Productivity Estimation of Earthmoving Operations Using a Discrete-Event Simulation Model

Simon D. Smith

PhD Thesis - University of Edinburgh 1995

Abstract

The earthmoving industry in the UK is described with reference to the major changes in emphasis over the past few decades. The sites that were chosen for study are described and the method of data collection outlined.

A determinate model is developed. This is solved using a spreadsheet and the term match factor, indicating the efficiency of an earthmoving operation, is introduced. The disadvantages of the determinate model are discussed with the main conclusion that the variability in the earthmoving plant's working rates means that the production cannot be calculated accurately by determinate methods.

A model is then developed based on the discrete-event models used to evaluate queueing systems. This model was initially tested on a computer spreadsheet but a dedicated program, in C, was later written. Discrete-event simulation requires the generation of random numbers and emphasis is placed on the determination of the probability distributions that accurately represent the real distributions of plant cycle times.

A three stage validation process was used which involved both the acceptance from 'experts' of the modelled system (which was obtained both in the UK and the United States) and also the ability of the model to determine accurately the production of observed operations.

The simulation model could then, once validated, be used as an experimental platform. Experiments on actual operations have indicated which of the input factors to an operation the output is most sensitive to. They have shown that the correct number of trucks and maximisation of the loader working rate is essential to efficient earthmoving.

Acknowledgements

The author wishes to acknowledge the contributions, help and encouragement of the following people.

The University of Edinburgh

Professor Mike Forde	For his supervision throughout the duration of the project. Mike's additional duties as the head of the Department of Civil Engineering and the Tarmac Professor of Civil Engineering Construction in no way detracted from his commitments to his students, and the encouragement, constructive criticism and belief in the author's work were invaluable.
Steve Dunn and Graeme Wood	As partners in the earthmoving project with the author, Steve and Graeme helped in the collection of data as well as providing support and advice during the analysis and writing of the thesis.

Tarmac Construction Limited

Tarmac provided part of the funding for the whole of the earthmoving project, which was carried out under the Teaching Company Scheme. The author is indebted to the whole company for this but in particular to the following people and sites for their help.

M1/A1 Contracts 2&3	Including Roger Reid and Nick Brooks.
M3 Bar End to Compton	Including John Campbell and Adam Green, and also to Blackwell Earthmoving Limited, the subcontractors on this site.
The Earthmoving Division	Especially George Daville, whose belief in the project was very encouraging amongst a few traditionalists.

M65 Blackburn Southern Bypass	Including Brian Griffiths, John Wainwright and Martin Gould. Thanks also to John Jones (Excavation) Limited, the subcontractors on this site.
John Osborne	John, as the industrial supervisor, provided much help and encouragement for all the partners on the earthmoving project, and has also been very helpful in the implementation of the research at a site level.

Table of Contents

List of Figures

List of Tables

Notation

Each equation is presented with the relevant notation. All units are SI unless otherwise stated.

Chapter 1

Introduction

The construction industry, including the road building industry, has seen many changes over the past few years, with contractors winding down unprofitable and uncertain businesses such as earthmoving divisions. However, the earthworks part of a major road building contract will cost many millions of pounds and be heavily reliant on a number of very expensive items of plant. If the accuracy of earthmoving estimations could be improved, operations could be carried much more efficiently and cost effectively.

The earthworks part of a major road building or widening scheme can involve the movement of millions of cubic metres of earth to a large number of permanent embankments or unsuitable fill tip sites. On very large contracts the total number of separate earthmoving operations can be very high and this, coupled with tight budgets and completion target dates, means that the earthworks should be planned as well as possible before site works start - not only to obtain the tender price (which the earthworks subcontractor will have a very short time to calculate) but also to determine the amount of plant required in total on the site. If the amount is incorrectly calculated large additional costs can be incurred once work has begun due to the hiring in of additional vehicles to complete an operation on time. Unfortunately, the contractor rarely has the time to do this kind of pre-planning work to the depth that he would like and may, therefore, rely on outdated and inaccurate estimation methods.

All earthmoving operations will involve the interaction of many items of plant that will affect the possible output from the operation. The dumptruck - backacter type of operation, on which the work presented herein is based, consists of in its simplest form, a backacter excavator that loads in turn each truck in a fleet of dumptrucks. The dumptrucks haul the loaded material to a tip or fill area, dump the material and return to the loading area where they are reloaded. The interactions between the items of plant need to be fully understood by the planner and estimator to enable them to confidently assess the probable productivity and required plant of an operation that has not yet been started. For example, the size of the bucket on the excavator will affect the time it takes to load the dumptrucks which in turn will affect the number of trucks needed to maintain the maximum possible production.

It is with this problem in mind that the work presented in this thesis has been undertaken. The work has been presented on what are classical problem solving lines: firstly, to identify the problem, then propose possible solutions, test and redevelop these ideas and, finally, implement and maintain a proposed method of solution. The initial step in identifying the problem is perhaps not as easy as at first may seem. At this stage all that is known about the situation is that the productivity and resources required for an earthmoving operation need to be estimated with greater accuracy. To understand the situation as fully as possible, the first step was therefore to simply watch and record real earthmoving operations. To do this, work study techniques were used which required full consultation with all the site personnel involved. An operator who is being watched without knowing why may understandably feel uncomfortable about the situation and work in a different way to normal. If this happens we cannot draw accurate conclusions about the existing situation. Four earthmoving contracts have been studied to various degrees that have also supplied the raw data on which the proposed solutions are based. These sites and the initial work undertaken are described fully in chapter 3.

The next stage is to develop possible solutions to the identified problem. This has been an iterative process with the first solution (model 1) being a simple but inadequate method of estimating earthmoving productivity. It does however provide a reliable vehicle for understanding the earthmoving situation further and has become the basis for a much more complex earthmoving model (model 2) based on simulation techniques.

As with all problems, any solutions proposed have to be tested fully to ensure results obtained are reliable. In this situation, a further problem is that a simulation model is very good at providing many results and hence the user soon becomes to believe that the data obtained is accurate, which could be very damaging if too many conclusions are drawn. A major part of the work undertaken was to *validate* the simulation model to ensure the model represents as close as possible real life situations and that any decisions made through using the model are the correct ones. Chapter 6 discusses various validation techniques and presents the results of a validation exercise carried out on the simulation model developed in chapters 4 and 5.

Finally, it is important that any weaknesses in the system of estimation proposed are fully understood. This will enable any users to allow for those results that may not be as accurate as required. An understanding of these areas will also provide the basis for future work in a subject that has rarely been looked at by many researchers before. Experimentation of the proposed simulation model has therefore been undertaken to determine how the output from the system is influenced by different input levels. The experimental analysis is outlined in chapter 7.

Chapter 2

Earthmoving in the UK An Overview

The background of the earthmoving industry is presented: the history, planning methods (including mass haul and linear optimisation), types of earthmoving plant and the importance of a knowledge of the ground conditions when undertaking an earthmoving project.

2.1 Introduction

This chapter is essentially an extended introduction to chapter 1; it sets the scene for the rest of the thesis and serves as a place for topics that, although important, are not big enough for their own chapter or have no other 'heading' to which they belong.

To start with, a brief history of earthmoving on road-building contracts in the UK will be given and a few comments on the likely state of the industry in the future. Earthworks planning will then be covered, including an overview of past methods of earthmoving planning. (In this thesis, the words *earthworks* and *earthmoving* have separate meanings.) A large section of this chapter concerns the long running debate over what type of plant should be used in earthmoving operations. Finally, soil properties relating to earthworks and earthmoving will be reviewed, especially those applicable to the geology of the UK.

2.2 A Brief History of Earthmoving in the UK

Although earthmoving operations undoubtedly existed in many early roads in Britain, it was not until the advent of the motorway network that major earthworks contracts were undertaken. The amount of earthmoving carried out escalated in the sixties and seventies until its decline during the eighties. Today, although the total volume of earth to be moved on construction sites may be high, it is spread over many minor road improvements, by-passes and motorway widening schemes.

2.2.1 Motorways

It was Italy who started the motorway building business; in 1923, the state formed several subsidised companies to construct and operate national *autostrada*. These roads were unlike present motorways for they consisted of single carriageways only but they set the standard for all future works, especially for earthmoving, as they were designed to permit high speed travel. This entailed limiting the gradients and curves of the road and intersections had to be grade separated. Such roads could therefore not follow the natural contours of the land, but had to be cut into and elevated above the natural landscape.

In the 1930s both Germany and the United States began forming their national motorway networks. In 1934 Hitler set up the *Reichsautobahnen*, a subsidiary of the state railway agency, with responsibility for the design and construction of the *autobahnen*. By 1937, almost 1,000 miles had been open to traffic with a further 1,100 miles under construction. The length open to traffic had increased to 2,300 miles just before the start of the war.

Britain's first motorway opened in December 1958 and can thus be considered as late starters in the motorway construction business. However, plans for such roads had, in perhaps typical British fashion, been considered many years before: in 1900 Balfour, the Prime Minister proposed “great highways constructed for rapid motor traffic and confined to motor traffic.” (Drake, et al, 1969). In 1905 the Road Improvement Association was formed. Time passed until in 1936, the Institution of Highway Engineers proposed a motorway network to the Ministry of Transport consisting 51 lengths of road covering 2,826 miles. This proposed network can be seen in figure 2.1 and it is interesting to see that it is very similar to the actual motorway and primary route network of today, some 60 years later. It wasn't until 1949, when the Special Roads Act was passed, that motorways could finally be built in Britain. The first motorway began construction on 12 June 1956 and in the following 10 years only 400 miles were constructed and opened to traffic.

2.2.2 Motorways and Earthmoving

The seventies saw a boom in motorway construction in Britain with a consequent rise in the amount of earth moved. In 1969, the total quantity of earthworks on major British contracts amounted to 34 million m^3. In 1972, this figure peaked at approximately 48 million m^3 and by 1984, with the majority of the motorway network in place, the total quantity had dropped to below 30 million m^3. (Barker, 1988.) The late eighties saw a change in the type of work being undertaken, with the volume of earthworks being spread over many smaller road building and improvement contracts. This change was mirrored also by a shift in the type of plant used to carry out the earthmoving.

