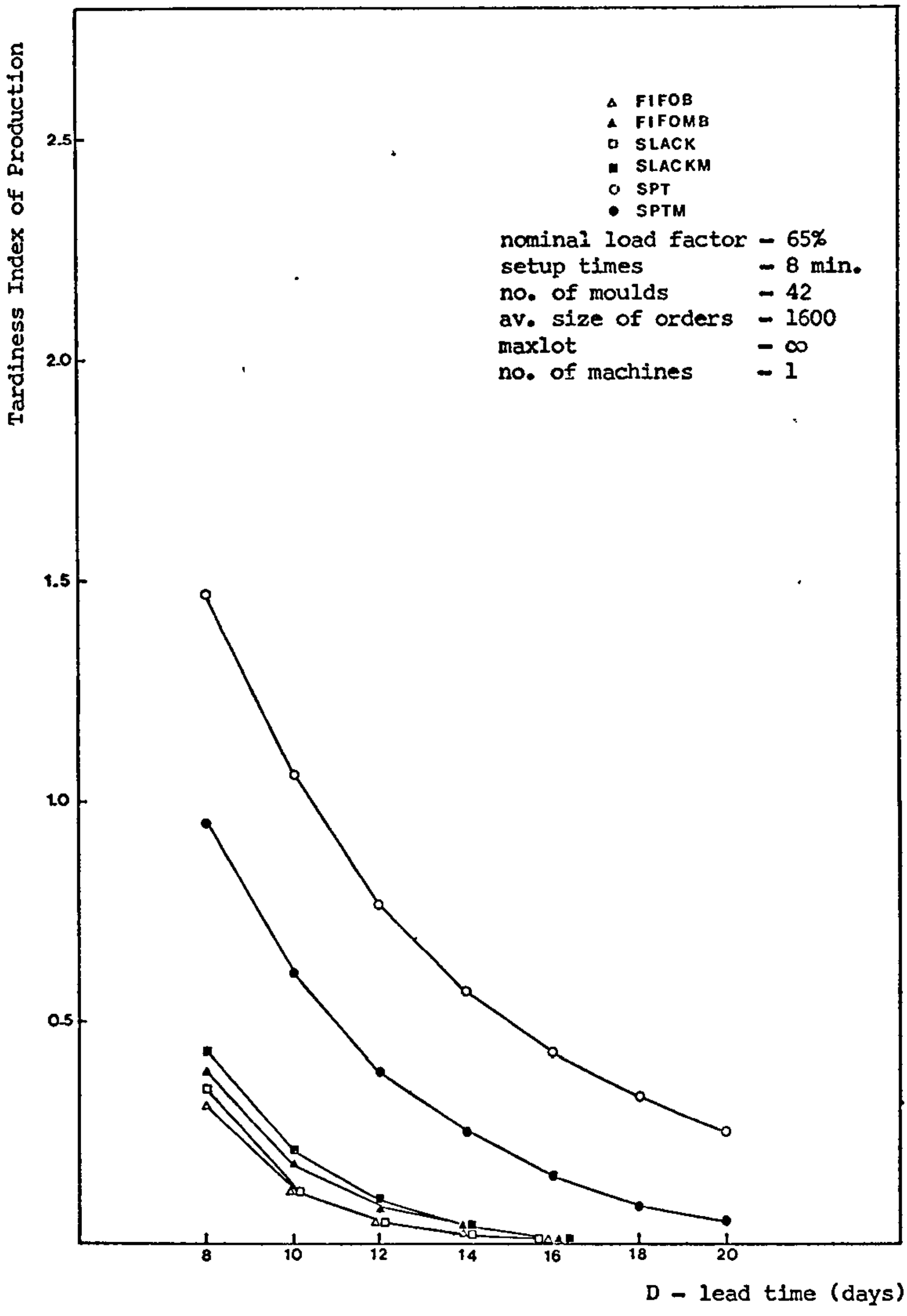


FIGURE 6.13

SYSTEM CONFIGURATION ace

PERCENTAGE OF LATE ORDERS

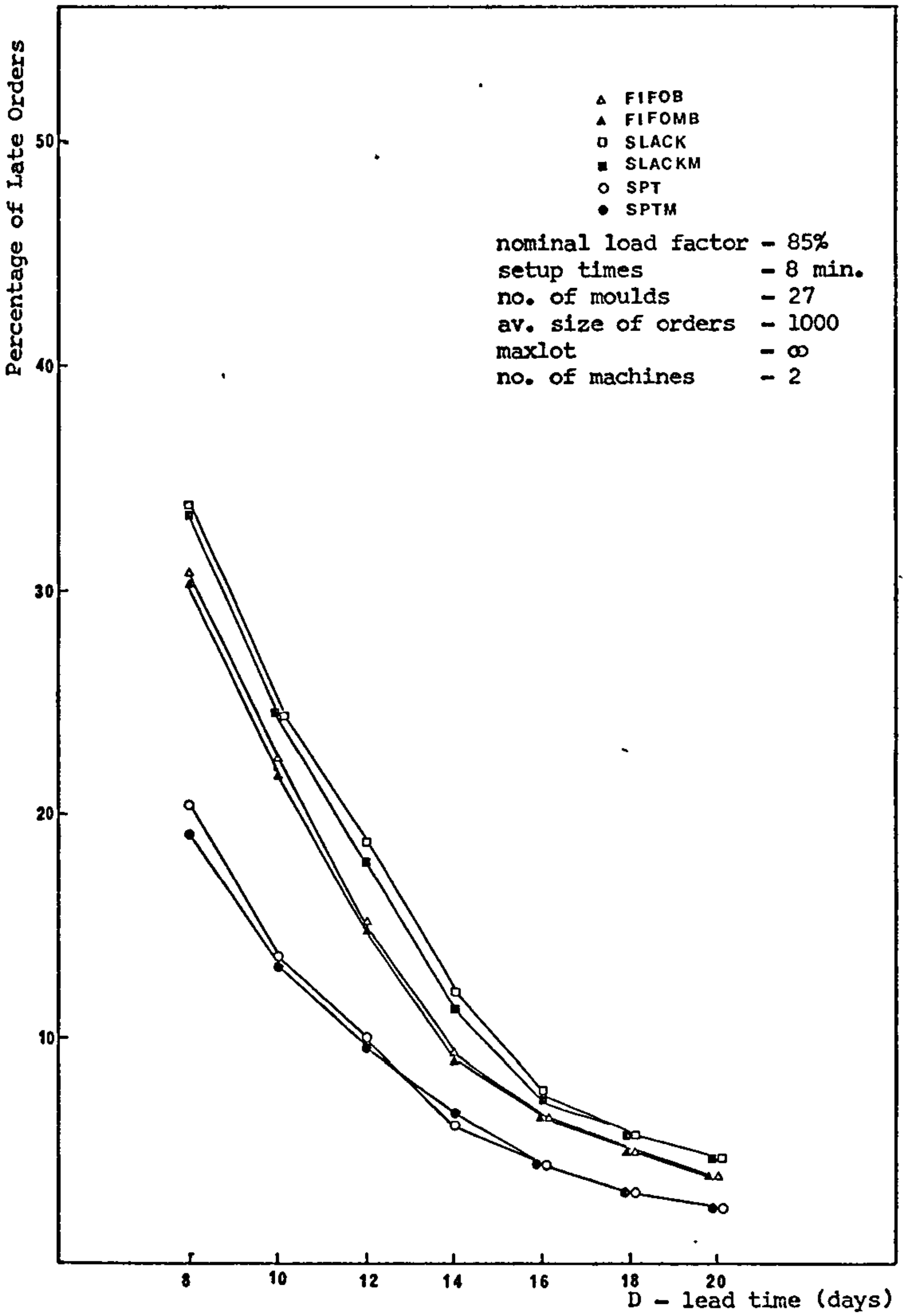


FIGURE 6.14

SYSTEM CONFIGURATION ace

TARDINESS INDEX OF ORDERS

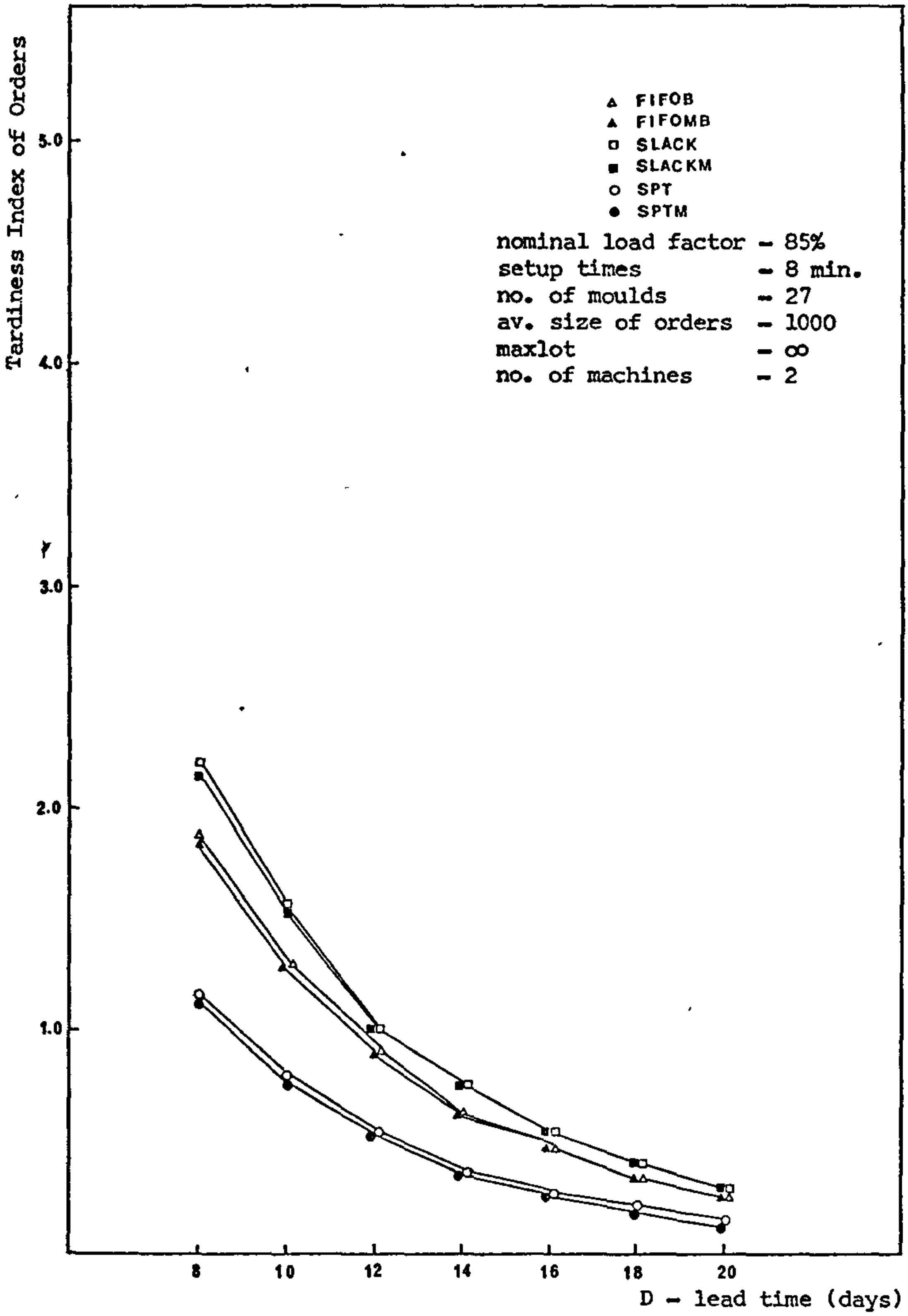


FIGURE 6.15

SYSTEM CONFIGURATION ace

PERCENTAGE OF PRODUCTION DELIVERED LATE

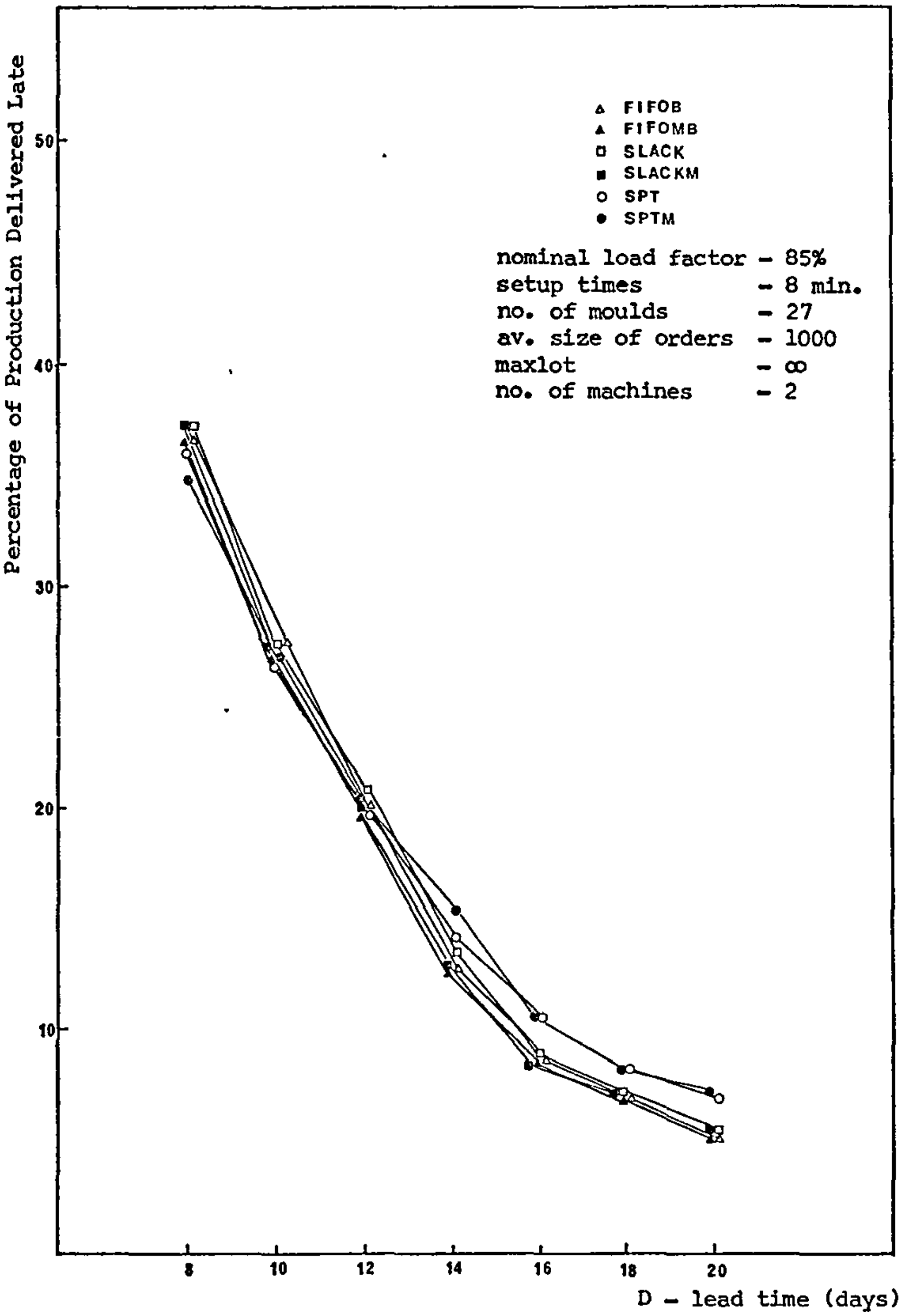


FIGURE 6.16

SYSTEM CONFIGURATION ace

TARDINESS INDEX OF PRODUCTION

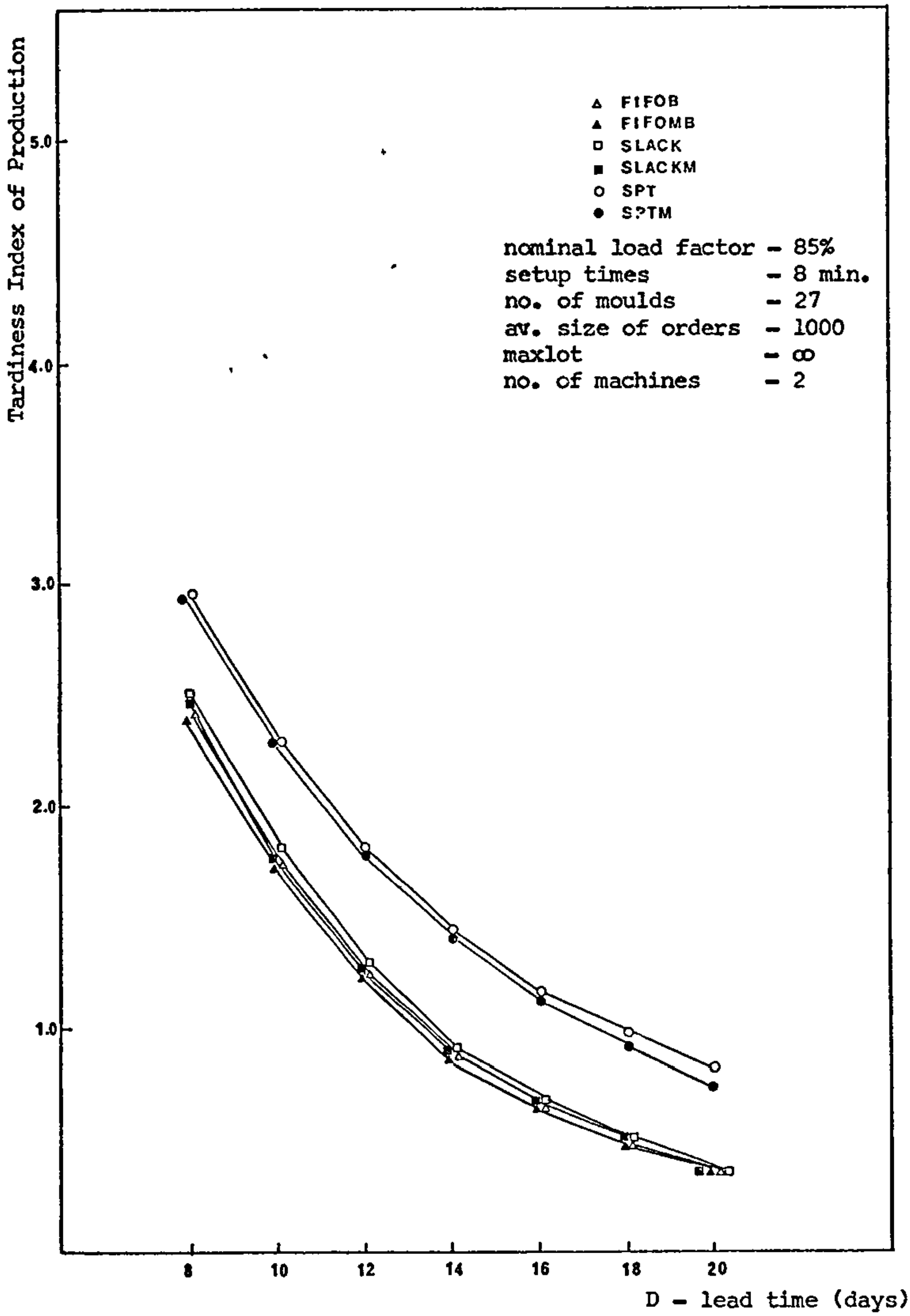


FIGURE 6.17

SYSTEM CONFIGURATION bdf

PERCENTAGE OF LATE ORDERS

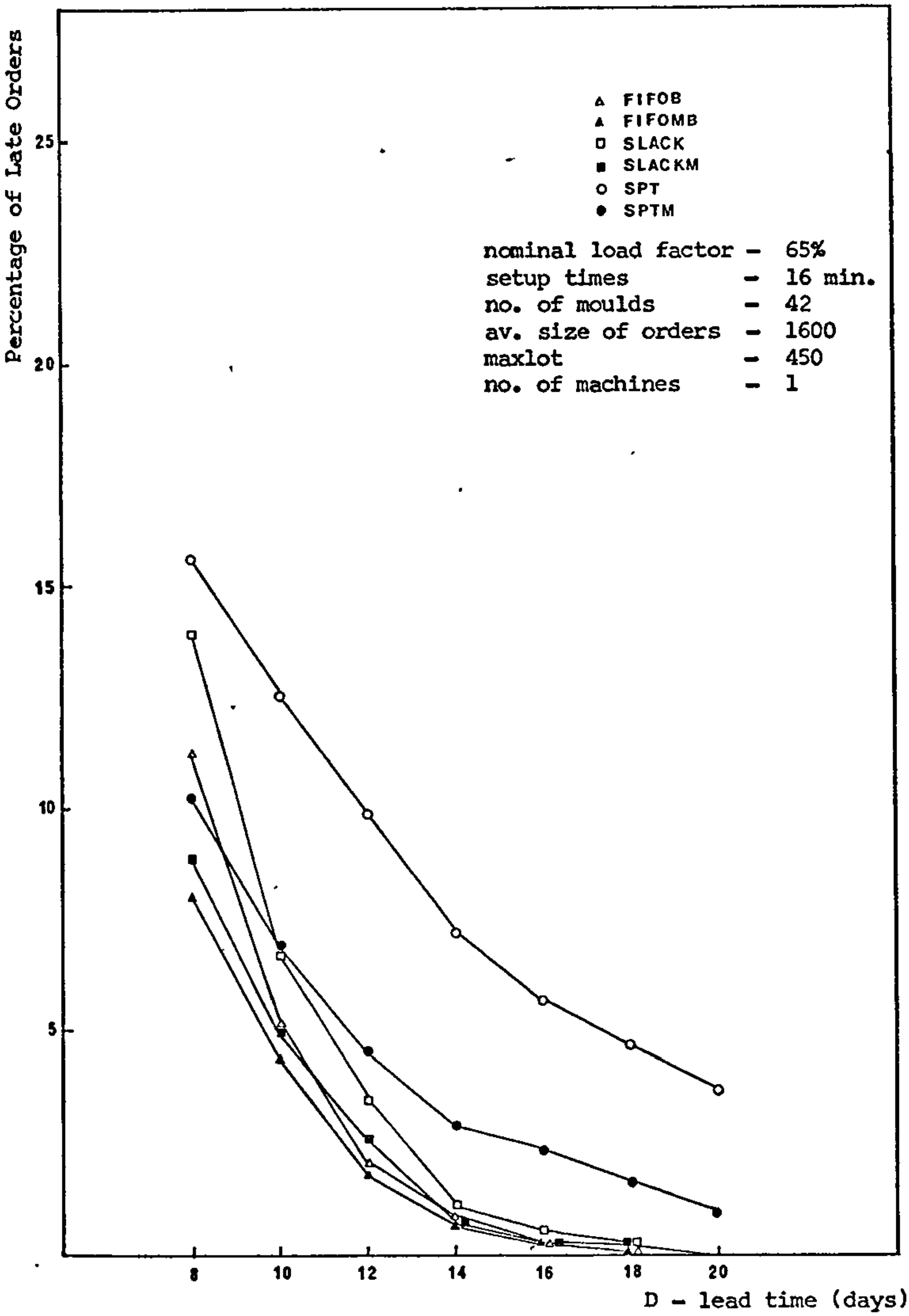


FIGURE 6.18

SYSTEM CONFIGURATION bdf

TARDINESS INDEX OF ORDERS

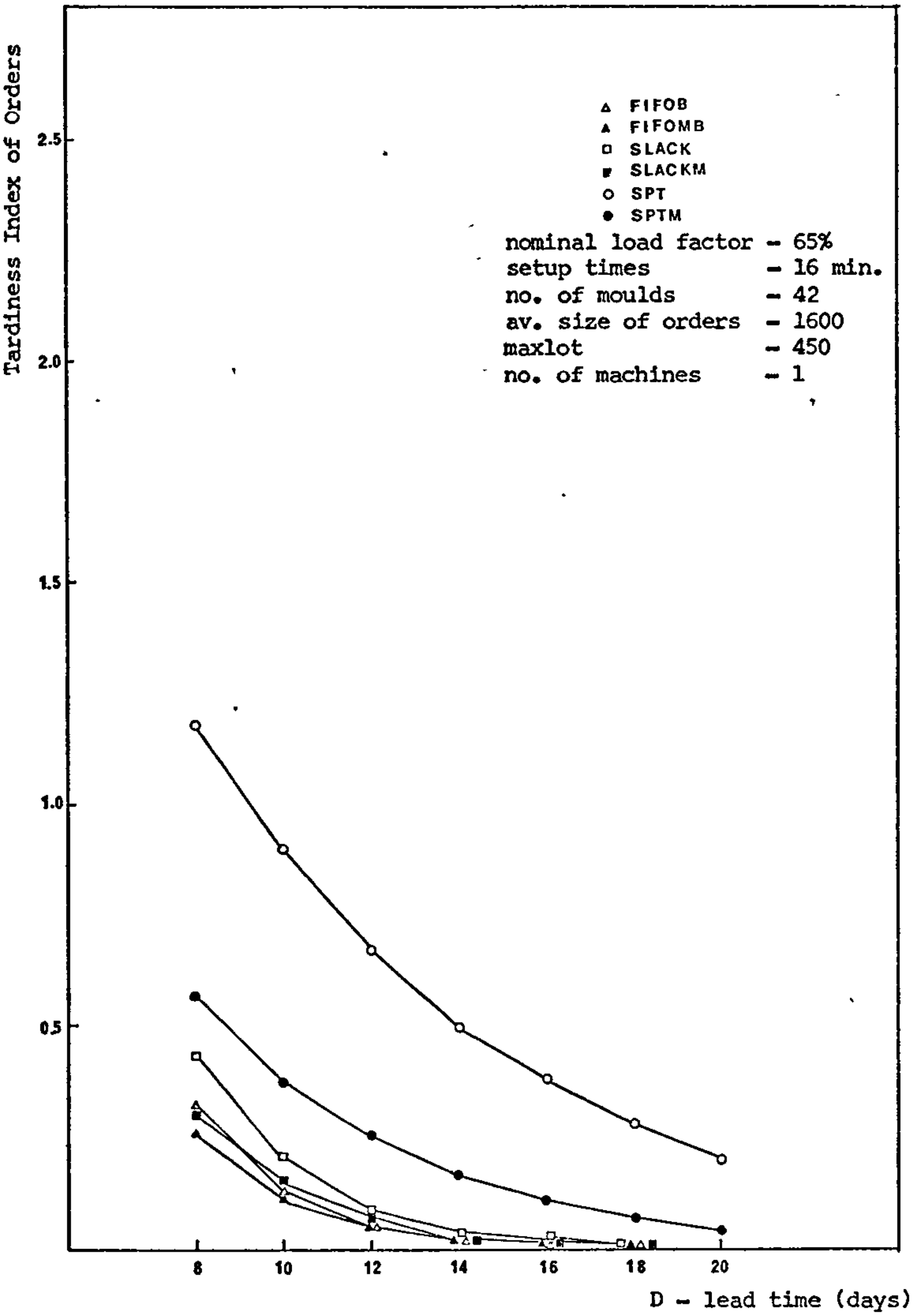


FIGURE 6.19

SYSTEM CONFIGURATION bdf

PERCENTAGE OF PRODUCTION DELIVERED LATE

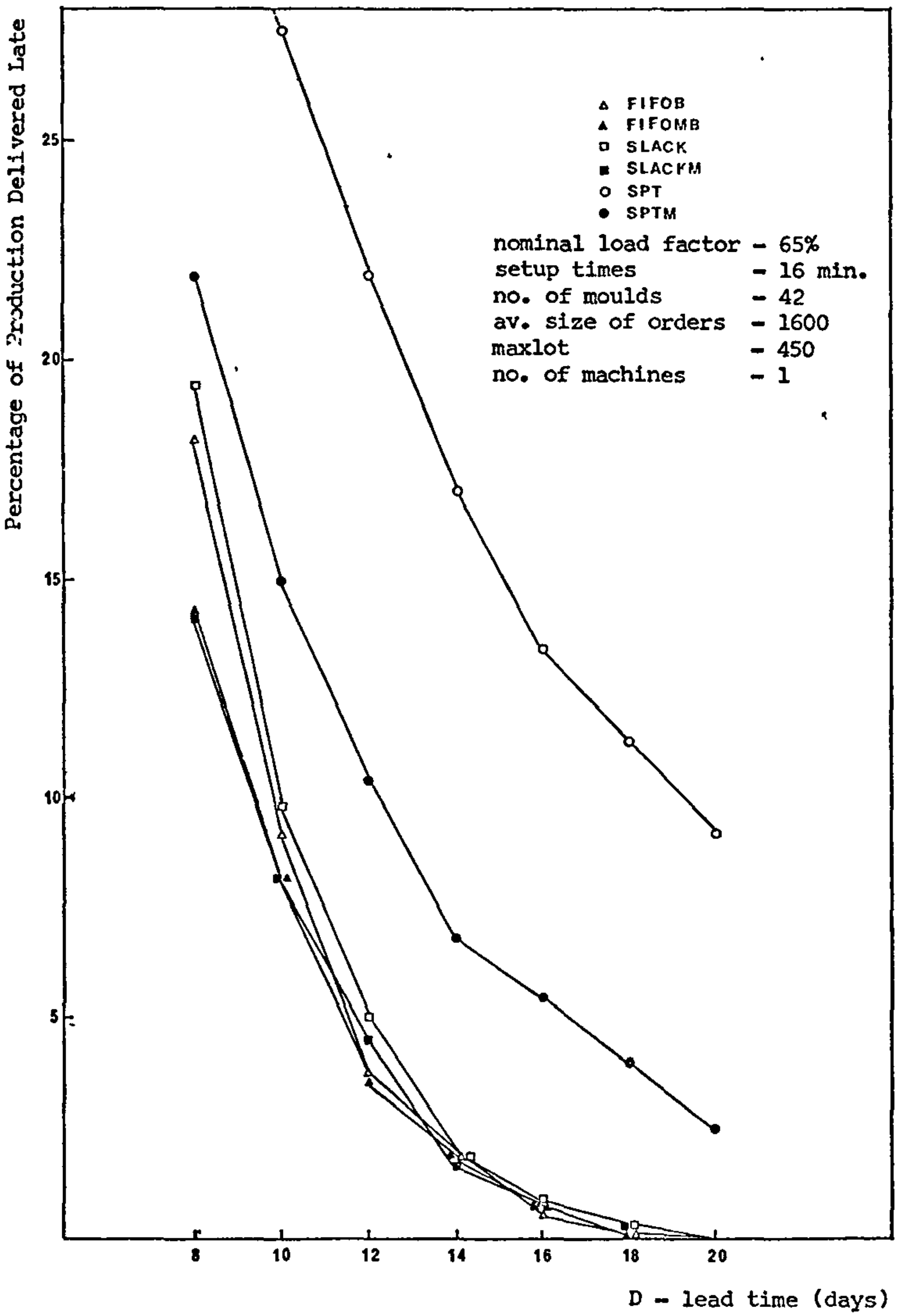


FIGURE 6.20

SYSTEM CONFIGURATION bdf

TARDINESS INDEX OF PRODUCTION

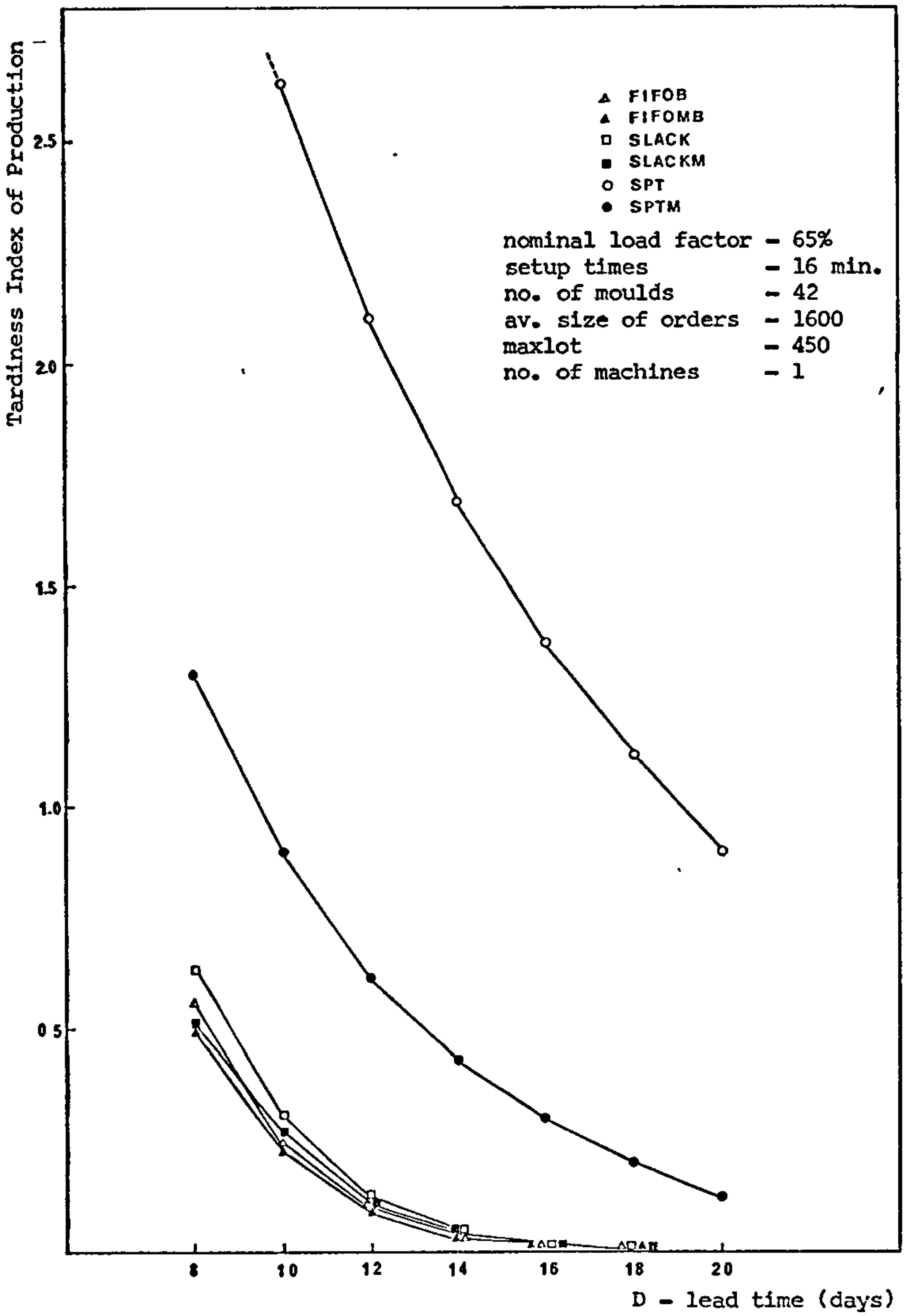


FIGURE 6.21

SYSTEM CONFIGURATION abcdef

PERCENTAGE OF LATE ORDERS

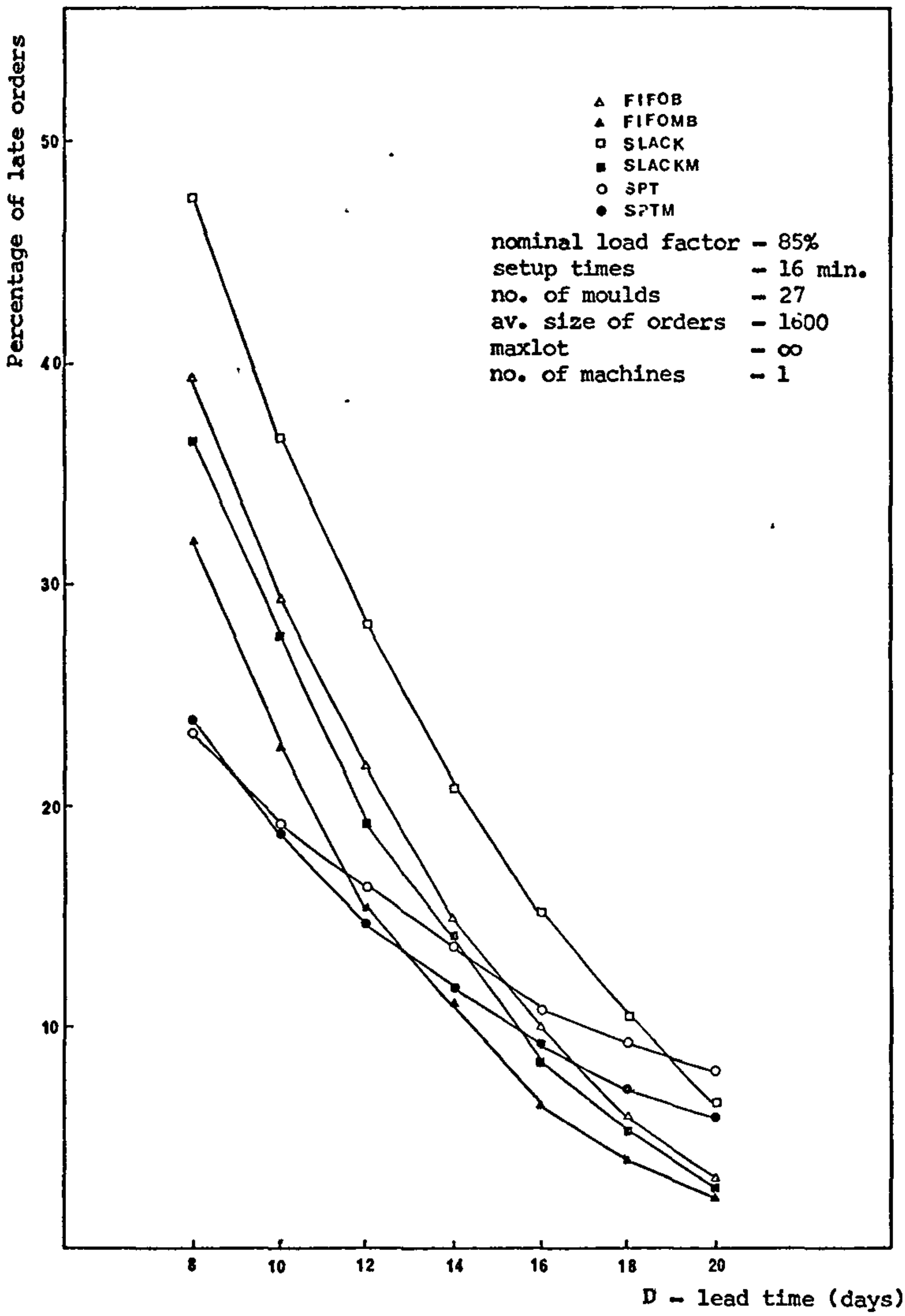


FIGURE 6.22

SYSTEM CONFIGURATION abcdef

TARDINESS INDEX OF ORDERS

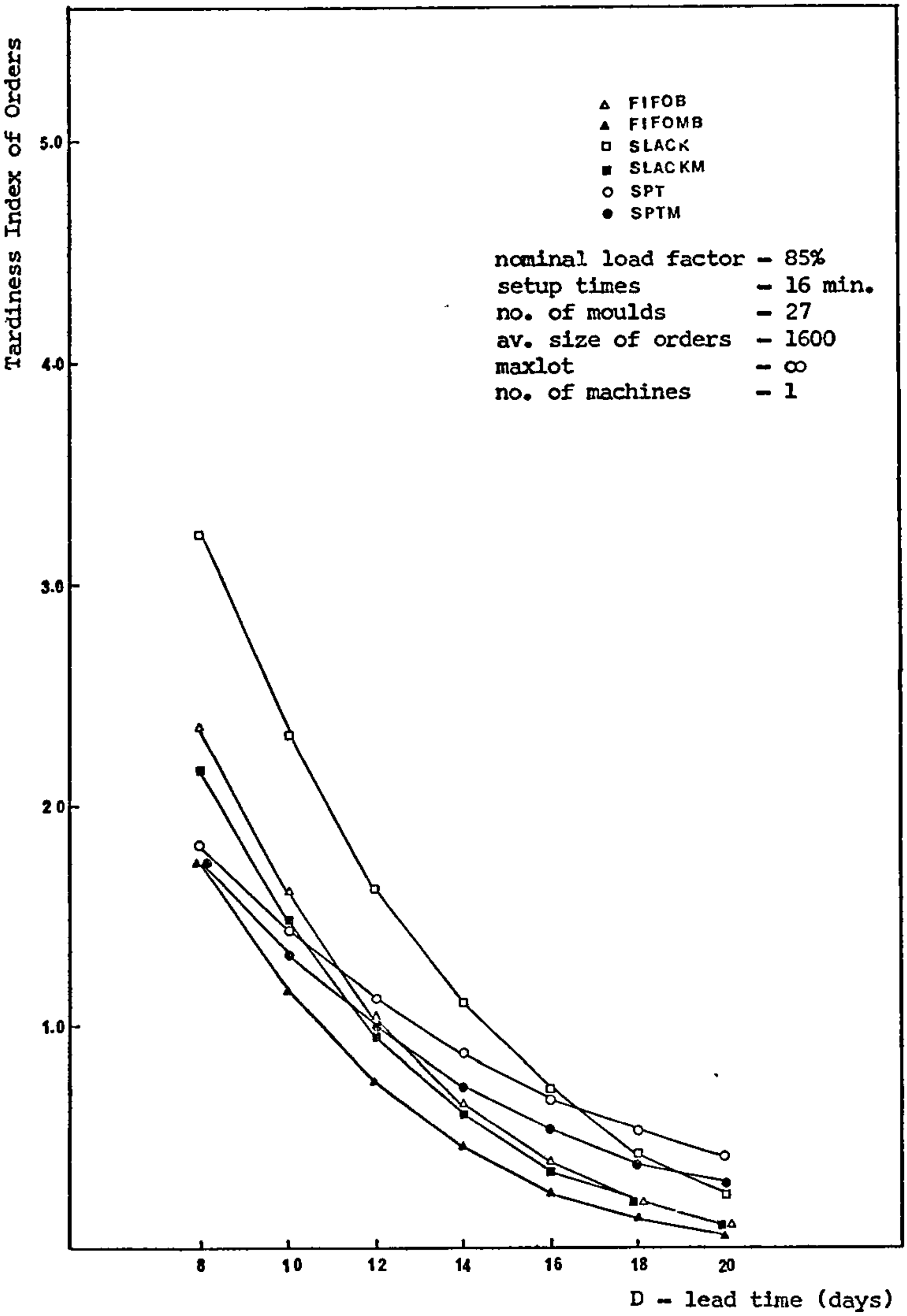


FIGURE 6.23

SYSTEM CONFIGURATION abcdef

PERCENTAGE OF PRODUCTION DELIVERED LATE

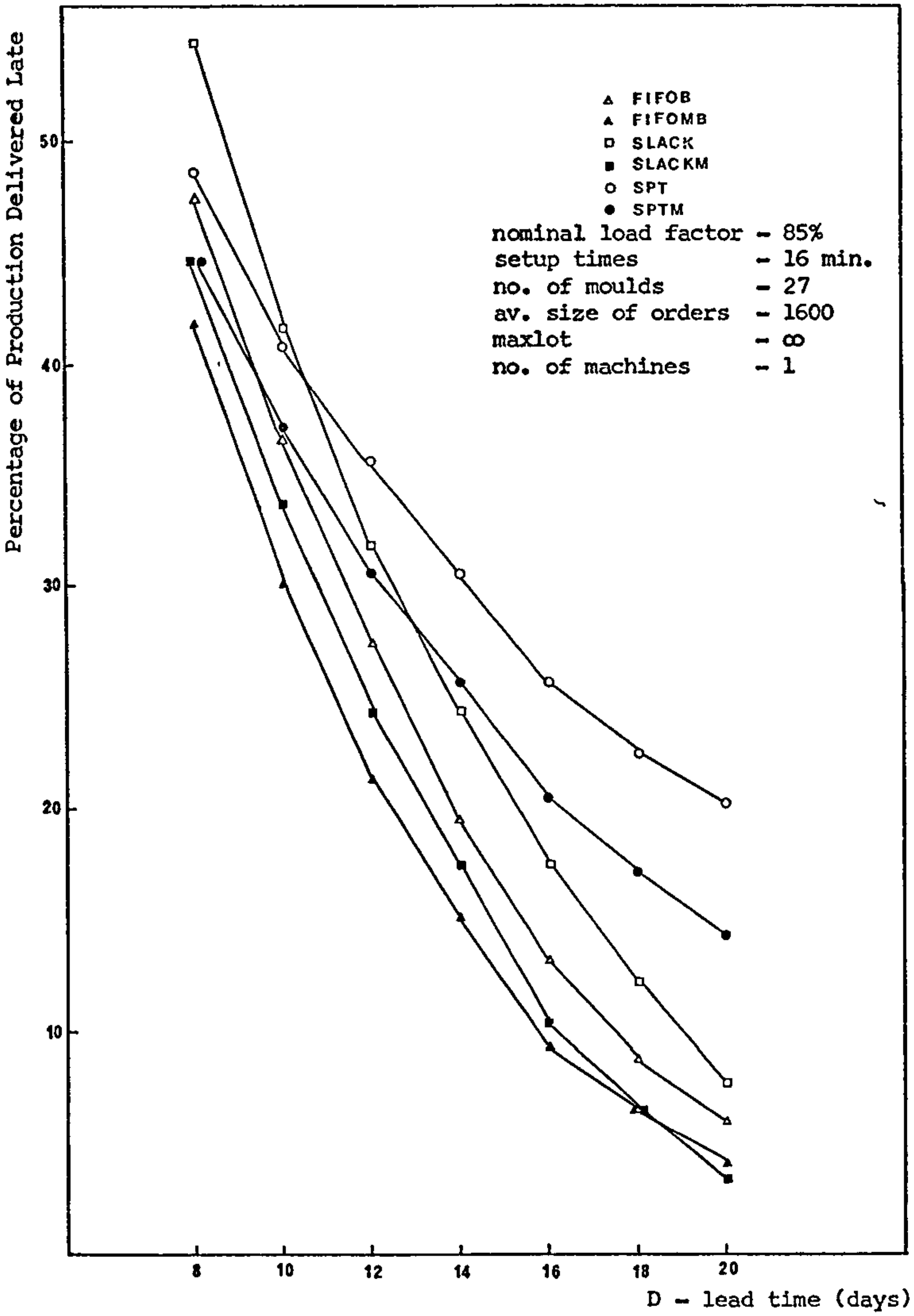
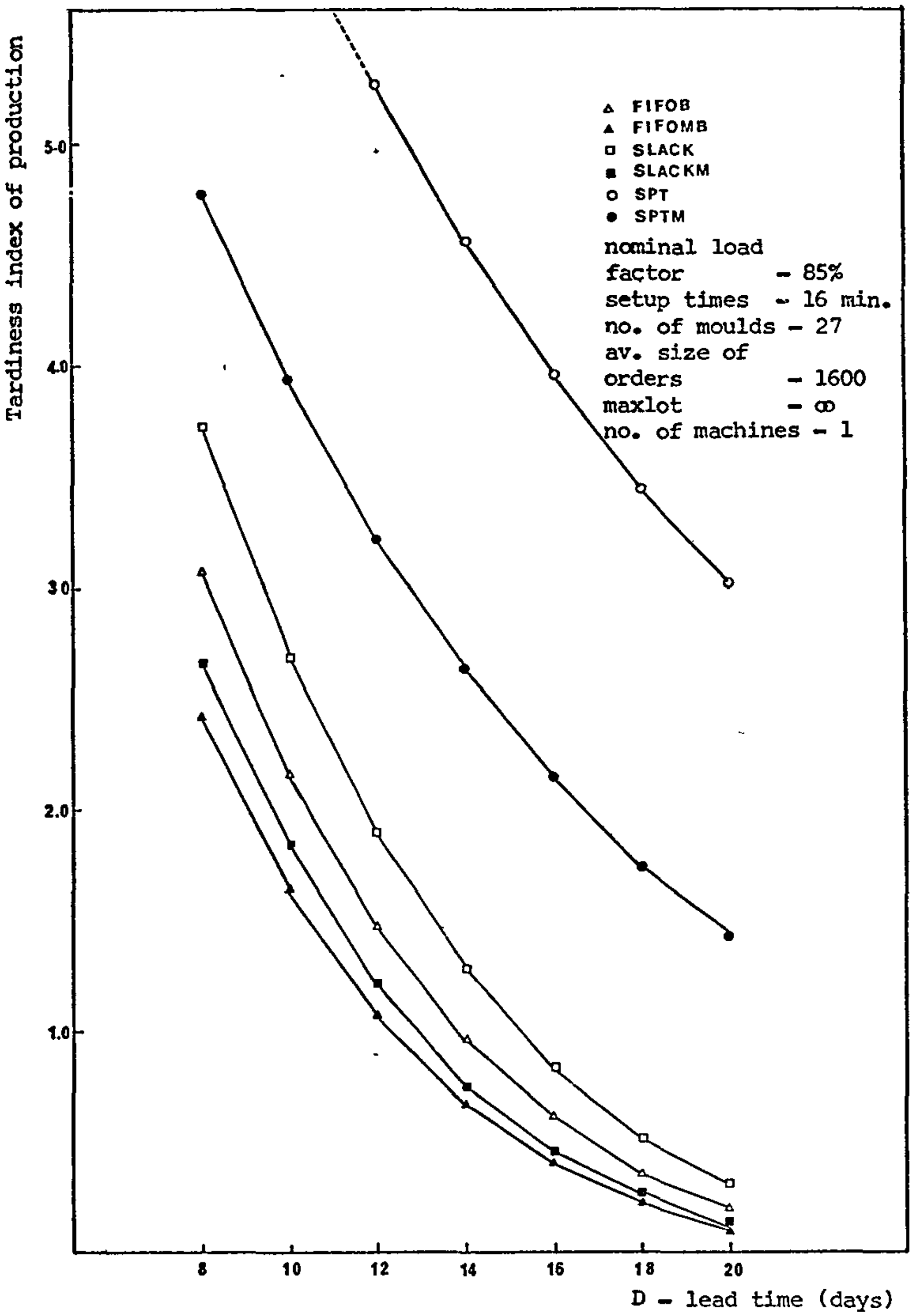


FIGURE 6.24

SYSTEM CONFIGURATION abcdef

TARDINESS INDEX OF PRODUCTION



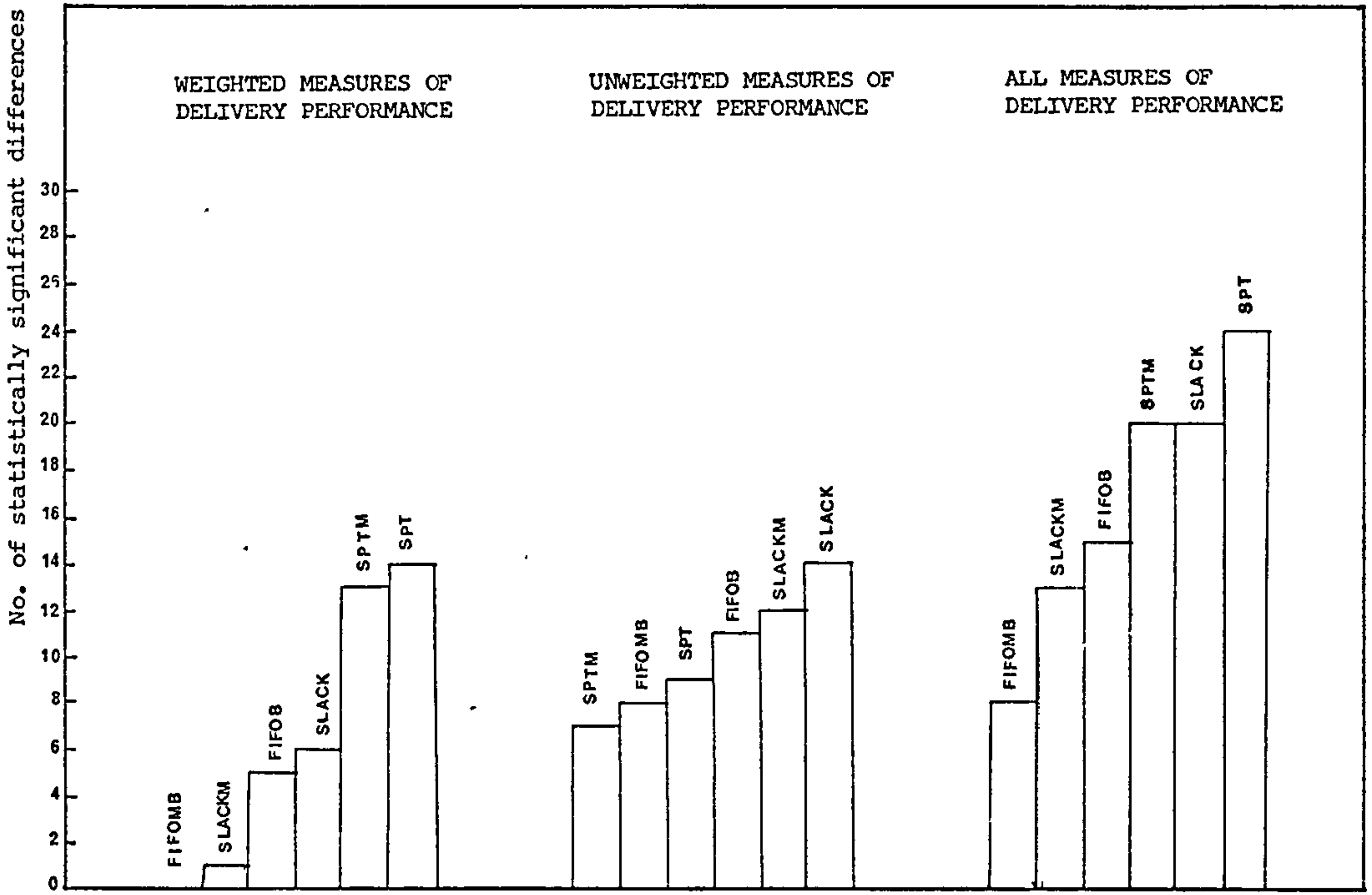


FIGURE 6.25

NUMBER OF STATISTICALLY SIGNIFICANT DIFFERENCES FOR EACH OF THE SIX PRIORITY RULES

CHAPTER 7

STUDY OF MAIN EFFECTS AND INTERACTIONS OF SOME VARIABLES ON THE SYSTEM BEHAVIOUR

7.1 - Introduction

The main objective of this part of the investigation is to study the effects on this class of production systems of variations in a number of important parameters of the system. In order to achieve this objective, an experimental design consisting of a 'half-replicate' factorial was organized as described in paragraph 4. 3.2.

As described in paragraph 4.3.1, the various experiments will be generated by assigning one of two values to each of the six major parameters, such that their effects can be analysed through the model's output variables. In paragraph 4.3.1 a table summarizing the six variables, their values and their notation was presented. That table is reproduced below, for ease of reference.

VARIABLES	SYMBOL	Value of parameters: Standard value 0	Value of parameters: Alternative value 1
A - nominal load factor	a	65%	85%
B - setup time	b	8 min.	16 min.
C - number of moulds	c	42	27
D - size of orders	d	1000	1600
E - job splitting (MAXLOT)	e	450	∞
F - number of machines	f	2	1

To measure the main effects and interactions caused by the above six variables on the system's behaviour, seven output variables will be used. Three of the output variables, viz. 'production delay';

'production late'; and 'tardiness of production'[1], are measures of delivery performance, while the other four, viz. 'process cycle time'; 'actual average load factor'; 'idle time due to setup'; and 'average number of jobs in queue' (queue size), are measures of internal behaviour. The effects on 'production late' and 'tardiness of production', are calculated for a lead time equal to eight days, (D = 8).

For each one of the seven output variables it is possible to measure six main effects (each representing the average independent effect of a change in one of the six variables) and fifteen first order interactions (each representing the interaction effects of two simultaneous changes, or, in other words, the interactions betwen pairs of variables), giving a total of 147 values which can be examined.

As shown in table 4.6, each main effect is represented by a capital letter corresponding to the variable which causes that effect, and each interaction is represented by a pair of capital letters, corresponding to the variables which are interacting. In this way, if A represents the main effect of changing the value of the nominal 'average load factor' from 65% (A_o) to 85% (A_1), and B represents the main effect of changing the mean value of setup times from 8 min. (B_o) to 16 min. (B_1), AB (or BA) will represent the interaction of 'average load factor' and 'mean value of setup times'.

(1) for definition of variables see paragraphs 3.6.1 and 3.6.2

There are several ways of calculating the main effects and interactions. The most straightforward way is to write the yields of the experiments in standard order (in accordance with table 4.5) and use the signs on the same table to produce an algebraic sum of all the yields. The result of such a summation will measure a given effect. The sign for the interactions are obtained by multiplying the signs of the corresponding variables. However for situations in which the number of yields is relatively large it is better to use a method described in Davies (1967) (4), and known as the 'Yates' method. This method was used in this study. The results of the mean values for each individual yield corresponding to the 32 experiments are presented in appendix 5.

The simple calculation of the main effects and interactions however is not sufficient, as they could be a consequence of random components. Therefore, in order to be able to make more accurate statements in respect of the results, an 'F test' of analysis of variance will be conducted on each one of the 147 values which will be obtained by measuring the 21 effects (6 main effects plus 15 interactions) over 7 output variables. The 'F tests' will be based on the null hypotheses that the changes on the systems variables have no significant effects on the output variables. Of course it is expected that some of these hypotheses will be accepted and others rejected with a certain probability, which will be given by the confidence level of 95% or 99% (0.05 and 0.01 respectivelly).(1)

(1) Bonini (1963) (3) points out that the confidence levels are correct only if it is agreed, before running the experiment that only one test is to be performed on the data. For this reason he suggests that the analysis of variance should be relied upon only to indicate which factors are important.

Davies (1967) (5) suggests different methods for assessing the experimental error to be used in the analysis of variance. In the case of this study, because of the sampling procedure used (paragraph 5.3.3) which generates six samples for each yield it is possible to assess the experimental error direct from the six samples. Details of this method can be found in Davies (1967) (6), and Guenther (1964) (4).

Finally it should be pointed out that in accordance with the conclusions of Chapter 6, the FIFOMB rule was selected to be used throughout the 32 experiments which comprise the 'half-replicate' factorial.

7.2 - Presentation of results