When the motorway network was first being set up, motorscrapers were the standard item of plant, compared with the dumptruck - excavator method favoured today. In those days, the road-building business was perfect for these large, highly productive work-horses: the roads being built were wide, allowing plenty of room for manoeuvring, profit margins were high (compared to the situation today) which meant that contractors paid comparatively less attention to cost control and also, there were fewer time restrictions. The earthmoving season is much shorter with a scraper compared

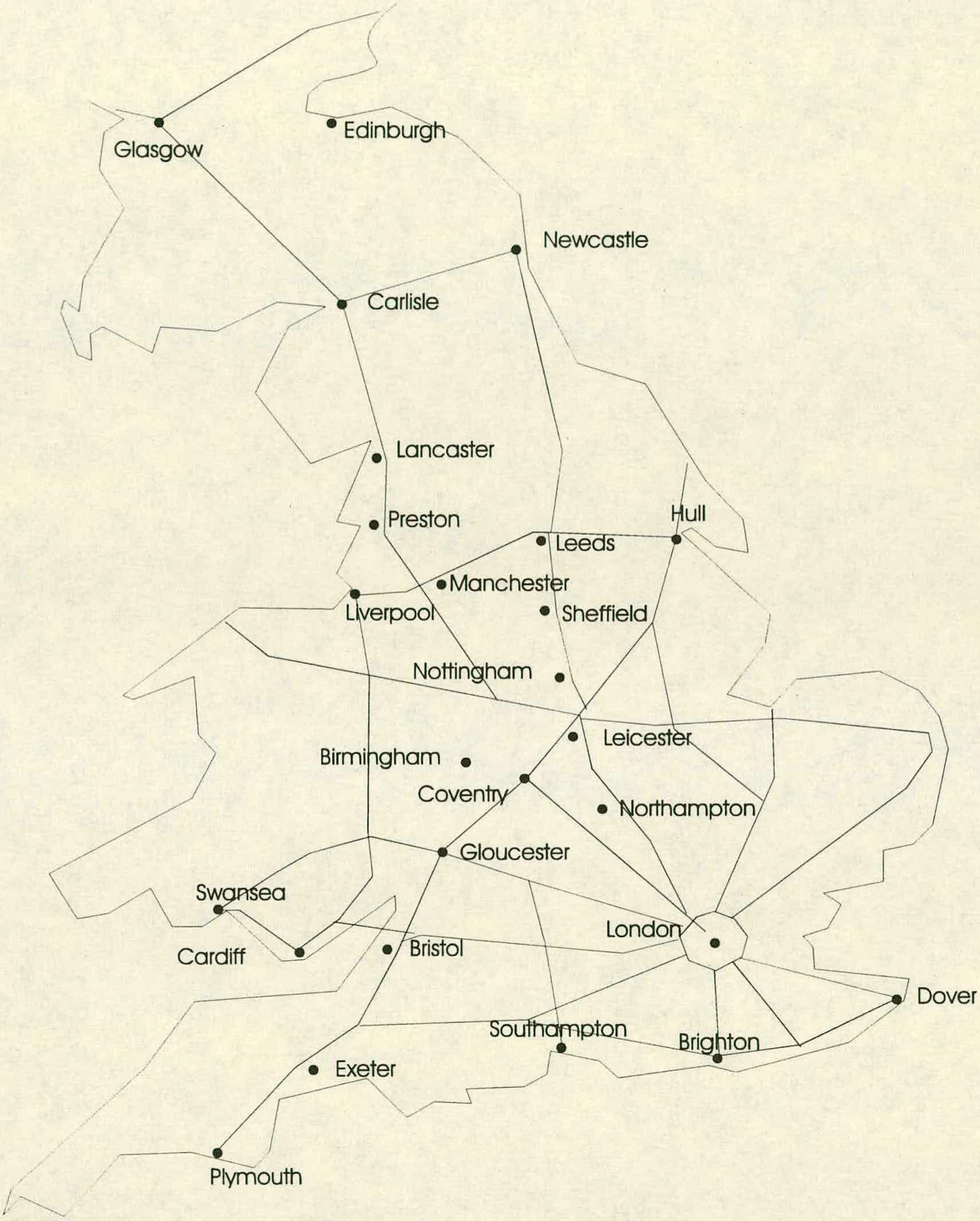


Figure 2.1 Motorway Network Proposed by Institution of Highway Engineers, 1936

with a dumptruck but this was not much of a problem when completion time was less important.

However, in the early eighties, the motorscraper lost favour with the earthmoving contractors and this is reflected in the number of units sold. In 1979, 90 were sold in the UK; in 1983, this number had dropped to 36 and by 1988, less than 20 units were bought by British earthmoving contractors (Byles, 1984; NCE, 1989.) The main reason for this had to be the near completion of the motorway network; with fewer large contracts available, suitable work could not be guaranteed for scrapers that, at today's prices, can cost up to £400,000. The dumptruck and in particular, the articulated dumptruck, was a much more flexible and practical machine; although not as productive as a scraper, the units are far cheaper and more fuel efficient than their predecessors. Dumptrucks could, due to their lower weight, reduce the damage to haul roads and thus extend the earthmoving season into the winter months. Consequently, contractors can now find work for their plant nearly all year round.

2.2.3 The Future of British Earthmoving

As has already been stated, there has been a shift in recent years from new road schemes to improving or widening contracts. For the earthmoving industry, this means that there are going to be fewer large contracts with substantial amounts of earthmoving involved. This is one reason why the project, for which this thesis is a part of , was set up: an over capacity of British earthmoving contractors, with lower profit margins and tighter time constraints means that they can no longer rely on old rules of thumb and must be more exact in their estimation methods. Barker noted (1988) that “in todays high-tech world, the average muckshifter relies still on the rules of thumb that remain unchanged for the past 50 years” and even now, 6 years later on from that comment, a PC is not a common tool in many earthmoving contractors' offices.

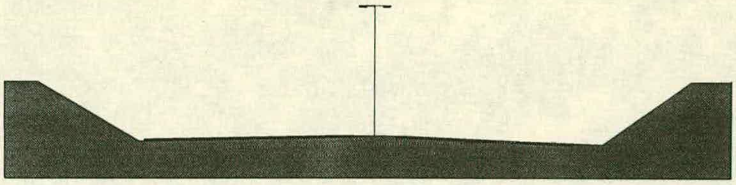
It is clear, therefore, that the way ahead is to approach earthmoving in a more scientific fashion and that the contractors that survive the next ten years will be the ones that have the correct attitude to the work. It is also clear that the workload of the seventies will never return: the near completion of the motorway network and the pressure from the environmental lobby has guaranteed that. It is unclear, however, exactly what size and shape the earthmoving industry will take as it approaches the year 2000.

In May 1989, a respite was offered by Paul Channon, the Secretary of State for Transport, when he announced that the value of proposed road schemes would be raised to £12 billion (November 1987 prices.) Of this (which is intended to cover a 10 year period) only £400 million is allocated to new motorways. Four billion, however, is to be spent on widening contracts and it is here where the majority of the earthmoving contractors will concentrate their work. Although not as intensive as new build, widening schemes have to keep disruption of traffic on the existing roads to a minimum; the best way to do this is to carry out parallel widening. As can be seen in figure 2.2, upgrading a motorway from dual three to dual four by this method is equivalent to building a new road with a width of five lanes (approximately 18m). This can be compared with symmetrical and asymmetrical widening that have only approximately 7m of extra land take. As a comparison, a very broad estimate of earthworks volume can be calculated. If the average motorway in the UK has, say, half of its length as cut with an average depth of 3m then the earthworks volumes are as follows:

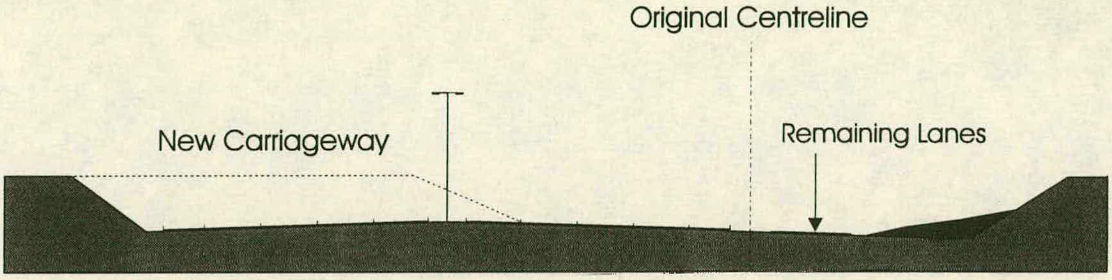
New dual 3 lane:	54,000 m^3/km
Dual 3 to dual 4,	
Parallel:	30,000 m^3/km
Symmetrical:	10,500 m^3/km
Asymmetrical:	10,500 m^3/km

Parallel widening is therefore very earthworks intensive. It cannot be said, however, how much of the planned widening scheme will be parallel or symmetrical as this will depend on local land use and public opinion.

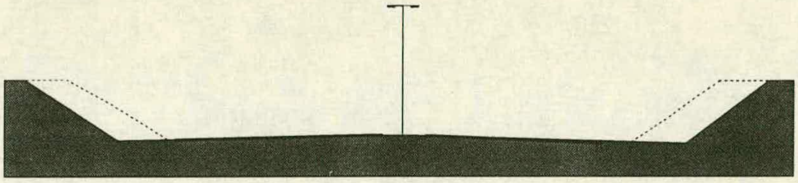
At the time of writing, the future of earthmoving associated with the road-building programme is even more uncertain. Highly unpopular road schemes such as the M3 Twyford Down (which is currently being built), the south London Oxleas project (which was abandoned) and the M11 extension to the east of London (still in consideration) have caused the present government to think again about the planned programme announced in May 1989. In March 1994, the Secretary of State for Transport, John MacGregor announced a major restructuring of the planned programme with nearly a third of the budget cut (The Guardian, April 1, 1994). Many planned projects were abandoned with more postponed. While this is a major step forward for environmental aspects, it is very bad news for the road-building contractors and earthmoving. For this reason, it is even more important that existing contractors are even more accurate with their estimates.