To facilitate presentation, the results of main effects and interactions will be presented independently for each one of the six variables which are being changed. During the presentation a few comments will be made, but there will be no attempt to explain some of the effects, as the explanation might be so complex as to require reference to results of other effects which have not been presented yet. However after all the results have been presented, a special section will deal with the discussion of the general behaviour of the system and explanations will be offered for some of the more unexpected effects.

When analysing the results attention should be paid to the fact that although some effects are shown to be statistically significant, their practical significance might be small because of the relatively small change that they cause in the average behaviour of the system. For this reason, before the results of effects are presented it is important to analyse the values of the output variables for the case of the 'average system'. The 'average system' is the system obtained by calculating the mean value of each one of the output variables over the 32 configurations generated by the half-replicate factorial.

The table below presents the mean ($\overline{X}$), the standard deviation (S) and the relative standard deviation ($S/\overline{X}$), for each of the seven output variables, when they are averaged over the 32 system configurations.

Values of the output variables for the 'average system'

	$\overline{X}$	S	$S/\overline{X}$
Average delay of production	5.72 (days)	1.60	0.28
Percentage of production late	20.35%	11.90	0.58
Tardiness index of production	0.999	0.824	0.82
Average process cycle time	5.08 min.	0.54	0.11
Average load factor	74.00%	12.52	0.17
Idle time due to setup (%)	6.61%	3.24	0.49
Average no. of jobs in queue	34.11	18.29	0.54

The results for 'relative standard deviation' ($S/\overline{X}$) show that 'tardiness of production' is the most sensitive of all the measures with respect to changes in the system's configuration, while the 'average process cycle time' is the least sensitive.

Before the presentation of results, a final point should be made about the consequences of interactions between variables and on the interpretation of main effects. When the interaction is large in relation to the value of the main effect, then the main effect ceases to have much meaning. In this case it is no advantage to know, for example, that on the average (i.e. averaged over all levels of other factors) A_1 differs from A_0. The existence of a large interaction means that the effect of one factor is markedly dependent on the level of the other, and when quoting the effect of one factor it is necessary to specify the level of the other. However when the interaction is small it may be inferred that the factors operate independently, and general conclusions on the main effect may legitimately be drawn.

7.2.1 - Results of effects caused by a change in the nominal average load factor: (low 65% vs. high 85%) (A_o vs. A_1)

Table 7.1 presents the results of main effects and first order interactions caused by a change in A (nominal load factor) from 65% to 85%. The results of the 'F test' are indicated by symbols where * means a statistical significance at the 0.05 level and ** means a significance at the 0.01 level.

a) Results of main effects

The results of table 7.1 show that, on average, an increase in the nominal value of the load factor from 65% to 85% has a very large impact on the general behaviour of the system, and in particular on its delivery performance. The 'F tests' indicated statistically significant main effects for all the seven output variables. This is shown by increases of 2.53 days on the 'average delay of production', 18.73% on the 'percentage of production late', and 1.25 on the 'tardiness index of production'. If these effects are compared to the results for the 'average system', presented in the last paragraph, the practical significance of such effects can be clearly seen. In numerical terms they indicate that on average a change from A_o to A_1 caused the 'average delay of production' to increase from 4.47 days to 7.00 days, the 'production late' to increase from 10.99% to 29.72% and the 'tardiness index' to increase from 0.37 to 1.62. In relation to the measures of internal behaviour, the main effects were -0,13 min. for the 'average cycle time', 18.99% for average actual load factor, 0.32% for the 'idle time due to setup', and 29.74 for the queue size. The result for the 'average process cycle time' is surprising as it would be expected that an increase in 'idle time due to setup' would cause an increase in 'process cycle time'. However the opposite has happened, and the possible reasons for this effect will be discussed in a later section. It should be pointed out however that this effect must be analysed in comparison to the results produced by the 'average system', and by the value of the interactions which are large as far as those two variables are concerned. As far as the average system is concerned, the values for 'average cycle time' and 'setup time' were equal to 5.08 min. and 6.61% respectively. This means that they have changed from 5.17 min. and 6.45% to 5.02 min. and 6.77% respectively. These changes although statistically significant, are very small. The effects on the actual average load factor, which increased from 64.61% to 83.50%, and on the average 'queue size', which also increased from 19.24 to 48.98, are large. These two effects were expected, because the variation in the nominal load factor is brought about by an increase in the arrival rate, which means more jobs arriving at the queue, and consequently a larger queue and larger throughput times caused by higher waiting times.

b) <u>Results of interactions</u>

In relation to the measures of delivery performance the only statistically significant interaction effects are with F (ratio no. style/ no. of machines). These interactions are however small in relation to the main effects. Their values are respectivelly 0.41 days for 'production delay', 3.56% for 'production late' and 0.30 for 'tardiness of production'.

These results indicate that the effect on delivery performance caused by an increse on the 'nominal load factor' is bigger when the ratio F is less favourable (3:1).
In numerical terms it means, for example, that the main effect of the 'load factor' on the 'percentage late' is equal to 22.29% when F is 3:1 and 15.17% when F is 3:2.