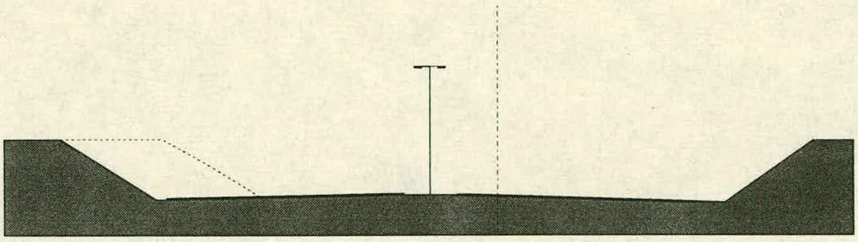
Original Section



Parallel Widening



Symmetrical Widening



Asymmetrical Widening

Figure 2.2 Motorway Widening Methods

This section has concentrated on road-building as it is a major part of the present government's transport policy. What if, however, even more cuts were announced in road-building or the government were to change? The main alternative to roads are railways and if a major investment were to be allocated to new rail schemes (such as the planned Channel Tunnel rail link through Kent) then this would bring work for civil engineering contractors, but the amount of earthmoving would be far less. If the above assumptions regarding road cut size are used, then for rural railways, with a track width of, say, 8m, the earthworks volume would be approximately 21,000 m^3 per kilometre. At the present time, however, such an investment seems unlikely and to try and predict the state of the British transportation network even further is beyond the scope of this thesis.

2.3 Earthworks Planning

The ideal road-building contract, as far as the earthworks is concerned, is one in which all the material cut out of the ground is used somewhere else in the contract and that the haul lengths for these cut and fill operations are not too great (the longer the haul, the greater the cost to move the earth). This is rarely achieved in practice, but the road-building contractor, when faced with the design earth quantities must minimise his cost by allocating the fill material carefully. This has traditionally been done by hand using mass-haul diagrams but more recently, computer methods have been developed which can do this task quicker and more efficiently.

Once the earthworks have been split into individual cut and fill activities, the contractor must determine the cost, duration and resources required to carry out the earthworks. It is here where the distinction between earthworks and earthmoving should be made. For the purpose of this thesis, *earthworks* shall mean the design and position of the finished product, i.e. the completed cuttings and embankments. *Earthmoving* is the process of carrying out the work to complete the earthworks. More specifically, earthmoving is merely the excavate, load and haul part of an earthworks contract. This thesis is primarily concerned with earthmoving but this section shall briefly describe earthworks planning for completeness.

2.3.1 Mass Haul Diagrams

When the contractor receives the documents to tender for an earthmoving contract, a schedule of cut and fill volumes for the various sections of the project is included. The traditional method of allocating what excavated material goes to which embankment is to construct a mass haul diagram (Ashworth, 1966; Oglesby and Hicks, 1982).

As an example, consider the mass haul shown in table 2.1. The first column is the chainage (distance from one end point of the road in metres) of the end points of each separate area of cut or fill. For each area there is an acceptable cut volume and a fill volume - the area is classified cut or fill depending on the what the overall volume of that area is. Cut is always denoted as positive while fill is negative. Column five shows the volume of the excavated material after shrinkage: soils generally swell (i.e. are looser) after excavation but are compacted to a smaller volume than the undisturbed state (or *bank volume)* in the embankment. In this case a shrinkage factor of 0.9 has been taken but this should be determined individually for each contract. If the material is rock then a swell factor should be used as rock takes up a greater volume in the recompacted state. In general, the swell (shrinkage) factor is:

$$s = \frac{V_c}{V_b} \qquad (2.1)$$

where V_c is the compacted volume in the embankment and V_b is the bank volume.

The overall cut or fill volume is then calculated and shown in column 7. The cumulative volumes, column 8, are then the ordinates for the mass haul diagram, plotted against the chainage. This can be seen in figure 2.3 which can then be used to allocate which items of cut go to which fills. The mass haul diagram has certain properties:

- If the mass haul line at any chainage is above the base line (i.e. the *x*-axis) then the net volume at that point is positive, indicating a surplus of material. Below the base line indicates a net deficit of earth. Figure 2.3 indicates that the there is a net deficit for the whole contract (chainage 8900) which must be filled from a borrow pit.
- Positive gradients indicate that the road is in cutting at that point, negative gradients represent embankment.

1 End Chainage	2 Cut/ Fill	3 Number	4 Acceptable Material	5 After Shrinkage	6 Fill	7 Cut (Fill)	8 Cumulative Cut (Fill)
0	Fill	1	1,750	1,575	3,500	-1,925	-1,925
100	Fill	2	1,400	1,260	4,500	-3,240	-5,165
500	Fill	3	50	45	2,600	-2,555	-7,720
765	Cut	1	17,022	15,320	2,000	13,320	5,600
1075	Cut	2	17,000	15,300	1,300	14,000	19,600
1200	Fill	4	200	180	5,960	-5,780	13,820
1435	Cut	3	14,222	12,800	0	12,800	26,620
1750	Fill	5	300	270	26,470	-26,200	420
2205	Fill	6	280	252	13,700	-13,448	-13,028
2400	Fill	7	1,500	1,350	8,502	-7,152	-20,180
2960	Cut	4	2,000	1,800	100	1,700	-18,480
3100	Cut	5	3,600	3,240	750	2,490	-15,990
3215	Fill	8	3,500	3,150	12,140	-8,990	-24,980
3765	Cut	6	27,000	24,300	2,500	21,800	-3,180
4900	Cut	7	17,500	15,750	50	15,700	12,520
5090	Cut	8	16,000	14,400	1,200	13,200	25,720
5470	Cut	9	22,500	20,250	2,500	17,750	43,470
5650	Fill	9	1,200	1,080	8,180	-7,100	36,370
6150	Cut	10	28,278	25,450	1,300	24,150	60,520
6720	Fill	10	2,500	2,250	25,000	-22,750	37,770
7265	Fill	11	3,000	2,700	13,000	-10,300	27,470
7880	Fill	12	2,700	2,430	14,500	-12,070	15,400
8390	Fill	13	1,250	1,125	17,500	-16,375	-975
8900	Fill	14	750	675	11,255	-10,580	-11,555
		Totals:	185,502	166,952	178,507		

Table 2.1 Example of Mass-Haul (Volumes in cubic metres)

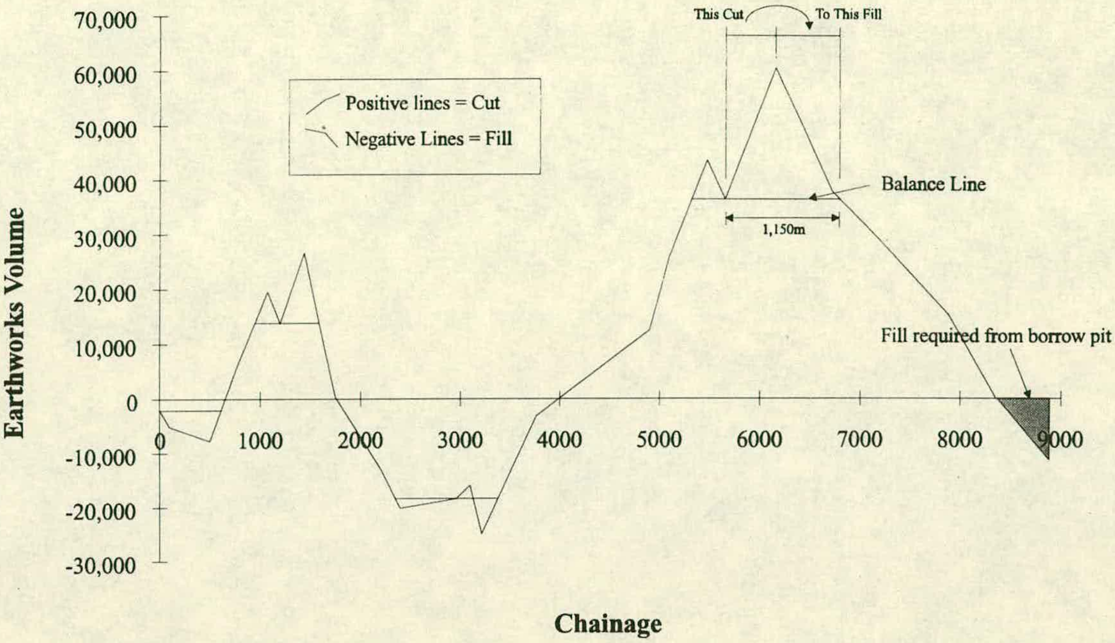


Figure 2.3 Mass Haul Diagram

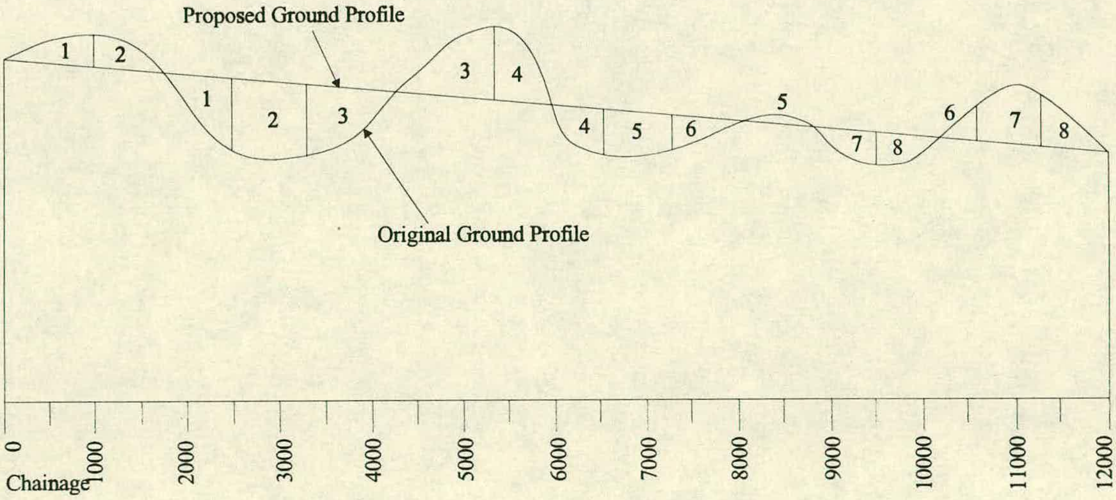


Figure 2.4 Longitudinal section for earthworks allocation example

- Maxima occur at the end of cuttings and minima at the end of embankments.
- The vertical distance between any two points on the diagram is the net difference in volume between the two points.
- Any horizontal line, including the base line, cuts the diagram at chainages between which the earthworks volume is balanced. These lines are balance lines.
- The length of each balance line is the maximum haul distance between two points.

Thus from the mass haul diagram, the contractor can determine the separate cut and fill operations for the contract.