As far as the measures of internal behaviour are concerned, there are no statistically significant interaction effects for the 'actual load factor', but many statistically significant interactions for the other four measures. For the 'average number of jobs in queue' the interations are negligible in relation to the main effects, with one exception, viz. interaction AD, which is equal to -6.15. This interaction means that the effect on the queue size of an increase in the load factor (arrival rate) is greater when the average size of orders is smaller. In strictly numerical terms it means that the main effect is equal to 23.59 when the average size of orders is 1600, and 35.89 when the average size of orders is equal to 1000. The interactions for 'average cycle time' and 'idle time due to setup' on the other hand are relatively large in relation to the main effects, but still not large in relation to the values for the 'average system'. These relatively large interactions (in relation to the main effects) mean that the main effect on its own does not have much meaning.

7.2.2 - Results of effects caused by a change in the mean value of setup times (low : 8 min. vs. high: 16 min) (B_o vs. B_1)

Table 7.2 presents the results of the main effects and interactions caused by a change in the value of B (mean value of setup time) from

8 min. to 16 min.

a) Results of main effects

The results of the 'F tests' on the main effects indicated statistically significant effects in all the seven output variables. In relation to the measures of delivery performance the increase of the value of B from 8 min. to 16 min. resulted in average increases of 0.52 days for 'production delay', 2.95% for the 'production late' and 0.35 for the 'tardiness index'. When compared with the results of the 'average system' these effects are not large. In numerical terms it means that a change from B_o to B_1 has caused the 'average delay of production' to increase from 5.47 days to 5.99 days, the 'production late' to increase from 18.88% to 21.83% and the 'tardiness index' from 0.88 to 1.10.

All these effects are to be expected. All measures of internal behaviour suffered increases in their average value. In numerical terms this means that on average, the increase in B has caused the 'average cycle time' to increase from 4.91 min. to 5.26 min; the 'actual load factor' to increase from 71.55% to 76.46%; the 'idle time due to setup' to increase from 4.90% to 8.33%, and the number of jobs in queue to increase from 32.42 to 35.80. It can be seen that, in relative terms, the largest effect was on the 'idle time due to setup', a result which should be expected, as the variable changed was the mean value of setup times. All the other results were also expected, as an increase in the amount of idle time should cause an increase in the 'average cycle time', and 'actual load factor', with a consequent increase in the queue size and deterioration on delivery performance.

b) Results of interactions

The 'F tests' failed to detect any statistically significant interactions for the measures of delivery performance, but indicated a few significant interactions for the measures of internal behaviour. For the 'number of jobs in queue' and 'actual load factor' there is only one statistically significant interaction for each, viz. BA and BC, respectively. The BA interaction is the same as AB and, as mentioned before, is small. Interaction BC is also small (-1.41%), but indicates that an increase in the mean value of setup time has a bigger effect on the 'actual load factor' when the number of moulds is larger (42). This interaction might be expected, as a larger number of moulds increases the chances of mould changeovers with a consequent increase in the amount of time spent in setting up. This is reflected in a higher 'actual load factor'. It should however be pointed out that in relative terms (compared with the 'average system') the effects are small. In relation to the 'average cycle time' and 'idle time due to setup', the 'F test' indicated, for both, statistically significant negative interactions with A, C and D, and positive interactions with F. Those interactions are however small as compared to the main effects and do not seem to have any influences on the delivery performance.

7.2.3 - Results of effects caused by a change in the number of moulds available (42 vs. 27) (C_o vs. C_1)

Table 7.3 presents the results of the main effects and interactions caused by a change on the number of moulds from 42 to 27. In order to simplify discussions, the comments will refer to the effect of increasing the number of moulds from 27 to 42, instead of the opposite. This simply means that the signs of all effects on table 7.3 should be reversed.

a) Results of main effects

The results of the 'F tests' on the data indicate statistically significant main effects in all the seven output variables. These effects are considerable, although large interactions with the variable F make them a meaningless average, whose values depend heavily on F. In strictly numerical terms it means that on average, the increase in the number of moulds from 27 to 42 causes an improvement in the delivery performance, which is expressed by a reduction of 0.86 days on the 'average delay of production', a reduction of 7.36% on the 'production late', and a reduction of 0.53 on the 'tardiness index'. It is interesting to note that this improvement in performance occurred in spite of considerable increases in the 'idle time due to setup', 'actual load factor' and 'average cycle time', which on average have increased from 4.93% to 8.30%; 68.85% to 79.15% and 4.71 min. to 5.45 min. respectively. These effects indicate that an increase in the number of moulds tends to generate an increase in the number of mould changeovers, with a consequent increase in the 'idle time due to setup'. 'Number of jobs in queue' was reduced by the increase in the number of moulds. This effect is in accordance with the improvement in delivery performance, which must have been caused by a reduction in the waiting times in queue for the jobs (which means smaller queues).

b) <u>Results of interactions</u>

In relation to the measures of delivery performance the 'F tests' have indicated a highly significant interaction between C and F, for all the three measures.

In comparison with the main effects these interactions are very large which means that the main effect does not have much meaning on its own. In general terms these interactions mean that the effect of increasing the number of moulds from 27 to 42 has a large effect on the delivery performance when the ratio 'no. of style/ no. of machine' is low (3:2), but almost no effect when the ratio is high (3:1). In numerical terms it means, for example, that the reduction in the 'percentage of production late' caused by an increase in C from 27 to 42, is, on average, equal to 12.52% when the ratio F is 3:2, but only 2.20% when the ratio F is 3:1. The same sort of conclusions are valid for 'average delay of production', and 'tardiness of production'. This confirms the results obtained during the preliminary investigation and reported in paragraph 4.2.4, which indicated a very small improvement in delivery performance when the number of moulds was increased from 27 to 45, for a system configuration having 3 product styles and 1 machine.

In relation to the measures of internal behaviour, the 'F tests' indicated various significant interactions apart from CF. Among those interactions are CA and CB, which were reported when the effects of A and B were discussed. In relation to the interactions CF, the most important effect is on the 'average number of jobs in queue'. For an average main effect of -8.72, the interaction was equal to 8.03, which is in complete agreement with the results for delivery performance. The CF interactions for 'idle time due to setup', 'actual load factor', and 'average cycle time', are relatively small in relation to the main effects. However they indicate that the increase in the values of those variables, caused by an increase in the number of moulds, is higher when the ratio F is 3:2, than when the ratio is 3:1. It should be noted that when the system has 27 moulds and 2 machines, the ratio of the no. of moulds to the no. of stations is 27:24, meaning that only 3 moulds are usually available for changeover. When the number of moulds increases to 42, the ratio becomes 42:24, meaning that 18 moulds are constantly available for changeover. The increasing factor for the number of moulds available for changeover is therefore 6 (18/3). However when the number of machines is 1, the number of moulds available for changeover is 15 (27 - 12) for the case of 27 moulds and 30 (42 - 12) in the case of 42 moulds. This means that the factor of increase in number of moulds available for changeover is only 2. Hence, in proportional terms, the number of possible changeovers increases more sharply when F is 3:2, than when F is 3:1.

7.2.4 - Results of effects caused by a change in the average size of orders (1000 vs. 1600) (D_o vs. D_1)

Table 7.4 presents the results of the main effects and interactions caused by a change in the average size of orders from 1000 items (D_o) to 1600 items (D_1).

a) Results of main effects

The 'F tests' indicate statistically significant effects for all the seven output variables. In relation to the measures of delivery performance the results indicate that, on average, the increase in the average size of orders causes a deterioration of the delivery performance of the system. In numerical terms it means that the variation in the value of D from 1000 to 1600 has caused the value of 'average delay of production' to increase from 5.24 days to 6.22 days, the 'percentage of production late' to increase from 17.63% to 23.08%, and the 'tardiness index' to increase from 0.86 to 1.14.

In relation to the measures of internal behaviour, the variation in D has resulted in an improvement in the internal performance which is indicated by smaller values for all the measures of internal behaviour. In numerical terms it means that, on average, the increase in the size of orders has caused the 'average cycle time' to decrease from 5.19 min. to 4.97 min., the average actual load factor to decrease from 77.78% to 70.23%, the idle time due to setup to decrease from 7.84% to 5.38%, and the average number of jobs in the queue to decrease from 39.64 to 28.58. These effects are understandable if one considers that the increase in the average size of orders corresponds to a decrease in the arrival rate (to maintain the same nominal load factor), which means less jobs arriving in the system. This decrease in the arrival rate must have caused a drop in the number of mould changeovers, with the consequent reduction in the 'idle time due to setup'. It has also caused a decrease in the number of jobs in queue.

These improvements in the measures of internal behaviour however have been outweighed by the increase in the batch sizes of jobs, which means longer production time for the batches, with a consequent deterioration in the delivery performance. It should be pointed out that the effects on 'tardiness' and 'lateness' are due mainly to the method used to fix due dates, which does not take account of the size of orders when establishing due dates (see paragraph 3.6.2).

b) Results of interactions

In relation to the measures of delivery performance there are no statistically significant interaction effects. For the measures of internal behaviour there are statistically significant interactions with A, B and C for the 'idle time due to setup'. These interactions are relatively small, however, in relation to the main effects, and were reported when the results of effects A,B and C were presented. For the average number of jobs in queue, there is only one interaction (with the load factor A) which means that the reduction of the queue size due to the increase in the average size of orders, is larger when the load factor is at its higher level (85%).

7.2.5 - Results of effects caused by splitting large jobs into smaller batches (Splitting vs. no splitting) (E_o vs. E_1)

Table 7.5 presents the results of the main effects and interactions caused by changing the procedure for splitting large jobs into smaller batches. In other words it shows the results of effects on the system behaviour caused by a change in the value of MAXLOT, from 450 to ∞. This change means that for the first case (A_o), all jobs with a batch size of 450 or more are split into smaller batches while on the other case (A_1) there is no splitting of jobs at all.

a) Results of main effect

The 'F tests' failed to find any statistically significant effects on all but one of the output variables. The only variable significantly affected by the splitting of jobs was the average number of jobs in queue. This effect, which is equal to - 2.70, indicates that the splitting of jobs into smaller batches tends to increase the queue size. This is to be expected, and indicates that a significant number of jobs were split into smaller jobs for a value of MAXLOT equal to 450. The fact that all the other variables were not significantly affected by the splitting procedure only means that for a value of MAXLOT equal to 450, there is, on average, no effect on the behaviour of the system.

b) Results of interactions

The 'F tests' failed to find any statistically significant interactions either at the 0.01 or 0.05 levels.

7.2.6 - Results of effects caused by a change in the ratio 'no. of styles/no. of machines' (a more favourable ratio (3:2) vs. a less favourable ratio (3:1) (F_o vs. F_1)

Table 7.6 presents the results of the main effects and first order interactions caused by a change in the ratio between the number of product styles and the number of machines, from 3:2 to 3:1.

a) Results of main effects

The 'F tests' applied to the data indicated statistically significant main effects on all the seven output variables. The results indicate that on average, the increase in the ratio 'no. style/no. machine' (more product style per machine) tends to cause a deterioration in the performance of the system. The only variable which actually decreased in value was the average number of jobs in queue. This is a reflection of the fact that in order to maintain the same nominal load factor, the arrival rate had to be halved when changing from two machines (ratio 3:2) to one machine (3:1). Considering that there is a single queue for both machines, this decrease in arrival rate should be reflected in a smaller queue. In relation to the 'average system', these effects are considerable, both for the measures of delivery performance and for the measures of internal behaviour. However, the interactions with A and C for the three measures of delivery performance and for the average number of jobs in queue are so high as to make the main effects for those variables meaningless. For the other variables of internal behaviour, although there are many statistically significant interactions, their values are not very high in relation to the main effects.

In numerical terms these main effects indicate that an increase in the ratio F from F_0 to F_1 causes the 'average cycle time' to increase from 4.81 min. to 5.35 min., the 'actual load factor' to increase from 70.77% to 77.24%, and 'idle time due to setup' to increase from 5.26% to 7.96%. These main effects can be explained by the fact that when the number of machines changes from 2 to 1, the ratio 'no. of moulds/ no. of stations' is multiplied by a factor of 2, with a consequent increase in the number of mould changeovers.

b) Results of interactions

As far as the measures of delivery performance are concerned the 'F tests' have indicated statistically significant interactions with A and C. The interactions with A indicate that the deterioration in delivery performance caused by an increase on the ratio F from 3:2 to 3:1 is higher when the nominal load factor is at a higher level. The interactions with C, on the other hand, are very high in relation to the main effects, and indicate that if the number of moulds is small (27), there is practically no effect on the delivery performance (the effect on 'production delay' is 0.01 day) when the ratio F is changed from 3:2 to 3:1. However if the number of moulds is large (42), the same variation in F causes a large deterioration in the delivery performance (the effect on the 'production delay' would be equal to 1.43 days)

In other words the FC interaction indicates that on average, a better ratio F (3:2) has no effect on delivery performance, if the number of moulds is small. However, if the number of moulds is large enough, a better ratio will result in considerable improvement in delivery performance.

In relation to the measures of internal behaviour the interactions AF, BF, and CF for the 'average cycle time' and 'idle time due to setup , and the interaction CF for the ' actual load factor' are all statistically significant but not as large as the interaction for the measures of internal behaviour. These interactions have been already reported in previous paragraphs. For the 'average number of jobs in queue', the interactions AF and CF are also statistically significant, with CF being very high in relation to the main effect (- 8.15 as compared to a main effect of - 7.38). This large interaction means that a change in the ratio F from 3:1 to 3:2 would cause an increase (on average) of 15.41 in the average queue size if the number of moulds is equal to 27. However if the number of moulds is equal to 42, the same change in F would actually produce a decrease of 0.70 in the average queue size.

7.3 - Discussion of results

For the sake of discussion the results of the experiments can be divided into two major groups. The first group refers to the measures of delivery performance. Considering that one of the main objectives of this class of production systems is to maintain an effective and reliable delivery performance, one of the prime objectives of this investigation should be to identify which of the system's parameters can influence its delivery performance, and how significant their influences are in relation to the other parameters. The second group refers to the variables of internal behaviour. While the first group of variables is used in order to evaluate the variations in the external performance of the system, the second group is used in order to understand the mechanism behind those variations.

7.3.1 - Effects on the delivery performance of the system

The results indicated that the variation in the 'nominal load factor' brought about by variations in the mean arrival rate is by far the most important single factor among the six examined, as far as their ability to influence the delivery performance is concerned. It should be pointed out that such an effect is highly dependent on the extent and range of the variation. For example, a variation in the nominal load factor from forty to sixty percent (a relative increase of 50%) might have less influence on the behaviour of a system than a variation from ninety to ninety five percent (a relative increase of only 5.6%).

Other authors (Eilon (1967); Conway et. al (1962); Hollier (1968)) have used higher loads than the ones used in this study (usually around 90% and 95%), but for systems with characteristics quite different from this one. Furthermore the choice of the parameters for this study have already been discussed in paragraph 4.3.1 and in the light of the results and system characteristics seems quite reasonable.

This capacity of the nominal load factor to influence the delivery performance was shown to be to a great extent independent of the other five variables. The only statistically significant interaction effect was with the ratio 'number of product styles/no. of machines', but even this interaction was relatively small in comparison with the magnitude of the main effect.

The second largest independent effect on the delivery performance of the system was caused by a variation in the 'average size of orders'. It should be pointed out that although the magnitude of the change made was quite large (an increase of about sixty percent on the average size of orders) in relation to the relative variation on the nominal load factor (about thirty percent), the effect on the delivery performance was much smaller than the one caused by the variation in the load factor. One interesting point to note is that the ability of the order size to influence the delivery performance is largely independent of the parameters of the other five variables. The analysis of variance failed to find any statistically significant interactions.

The only other variable which was able to influence the delivery performance of the system independently (as far as the significance test is concerned) of the parameters of other variables, was the mean value of setup time. The magnitude of the effect however was small when compared to the effects caused by the previous two variables, as can be seen by comparing the results of row 1, columns 1, 2 and 3 for tables 7.1, 7.2 and 7.4.

The other three variables have either not shown any significant effect (splitting of jobs), or have produced effects which are highly dependent on the parameters of other variables. This latter case happened for variables C (number of moulds) and F (ratio between number of styles and number of machines). The main effects of C and F on the delivery performance of the system are both considerable in numerical terms. However the interaction CF is so large, that these main effects lose all their meaning. When considering the influence of C or F on the delivery performance of the system, one must refer to the parameter of F and C respectively. The effect CF is however important, because of its magnitude in relation to the other effects. The CF effect has already been discussed in paragraphs 7.2.3 - b and 7.2.6 - b, and it should only be added that a better ratio F (3:2) can cause a considerable improvement in the overall performance of the system. This is shown by the fact that an arrival rate twice its original value (which followed the increase in the number of machines from 1 to 2), failed (on average) to cause any deterioration of the delivery performance even though the number of moulds was maintained constant at 27. However when the number of moulds was maintained constant at 42 the joint increase in the arrival rate and the number of machines (represented by a F ratio 3:2) has actually caused a considerable improvement (on average) in the delivery performance of the system. It should be pointed out that the increase in the number of moulds did not result (on average) in any improvement in delivery performance for the cases in which the number of machines was 1.

Finally some comments should be made about the results related to variable E (value of MAXLOT for 'job splitting'). The fact that the 'F tests' failed to find any significant effects in the measures of delivery performance, only means that for a value of MAXLOT equal to 450 the splitting of large jobs into smaller batches does not have any statistically significant effect on the delivery performance.

Although there was some evidence from experiments of paragraph 4.2.4 to suggest that 450 was a reasonable value for MAXLOT, the lack of effect does not allow any firm conclusion to be drawn. The problem of choosing a correct value for MAXLOT is complicated by the existence of many different system configurations. This means that a value which is good for one system configuration might not be so for another. An absolute answer to this problem would require a large number of experiments, which might not bring any important conclusions. A partial answer was obtained however by having two extra series of experiments, in which the value of MAXLOT was varied while all the other variables were maintained constant. The two series differ from each other only in respect of the system configurations used. The first series was executed with system configuration abcdf and the second with system configuration d. The choice of those two configurations was based on the idea of having two extreme cases (most 'tight' and most 'loose' configurations) for which the average value of orders was at its higher level (1600). System configuration (I) would be even more 'loose' than d, but it was thought that with an average size of orders already small, the possible influence of MAXLOT value (whose objective is to split large jobs) would not be shown as well as with configuration d.

Each series therefore, consisted of having a total of 5 experiments corresponding to values of MAXLOT equal to 250, 350, 450, 550 and ∞ (no splitting at all). After each series a multiple comparison test (the same test described in paragraph 6.1) of analysis of variance was applied to the five results in order to detect any statistically significant differences caused by variations in the value of MAXLOT. The results of the experiments and multiple comparison test are shown in tables 7.7 and 7.8, for system configurations abcdf and d, respectively.

For system configuration abcdf the lowest value of all three measures of delivery performance ('average delay of production', 'percentage of late production' and 'tardiness index of production') were obtained for MAXLOT = 450. The multiple comparison test indicated significant differences in favour of MAXLOT = 450 in relation to MAXLOT = 350 and MAXLOT = 250, but in general no significant differences for MAXLOT = 550 and MAXLOT = oo. The results for the 'average number in queue' indicated (as expected) significant increases in the queue size, created by the splitting procedure. As the value of MAXLOT gets smaller, the queue size gets larger.

The results for system configuration d (table 7.8) indicated the same effect on the queue size, but different results for the variables of delivery performance. The lowest value of 'average delay of production' was produced by MAXLOT = 250, and the multiple comparison test indicated statistically significant differences for all the other values of MAXLOT. For the 'percentage of production late' and 'tardiness index of production', the lowest values were produced by MAXLOT = 350. However the multiple comparison test failed to find any statistically significant differences between MAXLOT = 350 and MAXLOT = 250 or MAXLOT = 450. It should be pointed out that in relation to the measures of delivery performance, even the statistically significant differences were in absolute terms very small. The same is not true in relation to the effect on the queue size. As the value of MAXLOT decreases the increases in queue size becomes considerable.

Based on the above evidence it is therefore possible to say that there are strong indications to suggest that the procedure for splitting large job into smaller batches does not bring, on average, any significant advantages to the performance of the system.

7.3.2 - Effects on the internal behaviour of the system

The results for the measures of internal behaviour presented a different picture from the measures of delivery performance. While for the latter the results were very uniform for the three variables, for the former there was a more complex picture. One interesting point to note was that depending on the effect being analysed, an improvement in the internal performance (smaller 'process cycle time', smaller 'idle times due to setup', smaller 'queue size'), could in fact correspond to a deterioration in the external performance (delivery performance). This was certainly the case for effects D and C. In the case of D, for example, the increase in the size of orders caused, as expected, a reduction in the value of 'idle time due to setup' and in 'average queue size'. This was reflected in the 'actual load factor', and 'process cycle time' which also had their values reduced. However these improvements (which represent smaller processing time per item) were outweighed by the increase in batch sizes, resulting in larger throughput times and consequently poorer delivery performance. The same kind of phenomenon happened when the number of moulds was reduced from 42 to 27 (effect C). The only difference being that this time the deterioration in delivery performance was followed by an increase in the average queue size.

In relation to the mechanisms governing the internal behaviour of the system, the results indicated a direct relationship between the 'idle time due to setup', and two of the other variables viz.' actual load factor' and 'process cycle time'. The whole mechanism seems to be governed by the effect that the system's parameters have on the amount of time spent in setting up. A variation in the 'idle time due to setup' (in one direction or another) is immediately reflected in a variation in the same direction of the 'actual load factor' and 'average process cycle time'. This mechanism is no surprise, and should be expected. However the results of table 7.1 indicated an exception. On that occasion an increase in the 'idle time due to setup' actually corresponded to a reduction in the 'average process cycle time'. At first this result seems quite dubious, but a close examination of the circumstances generated a reasonable explanation. Firstly, it should be recorded from paragraph 3.6.1 that the 'process cycle time' is obtained by dividing the total processing time of each 'job' by its batch size, where processing time includes the actual production time plus all the idle times (due to setup) suffered by each job. For this reason the value of process cycle time is a direct function of the amount of idle time suffered by each job. Now, if the results of table 7.1 are examined, it is possible to see that although the total time spent with setup has increased (from 6.45% to 6.77%), the amount of setup time per job has actually been reduced. This is because the increase in the nominal load factor from 65% to 85% was brought about by an increase of 30% in the arrival rate. Now, if the two values of setup time are divided by 1.0 and 1.3, respectively, the results are 6.45% and 5.21%, which indicate that the amount of setup time per 'job' actually drops, with a consequent reduction in the 'process cycle time'.

A final point about the mechanisms governing the behaviour of the system relates to the relationship between the average number of jobs in queue and the delivery performance. From the results it can be seen that in general a deterioration (or improvement) in delivery performance corresponds to an increase (or reduction) in the queue size. There were however two exceptions. The first happened when the average size of orders was increased from 1000 to 1600 (table 7.4). On that occasion a significant drop in the queue size was followed by a significant deterioration in delivery performance. This is understandable as the increase in the average size of orders corresponds to a reduction in the arrival rate of orders, which means less jobs in queue but larger batch sizes (larger processing times per batch).

The second exception happened when the value of MAXLOT (table 7.5) was changed from 450 to ∞. This meant that jobs were not split into batches, which means less batches (jobs) and so, smaller queues. The reduction in queue size however was not followed by any significant change in delivery performance.

7.4 - Summary

In this chapter the results of a fractional factorial design corresponding to six factors at two levels each, was presented and discussed.

The objectives of these experiments were to study the main effects and possible interactions of those six variables on the behaviour of the system, both in relation to its internal and external (delivery) performance.

The results indicated that on average the largest main effect on delivery performance was caused by a variation in the nominal load factor brought about by an increase in the arrival rate. The value of the average size of orders was also shown to have a considerable influence on the delivery performance and internal behaviour of the system.

There was a strong indication to suggest that the procedure of splitting large jobs into smaller batches had, on average, a very small effect on the delivery performance of the system, but a significant effect on its queue size.

The results have also indicated a large interaction between the number of moulds and the ratio F (number of product style/ number of machines) for the measures of delivery performance. These interactions were shown to be highly significant and of considerable practical importance due to their magnitude in relation to the other effects.

A comparison between the measures of internal behaviour and the measures of delivery performance indicated that in some circumstances an improvement in the internal performance can correspond to a deterioration of delivery (external) performance.

Finally a brief discussion was carried out in order to explain the working mechanism of the system, and to justify some of the more unexpected effects.

TABLE 7.1

EFFECTS OF A CHANGE IN THE NOMINAL LOAD FACTOR OF THE SYSTEM

(low: 65% vs. high: 85%) (A_o vs. A_1)

	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE PROCESS CYCLE TIME	AVERAGE ACTUAL LOAD FACTOR	IDLE TIME DUE TO SETUP (%)	AVERAGE NO. OF JOBS IN QUEUE
MAIN EFFECTS	** 2.53	** 18.73	** 1.25	** – 0.13	** 18.99	** 0.32	** 29.74
Interactions with:							
B: mean setup time	0.17	1.78	0.15	** – 0.06	– 0.28	** – 0.16	** 1.39
C: no. of moulds	0.20	0.68	0.22	** 0.15	1.02	** 0.21	** 2.31
D: average size of orders	– 0.07	– 0.03	0.09	** 0.06	0.00	** 0.27	** – 6.15
E: splitting of jobs	– 0.04	– 0.24	0.00	0.00	– 0.08	– 0.06	– 0.44
F: ratio no.style/no. mach.	** 0.41	** 3.56	** 0.30	** – 0.13	– 0.75	** – 0.25	** – 2.95

** Significant at 0.01 level
* Significant at 0.05 level

TABLE 7.2

EFFECTS OF A CHANGE IN THE MEAN VALUE OF SETUP TIMES

(low: 8 min. vs. high: 16 min.)