2.3.2 Earthworks allocation by linear optimisation

Constructing a mass haul diagram can be very time consuming, although the use of a spreadsheet improves efficiency, and under the present conditions of contract (ICE, 1991, cl.14) if a contractor is awarded the contract, he must, within 21 days, submit a programme showing the order in which he intends to carry out the works. Therefore, the earthworks planning must be carried out efficiently, in as short a time as possible. For many larger contracts, where the total volume may be up to 3 or 4 million cubic metres with up to seventy separate operations, this is not a quick task. The mass haul application is also limited in other occasions, such as when the haul costs are not proportional to haul distance (which can occur when plant crossings, Bailey bridges or soft spots are present); different strata occur at different layers; alternative borrow or dump sites exist; shrinkage and swell factors change over the length of the road or construction is staged. One way to overcome this problem is to apply linear optimisation techniques to minimise the cost of earthworks or maximise profit.

A linear programming model was first proposed by Stark and Nicholls (1972) and has been developed over the intervening years by Mayer and Stark (1981), Nandgaonkar (1981), Hander and Barcia (1986), Easa (1987, 1988) and Jayawardane and Harris (1990). More recently, Jayawardane and Price (1994 a and b) have combined the techniques of simulation and linear programming to optimise earthworks allocations. If the objective is to allocate the earthworks quantities for minimum cost, then the unit costs for excavation and load (in £/m^3), haul (in £/m^3/m) and compaction (in £/m^3)

must be known. These can be represented by u_e, u_h and u_c respectively. Therefore, the cost to move a cubic metre from cut *i* to fill *j* is

$$C(i,j) = u_e + u_h d_{ij} + u_c \tag{2.2}$$

where d_{ij} is the distance between the centres of mass of cut *i* and fill *j*. Costs for disposal to tip, $C_D(i,k)$ for each tip *k*, and borrow, $C_B(p,j)$ for each borrow pit *p* can be constructed in a similar way. These are:

$$C_D(i,k) = u_e + u_h d_{ik} + u_c \tag{2.3}$$

and

$$C_B(p,j) = u_p + u_e + u_h d_{pj} + u_c \tag{2.4}$$

where u_p is the additional unit cost of purchase (£/m^3).

For each cut area *i*, let *X*(*i,j*) be the (unknown) quantity of earth to be moved to fill area *j* and $X_D(i,k)$ be the quantity of earth to be disposed in dump area *k*. If material is to be taken from borrow pits, then $X_B(p,j)$ is the quantity to be taken from pit *p* to fill area *j*. The total cost for excavate, load, haul and compact for the whole contract, *Z*, is therefore:

$$Z = \sum_i \sum_j C(i,j) X(i,j) + \sum_i \sum_k C_D(i,k) X_D(i,k) + \sum_p \sum_j C_B(p,j) X_B(p,j) \tag{2.5}$$

which can be solved, using linear programming techniques, for all *X*.

The variables $X(i,j)$, $X_D(i,k)$ and $X_B(p,j)$ are all limited by the designed quantities of cut and fill for each area; the quantity constraints for (2.5) are:

$$\sum_j X(i,j) + \sum_k X_D(i,k) = Q_c(i) \tag{2.6}$$

where $Q_c(i)$ is the quantity of cut available from each area *i*, and:

$$\sum_i X(i,j) + \sum_p X_B(p,j) = Q_f(j) \tag{2.7}$$

where $Q_f(j)$ is the amount of fill required in each embankment j. There may also be constraints for each tip and borrow pit of the form:

$$\sum_i X_D(i,k) \le Q_D(k) \tag{2.8}$$

and

$$\sum_j X_B(p,j) \le Q_B(p) \tag{2.9}$$

Finally, all volumes must be positive and so the conditions:

$$X(i,j) \ge 0,\ X_D(i,k) \ge 0 \text{ and } X_B(p,j) \ge 0 \tag{2.10}$$

must also apply.

The above optimisation model does not take into account any swell or shrinkage of the material on excavation and compaction. This must be done if volumes calculated are to 'fit' on site and therefore the above equations must be revised to take this into account. For example, the cost coefficient for excavation, load, haul and compact between cut and fill (2.2) becomes:

$$C'(i,j) = u_e + s_i^h(u_h d_{ij} + u_c) \tag{2.11}$$

where s_i^h is the swell or shrinkage factor (calculated from 2.1) for the material in haul. The cost coefficients for C_D and C_B must be similarly changed. Constraints for fill (2.7 and 2.8) must also allow for swell or shrinkage of the material upon compaction although cut constraints (2.6 and 2.9) need no modification as the quantities in these equations refer to material in the unexcavated state. The revised cost equation to be optimised is therefore:

$$Z = \sum_i \sum_j C'(i,j)X(i,j) + \sum_i \sum_k C'_D(i,k)X_D(i,k) + \sum_p \sum_j C'_B(p,j)X_B(p,j) \tag{2.12}$$

and this, along with the constraints and conditions defined above can be solved using a computer package designed for the task. Such examples can be found in Mayer and Stark, 1981.

2.3.3 Example of Optimum Earthworks Allocation Problem

In addition to dedicated linear programming packages, earthworks allocation problems can also be carried out using a computer spreadsheet package. An example has been set up using the *Solver* tool in the *Microsoft Excel* package. (Microsoft, 1992.)

Figure 2.4 shows a longitudinal section through an imaginary road construction project with 8 cut sections and 8 fill sections. Table 2.2 shows the design volumes for the cuttings and embankments with the chainages of the centres of mass of each section. The problem is: what volume of material from each cut area is required in each fill area for the minimum cost? For simplicity, the total cut has been made equal to the total fill so no material needs to be disposed nor is any extra required. In addition, the swell or shrinkage factors of the material are ignored. The cost equation is therefore:

$$Z = \sum_{i=1}^{8}\sum_{j=1}^{8} C(i,j)X(i,j) \tag{2.13}$$

The cost coefficients for each *(i,j)* are calculated using haul lengths determined from table 2.2 and the unit costs, which are:

$u_e = 0.3\ £/\mathrm{m}^3$

$u_h = 0.001\ £/\mathrm{m}^3/\mathrm{m}$

$u_c = 0.15\ £/\mathrm{m}^3$

Using the constraints that the total cut and fill volumes cannot exceed those stated in table 2.2, equation 2.13 was solved for a minimum answer using Solver. Using initial values of 1000 m^3 for each possible cut to fill, a solution, to an accuracy of 0.1 m^3 was found after 51 iterations. The results can be seen in figure 2.5, which shows that there are 21 separate operations to give a minimum total cost of £280,050.

This procedure can easily be extended for bigger contracts and for different swell or shrinkage factors.

2.4 Earthmoving Plant

Section 2.1 briefly mentioned the types of plant that can be used to carry out earthmoving operations, and also that the motorscraper has been gradually superseded in recent years by the articulated dumptruck. This section will expand this

Cut	Design Volume	Chainage of Centroid	Fill	Design Volume	Chainage of Centroid
1	20,000	600	1	17,000	2,200
2	17,500	1,300	2	20,000	2,900
3	15,000	4,950	3	16,500	3,600
4	16,500	5,600	4	13,000	6,350
5	7,500	8,500	5	21,000	6,950
6	8,000	10,450	6	14,500	7,500
7	17,000	10,950	7	8,000	9,300
8	14,500	11,500	8	6,000	9,750
	116,000			116,000	

Table 2.2 Volumes and Locations for Earthworks Allocation Example

Unit costs:

excavate	0.3	£/m3
haul	0.001	£/m3/m
compact	0.15	£/m3

Volumes:

Fill	Cubic Metres								
Cut	1	2	3	4	5	6	7	8	Totals:
1	4,641	15,359	0	0	0	0	0	0	20,000
2	12,359	4,641	500	0	0	0	0	0	17,500
3	0	0	15,000	0	0	0	0	0	15,000
4	0	0	1,000	13,000	2,500	0	0	0	16,500
5	0	0	0	0	7,500	0	0	0	7,500
6	0	0	0	0	1,716	5,378	906	0	8,000
7	0	0	0	0	5,045	4,349	4,115	3,491	17,000
8	0	0	0	0	4,239	4,773	2,979	2,509	14,500
Totals:	17,000	20,000	16,500	13,000	21,000	14,500	8,000	6,000	

Cost Coefs.:

Fill	£ per cubic metre							
Cut	1	2	3	4	5	6	7	8
1	2.05	2.75	3.45	6.20	6.80	7.35	9.15	9.60
2	1.35	2.05	2.75	5.50	6.10	6.65	8.45	8.90
3	3.20	2.50	1.80	1.85	2.45	3.00	4.80	5.25
4	3.85	3.15	2.45	1.20	1.80	2.35	4.15	4.60
5	6.75	6.05	5.35	2.60	2.00	1.45	1.25	1.70
6	8.70	8.00	7.30	4.55	3.95	3.40	1.60	1.15
7	9.20	8.50	7.80	5.05	4.45	3.90	2.10	1.65
8	9.75	9.05	8.35	5.60	5.00	4.45	2.65	2.20

Costs:

Fill	£								
Cut	1	2	3	4	5	6	7	8	
1	£9,514	£42,237	£0	£0	£0	£0	£0	£0	£51,751
2	£16,684	£9,514	£1,375	£0	£0	£0	£0	£0	£27,574
3	£0	£0	£27,000	£0	£0	£0	£0	£0	£27,000
4	£0	£0	£2,450	£15,600	£4,500	£0	£0	£0	£22,550
5	£0	£0	£0	£0	£15,000	£0	£0	£0	£15,000
6	£0	£0	£0	£0	£6,778	£18,286	£1,449	£0	£26,514
7	£0	£0	£0	£0	£22,449	£16,960	£8,642	£5,760	£53,812
8	£0	£0	£0	£0	£21,196	£21,240	£7,894	£5,519	£55,849
								Total Cost =	£280,050

Figure 2.5 Results from analysis of earthworks allocation example

further by making a quantitative comparison between the two methods of earthmoving.

2.4.1 The Demise of the Motorscraper

For many, earthmoving is synonymous with the motorscraper. In the motorway construction boom of the seventies, hundreds of these machines could be found up and down the country eating away at the landscape, belching huge black clouds of exhaust fumes with the constant drone of huge diesel engines toiling for up to 14 hours a day. However, as has already been mentioned, the slowing down of the motorway building programme and the consequent overcapacity of earthmoving contractors, making a profit from any contract became more and more difficult. By the early eighties, cheaper methods of moving the earth, with plant that could be used all year round, had to be found.