(B_o vs. B_1)

	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE PROCESS CYCLE TIME	AVERAGE ACTUAL LOAD FACTOR	IDLE TIME DUE TO SETUP (%)	AVERAGE NO. OF JOBS IN QUEUE
MAIN EFFECTS	** 0.52	** 2.95	* 0.22	** 0.35	** 4.90	** 3.43	** 3.38
Interactions with:							
A: nominal load factor	0.17	1.78	0.15	** – 0.06	– 0.28	** – 0.16	** 1.39
C: no. of moulds	– 0.12	– 0.67	– 0.01	** – 0.11	** – 1.41	** – 1.15	– 0.76
D: average size of orders	– 0.02	– 0.04	0.01	* – 0.04	– 0.47	** – 0.54	– 0.79
E: splitting of jobs	– 0.07	– 0.29	– 0.05	0.02	0.39	0.07	– 0.66
F: ratio no. style/no.mach.	0.20	1.20	0.11	** 0.08	1.12	** 0.95	0.39

** Significant at 0.01 level
* Significant at 0.05 level

TABLE 7.3

EFFECTS OF A CHANGE IN THE NUMBER OF MOULDS AVAILABLE

(42 vs. 27) (C_o vs. C_1)

	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE PROCESS CYCLE TIME	AVERAGE ACTUAL LOAD FACTOR	IDLE TIME DUE TO SETUP (%)	AVERAGE NO OF JOBS IN QUEUE
MAIN EFFECTS	** 0.86	** 7.36	** 0.53	** – 0.74	** – 10.30	** – 3.37	** 8.72
Interactions with:							
A: nominal load factor	0.20	0.68	0.22	** 0.15	1.02	** 0.21	** 2.31
B: mean setup time	– 0.12	– 0.67	– 0.01	** – 0.11	* – 1.41	** – 1.15	– 0.76
D: average size of orders	0.19	1.80	0.15	0.03	0.81	** 0.51	– 0.22
E: splitting of jobs	– 0.06	– 0.50	– 0.02	0.00	– 0.08	0.09	– 0.06
F: ratio no.style/no. mach.	** – 0.71	** – 5.16	** – 0.45	** 0.12	** 2.22	** 0.78	** – 8.03

** Significant at 0.01 level
* Significant at 0.05 level

TABLE 7.4

EFFECTS OF A CHANGE ON THE AVERAGE SIZE OF ORDERS
(small: 1,000 vs. large: 1,600) (D_o vs. D_1)

	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE PROCESS CYCLE TIME	AVERAGE ACTUAL LOAD FACTOR	IDLE TIME DUE TO SETUP (%)	AVERAGE NO OF JOBS IN QUEUE
MAIN EFFECTS	** 0.98	** 5.45	** 0.28	** – 0.22	** – 7.55	** – 2.46	** – 11.06
Interactions with:							
A: nominal load factor	– 0.07	– 0.03	0.09	** 0.06	0.00	** 0.27	** – 6.15
B: mean setup time	– 0.02	– 0.04	0.01	** – 0.04	– 0.47	** 0.54	– 0.79
C: no. of moulds	0.19	1.80	0.15	* 0.03	0.81	** 0.51	– 0.22
E: splitting of jobs	0.05	0.40	0.00	0.00	0.23	– 0.03	– 0.44
F: ratio no.style/no. mach.	– 0.18	– 0.70	0.09	– 0.02	– 0.17	– 0.39	0.27

** Significant at 0.01 level
* Significant at 0.05 level

TABLE 7.5

EFFECTS OF A CHANGE IN THE PROCEDURE OF SPLITTING JOBS
(splitting vs. no splitting) (E_o vs. E_1)

	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE PROCESS CYCLE TIME	AVERAGE ACTUAL LOAD FACTOR	IDLE TIME DUE TO SETUP (%)	AVERAGE NO OF JOBS IN QUEUE
MAIN EFFECTS	0.11	0.48	0.02	0.02	0.29	− 0.04	** − 2.70
Interactions with:							
A: nominal load factor	− 0.04	− 0.24	0.00	0.00	− 0.08	− 0.06	− 0.44
B: mean setup time	− 0.07	− 0.29	− 0.05	0.02	0.39	0.07	− 0.66
C: no. of moulds	− 0.06	− 0.50	− 0.02	0.00	− 0.08	0.09	− 0.06
D: average size of orders	0.05	0.40	0.00	0.00	0.23	− 0.03	− 0.44
F: ratio no. style/no.mach.	− 0.03	− 0.06	0.00	0.00	0.04	− 0.07	0.40

** Significant at 0.01 level
* Significant at 0.05 level

TABLE 7.6

EFFECTS OF A CHANGE IN THE RATIO NO. OF STYLES/ NO. OF MACHINES

(more favourable (3:2) vs. less favourable (3:1)) (F_o vs. F_1)

	AVERAGE DELAY OF PRODUCTION	PERCENTAGE OF PRODUCTION LATE	TARDINESS INDEX OF PRODUCTION	AVERAGE PROCESS CYCLE TIME	AVERAGE ACTUAL LOAD FACTOR	IDLE TIME DUE TO SETUP (%)	AVERAGE NO OF JOBS IN QUEUE
MAIN EFFECTS	** 0.72	** 6.64	** 0.36	** 0.54	** 6.47	** 2.70	** – 7.38
Interactions with:							
A: nominal load factor	** 0.41	** 3.56	** 0.30	** – 0.13	– 0.75	** – 0.25	** – 2.95
B: mean setup time	0.20	1.20	0.11	** 0.08	1.12	** 0.95	0.39
C: no. of moulds	** – 0.71	** – 5.16	** – 0.45	** 0.12	** 2.22	** 0.78	** – 8.03
D: average size of orders	– 0.18	– 0.70	0.09	– 0.02	– 0.17	** – 0.39	0.27
E: splitting of jobs	– 0.03	– 0.06	0.00	0.00	0.04	0.07	0.40

** Significant at 0.01 level
* Significant at 0.05 level

FIGURE 7.7

EFFECT OF MAXLOT VALUE ON THE BEHAVIOUR OF THE SYSTEM
(SYSTEM CONFIGURATION abcdf)

VALUE OF MAXLOT	AVERAGE NUMBER OF JOBS IN QUEUE		AVERAGE DELAY OF PRODUCTION		PERCENTAGE OF LATE PRODUCTION		TARDINESS INDEX OF PRODUCTION	
	Mean	difference from lowest mean (control)	mean	difference from lowest mean (control)	mean	difference from lowest mean (control)	mean	difference from lowest mean (control)
∞	35.28	0.00	8.38	* 0.16	40.65	0.97	2.37	0.10
550	37.92	** 2.65	8.26	0.04	40.27	0.59	2.27	0.01
450	38.65	** 3.37	8.22	0.00	39.68	0.00	2.26	0.00
350	40.87	** 5.59	8.42	** 0.20	40.79	1.11	2.42	* 0.15
250	46.68	**11.40	8.46	** 0.24	42.57	** 2.89	2.39	* 0.13

** = significant at 0.01 level
* = significant at 0.05 level

FIGURE 7.8

EFFECT OF MAXLOT VALUE ON THE BEHAVIOUR OF THE SYSTEM
(SYSTEM CONFIGURATION d)

VALUE OF MAXLOT	AVERAGE NUMBER OF JOBS IN QUEUE		AVERAGE DELAY OF PRODUCTION		PERCENTAGE OF LATE PRODUCTION		TARDINESS INDEX OF PRODUCTION	
	mean	difference from lowest mean (control)	mean	difference from lowest mean (control)	mean	difference from lowest mean (control)	mean	difference from lowest mean (control)
∞	11.17	0.00	4.24	** 0.73	6.07	** 3.36	0.13	** 0.08
550	12.19	** 1.02	3.81	** 0.30	3.98	* 1.27	0.07	0.02
450	12.73	** 1.56	3.76	** 0.25	3.79	1.09	0.06	0.01
350	14.61	** 3.44	3.61	** 0.11	2.71	0.00	0.05	0.00
250	18.58	** 7.41	3.51	0.00	2.76	0.05	0.06	0.01

** = significant at 0.01 level

* = significant at 0.05 level

CHAPTER 8

STUDY OF OPERATION STRATEGIES FOR CAPACITY MANIPULATION

8.1 - Introduction

In paragraph 4.3.3 a brief discussion was conducted on the objectives of this series of experiments. It was however indicated that a more comprehensive discussion would follow in this chapter.

The two previous series of experiments, reported in chapters 6 and 7, have presented a comprehensive study of the effects of operation rules and some of the system's parameters on its internal behaviour and delivery performance. The information obtained helped to improve the understanding of the system and to indicate the effects and interactions between parameters. However it did not relate those effects to the economic consequences of changes in the parameters.

Cantelow et al.(1973) point out that two measures are of real importance when judging the performance of a manufacturing system. Those are the delivery performance to promised dates, and the cost per unit of output. The aim of a company's policy should be to meet demand on time at minimum cost - within such practical constraints as limit of available cash, floor space and manning policies. The fulfilment of these aims however depends to a large extent on the decisions taken by management with respect to operation procedures and capital investment.

In order to take such decisions management need to assess the consequences of the possible alternative actions. In situations where due dates are viewed as critical, management can usually take decisions using a three dimensional decision grid. The first dimension represents cost per unit of output. This cost can be changed by modifying its variable component, which is represented by such variables as capital costs, cost of overtime and extra shift, inventory costs, etc. As Folie (1974) pointed out, 'Firms can generally substitute variable inputs, such as additional shifts, overtime and inventory, for capital inputs in the form of plant capacity, which consists of buildings and production equipments'. It is also possible to monitor the utilization factor of the production facilities by increasing or decreasing the load factor.

The second dimension on the decision grid represents the 'lead time' used for quoting delivery dates. Eilon and Chowdhury (1976) suggest that instead of confining the scheduler to a given array of due dates, the due dates should be specified so as to take account of individual jobs and the level of congestion in the shop. They then add : 'In practice, of course, the scheduler is not free to assign due dates on his own, and the wishes of the customer in this respect undoubtedly play a significant part'.

The third dimension on the decision grid represents the delivery performance. Delivery performance can be measured by the number of tardy jobs, or any other tardiness-related criterion. Having decided on a given 'lead time' parameter, it is always possible to trade extra capacity, which might be represented by cost per unit of output, for improved delivery performance. Littlechild (1974) expressed this point well: 'Queueing theorists have long argued that a less than

100% utilization factor does not represent inefficient idle capacity. Investment in additional facilities (capacity) to service a given demand is warranted as long as the incremental cost is more than outweighted by reduction in waiting cost'.

One way of providing management with information to help in making the above decisions would be to construct trade-off curves relating capacity cost per unit produced, to delivery performance criteria, for different lead time parameters. As far as this particular model is concerned, capacity can be manipulated through the following variables:

i) the number of machines

ii) the number of moulds

iii) the number of working hours/week

iv) the use of inventory of finished goods

Capacity cost per unit produced can be expressed by:

$$C_u = C_t/Q_t \qquad \dots 8.1$$

where

C_u = capacity cost per unit produced

C_t = total capacity cost/period

Q_t = total production/period

C_t can be measured by:

$$C_t = M \cdot C_m + N \cdot C_n + L_c + I_c \qquad \dots 8.2$$

where

M = number of machines in the shop

C_m = machine depreciation cost/period

N = number of moulds held

C_n = mould depreciation cost/period

L_c= labour cost/period

I_c= total inventory cost/period

Delivery performance can be measured by the 'percentage of production (or orders) late' and/or the 'tardiness index of production(or orders)', and due dates can be varied by having different values for the lead time, D.

In this chapter the results obtained from a series of experiments designed to produce trade-off curves relating delivery performance to capacity cost per unit, are presented. As explained briefly in paragraph 4.3.3, the experimental design for this part of the study consists of making single changes to the model's capacity parameters and of measuring the consequences of those changes on both the delivery performance and the capacity cost per unit produced. Delivery performance (measured by the 'percentage of production late') is obtained directly from the model output, but the capacity cost per unit is calculated afterwards, outside the model, through the use of equations 8.1 and 8.2. The calculation of the capacity cost per unit outside the model gives more flexibility, as it allows the use of different cost parameters on the same set of results obtained by a single simulation experiment.

The cost parameters used in the results presented in this chapter are based on typical costs of the industrial company mentioned in paragraph 4.2. These costs are presented in appendix 2 together with data about the company's production unit. It should be pointed out that the results of the trade-off curves are very much dependent on the cost structure as well as the system configurations of a particular company (as indicated by some of the results of chapter 7). For this reason it was decided to use a system configuration resembling the situation found in the industrial company referred to above. The configuration used has therefore; a mean setup time equal to 8 minutes, three product styles, one machine, average size of orders equal 1600, and a value for the mean arrival rate sufficient to produce a load factor of around 75% for a single shift of 45 hours per week. The procedure for splitting jobs was also used. The number of moulds held by the company, the number of working hours per week and the parameters for inventory control are the variables changed during the experimentation. The number of machines was not varied, as it would produce too high a capacity cost at this level of demand.

8.2 - Discussion of experiments and presentation of results

For the sake of discussion and presentation, the experiments were divided into two groups. The first group relates to the experiments involving variations in the number of moulds and number of working hours per week, while the second group relates to the experiments involving changes in the inventory control parameters.

8.2.1 - Experiments involving changes in the number of moulds and working hours per week

The results for this first group of experiments are presented in tables 8.1 to 8.3, and in figures 8.1 to 8.5. Table 8.1 presents the numerical results of 'capacity cost per unit' and 'percentage of production late' which resulted from variation in the number of moulds from 18 to 45. The variations in the number of moulds were made in steps of three, in order to take account of the three product styles. This change could have been made in steps of 1, but in this particular situation, in which demand is assumed to be the same for all three styles, the selection of an extra mould for a particular product would tend to create an imbalance in the delivery performance among the styles. The extra moulds were selected in accordance with the procedure described in paragraph 3.6.1.[1] It should be pointed out that experiment 1 of table 8.1 (1 machine, 18 moulds) represents the minimum feasible capacity for a three-product-styles situation (see paragraph 3.5.3). The percentage of late production was calculated

(1) Both the waiting time in queue and processing time distributions were assumed to be exponential, when calculating the probabilities described in paragraph 3.6.1

for different values of lead time, D, ranging from 8 to 20 days, as shown in table 8.1. In figure 8.1 the capacity cost per unit is plotted against the corresponding percentage of production late, for three different values of lead time (D = 8, 12 and 16 days). This gives rise to the trade-off curves, which represent the three dimensional decision grid discussed in paragraph 8.1. Figure 8.1 confirms the results of the preliminary investigation concerning the effect of extra moulds on the delivery performance of the system (see paragraph 4.2.4.). It is possible to see that, irrespective of the value of lead time, the trade-off curve has an interesting shape, which decreases very sharply when the number of moulds increase from 18 to 27, but which tends to flatten out after 27 moulds.(See table 8.1)

For example, for D = 12 days and the number of moulds set at 18, the result for the percentage of production late is equal to 49.19%, at a capacity cost of 3.97 m.u. (monetary unit) per unit produced. When the number of moulds was increased to 27, there was a trade-off between costs and delivery performance which is indicated by a sharp drop in the percentage of production late, which is reduced to less than a third of its original value (from 49.19% to 15.13%), against an increase of only 4.79% in the capacity cost per unit produced (from 3.97 m.u. to 4.16 m.u.). On the other hand, when the number of moulds was increased from 27 to 45, the increase in cost was equal to 18.51% (from 4.16 m.u. to 4.93 m.u.) compared with an absolute drop of only 0.59% (3.90% in relative terms) in the percentage of production late.

Another interesting point to note is the influence of D on the percentage of orders late. For example, by increasing the number of moulds from 18 to 24, and the value of D from 8 to 12 days, it would be possible to reduce the percentage of production late from 68.19% to 21.60% at an extra capacity cost per unit of only 0.05 m.u. or 1.26%. It is true that increasing the lead time means reducing customer's satisfaction in terms of delivery dates. On the other hand, by increasing the lead time the company will be more reliable in respect to its promise, without having to increase the cost of the product. As long as it is arranged before hand, it might be preferable to a customer to have slightly longer but more reliable delivery times.

Table 8.2 presents similar data to that shown in table 8.1, but this time it represents the results obtained by varying the number of working hours/week. The variations were made in steps of 5 hours per week, representing one extra hour per day in a five-day-week. It should be pointed out that eighty hours per week represents two normal shifts of 40 hours each, and for this reason labour cost is charged at the appropriate rate indicated in appendix 2. The number of moulds was maintained constant at 18 during this series of experiments. In figure 8.2 the capacity costs per unit given by table 8.2 are plotted against the corresponding 'percentage of production late', for three different values of lead time (D = 8, 12 and 16 days). Figure 8.2 shows an interesting pattern for the trade-off curves, which is a characteristic of the cost structure. For example, for a lead time equal to 8 days it can be seen that the trade-off curve has three distinct sections. The first section, corresponding to the increase in the number of working hours per week from 45 to 70 hours, presents an almost linear pattern with a sharp decline. The second section, which corresponds to the increase in the number of working hours per week from 70 to 80, presents a vertical drop. This means an improved delivery performance without any corresponding increases in costs. This phenomenon happens only because of the cost structure, which makes it less expensive to have two shifts of 40 hours (80 hours in total) than one shift of 40 hours plus 30 hours overtime. For this reason, when calculating labour cost, it was assumed that the company would rather have a two shift system than 30 hours overtime, and this causes the vertical drop on the trade-off curve. The third section of the curve, which corresponds to the variation in the number of working hours per week from 80 (two shifts) to 100 (two shifts plus 20 hours overtime) assumes an assymptotic shape indicating that a large increase in costs would be needed in order to obtain a small improvement in delivery performance. If the three sections of the curve are compared numerically it can be seen that for the first section an increase of 17.63% (0.70 m.u. in absolute terms) in the capacity cost per unit resulted in a drop of 52.03% (from 68.19% to 16.16%) in the percentage of production late. In the second section the use of two shifts of 40 hours, instead of one shift of 40 hours plus overtime, resulted in a drop of 8.32% (from 16.16% to 7.84%) in the percentage of production late at no extra cost. For the third and last section of the curve an increase of 15.20% (0.71 m.u. in absolute terms) in the capacity cost per unit resulted in a reduction of only 6.66% (from 7.84% to 1.18%) in the percentage of production late.

The variation in the value of D caused the trade-off curves to shift to the left, and consequently to produce a lower 'percentage of production late'. For example, for a lead time of 16 days and a capacity cost per unit of 4.38 m.u., the system would deliver only 6.84% of its production late, while for the same cost and a lead time of 8 days, the system would deliver on average 36.20% of its production late.

Table 8.3 presents the results of capacity cost per unit and 'percentage of production late' for different values of lead time, D, which were obtained by varying the number of working hours per week from 45 to 80 hours, in steps of 5 hours. The number of moulds was maintained constant at 27 during this series of experiments. This 'strategy', a combination of the previous two, increases both the number of moulds (from 18 to 27) and the number of hours per week.

Figure 8.3 shows the values of capacity cost per unit (of table 8.3) plotted against their corresponding 'percentage of production late' for three different values of lead time (D = 8, 12 and 16 days). The pattern of the trade-off curves of figure 8.3 are similar to the corresponding curves of figure 8.2. However the curves of figure 8.3 show lower values for the 'percentage of production late', than the corresponding curves of figure 8.2 for the same unit costs.

In order to facilitate comparisons between the different strategies, the trade-off curves of figures 8.1, 8.2 and 8.3 were plotted together in figures 8.4 and 8.5. Figure 8.4 presents the trade-off curves of figures 8.1, 8.2 and 8.3, for D = 8 days, and figure 8.5 presents the same three curves but for D = 16 days. An analysis of figures 8.4 and 8.5 indicates that the relative performance of the different strategies for capacity manipulation depends on the value of lead time, D, and the level of delivery performance (or capacity cost per unit) that a company chooses to have.

For example, if it is decided that a company should fix its due date based on a lead time of 8 days, and that it intends to deliver between 90% and 95% of its production within the due date, then it would not make much difference (as far as the capacity cost per unit is concerned), whether 18 moulds and 2 shifts of 40 hours, or 27 moulds, one shift of 40 hours and 15 or 20 hours of overtime per week were used. It seems however that the second option would produce a smaller tardiness as shown by the analysis of the two curves. On the other hand, if a company decides that it would be worthwile to sacrifice performance in terms of 'production delivered late' in exchange for lower costs for the same lead time of 8 days, then the use of 27 moulds and some amount of overtime seems to be the best alternative. For example, for a capacity cost of 4.33 m.u., it would be possible to obtain on average more than 80% of all production delivered inside the due date if 27 moulds and 10 hours of overtime per week are used. On the other hand, if 18 moulds and 20 hours of overtime are used, the same capacity cost per unit would be incurred but only 61% of the production would be delivered on time.

Another alternative would be to increase the lead time D. As shown in figure 8.5 if the lead time is fixed at 16 days then the use of 18 moulds and overtime appears to be a worse alternative than either increasing the number of moulds to 27 or having 27 moulds and overtime. For example, it would be possible to delivery 100% of the production inside the due date at a capacity cost per unit of 4.40 m.u. if 27 moulds and 15 hours overtime are used. To obtain the same 100% performance with only 18 moulds would require nearly 30 hours overtime(or two shifts of 40 hours) at a cost of 4.75 m.u. per unit produced.

Finally, it should be pointed out that the addition of a new machine to the system would result in a capacity cost per unit of 6.92 m.u., which, irrespective of the delivery performance, would be far more expensive than any other strategy.

8.2.2. - Experiments involving changes in the parameters of inventory control

Before the presentation of results, some comments have to be made about the limitations of this series of experiments. The first limitation concerns the problem of interference between inventory replenishment orders and customer's orders. In paragraph 3.4 it was pointed out that when switched on, the inventory subsystem would tend to interfere with the priority scheduling rules and the whole of the shop's scheduling procedure, which would have to handle both customers and inventory replenishment jobs, as they would be competing for the same production facilities. This possible interference means that a priority scheduling rule which is efficient for a non-inventory situation might not be so, when the inventory is switched-on. A full study of the problem would require a lengthy investigation, which was not carried out in this study for reasons already discussed in paragraph 3.4. Instead it was decided to use the FIFOMB rule together with a procedure which separates customer's 'jobs' from inventory replenishment 'jobs', with the former getting absolute priority over the latter, as described at the end of paragraph 3.5.1.

Secondly there is the problem of selecting the parameters of control for the inventory subsystem. As discussed in 3.4., the inventory subsystem is controlled by three sets of variables. The first set (STOCK (i,j)) is used to specify whether or not product (i,j) is manufactured for stock; the second set (RPOINT (i,j)), is used to specify the reorder point of product (i,j); and the third set (EBQ (i,j)) is used to specify the batch sizes for the replenishment orders of product (i,j). In paragraph 1.2.3 the analytical approaches to the problem of determining reorder point and reorder batch quantities, have been discussed. It is evident from the characteristics of this class of production systems, that the economic batch quantity approach would be completely inadequate, and a mathematical approach of the (s,S) type too complex. It was therefore decided to use an experimental approach based on a simple heuristic search procedure, to determine effective values for RPOINT (i,j) and EBQ (i,j). This procedure will be discussed later in this chapter.

Finally there was a need to decide which products should be manufactured for stock. As described in paragraph 3.4, the objective of using inventory in this study was not to eliminate or decrease the typical values of lead time (delivery delay promises), but instead to improve the efficiency in meeting those promises. Following this objective, it was decided that the decision on which products to produce for inventory should be based on the same procedure used for selecting extra moulds. This procedure has been described in paragraph 3.6.1. The idea was to examine the outputs from the previous series of experiments and to determine which were the product sizes with substantial probabilities of delaying orders associated with their style. As the objective is to minimize both capacity costs and percentage of production late, it would be desirable to hold stock for the minimum possible number of products, such that products with very small demand (the extreme sizes in the range, both large and small) and, consequently, small throughput times, could be excluded from the list of inventoriable items without causing any delay on delivery (above the lead time). Examination of the outputs of previous experiments indicated that the two largest and three smallest product sizes in each range had probabilities of less than 0.03 of having throughput times larger than 8 days. Therefore it was decided that they should not be produced for stock. It should be pointed out that this analysis was made for a system configuration with 18 moulds, 45 working hours per week and no inventory. The introduction of inventory would most certainly alter those probabilities, which should be reduced as the amount in stock for the other products is increased. This is a consequence of the fact that customer's orders have absolute priority over stock replenishment orders. Later checks on the results obtained have confirmed that the throughput times of non inventoriable products are indeed reduced when inventory is introduced into the system.

Having decided which products to keep in stock, there was a need to choose the reorder point and reorder batch size parameters. As said before, this was done by a simple heuristic search procedure which consisted of the following steps:

i) Choose a value for the maximum amount to be kept in stock (at any time) for each product style. This maximum is given by the relationship:

$$MAX(i) = \sum^{j} RPOINT\ (i,j) + \sum^{j}\ EBQ(i,j)$$

where i = product style and j = product size

ii) After deciding on the value of MAX (i) calculate the values MAXS(i,j) for each product size, where MAXS (i,j) = MAX (i) X p(j), where p(j) represents the distribution of proportions of demand, for individual product sizes, as described in paragraph 3.2 , and presented in appendix 3.

iii) Try different combinations of RPOINT (i,j) and EBQ (i,j), maintaining the constraint that RPOINT (i,j) + EBQ (i,j) = MAX (i). For most cases only three combinations were tried:

a) RPOINT (i,j) = EBQ (i,j)

b) RPOINT (i,j) = 3 X EBQ (i,j)

c) RPOINT (i,j) = $\frac{1}{3}$ X EBQ (i,j)

Each combination represents an experiment with the model.

iv) Increase the value of MAX (i) and repeat the procedure as from (ii).