In 1981, *Contract Journal* asked “is the motorscrapers' supremacy ... being challenged by the dumptruck/backacter combination?” (Marshall, 1981.) Byles (1984) noted that the “gradual decline of larger contracts and near completion of the motorway network has had a dramatic effect on UK motorscraper sales.” It was commented by a director of one of the countries largest earthmoving contractors that a new £400,000 motorscraper could cost £40/hr just in interest charges to the bank (Byles, 1988). Another earthmoving contractor noted that “motorscrapers entail all kinds of problems, not only in terms of high initial cost but also in heavy fuel bills and maintenance problems. Truck and Shovel gives flexibility to work in bad conditions” (NCE, 1989).

Unfortunately, the plant manufacturers responded perhaps too quickly to this change of attitudes and early versions of articulated dumptrucks had many problems, particularly with the suspension: some trucks, especially when empty, bounced so much when moving that the haul roads quickly deteriorated (Byles, 1984.)

2.4.2 Scraper or Dumptruck: a Quantitative Comparison

In terms of production, the main difference between a scraper and a dumptruck is the time taken to load. A scraper, because the loading action is continuous, takes much less time for a full load than a dumptruck of equivalent size - especially if it is push loaded with a bulldozer. The consequence of this is that an earthmoving operation that uses scrapers can utilise more items of plant, due to a quicker turnover rate, than an operation that uses dumptrucks and hence the production can be much higher. However, this can be offset by the fact that the unit cost per cubic metre is much higher.

In an attempt to quantify the difference in production and cost between the two types of operation, an analysis has been conducted for an operation that has the same parameters apart from the type of plant used. (Note that the operation depicted is not a real operation.) Table 2.3 shows the parameters for the two alternatives: Terex TS 24 motorscrapers coupled with a Cat D9 pusher and Cat D400 dumptrucks loaded with a Cat 245 backhoe excavator. Using the simple model that will be detailed in chapter 3 (3.7) the production rate for the two operations can be determined for different haul lengths. As can be seen from figure 2.6 the production of an individual hauler falls logarithimicaly as the one way haul length increases. Although the load time of a scraper is 30 seconds quicker than for the dumptruck, the lighter Cat D400 travels at a greater speed (20 km/h) than the Terex TS 24 (14km/h) and so the overall cycle time and hence the number of cycles per hour is not much different for short hauls. As can be seen however, as the haul length increases the production for the scraper falls at a greater rate than that of the dumptruck due to the difference in speeds having a greater effect.

Figure 2.7 depicts the total operation production for the two alternatives and it can be clearly seen that the scraper fleet can produce more earth per hour and this is solely due to the quicker load cycles: more scrapers can be accommodated by the pusher and so the service rate is higher. The number of hauler units required for these production rates can be seen in figure 2.8. As the haul length and cycle time increases, the number of haulers increases in steps to keep the efficiency of the operation as high as possible. The shorter increase in haul length before an increase in resource level is required by the pusher in the scraper operation indicates the higher service rate of the pusher; that is, the pusher is more susceptible to falling idle than the excavator would be.

Finally, figure 2.9 shows the difference in cost per cubic metre for the two alternatives. It can be easily concluded therefore that the dumptruck/excavator operation, while not as productive, is far more cost effective than the scraper/pusher method. This production comparison alone should convince earthmoving contractors which method is better but it is fair to say that scrapers in some instances can be the best alternative. Some of these are:

	Terex TS 24/Cat D9	Cat D400/Cat 245
Minutes per hour	50	50
Volume per cycle (m3)	14.0	14.0
Dump Time (s)	60	60
Load Cycle Time (s)	80	120
Average speed of hauler (km/h)	14.0	20.0
Cost of Hauler (£/hr)	43.00	25.00
Cost of Loader/Pusher (£/hr)	41.00	46.00