A further problem with this series of experiments concerns the question of stabilization and starting conditions. As pointed out in paragraph 5.2.1, the discussions of chapter 5 about initial conditions and stabilization period assumed a system without inventory of finished goods. The existence of inventory tends to create a buffer in the system, which might bias the final results if the initial conditions, the stabilization period and the length of runs are not well considered.

In order to made comparisons with other experiments easier, it would be ideal to have identical samples sizes for all experiments. This would be possible for small values of MAX (i). However when the value of MAX (i) gets larger, there is a need to increase the length of the stabilization period and/or the sampling period. The best way to determine the proper parameters would be to have a pilot study, but unfortunately because of time restriction such a study could not be made. Instead it was decided to use a technique described in paragraph 5.3.3. This technique consists of having two sets of runs, in which the length of the runs for the two sets is varied so that the results can be compared to see whether or not there is any considerable difference between the two sets of runs. If the answer is negative, then the original (smaller) length is good enough, otherwise a longer run should be tried. It would not be feasible to use this procedure for all experiments, and therefore an intermediary value of MAX (i) was tested in order to determine typical values for the stabilization and sampling periods. Values of MAX (i) above that would have proportionately longer runs, while values below it would have smaller runs.

In tables 8.4 and 8.5 the lengths of the stabilization and sampling periods for each experiment are presented, together with results of costs and delivery performance. Experiments 42, 43 and 44 of table 8.4 are the experiments for which the test of stabilization was made. It can be seen from those three experiments that the increase of the stabilization period from 20 to 40 and the sampling period from 130 to 180 caused a considerable change in the outputs (percentage of production late), while a further increase from 40 to 50 in the stabilization period and from 180 to 260 in the sampling period, failed to produce any considerable changes in the output. The initial conditions for all the experiments were the same, as discussed in paragraph 5.2.1.4, but with the amount in stock for each product set arbitrarily to the value of the reorder point, such that the arrival of the first order would generate an issue of stock replenishment orders.

Considering all the limitations discussed above, it is clear that these series of experiments should be seen more as an exploratory exercise, whose main objective has been to lift the veil of what looks to be an interesting point for further research.

The whole series consisted of 27 experiments which can be divided into three groups. The first and second groups, whose results are presented in table 8.4, represent the experiments made with a system configuration having 18 moulds, while the third group consists of 4 experiments, in which two correspond to a system configuration having 24 moulds, and the other two correspond to a system configuration having 27 moulds. The results of those experiments are presented in table 8.5.

The first group of experiments of table 8.4 consists of 17 experiments (from number 29 to 45) in which MAX (i) have taken the values of 1600; 3200; 4800; 6400 and 9600. For each value of MAX (i) three combinations of RPOINT (i,j) and EBQ (i,j) were tried, as shown in table 8.4 and indicated previously. It should be pointed out that the values of reorder point and reorder quantity of table 8.4 represent respectively the summation of all RPOINT (i,j) and EBQ (i,j) for a given product style I. The individual values of RPOINT (i,j) and EBQ (i,j) for each product size can be determined by multiplying the values of reorder point and reorder quantity by the corresponding distribution of proportions of demand p(j) presented in appendix 3.

The second group of experiments consisted of 6 experiments in which the value of MAX (i) was fixed at a value of 24000, and six different combinations of RPOINT (i,j) and EBQ (i,j) were tried in order to have a more critical analysis of the effects of those two variables on the cost and delivery performance.

In order to facilitate the analysis, the results of experiments from groups 1 and 2 have been plotted in figure 8.6. Figure 8.6 relates the 'percentage of production late' (for a lead time D = 8 days), to capacity cost per unit produced, for different values of MAX (i) and combinations of RPOINT (i,j) and EBQ (i,j). An analysis of this figure indicates the effects of the relationship RPOINT (i,j) vs. EBQ (i,j) on the delivery performance. This can be seen more clearly in the second group of experiments (46 to 51).

The results indicate that for a value of MAX (i) = 24000, the variation in the ratio RPOINT : EBQ, from 11:1 to 1:1, caused the percentage of production late first to drop (up to the point where the ratio is 2:1) and then to start rising as the ratio gets smaller. It is interesting to note the corresponding variation in costs. Independently of the ratio between reorder point and reorder quantity, as the delivery performance gets worse, costs get smaller. This is caused by variations in the average stock level, which gets smaller as the delivery performance gets worse. This is however not the case for smaller values of MAX (i). For example, from experiments 31 to 44, costs tend to decrease as the delivery performance improves. This can be seen by the dotted line of figure 8.6. For example, in experiment 31, capacity cost per unit was equal to 3.99 m.u. and the percentage of production late 64.54%. In experiment 44 on the other hand, costs were reduced to 3.89 m.u. together with a reduction in the percentage of production late, which came down to 26.61%. This reduction in costs happened in spite of increases in the average stock level (column 4 of table 8.4). This means that the increase in stock levels has been outweighed by the increase in the total amount produced by the system, which by holding stocks has increased its actual capacity of production.

The results of the third series of experiments, presented in table 8.5, show that for a value of MAX (i) =16000 the system would deliver 97.71% of its production inside the 8 days lead time, if the system was working with 27 moulds. The capacity cost per unit would be equal to 4.43 m.u. Similar results in terms of delivery performance would be obtained for a system configuration having 24 moulds. The cost however would be only 4.30 m.u.

An interesting point to note in all these results is that a considerable amount of stock would be needed in order to obtain a good delivery performance (above 90% of production delivered within the due dates). For the case of 18 moulds the average stock level would be around 30000 units (experiment 51) for a percentage of production late equal to 9% or 42000 for a 3% production late. This amount of stock represents between 6 and 9 weeks of demand. The same level of delivery performance (3%) for the case of 24 or 27 moulds would require an average stock level of 26000 units, which represents about 5.5 weeks of demand.

8.3 - Discussion of results

One of the aims of this part of the investigation was to compare different strategies for capacity manipulation. In accordance with this objective it would be desirable to compare the results of the first series of experiments (modifications in the number of moulds and amount of working hours per week) with the results of the second series (modification in the parameters of inventory control). This comparison is made in figures 8.7 and 8.8 where trade-off curves for inventory (with 18 moulds) is represented by the profile of minimum cost shown by the dotted line of figure 8.6. Figure 8.7 makes comparisons between the trade-off curves for a lead time D = 8 days, and in figure 8.8 the curves are compared for a lead time D = 16 days. Both figures indicate that the use of inventory tends to produce smaller values for the percentage of production late than the other strategies, at a comparable cost. For example, if the aim is to obtain a delivery performance corresponding to 90% of all production delivered inside an 8 days lead time, it would be possible to achieve this by having 24 moulds and an average stock level of 12031 items (or 2.5 weeks of stock), at a capacity cost per unit of 4.13 m.u. This compares with a cost of 4.55 m.u. for a strategy of 27 moulds and 10 hours overtime per week.

It should be pointed out that the relative performances of the various strategies are very much dependent on the cost structure used. For example, in these experiments data from an industrial company was used, in which the inventory cost was calculated at a 25% flat rate of the total production cost, on the average stock level. This is a crude estimate of costs and therefore conclusions should be carefully judged against this background. On the other hand, the fact that the calculation of costs is made afterwards, allows the analyst to try different costs structures on the same data.

Finally, it should be pointed out that statistical tests could be made on the data if desired. It is possible, for example, by the use of Tukey's multiple comparison test, to test the statistical significance of differences between delivery performance of different strategies which have statistically equivalent costs, or otherwise to test the statistical significance of differences in costs for strategies which have equivalent delivery performances. However in view of the particular nature of the data costs and system configuration, it was thought that statistical tests would bring no additional relevant information to the conclusions, and therefore they have not been applied to these results.

8.4 - Summary

In this chapter the relationship between production capacity and delivery performance was discussed, with the objective of demonstrating how production capacity can be trade-off against delivery performance.

The results of a series of experiments were presented and trade-off curves relating capacity cost per unit produced, to percentage of production late, were drawn for different values of lead time D, used to fix due dates.

The variation in production capacity was brought about by changes in some of the models' parameters, viz. the number of moulds, the number of working hours per week, and the amount of items kept in stock.

Results indicated that the effectiveness of a given strategy depends on the level of delivery performance desired, and on the value of lead time, D, used to quote due dates. Discussions have indicated that management can use the trade-off curves in order to make strategic decisions in a three dimensional grid, which has costs, delivery performance and the length of delivery promises as the decision parameters.

Comparisons of different strategies indicated that, for the particular cost parameters used, the utilization of finished goods inventory produces better results than the utilization of overtime and extra moulds. It was however pointed out that a considerable amount of stock would be needed in order to obtain good performances, and therefore, when deciding on a strategy, consideration should be given to this aspect of the inventory strategy.

TABLE 8.1

RESULTS OF TRADE-OFF BETWEEN CAPACITY COST/UNIT AND DELIVERY PERFORMANCE

(VARIATION IN THE NUMBER OF MOULDS)
(NO. OF HOURS/WEEK CONSTANT AT 45)

EXPERIMENT NO.	NO. OF MOULDS	CAPACITY COST/UNIT	PERCENTAGE OF PRODUCTION LATE — LEAD TIME (DAYS) 8	10	12	14	16	18	20
1	18	3.97	68.19	59.20	49.19	42.49	35.73	31.04	24.63
2	21	4.02	64.12	52.86	44.26	34.62	28.32	24.09	19.64
3	24	4.02	44.78	32.85	21.60	14.98	9.50	5.79	3.80
4	27	4.16	34.94	24.50	15.13	9.75	6.26	3.88	2.77
5	30	4.29	32.91	22.74	14.62	9.97	6.48	2.98	1.32
6	33	4.42	34.01	23.11	15.44	9.86	6.40	3.26	2.16
7	36	4.55	29.41	20.07	14.07	7.96	5.82	3.26	2.70
8	39	4.67	29.14	20.03	14.15	8.15	6.22	3.15	2.60
9	42	4.79	28.90	19.77	14.23	9.14	6.52	3.34	2.08
10	45	4.93	29.56	20.78	14.54	9.50	7.49	4.01	2.26

System Configuration a = 85, b = 8, c = variable, d = 1600, e = 450, f = 1

TABLE 8.2

RESULTS OF TRADE-OFF BETWEEN CAPACITY COST/UNIT AND DELIVERY PERFORMANCE

(VARIATION IN THE NUMBER OF HOURS/WEEK)
(NO. OF MOULDS CONSTANT AT 18)

EXPERIMENT NO.	HOURS WEEK	CAPACITY COST/UNIT	PERCENTAGE OF PRODUCTION LATE — LEAD TIME (DAYS) 8	10	12	14	16	18	20
1	45	3.97	68.19	59.20	49.19	42.49	35.73	31.04	24.63
11	50	4.07	58.94	57.97	36.17	29.28	23.67	18.92	16.17
12	55	4.20	48.31	34.63	25.10	18.43	13.41	10.02	7.20
13	60	4.38	36.20	22.21	15.10	9.90	6.84	4.18	2.01
14	65	4.60	23.61	14.85	8.22	4.77	3.17	1.13	0.32
15	70	4.67	16.16	8.49	5.04	2.88	0.74	0.00	0.00
16	75	4.67	10.91	5.48	3.37	0.65	0.00	0.00	0.00
17	80	4.67	7.84	4.46	1.29	0.00	0.00	0.00	0.00
18	85	4.84	5.28	2.15	0.14	0.00	0.00	0.00	0.00
19	90	5.00	5.02	0.86	0.00	0.00	0.00	0.00	0.00
20	95	5.16	1.65	0.00	0.00	0.00	0.00	0.00	0.00
21	100	5.38	1.18	0.00	0.00	0.00	0.00	0.00	0.00

System Configuration $a = 85$, $b = 8$, $c = 18$, $d = 1600$, $e = \infty$, $f = 1$

TABLE 8.3

RESULTS OF TRADE-OFF BETWEEN CAPACITY COST/UNIT AND DELIVERY PERFORMANCE

(VARIATION IN THE NUMBER OF HOURS/WEEK)
(NO. OF MOULDS CONSTANT AT 27)

EXPERIMENT NO.	HOURS WEEK	CAPACITY COST/UNIT	PERCENTAGE OF PRODUCTION LATE — LEAD TIME (DAYS) 8	10	12	14	16	18	20
4	45	4.16	34.94	24.50	15.13	9.75	6.26	3.88	2.77
22	50	4.33	19.50	10.02	5.43	3.58	2.16	0.00	0.00
23	55	4.55	11.02	5.32	3.04	1.16	0.00	0.00	0.00
24	60	4.75	5.81	2.34	0.72	0.00	0.00	0.00	0.00
25	65	4.96	2.63	0.87	0.00	0.00	0.00	0.00	0.00
26	70	5.03	1.66	0.00	0.00	0.00	0.00	0.00	0.00
27	75	5.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	80	5.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00

System Configuration a = 85, b = 8, c = 27, d = 1600, e = 450, f = 1

TABLE 8.4

RESULTS OF TRADE-OFF BETWEEN CAPACITY COST/UNIT AND DELIVERY PERFORMANCE

(VARIATION IN STOCK CONTROL PARAMETERS)

System Configuration a = 85, b = 8, c = 18, d = 1600, e = ∞, f = 1   45 hours

EXPERIMENT NO.	REORDER POINT (*)	REORDER QTY. (*)	AVERAGE STOCK LEVEL	CAPACITY COST/UNIT	PERCENTAGE OF PRODUCTION LATE — LEAD TIME (DAYS)						
					8	10	12	14	16	18	20
29 (1)	400	1200	1151	3.97	55.98	47.86	40.04	33.68	29.34	23.76	20.50
30 (1)	800	800	970	3.98	58.74	48.67	41.63	34.98	30.91	24.67	21.26
31 (1)	1200	400	449	3.99	64.54	54.74	45.03	38.77	33.05	27.76	22.86
32 (1)	800	2400	2594	3.95	47.48	41.11	36.09	28.00	23.90	20.47	17.75
33 (1)	1600	1600	2493	3.98	47.66	39.47	34.22	28.88	24.20	20.63	18.54
34 (1)	2400	800	1149	3.99	57.44	47.41	40.26	34.83	29.80	24.08	20.32
35 (1)	1200	3600	4224	3.90	39.96	35.28	29.78	24.18	20.83	17.28	15.53
36 (1)	2400	2400	4282	3.96	38.13	32.87	28.40	22.49	19.30	18.01	15.77
37 (1)	3600	1200	2281	4.00	47.20	39.97	34.25	29.66	23.68	20.17	17.76
38 (1)	1600	4800	6047	3.91	35.85	31.13	24.81	21.10	18.76	16.07	14.76
39 (1)	3200	3200	6144	3.93	31.98	27.54	22.15	19.33	18.13	16.42	13.91
40 (1)	4800	1600	4161	4.01	37.70	32.79	29.08	23.92	19.94	16.37	15.33
41 (2)	2400	7200	10380	3.92	32.03	26.91	22.48	18.49	16.48	15.35	14.54
42 (1)	4800	4800	10480	3.92	20.98	18.21	15.74	14.48	12.69	10.82	8.51
43 (2)	4800	4800	10267	3.89	25.58	21.75	19.34	16.42	13.87	11.98	10.07
44 (3)	4800	4800	10243	3.88	26.61	22.78	19.90	16.91	14.55	12.33	10.71
45 (2)	7200	2400	10129	4.00	25.33	22.19	17.92	16.01	13.80	12.62	10.56
46 (4)	22000	2000	14880	3.97	28.21	24.51	21.19	18.10	15.71	13.47	11.52
47 (4)	20000	4000	38128	4.21	6.82	5.04	4.11	3.34	2.80	2.18	1.82
48 (4)	18000	6000	41967	4.22	3.86	2.91	2.34	1.62	1.38	1.16	0.91
49 (4)	16000	8000	42508	4.23	2.97	1.85	1.25	0.86	0.71	0.53	0.42
50 (4)	14000	10000	33876	4.16	6.80	5.12	3.68	2.63	1.91	1.59	1.35
51 (4)	12000	12000	31773	4.08	9.28	7.08	5.65	4.37	3.46	2.32	1.44

(1) Stabilization period = 20 orders, sampling period = 130 orders
(2) Stabilization period = 40 orders, sampling period = 180 orders
(3) Stabilization period = 50 orders, sampling period = 260 orders
(4) Stabilization period = 70 orders, sampling period = 450 orders
(*) The values for the reorder point and reorder quantity indicated in the table represent the addition of reorder point and reorder quantities of all product sizes belonging to a style.

$$\text{REORDER POINT} = \sum_{j} \text{RPOINT}(i,j)$$

$$\text{REORDER QUANTITY} = \sum_{j} \text{EBQ}(i,j)$$

TABLE 8.5

RESULTS OF TRADE-OFF BETWEEN CAPACITY COST/UNIT AND DELIVERY PERFORMANCE

(VARIATION IN STOCK CONTROL PARAMETERS)

System Configuration a = 85; b = 8; c = 24, 27; d = 1600; e = 450; f = 1 45 hours

EXPERIMENT NO.	NO. OF MOULDS	REORDER POINT	REORDER QTY.	AVERAGE STOCK LEVEL	CAPACITY COST PER UNIT	PERCENTAGE OF PRODUCTION LATE						
						LEAD TIME (DAYS)						
						8	10	12	14	16	18	20
52 (2)	24	4800	4800	12031	4.13	9.76	6.67	4.62	2.90	1.78	1.14	0.73
53 (3)	24	12000	4000	26230	4.30	2.35	1.49	1.00	0.78	0.53	0.37	0.36
54 (2)	27	4800	4800	12105	4.25	8.35	5.37	4.03	2.60	1.50	0.86	0.55
55 (3)	27	12000	4000	26409	4.43	2.29	1.37	0.90	0.49	0.25	0.17	0.00

(2) Stabilization period = 50 orders, sampling period = 260 orders
(3) Stabilization period = 60 orders, sampling period = 320 orders

FIGURE 8.1

TRADE-OFF CURVES BETWEEN CAPACITY COST PER UNIT AND DELIVERY PERFORMANCE

(VARIATION IN THE NUMBER OF MOULDS)

(WORKING HOURS PER WEEK = 45)

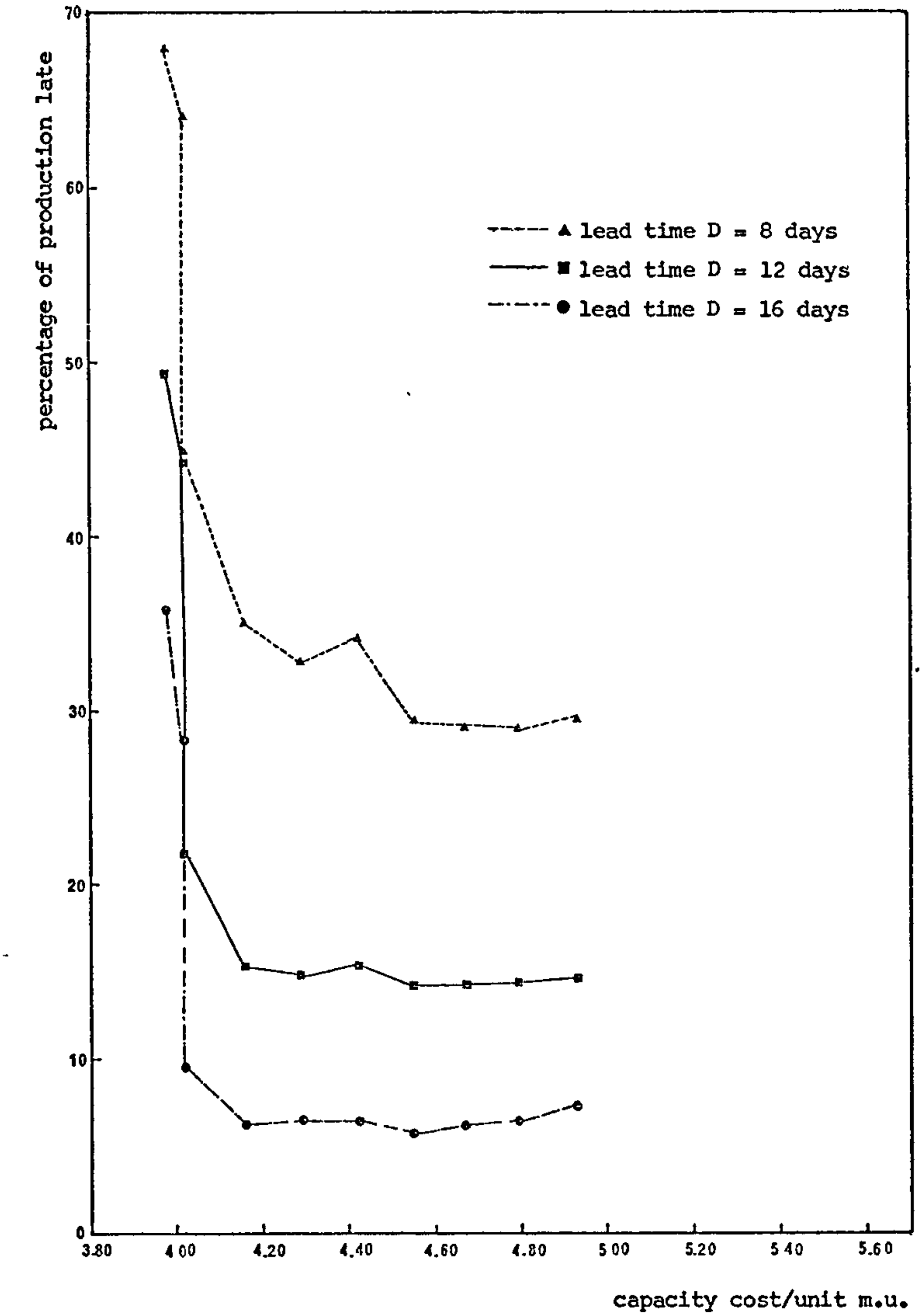


FIGURE 8.2

TRADE-OFF CURVES BETWEEN CAPACITY COST PER UNIT AND DELIVERY PERFORMANCE

(VARIATION IN THE NUMBER OF WORKING HOURS/WEEK)

(NUMBER OF MOULDS = 18)