Table 2.3 Operation parameters for dumptruck/scraper comparison

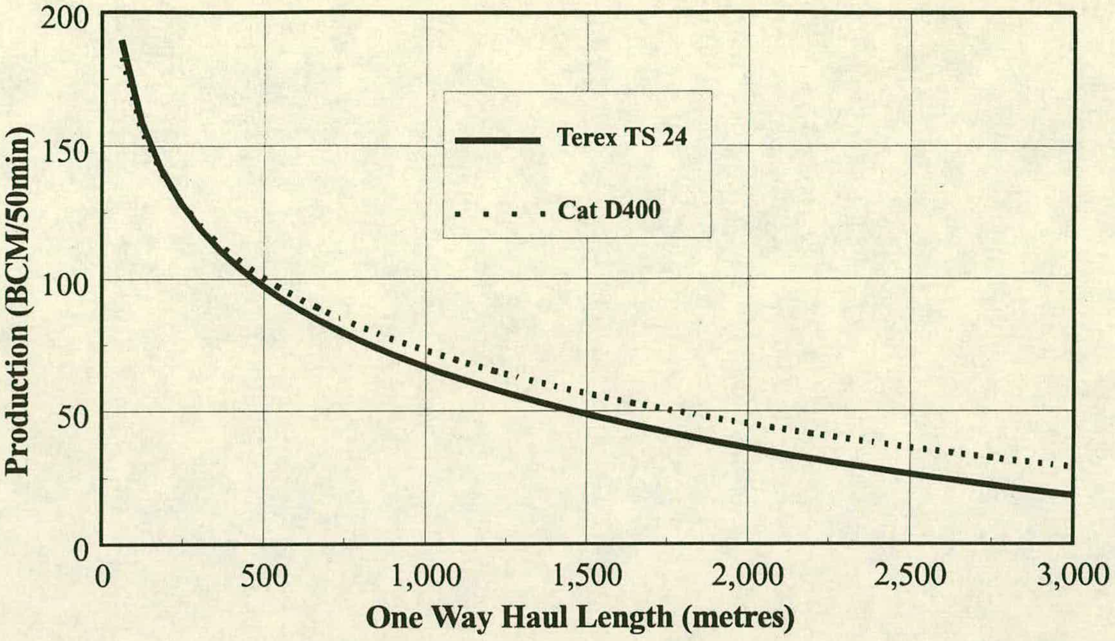


Figure 2.6 Individual Production Comparison

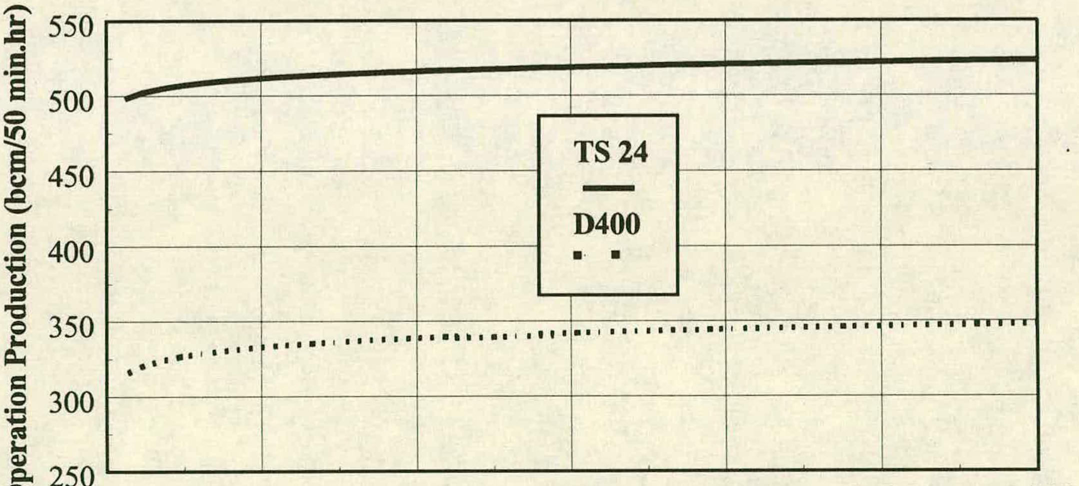


Figure 2.7 Total operation production comparison

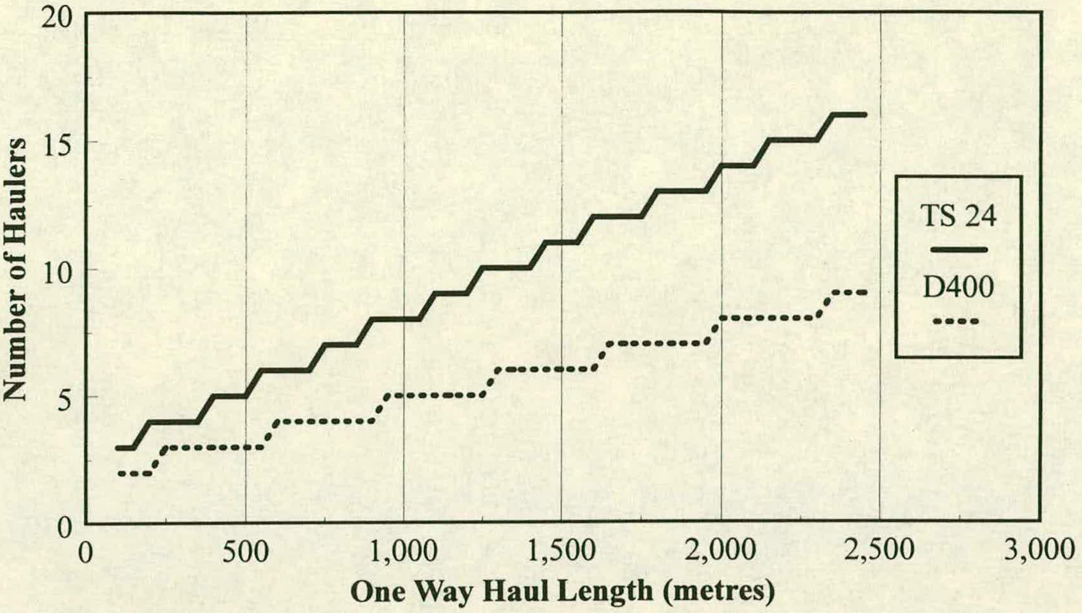


Figure 2.8 Comparison of Plant Numbers for Maximum Production

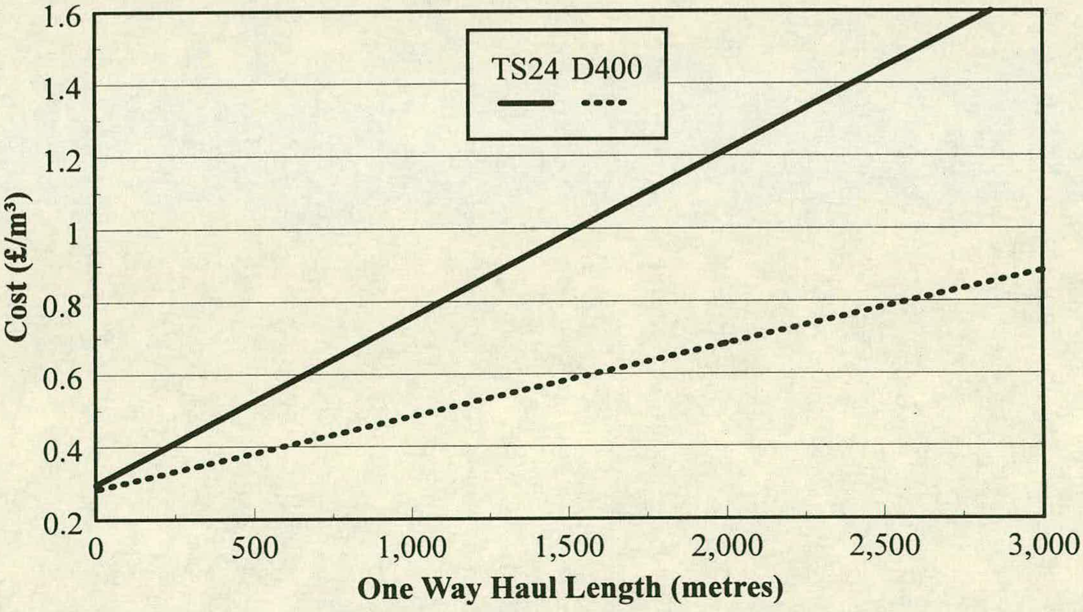


Figure 2.9 Cost Comparison

- Topsoil stripping. The motorscraper is far better at this type of operation than the excavator merely by the cutting action: scrapers have a 'planing' action than lends itself to cutting small depths while the excavator is at its most efficient when large depth cuttings are to be removed.
- Very short hauls. For hauls up to 500m, the scraper is ideal as the advantages in the higher production rate far exceed that of lower cost.
- When high production is required. In some instances, cost may not be the deciding factor in choosing plant. If an operation is required to be completed as quickly as possible then clearly a scraper should be used.

2.5 Site Investigation, Ground Conditions and Earthmoving

Like all projects that require ground engineering, earthworks operations require a detailed site investigation for all of the planned haul routes. Unfortunately, the requirements for road pavement design are different to those required if the unfinished ground is used as a haul route; most site investigations for road contracts, therefore, have few soil parameters which could be of use in determining the trafficability of the soil. The importance of site investigation is well recognised (Littlejohn et al. 1994; Institution of Civil Engineers, 1991; Peacock and Whyte, 1988) and the risks associated with poor site investigation can be studied by risk analysis (Peacock and Whyte, 1992; Whyte 1994). More specifically, probabilistic risk analysis assesses the probability of identified risks and the financial consequences thereof. For example, risk analysis could identify the probability that the cost of the works would be identified for a certain level of site investigation. The risks would decrease with better site investigation, but the level of confidence in the ground conditions affect the outcome of the permanent works: will the structure stay up or will the pavement not deform? In these cases, however, the risk is mainly the clients and the contractor will bear no costs if a structure fails due to inadequate site investigation. Unfortunately, earthmoving operations can be seen as temporary works and the contractor, if he chooses to use the ground as his haul road, will take some of the risks involved with the conditions of the ground. He may carry out additional site investigation, thus affecting tender price, or claim for extra costs involved; either way, the client will have to meet some of the costs.

Although this thesis is not a study of soil mechanics, the ground conditions and the relationship between soil strength and soil moisture content is very important in earthmoving and, in some cases, a misunderstanding of the soil characteristics of the contract haul roads can mean the difference between a successful or loss-making project. Essentially, as a soil's moisture content increases, its strength decreases. However, the rate at which the strength decreases is dependent upon the plastic range of the material, which means that some contracts will be more susceptible to the effects of rain than others.

The susceptibility of a soil to moisture content can be explained by considering the *consistency* of a soil. As the moisture content of a soil changes from its liquid limit to its plastic limit, the consistency of the soil changes from a liquid to a brittle material and so the term consistency index can be used to quantify the moisture content of a soil with respect to its plastic range:

$$\mathrm{I_c} = \frac{w_L - w}{w_L - w_P} \tag{2.14}$$

or

$$\mathrm{I_c} = \frac{w_L - w}{PI} \tag{2.15}$$

Where w is the moisture content of a soil, w_L is the liquid limit, w_P is the plastic limit and PI is the plasticity index. The liquid limit of a soil is determined using a cone penetrometer test (BS 1377:1990) and has an undrained shear strength ($\mathrm{C_u}$) of approximately 1.6 kN/m^2. The plastic limit is considered to be the moisture content at which the soil becomes brittle and has long been determined using a rolling bead method (BS 1377:1990). The shear strength of the soil at this point can be determined using a standard triaxial compression test (BS 1377:1990) and various researchers (Skempton and Northey, 1952; Dennehy, 1978 and Arrowsmith, 1978) have suggested that the value of $\mathrm{C_u}$ at the plastic limit lies between 85 and 320 kN/m^2. Whyte (1982) suggests that if the average of these values is taken then a value of 110 kN/m^2 should be used. Taking these values of $\mathrm{C_u}$ at liquid and plastic limit and using data provided by Skempton and Northey (1952) Whyte proposes that there is a log linear relation ship between strength and consistency summarised by:

$$\mathrm{C_u} = 1.6e^{4.24\mathrm{I_c}} \tag{2.16}$$

To show how the plasticity index of the soil affects its susceptibility, consider two soils, A and B with plasticity indices of 10% and 40%, and liquid limits of 30% and 60% respectively. From 2.15, it can be seen that:

$$\ln \mathrm{C_u} = 0.47 + 4.23\left(\frac{w_L - w}{PI}\right) \tag{2.17}$$

so for soil A:

$$\mathrm{C_{uA}} = e^{13.16 - 0.423w} \tag{2.18}$$

and for soil B:

$$\mathrm{C_{uB}} = e^{6.82 - 0.106w} \tag{2.19}$$

Therefore, for a 10% increase in moisture content (say from 25% to 27.5%) soil A shows a decrease of 65% in $\mathrm{C_u}$ while soil B only shows a 23% drop. Figure 2.10 shows how $\mathrm{C_u}$ changes for changes in moisture content for soils A and B. For low plasticity soils, such changes in shear strength for small increases in moisture content, such as after a rain shower, can have disastrous results, especially for heavier plant. This has consequences for scraper operations due to their heavier weight and has the effect of reducing the scraper earthmoving season: the moisture contents at the beginning and end of the dumptruck season may weaken the soil too much for a scraper.

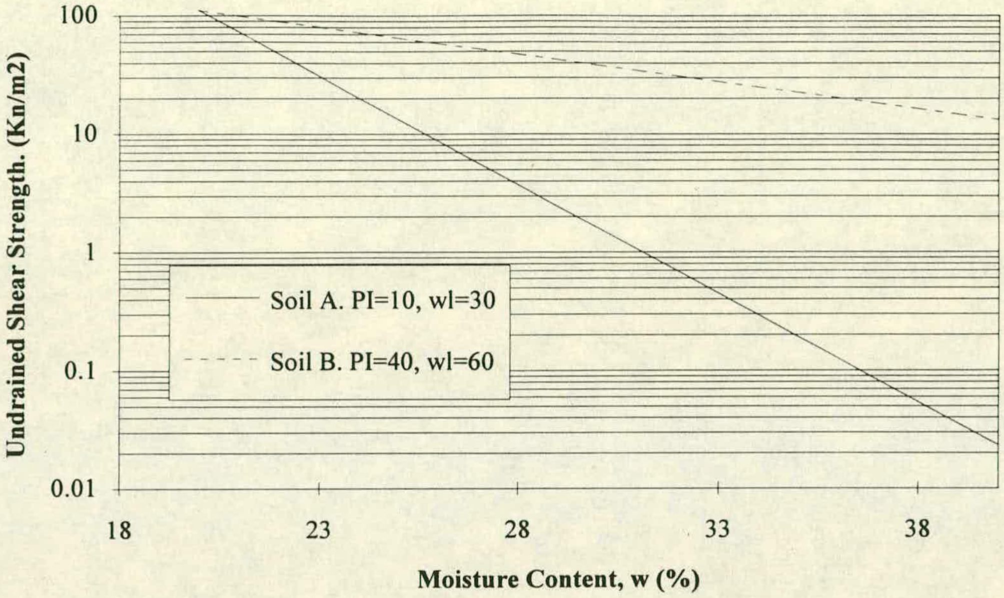


Figure 2.10 Effect of Moisture Content on Undrained Shear Strength of a clay soil.

Chapter 3

Field Work, Problem Identification and Initial Studies

The four road construction sites studied are detailed and the data collected from them is explained. The earthmoving system is described as a queueing system and the effect of plant match on the operation productivity is discussed. A deterministic model for the estimation of earthmoving productivity is introduced which works on a spreadsheet. It is found that a further loss of production occurs due to the variability of the plant cycle times

3.1 Introduction

The starting point in the investigation of earthmoving productivity estimation is to fully understand the situation that will be studied. The problem needs to be defined and data collected and analysed before any solutions are proposed. The best way to achieve this is to study actual earthmoving operations and this chapter describes the sites studied, describes and defines the different aspects of the situation that is being studied, the data obtained from observation of the existing situation and results from initial analysis of that data.

To collect data on the existing methods of working, the techniques of work study and more particularly, method study have been adopted and adapted to suit this particular situation.

3.2 Sites Studied

Four road construction projects were available for study at various times between December 1991 and October 1993 providing a wide variety of soil types, plant types, terrains and project size. These were, in order of study, the A30 Okehampton to Launcston, the M1\A1 link road contracts 2&3, the M3 Bar End to Compton and the A52 Ashbourne relief road. All contracts were undertaken by Tarmac Construction Ltd. with the client being the Department of Transport. Each site will now be discussed, paying particular attention to how the soil conditions affect the earthmoving.

3.2.1 A30 Okehampton to Launceston

This contract was the first to be studied and provided little data in the short time in which the site was visited. The project was a dual two lane trunk road through a green field site in Devon, England. The haul roads for the earthmoving plant were not constructed separately; the plant ran on the excavated cuts and fills. The haul road soil could therefore be studied before the site was visited by inspection of the site investigation reports. In general the soil was a silty, gravelly clay towards the surface that hardened to a sandstone/mudstone the deeper the cuts were taken.

The plant used on this site was a mixture of scrapers and dumptrucks. The scrapers were Terex motorscrapers (TS14, TS24 and TS24C) coupled with Caterpillar D8 and D9 tractor units with a pushing blade. Dumptrucks used were Volvo BM A25 articulated units with Caterpillar 235 and Komatsu PC210 backhoe excavator.

At this stage of the project, the problem was not formulated and so the data taken from this site was of little value. However, the site provided many ideas. The main conclusions drawn were that the dumptrucks were not utilised correctly. There seemed to be too many queuing for the loader at any one time and while they were hauling, the units congregated together in a line, especially on longer hauls.

3.2.2 M1/A1 Link. Contracts 2&3

This project provided the bulk of the data upon which the models of the system were based. The project consisted of two separate contracts let as one and as such was one of the longest road construction projects in Britain at the time. The job was approximately 28 kilometres of dual two lane carriageway through a green field site in Northamptonshire, England. The total earthworks volume was some 3 million cubic metres that was split into 58 separate cut operations. Like the A30 project, the haul roads were the cut and fill surfaces with the soil being fairly consistent along the whole of the 28 kilometres. The soil is best described as a silty clay with plasticity index ranging from approximately 10% to 43% (average 25%). Eighty-six percent of the plasticity indices fell in the range 15% to 35%. The fairly large plasticity range would indicate that the soil was not too susceptible to moisture changes, which is an important factor in earthmoving. (See section 2.5)

The plant used on this project was, like for the A30 contract, a combination of scraper and dumptruck based fleets, due to the wide range of haul lengths. For hauls of up to approximately 1000 m Terex TS24 and TS24C motorscrapers were used coupled with a Cat D9N tractor unit. A fleet of 4 scrapers on a haul of 300m could achieve approximately 400 cubic metres per hour. This rate decreased rapidly as the haul length increased necessitating the use of dumptruck based operations for hauls of over 1000m. The production rate of a dumptruck operation is very dependent on the haul length and the number of trucks; for example, an operation with a haul length of 1500m required 5 hauler units producing 225 cubic metres per hour.

3.2.3 M3 Bar End to Compton

This project was different to the first two for a number of reasons and as such gave a good opportunity to study earthmoving operations on a different scale.

The main feature of this motorway construction contract is that the cuttings were through chalk; a soft highly weathered limestone. This material gives much harder and rut free haul roads than clay and its friable nature made excavating and loading a quick operation. Because of the degradation effects of twin engined scrapers, this type of plant was not allowed under the contract; consequently, the earthmoving was done exclusively by excavator and dumptrucks.

This project had a total of 2.7 million cubic metres of earthworks with 1.7 million contained in one single cut. This cut was approximately 100m wide at its widest point with a depth of approximately 40m that gave the cut almost quarry like proportions. Backhoe excavators were working in ideal conditions and could produce up to 400 cubic metres per hour per unit. The hauling units were made up of fleets of Caterpiller D400D and Volvo BM A35 articulated dumptrucks and Caterpillar 769 rigid dumptrucks. The latter, although approximately 10 years older than the articulated units, were very efficient due to the hard running surfaces and large manoeuvring areas. Such dumptrucks could not be used on a fairly soft, rutted clay haul road.

3.2.4 A52 Ashbourne Relief Road

This project is very small in comparison to the previous three and gave the opportunity to study earthmoving on a small scale. Production rates turned out to be much lower than those achieved on the M3, which indicates that data cannot be reliably transferred between large and small projects. The total earthworks was only 120,000 cubic metres - 4% of the total on the M1/A1.

Surprisingly perhaps for such a small contract the soil conditions changed from a sandstone to a very sandy clay to a gravel over a length of 3 kilometres. The soil also had a fairly high silt content that gave a low plasticity index; the average over the length was approximately 15. This implied that the haul roads would be very susceptible to the effects of rain, as was discussed in the section for the M1/A1 (see figure 2.10); a short rain fall would rapidly turn the surface to a liquid state causing a drop in production and eventually a halting of the works. This assumption turned out to be very accurate as many days were lost to bad weather.

The plant used on this project was again a mixture of scraper and dumptruck fleets but on a smaller scale. TS14B motorscrapers were coupled with a Cat D8 pusher for topsoil stripping and short haul bulk cuts. For longer

> hauls, a Cat 235 backhoe excavator was used to load a fleet of Volvo BM A25 articulated haulers.

Data collected from these four sites formed the basis for the models that have been developed. The data is in the form of cycle times for the plant and production rates and was collected by simply observing the operations as they proceeded and timing the cycles with a stopwatch. This form of data collection is known as work and method study.

3.3 Work and Method Study

The British Standard *BS3138:1969 Glossary of Terms used in Work Study* defines work study as:

> A measurement service based on those techniques, particularly method study and work measurement, which are used in the examination of human work in all its contexts, and which lead to the systematic investigation of all the resources and factors which effect the efficiency and economy of the situation being reviewed, in order to effect improvement.

This is a very good description of what the whole project described here hopes to achieve. BS 3138 takes this definition further by defining method study as:

> The systematic recording and critical examination of the factors and resources involved in existing and proposed ways of doing work, as a means of developing and applying easier and more effective methods of reducing costs.

To fully utilise these techniques it is important to understand the system under study as it exists. Work and method study can then be used to gather data on the individual parts of the system to see how they individually affect the output. The next section will describe the existing earthmoving system studied.

3.4 The Earthmoving System

The earthmoving process can be described as a *system*. A system, as defined by Schmidt & Taylor (1970) is 'a collection of entities, e.g. people or machines, that act and interact together toward the accomplishment of some logical end'. A *mathematical model* represents such a system in terms of logical and quantitative relationships that can then be changed to see how the model reacts. If a model is *valid* we can find out what would happen to the output of a system with certain initial constraints without changing the real life system itself. In this case, a valid earthmoving model can be used to estimate the productivity for an operation that has not yet started. To build a model of the earthmoving system we can use work study techniques to investigate the existing situation. Figure 3.1 shows the earthmoving system schematically. Hauling units queue until the loading unit is free to service them. The hauler then manoeuvres into position and is loaded by the server. The loaded hauler travels to the dump area, deposits the load and returns to join the queue once more. The earthmoving system is a classical queuing system and can be compared to many other queuing systems. Every queuing system has customers, servers, an input process and an output process and these are discussed next with reference to the earthmoving system.

3.4.1 Customers

The customers in the earthmoving system are the hauling units: either dumptrucks or scrapers. They are both valid methods of moving material but used in different situations. Although the queuing system is applicable to both types of hauler, this work will from now on consider only dumptrucks. The reason for this is that scrapers are slowly being phased out as a piece of construction plant (see chapter 2.)

3.4.2 Servers

The servers in this situation are, in the case of dumptrucks, an excavator and for scrapers, a bulldozer is used to push load the scraper. (Twin engined scrapers can in fact load themselves; the rear engine provides the force required to push the unit. However, a pusher will enable the scraper to be loaded more quickly and fuller thus improving the efficiency of the operation.) Because they can load themselves, the scrapers will rarely form a queue and this is the major difference between a dumptruck and a scraper operation. When the pusher is free it will manoeuvre itself to push a scraper that has already started loading.

3.4.3 The Input Process

The input process in this situation is the arrival of a customer. A bank queuing system, for example, can have an infinite number of customers and

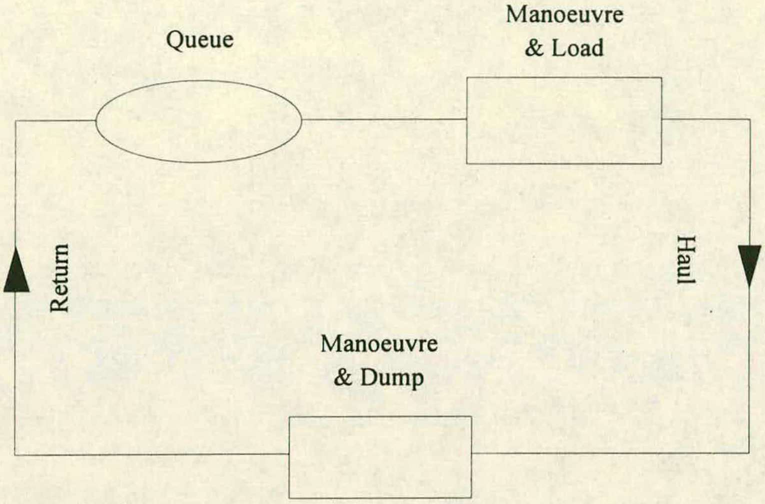


Figure 3.1 The Earthmoving System

thus an arrival of a customer at a bank is not affected by the number of customers already waiting. The earthmoving system is not like that of a bank because there are a finite number of customers. The probability of an arrival of a dumptruck at the queue is dependent on the number of trucks already waiting. If all trucks in the system are in the queue then the probability of an arrival is zero until a truck has been loaded. These ideas will be developed more in chapter 4 as the model of the earthmoving system is developed.

3.4.4 The Output Process

The output process is the service process. In this case it is the process of loading a truck. This process involves two separate parts: the manoeuvring of the truck into position (hereafter described as the *spotting* of the truck) and the loading of the truck with material. The time taken to actually load the truck depends on the number of full buckets of material that are deposited, which in turn will depend on the size of the bucket and the size of the dumptruck.

3.4.5 Queue Discipline

The order in which the trucks are loaded can have an effect on the output rate of the system. Usually, the dumptrucks are served in the order in which they join the queue. This is called FCFS (first come, first served). Other types of queuing discipline are LCFS (last come, first served) and SIRO (service in random order). Queue discipline may be difficult to maintain if more than one excavator is used in one operation; truck operators will then have to decide which server they will use.