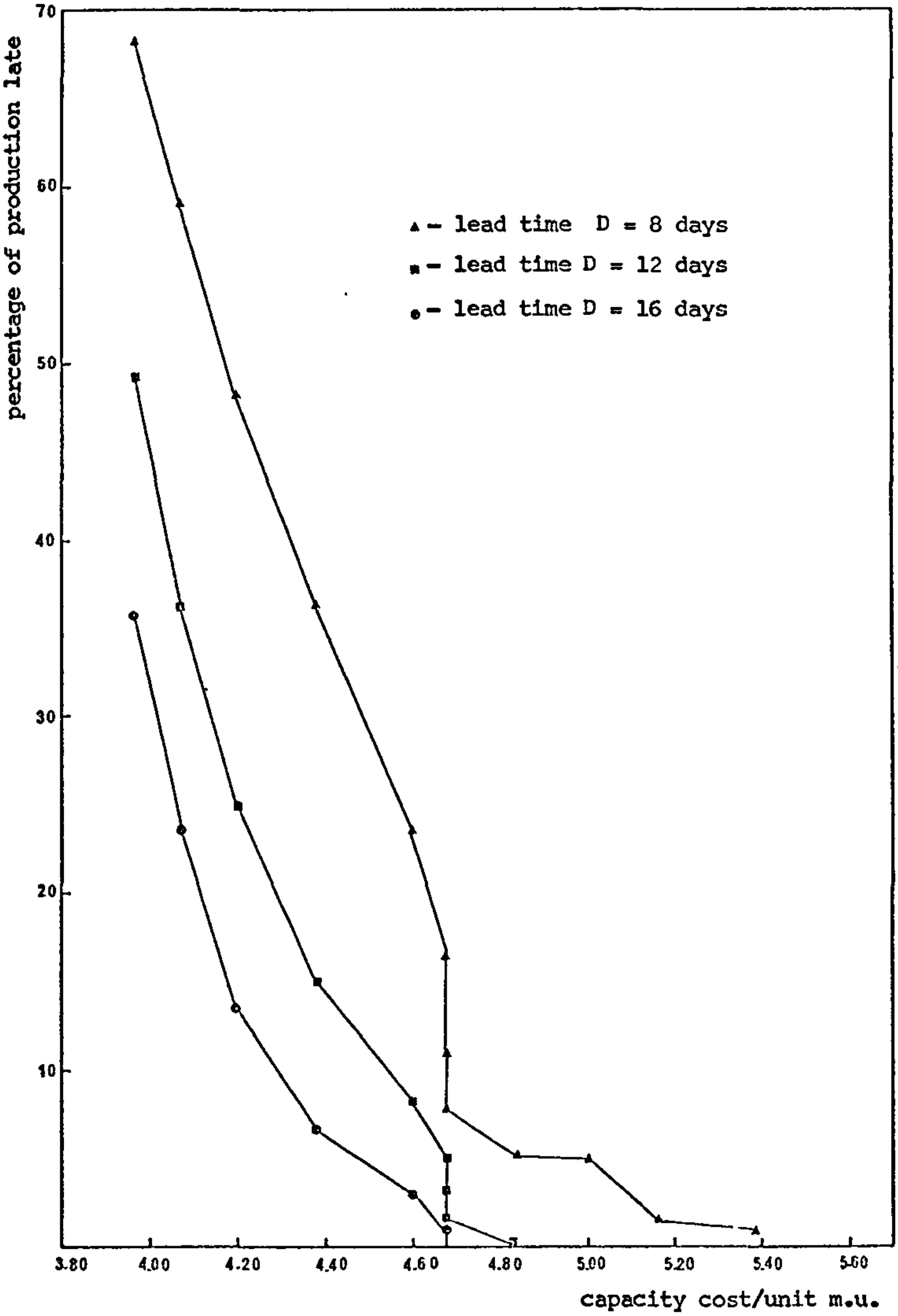


FIGURE 8.3

TRADE-OFF CURVES BETWEEN CAPACITY COST PER UNIT AND DELIVERY PERFORMANCE

(VARIATION IN THE NUMBER OF WORKING HOURS/WEEK)

(NUMBER OF MOULDS 27)

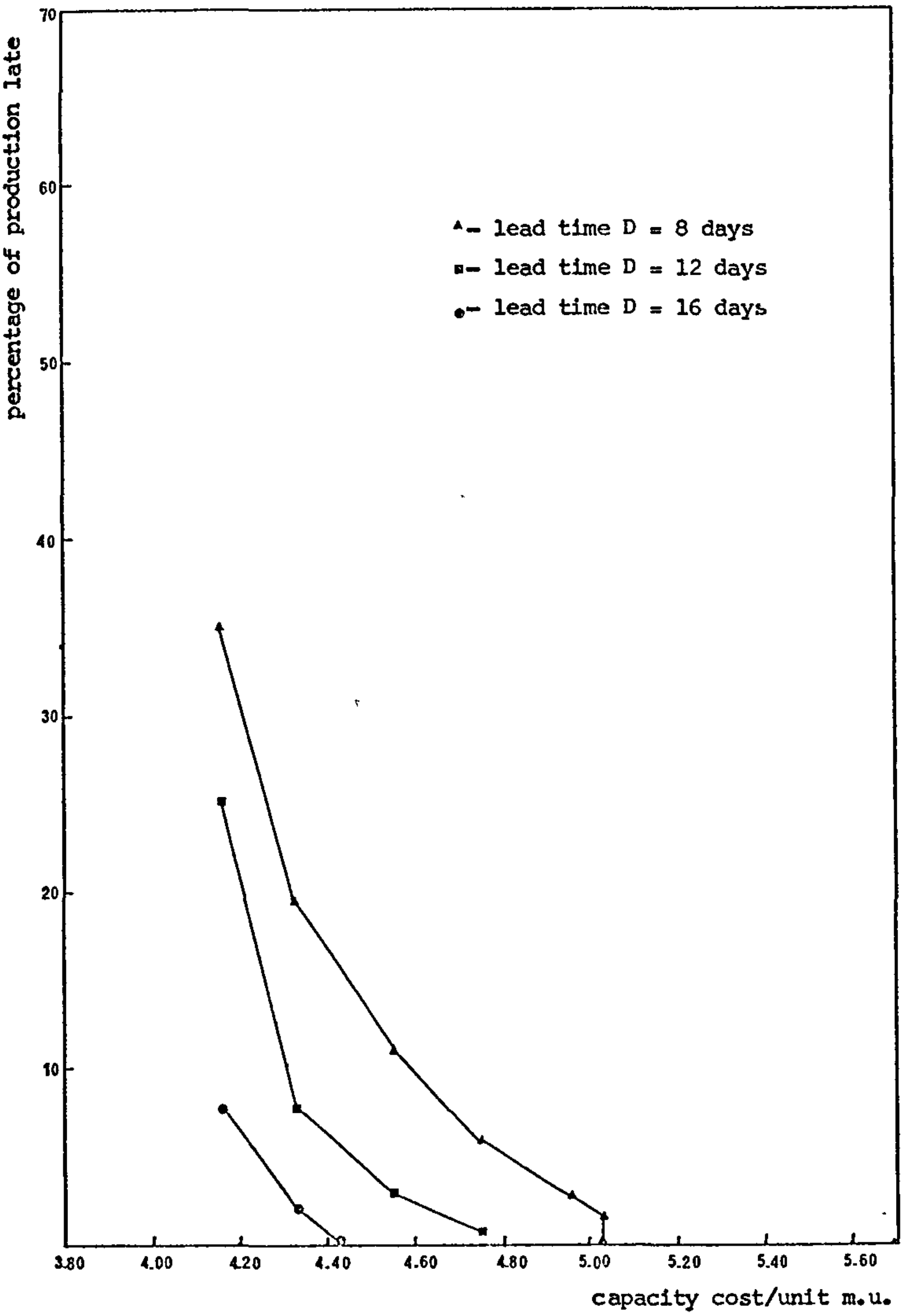


FIGURE 8.4

COMPARISON BETWEEN TRADE-OFF CURVES GENERATED BY THREE DIFFERENT STRATEGIES OF CAPACITY MANIPULATION

(LEAD TIME D = 8 DAYS)

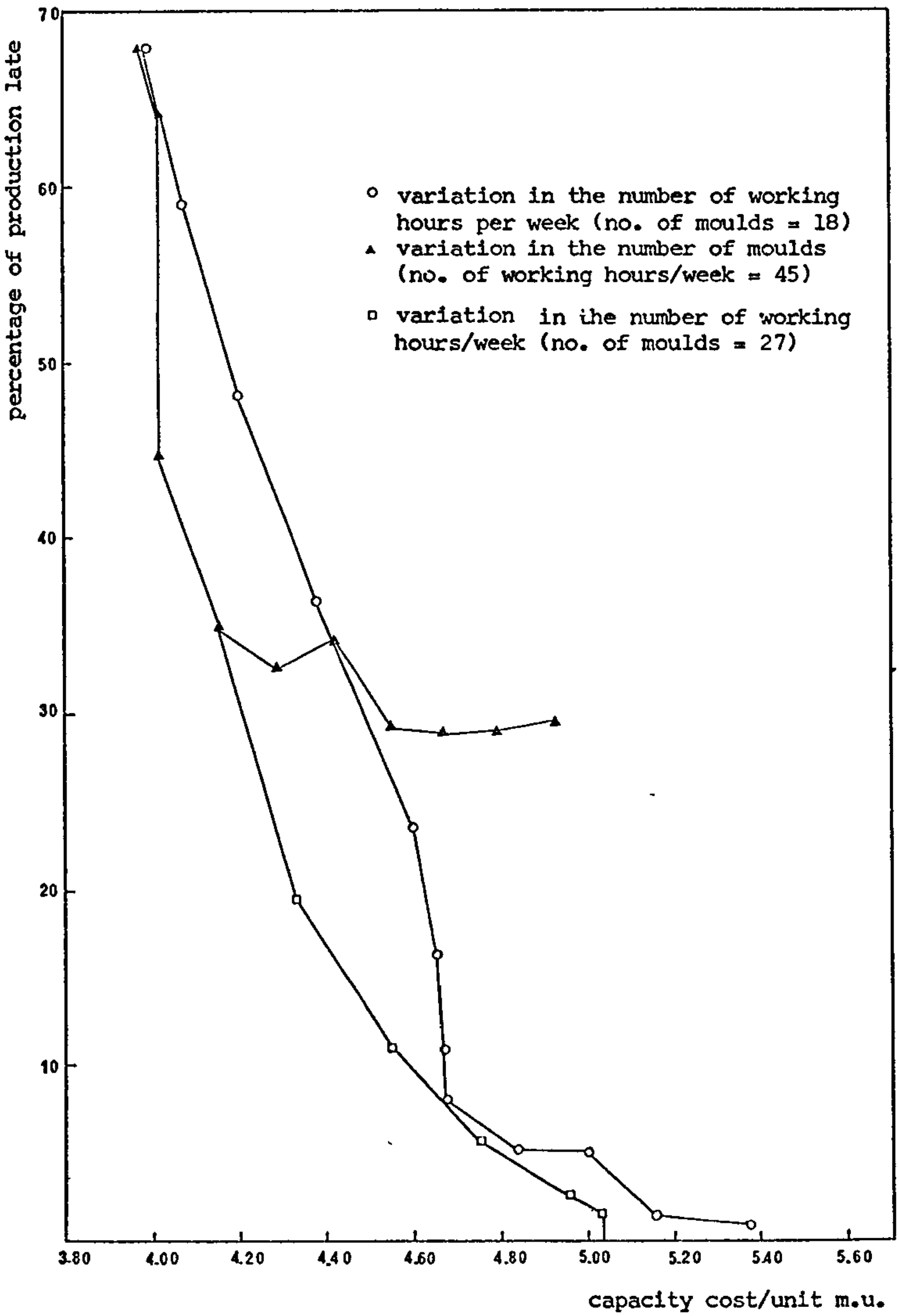


FIGURE 8.5

COMPARISON BETWEEN TRADE-OFF CURVES GENERATED BY THREE DIFFERENT STRATEGIES OF CAPACITY MANIPULATION

(LEAD TIME D = 16 DAYS)

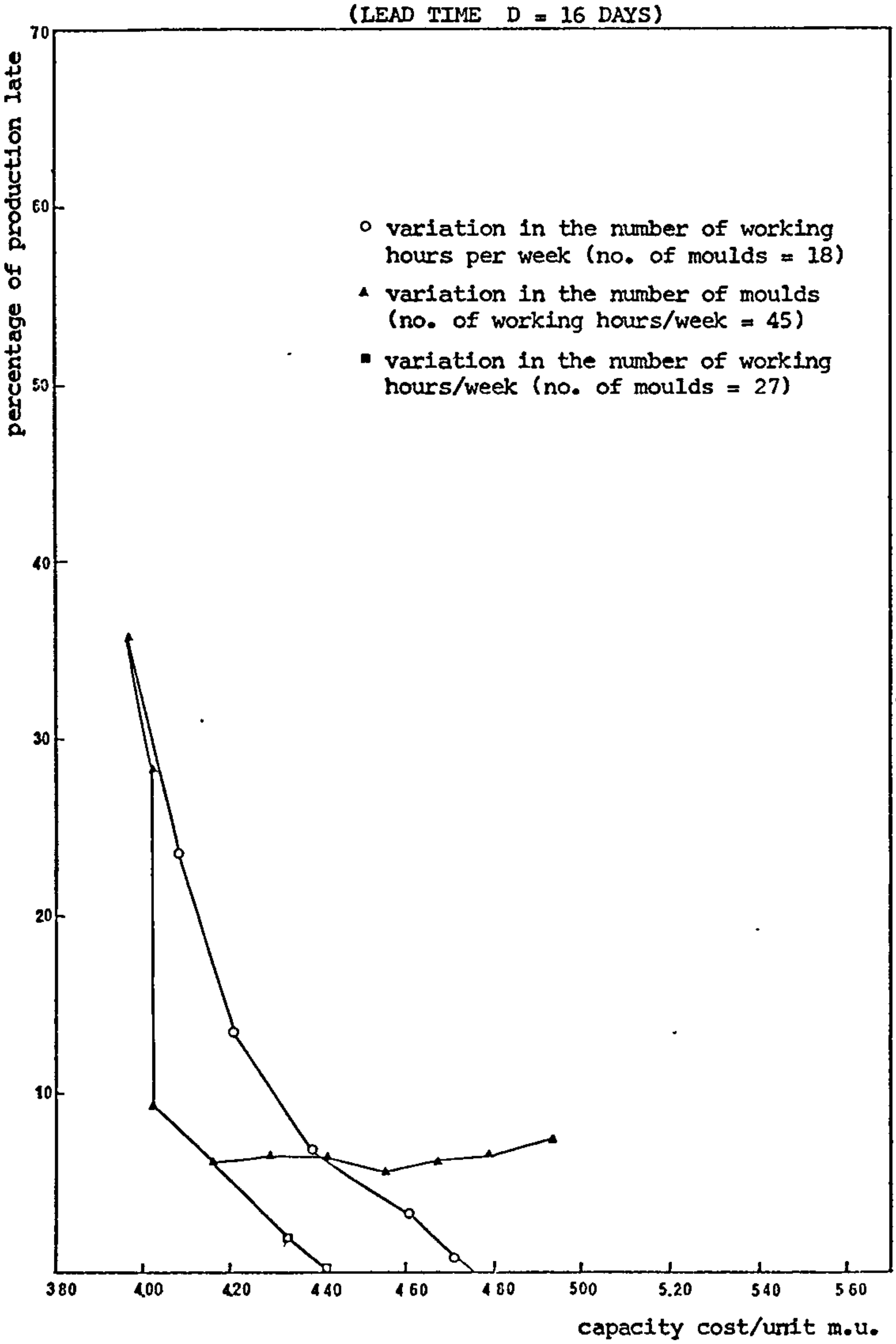


FIGURE 8.6

TRADE OFF CURVES BETWEEN CAPACITY COST PER UNIT AND PERCENTAGE OF PRODUCTION GENERATED BY MAKING CHANGES IN THE PARAMETERS FOR INVENTORY CONTROL

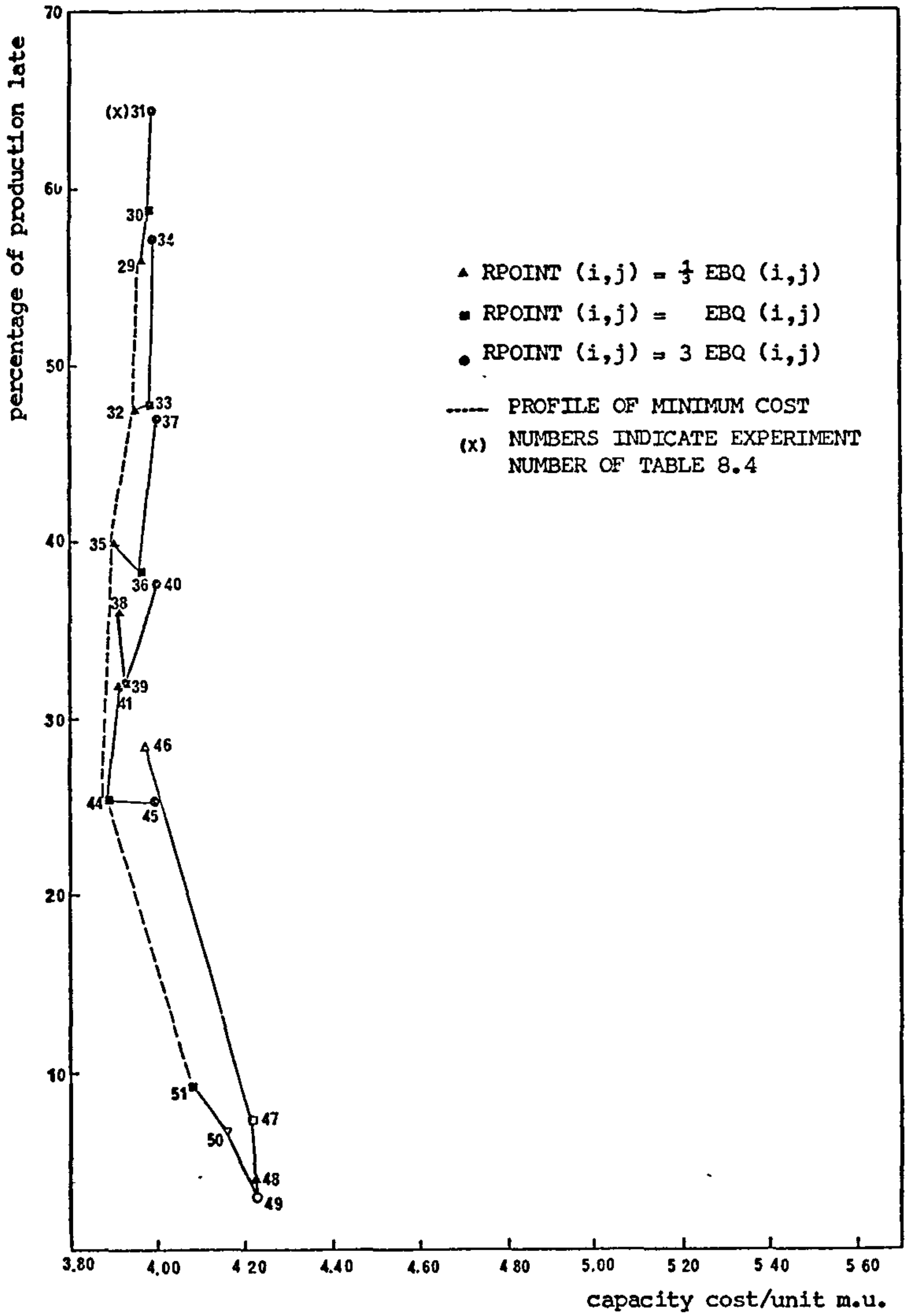


FIGURE 8.7

COMPARISON BETWEEN TRADE-OFF CURVES FOR DIFFERENT STRATEGIES FOR CAPACITY MANIPULATION

(LEAD TIME D = 8 DAYS)

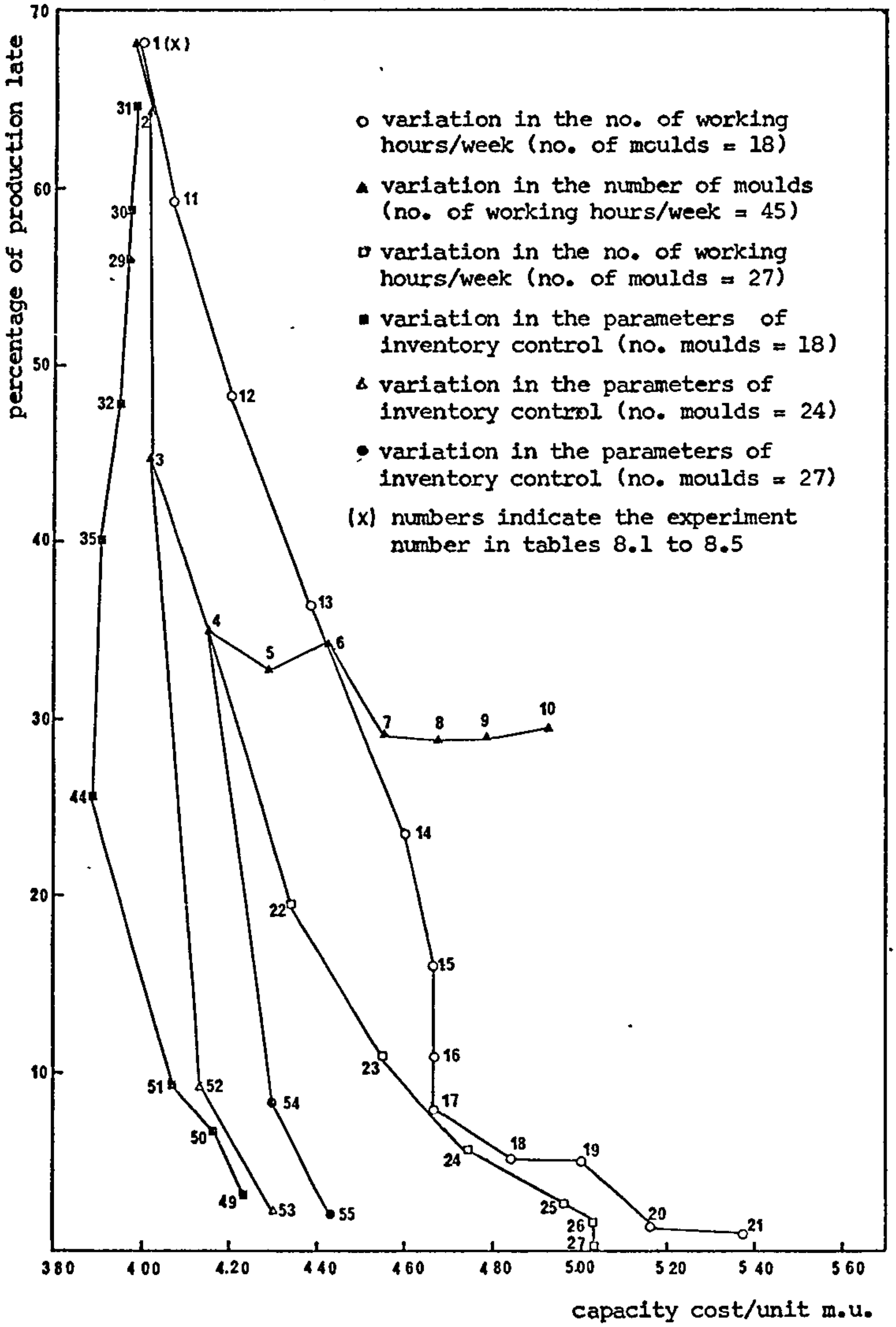
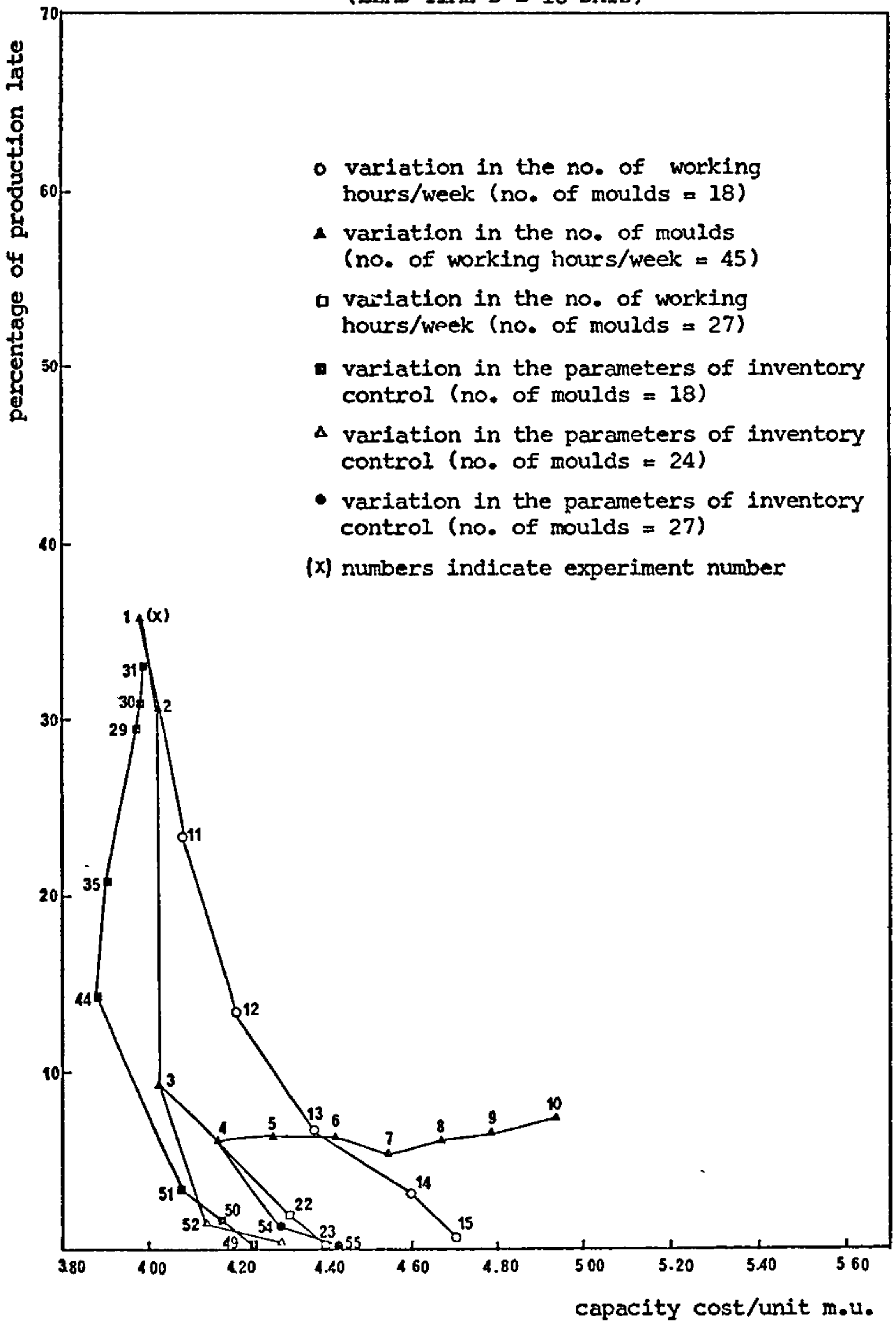


FIGURE 8.8

COMPARISON BETWEEN TRADE - OFF CURVES FOR DIFFERENT STRATEGIES FOR

CAPACITY MANIPULATION

(LEAD TIME D = 16 DAYS)



CHAPTER 9

CONCLUSIONS AND RECOMMENDATIONS

9.1 - Conclusions

The results of this investigation can be divided into four groups representing the four major sections in which the experiments with the model were organized. In chapters 4, 6, 7 and 8 those results were presented and discussed at length. Therefore, this section will be limited to presenting a summary of the main characteristics and conclusions of the study.

9.1.1 - Conclusions from the preliminary investigation

The first group of results, discussed in chapter four, refers to the preliminary investigation, in which empirical data from an industrial organization was used. The main objectives of that part of the investigation were to validate the model and gain an insight into the behaviour and characteristics of the system. The results of the experiments were utilized for the development of appropriate priority scheduling rules, and other control procedures, and for the determination of a formal experimental design.

9.1.2 - Conclusions from the study of priority scheduling rules

The second group of results, discussed in chapter 6, refers to the study of the behaviour and relative performance of six priority scheduling rules. The priority rules were compared over six system configurations, which were obtained by the joint variation of six of the model's parameters. Comparisons were made in respect of their ability to improve the delivery performance of the system, which was measured by 6 output variables, where the first 3 are 'unweighted' measures of delivery performance and the other 3 are 'weighted' measures of delivery performance. Five measures of the system's internal behaviour were also used in order to support the results of delivery performance. The results indicated the following main conclusions.

i) Both the absolute and relative performances of the priority rules are affected by the system configuration; by the way in which delivery performances are measured; and by the value of the lead time used to fix due dates.

ii) Overall, the priority rules designed to avoid setup times (FIFOMB, SLACKM and SPTM) were shown to be superior to equivalent rules (FIFOB, SLACK and SPT) which do not avoid setup times.

iii) The SPTM (and SPT) rule seems to perform best for the 'unweighted' measures of delivery performance, 'average delay of orders' and 'percentage of late orders' and for tight due dates. However it tends to lose its advantage over the other rules (FIFOMB, particularly) when the due dates gets less tight, and to perform badly in respect to the 'weighted' measures of delivery performance, 'average delay of production' and 'tardiness index of production', and for some of the measures of internal behaviour, viz. 'remaining content', and 'percentage of time spent with setting up' (for definition of measures of internal behaviour and performance see paragraphs 3.6.1 and 3.6.2, respectively).

iv) The performance of SLACKM and FIFOMB were very similar, but FIFOMB produced overall better results for the measures of delivery performance. The differences however are very small and in the majority of the cases are not statistically significant.

v) In view of the overall results it was concluded that FIFOMB seems to be the most appropriate of all the six priority rules, as far as this class of production system is concerned. (For description of the priority rules operation procedures see paragraph 3.5.1).

9.1.3 - Conclusions from the study of main effects and interactions

The third group of results, discussed in chapter 7, refers to the study of the effects on the system's internal and external (delivery performance) behaviour, caused by changes in the parameters of six of the system's variables, viz. average nominal load factor; mean value of setup times; number of moulds; average size of orders; splitting of jobs; and the ratio between the number of product styles and the number of machines. A half-replicate factorial design was used, in which all six factors (variables) had two levels (values of parameters). This experimental design has allowed the measure of all six main effects and all fifteen first order interactions. Listed below are the main conclusions.

i) By far the largest main effect on delivery performance was caused by the variation in the nominal load factor, which was obtained by changing the arrival rate of orders. The increase in the nominal load factor from 65% to 85% caused a considerable deterioration in all the measures of delivery performance. This effect has been observed in other studies of more traditional batch production systems.

Although there was a statistically significant interaction between the load factor and the ratio of the number of product styles to the number of machines, the magnitude of the interaction was small enough in relation to the main effect to allow the conclusions about the main effect to be valid, independently of the other parameters.

ii) The increase in the average size of orders caused the second largest independent main effect on the delivery performance of the system, as was indicated by considerable deteriorations in all the measures of delivery performance. The magnitude of this effect was however much smaller than the effect caused by the variation in the load factor.

iii) The increase in the mean value of setup time caused the delivery performance of the system to deteriorate. However, although this effect was shown to be statistically significant and independent of the other variables, its magnitude was very small in relation to the other two effects indicated above. It should be pointed out that although the relative increase in the mean value of setup time was large (100 %), the absolute increase was small. This must account for the small effect on delivery performance.

iv) There were strong indications to suggest that the procedure of splitting jobs into smaller batches has, on average, a very small effect on the delivery performance of the system, but a significant effect on the queue size.

v) There was a large interaction between the number of moulds and the ratio, number of product sizes/number of machines, in relation to the measures of delivery performance. This interaction is of considerable importance due to its magnitude in relation to the other effects. In practical terms, it indicates that the effect on the delivery performance caused by an increase in the number of moulds, is very much dependent on the ratio of the number of product styles to the number of machines.

vi) Independently of the interaction described above, a better ratio of the number of product style/ number of machines has a considerable effect on improving the performance of the system.

vii) A comparison between the measures of internal behaviour and the measures of delivery performance, indicated that in some circumstances an improvement in the measures of internal performance corresponds to a deterioration in the external (delivery) performance.

9.1.4 - Conclusions from the study of strategies for capacity manipulation

The fourth and last group of results, discussed in chapter 8, relates to the study of operation strategies for capacity manipulation. Three main strategies, viz. variation in the number of moulds; variation in the number of working hours per week; and the utilization of inventory of finished goods, were analysed and compared, together with combinations of these strategies. The main objective was to produce trade-off curves relating capacity cost per unit produced, to the percentage of production delivered late, for different values of lead time D, such that different strategies could be compared using those curves. Typical costs and other parameters from a particular industrial company were used. It should be pointed out that cost parameters have a major effect on the results obtained and so some of the conclusions should be seen with this constraint in mind. Below are some of the major conclusions.

i) The relative effectiveness of the various strategies depends on the level of delivery performance desired and on the value of lead time D, used to quote due dates.

ii) Overall, the strategy of using finished goods inventory seems to give better results than any of the other strategies. It was however pointed out that a considerable amount of stock would be needed in order to obtain good performances, and therefore consideration should be given to this aspect when choosing between different strategies.

iii) The trade-off curves generated by increases in the number of moulds confirmed some of the previous results from the preliminary investigation and from the study of main effects and interactions. They show that, up to a certain value, increases in the number of moulds have a sharp effect in reducing the percentage of production delivered late. However after a certain point the sharp effect dies down very quickly, and the trade-off curves flatten out, indicating that extra moulds have almost no effect in reducing the percentage of production delivered late.

iv) Discussions have indicated how management can use the trade-off curves in order to make strategic decisions in a three dimensional grid, which has costs, delivery performance and the length of delivery promises as the parameters of decision.

9.2 - Recommendations for further research

In accordance with the information obtained, four main points of research could be followed in further investigations.

i) A more critical study of the use of inventory as an optional strategy for capacity manipulation should be pursued. Among the points to be investigated are the interactions between priority scheduling rules and inventory control procedures, and the consequences of holding inventory for a restricted number of product sizes in a range, as compared to holding stock for all product sizes in the range.

ii) The effect on the relative performance of the priority scheduling rules, of different methods for establishing due dates, should be investigated. Instead of having a constant value for the lead time used to fix due dates, a range of lead times should be used, with due dates being fixed in accordance with the order batch size

iii) It seems worthwhile to study the effects on the system behaviour, of some of the variables which were maintained constant during this investigation. Examples of such variables are the pattern of demand for the different product sizes in a range, and the number of stations per machine. In particular the technological, economical and operational consequences of different numbers of stations per machine should be investigated. Such an investigation would require a co-ordinated effort between the manufacturers and users of those machines, such that the technological and economical constraints on the manufacturer's side could be matched to the economical and operational characteristics of the production system on the user's side.

iv) Finally, the results from the study of capacity manipulation, suggest that it might be of considerable help to management, when making strategic decisions on capacity, to have at their disposal the kind of information generated by the trade-off curves which relate cost, delivery performance and length of delivery promises. Particularly in situations where delivery delay and special tools are of vital importance, and in which the product lines change from time to time, it would be worthwhile to have tailor-made simulation models which could be used on a regular basis, for medium term planning involving strategic decisions of the kind discussed in this study.

LIST OF REFERENCES

LIST OF REFERENCES

AGGARWAL, S.C., WYMAN, F.P., and McCARL, B.A., 1973, An investigation of a cost-based rule for job-shop scheduling, Int. J. Prod. Res. Vol. 11, no. 3

AGGARWAL, S.C., 1974, A review of current inventory theory and its applications, Int. J. Prod. Res. Vol. 12, no. 4

BERRY, W.L., 1972, Priority scheduling and Inventory Control in Job lot Manufacturing systems, AIIE TRANSACTIONS, Vo. 4, no. 4

BERRY, W.L. and RAO, V., 1975, Critical ratio scheduling: An experimental analysis, Mgmt Sci. Vol. 22, no. 2

BONINI, C.P., 1963, Simulation of information and decisions systems in the firm, (USA, Prentice Hall), (1) p.75; (2) p.91; (3) p.105

BOWMAN, E.H., 1956, Production Scheduling by the transportation method of linear programming, Ops Res.,Vol. 4, no. 1

BOWMAN, E.H., 1959, The schedule sequencing problem, Ops. Res., vol. 7, no. 5

BROWN, A.P. and LOMNICKI, Z.A., 1966, Some Applications of the branch and bound algorithm to the machine Scheduling problem, Opl. Res. Q., vol. 17, no. 2

BUFFA, E.S., 1973, Operations management, problems and models, (USA, WILEY)

BURBIDGE, J.L., 1971, Production Flow Analysis, The production Engineer, Vol. 50, no. 4/5

CAMPBELL, H.G., DUDEK, R.A., SMITH, M.L., 1970, A heuristic algorithm for the n job m machine sequencing problem, Mgmt Sci. Vol. 16, no. 10

CANTELLOW, D.G., PITT, R.V., SAW, R.J. and HOUGH, J., 1973, Machine-shop problems: An operational research Approach, Opl. Res. Q., vol. 24, no. 4

CARRIE, A.S., 1973, Numerical taxonomy applied to group technology and plant layout, 2nd. International Conference on Production Research (Copenhagen)

COALES, J.F., 1971, The control of industrial processes, Int.J. Prod. Res. Vol, 9, no. 1

CONWAY, R.W., JOHNSON, B.M. and MAXWELL, W.L., 1960, An experimental investigation of Priority Dispatching, J. of ind. Engng., vol.11, p.221

CONWAY, R.W., and MAXWELL, W.L., 1962, Network dispatching by the shortest operation discipline, Ops Res. Vol. 10, p.51

CONWAY, R.W., 1963, Some tactical problems in digital simulations, Mgmt Sci., Vol. 10, no. 1

CONWAY, R.W., 1964, An experimental investigation of priority assignments in a job shop, RM;3789;PR (DCC no. AD 429970), Rand Corp., Santa Monica, Calif. Feb. 1964

CONWAY, R.W., 1965, Priority dispatching and work-in-process inventory in a job shop, J. ind. Engng., Vol. XVI, no.2

DAVIES, O.L., 1958, Statistical methods in Research and Production , (London, Oliver and Boyd), p. 117

DAVIES, O.L., 1967, The design and analysis of industrial experiments, (London, Oliver and Boyd), (1) p. 247; (2) p. 257; (3) p. 440; (4) p. 263; (5) p. 263; (6) p. 280

DAY, J.E. and HOTTENSTEIN, M.P., 1975, The impact of advancing due dates in a pure job shop, Int. J. Prod. Res.,Vol. 13, no. 6

DOLTON, B., and MOODY, P., 1975, Simulation as a routine tool for resource and work load planning, Operational Research Society Conference, Loughborough, England, September, 1975.

DUNNET,C.W., 1955, A multiple comparison procedure for comparing several treatments with a control, J. Amer. Stat. Assoc., no. 50, p. 1096

EASTMAN, W.L., 1964, Bounds for the optimal scheduling of n jobs on m processors, Mgmt. Sci. Vol 11, no. 2

EDWARDS, G.A.B., 1974, Group Technology - technique, concept or new manufacturing system? (part 2), CME, Vol. 21, no. 3

EHRENFELD , S. and BEN-TUVIA, S., 1962, The Efficiency of Statistical Simulation Procedures, Technometrics, Vol. 4, no. 2

EILON, S. and HODGSON, R.M., 1967, Job shops scheduling with due dates, Int. J. Prod. Res., Vol. 6, no. 1

EILON, S. and COTERILL, D.J., 1968, A modified SI rule in job shop scheduling, Int. J. Prod. Res., Vol. 7, no. 2

EILON, S. and ELMALEH, J., 1968, An evaluation of alternative inventory control policies, Int. J. Prod. Res. Vol. 7, no. 1

EILON, S. and CHOWDHURY, I.G., 1976, Due dates in job shop scheduling, Int. J. Prod. Res., Vol.14, no. 2

EL-RAYAH, T.E. and HOLLIER, R.H., 1969, A review of plant design techniques, Int. J. Prod. Res.,Vol. 8, no. 3

EL-RAYAH, T.E., 1973, Experimental investigation of performance of various production systems designs, Ph.D. Dissertation, University of Birmingham

ELVERS, D.A., 1974, The sensitivity of the relative effectiveness of job shop dispatching rules with respect to various arrival distributions, AIIE TRANSACTIONS, Vol. 6, no. 1

FAZAKERLEY, G.M., 1974, Group Technology, Social benefits and social problems, The Production Engineer, vol. 53, no. 10

FAZAKERLEY, G.M., 1976, A research report on the human aspects of group technology and cellular manufacture, Int. J. Prod. Res., Vol. 14, no. 1

FISHMAN, G.S., 1967, Problems in the Statistical Analysis of Simulation Experiments: The Comparisons of Means and the Length of Sample Records, Communications of the ACM, Vol. 10, no. 2

FISHMAN, G.S. and KIVIAT, P.J., 1967, The Analysis of Simulation - Generated Time Series, Mgmt Sci. , Vol. 13, no. 7

FOLIE, G.M., 1974, A simple model for computing the trade-offs between plant capacity and operating strategies, Opl. Res. Q., Vol. 25, no. 1

FRYER, J.S., 1975, Effects of shop size and labor flexibility in labor and machine limited production systems, Mgmt Sci., Vol. 21, no. 5

GALLAGHER, C.C., GRAYSON, T.J. and PHILLIPS, A., 1971, A Review of the development and application of group technology at the University of Birmingham, Int. J. Prod. Res., Vol. 9, no. 2

GORDON, G., 1969, System Simulation (USA, Prentice Hall, Inc.)

GOMBINSKI, J., 1967, Group Technology - An Introduction, The Production Engineer, Vol. 45, no. 9

GUENTHER, W.C., 1964, Analysis of Variance, (USA, Prentice Hall), (1) p. 54; (2) p. 65; (3) p. 15; (4) p. 112

HAMMERSLEY, J.M. and MORTON, K.W., 1956, A new Monte Carlo Technique Antithetic variates, Proc. camb. phil. soc. Vol. 52, p. 449

HANSSMANN, F. and HESS, S.W., 1960, A linear programming Approach to production and employment scheduling, Management Technology, Vol. 1, p. 46

HOLLIER, R.H. and CORLETT, E.N., 1966, Work flow in batch production, (London, HMSO, Problems of progress in industry, no. 18)

HOLLIER, R.H., 1968, A simulation study of sequencing in Batch Production, Opl. Res. Q., Vol. 19, no. 4

HOLLIER, R.H. and BHATTACHARYYA, S.K., 1974, Future of batch manufacturing, The Production Engineer, vol. 53, no. 7/8

HOLT, C.C., MODIGLIANI, F. and SIMON, H.A., 1955, A linear decision rule for production and employment scheduling, Mgmt Sci., Vol. 2, no. 2

HOSEIN, R.W., and ROSS, M., 1975, Job shop scheduling in an electroplating environment, The Production Engineer, Vol. 74, no. 5

IGNALL, L.E. and SCHRAGE, L.,1965, Applications of the branch and bound techniques to some flow shop scheduling problems, Ops. Res., Vol. 13, no. 3

JOHNSON, A.W.A., 1974, Plastic insole backparts, SATRA BULLETIN, Vol. 16, no. 4

JOHNSON, S.M., 1954, Optimal two and three stage production schedules with setup times included, Naval Research Logistics Quarterly, Vol. 1, no. 1

KENDALL, M.G. and STUART, A., 1966, The advanced theory of statistics, Vol. 3: Design and Analysis and time series, (Charles Griffin Co.)

KING, J.R., 1976, The theory-practice gap in job-shop scheduling, The Production Engineer, Vol. 75, no. 3

KLEIJNEN, J.P.C., 1973, Antithetic variates, common random numbers and optimal computer time Allocation in simulation, Mgmt Sci., Vol. 21, no. 10

KNIGHT, W.A., 1974, The economic benefits of Group Technology, The Production Engineer, vol. 53, no. 5

LAMPKIN, W., 1967, A review of Inventory Control Theory, The Production Engineer, Vol. 46, no. 2

LITTLECHILD, S.C., 1974, Optimal arrival rate in a simple queueing system, Int. J. Prod. Res., Vol. 12, no. 3

LOCKYER, K.G., 1974, Factory and Production Management, (London, Pitman), p. 113

LOMNICKI, Z.A., 1965, A branch and bound algorithm for the exact solution of the three machine scheduling problem, Opl. Res.Q., Vol. 16, no. 1

McKAY, A.T., 1929, Statistical study of the feet of the boot and shoe operatives in Northampton - Report no. 29 of the British Boot, shoe and Allied Trades Research Association.

MANNE, A.S., 1960, On the job-shop scheduling problem, Ops. Res., Vol. 8, no. 2

MOORE, F.G. and JABLONSKI, R., 1969, Production Control, (USA, McGraw-Hill), p. 14

OPITZ, H. and WIENDAHL, H.P., 1971, Group technology and manufacturing systems for small and medium quantity production, Int. J. Prod. Res., Vol. 9, no. 1

ORAL, M. and MALOUIN, J.L., 1973, Evaluation of the shortest processing time scheduling rule with truncation process, AIIE TRANSACTIONS, Vol. 5, no. 4

ORAL, M., MALOUIN, J.L., HOBBS, J.B., 1974, An inventory Control policy of the (Q,S,R) Type for manufacturing industries: Simulation Analysis Approach, AIIE TRANSACTIONS, Vol. 6, no. 4

PALMER, D.S., 1965, Sequencing jobs through a Multistage Process in the minimum total time - A quick method of obtaining a near optimum, Opl. Res. Q., Vol. 16, no.1

ROCHETTE, R., and SADOWSKI, R.P., 1976, A statistical comparison of the performance of simple dispatching rules for a particular set of job shops, Int. J. Prod. Res., Vol. 14, no. 1

SCHEFFE, H., 1959, The Analysis of Variance, (New York, John Wiley & Sons, Inc.)

STARR, M.K., 1972, Production Management - Systems and Synthesis (USA, Prentice-Hall) p. 224

TUKEY, J.W., 1953, The Problem of multiple comparisons, Dittoed manuscript, Princeton, N.J.

WAGNER, H.M., 1959, An integer linear programming model for machine scheduling, Naval Research Logistics Quarterly, Vol. 6, no. 2

WILBRECHT, J.K. and PRESCOTT, W.B., 1969, The influence of setup time on job shop performance, Mgmt Sci., Vol. 16, no. 4

WILSON, R.H., 1935, A scientific routine for stock control, Harvard Business Review, Vol. 13