3.5 Data Collection

Having described the earthmoving system, it is now known what the individual components that need to be investigated are:

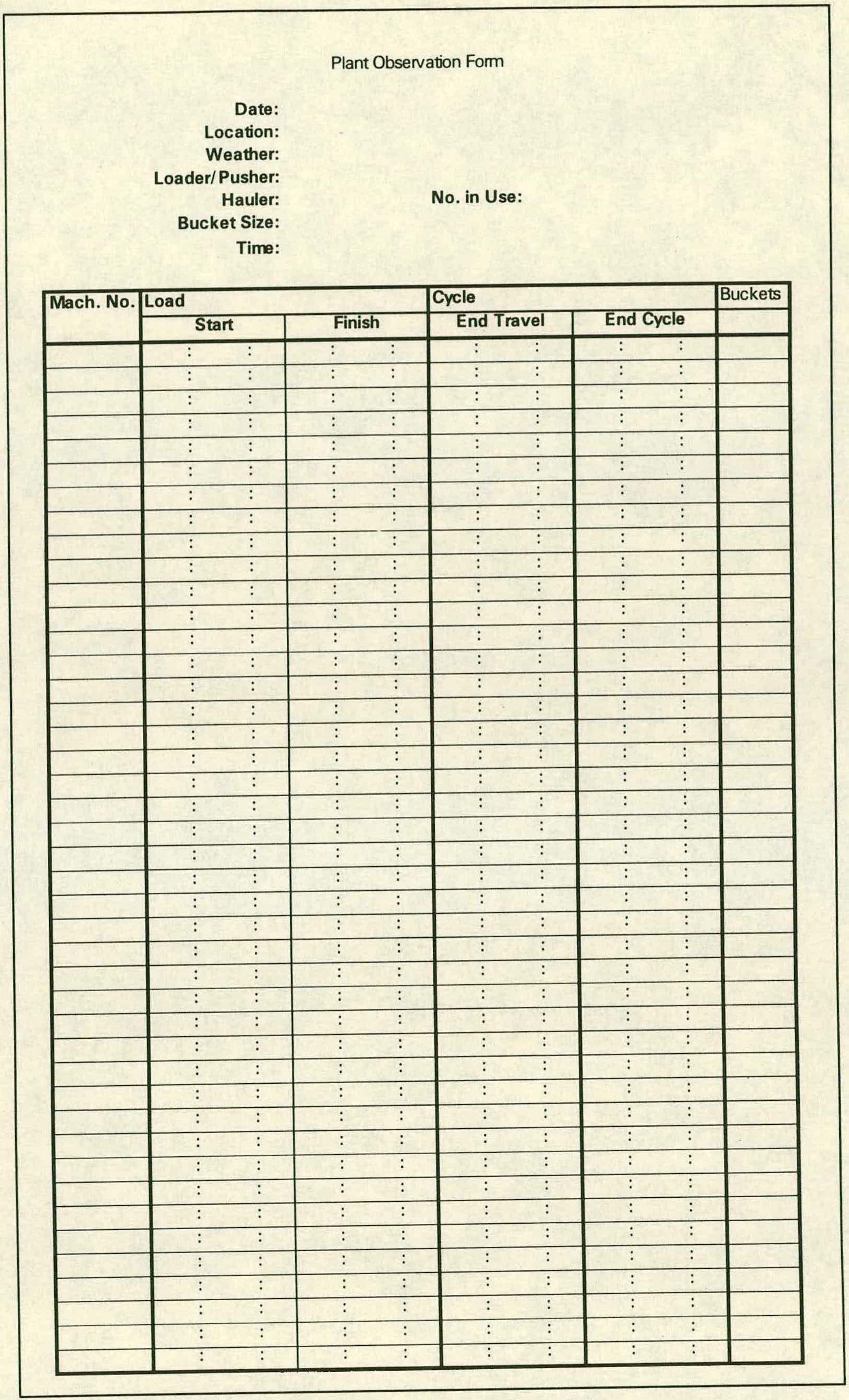
- The service time made up of the spotting time and the loading time. This is also referred to as the loading cycle time.
- The size of the excavator bucket or, more specifically the quantity of material contained in one full bucket.
- The number of full buckets required to fill the truck.

- The arrival time. Because this system is cyclical, the time a truck arrives at the queue depends on when it last left the service process and the time taken to haul, dump and return.

To get values for each of these components, actual operations on the sites described in 3.2 were observed. Essentially, the time to complete each task was timed with a stop watch. Average times for a particular situation can then be determined which can be used as input to a model of the system and provide a database for future work. If it is known how long a particular operation should take, these times can be considered targets for future operations.

The method for collecting the data is as follows. The observer is situated at the loading area, usually next to the fence line so as to be completely out of the way of operations (for safety reasons as well as avoiding interference of the operations), and able to see all arrivals and departures from service clearly. Actual times are recorded for each component, that is the times when a truck arrives at the queue, starts to spot, starts to load and completes the load. The number of buckets per load is also recorded. This data is recorded on a form such as that shown in figure 3.2. Recording the actual time an event happens means that more data can be recorded than would be possible if the components were individually timed. However, after the observation has finished, all the times need to be calculated. For long observations, the best way was to enter all the recorded times into a spreadsheet to avoid mistakes. The spreadsheet could then be used to calculate the production of the operation observed. Figure 3.3 shows a typical results summary sheet for the operations that were observed on the M1/A1. This summary sheet indicates all the relevant information about the operation, including the total volume of the cut being excavated, the haul length and the number of trucks and excavators being used. The productivity for the operation can be calculated from the cycle time of the trucks and the volume carried per cycle by each truck. The latter is estimated from the volume of earth contained in one bucket and the number of buckets in one truck full. The summary sheet also calculates how long the particular operation would take if the observed productivity could be maintained and compares this to the programmed duration of the operation.

The dump time for an operation could not usually be determined if the observer was situated at the load point. Therefore, certain operations were observed from the dump area to determine the average time a truck takes to dump a load.

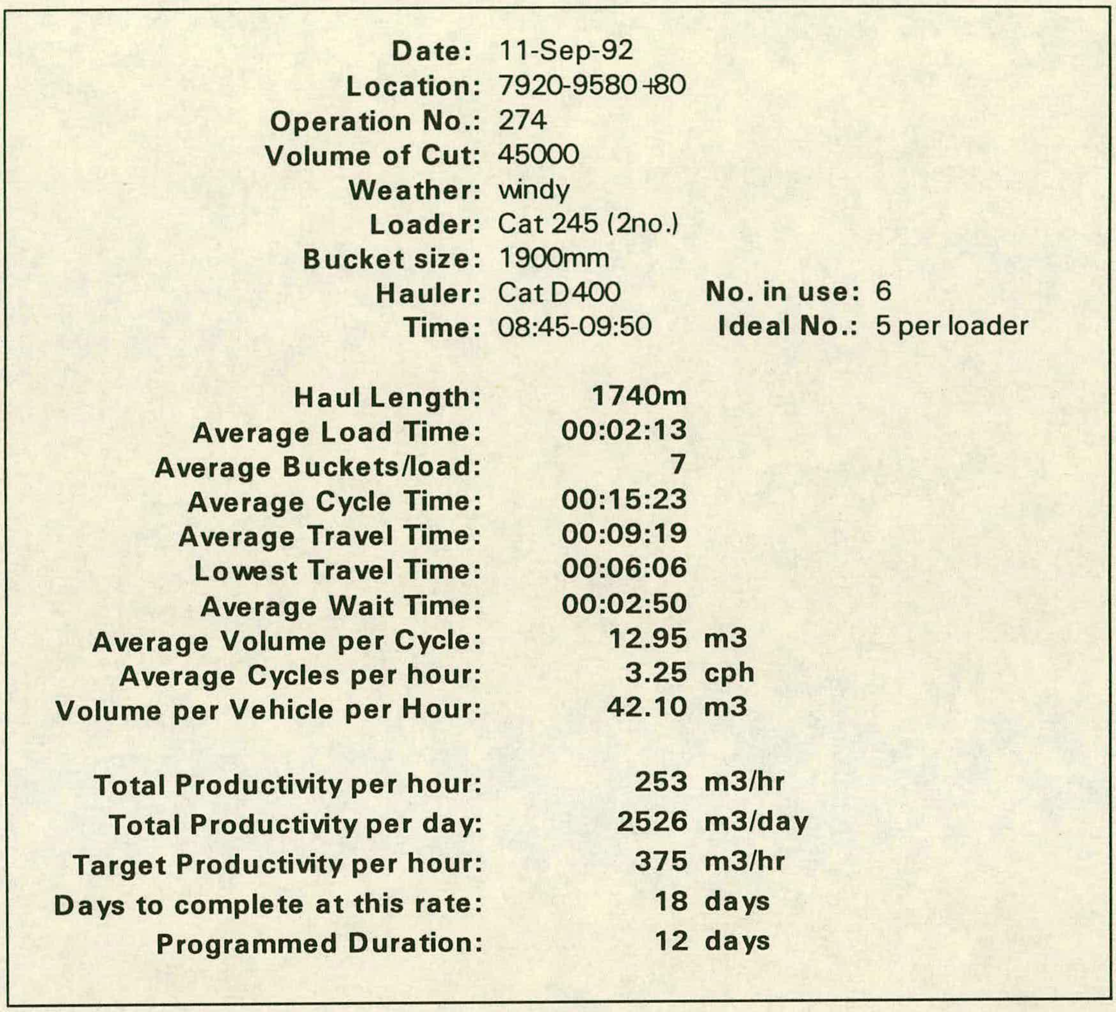


Plant Observation Form

Date:
Location:
Weather:
Loader/ Pusher:
Hauler: **No. in Use:**
Bucket Size:
Time:

Mach. No.	Load		Cycle		Buckets
	Start	Finish	End Travel	End Cycle	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	
	: :	: :	: :	: :	

Figure 3.2 Plant Observation Form for collection of data



Date: 11-Sep-92
Location: 7920-9580+80
Operation No.: 274
Volume of Cut: 45000
Weather: windy
Loader: Cat 245 (2no.)
Bucket size: 1900mm
Hauler: Cat D400 **No. in use:** 6
Time: 08:45-09:50 **Ideal No.:** 5 per loader

Haul Length: 1740m
Average Load Time: 00:02:13
Average Buckets/load: 7
Average Cycle Time: 00:15:23
Average Travel Time: 00:09:19
Lowest Travel Time: 00:06:06
Average Wait Time: 00:02:50
Average Volume per Cycle: 12.95 m3
Average Cycles per hour: 3.25 cph
Volume per Vehicle per Hour: 42.10 m3

Total Productivity per hour: 253 m3/hr
Total Productivity per day: 2526 m3/day
Target Productivity per hour: 375 m3/hr
Days to complete at this rate: 18 days
Programmed Duration: 12 days

Figure 3.3 Observation Results Summary Sheet

3.6 Productivity Estimation: Plant Match

Observations of earthmoving operations on the M1/A1 produced a substantial amount of data that can now be used as the basis for estimating productivity on future operations. Table 3.1 shows a summary of all the operations observed on this site involving dumptrucks and excavators. The dump times for many operations had to be assumed because the observer could not see the dump area. The average time of 1 minute 30 seconds was used as this had been determined from operations observed separately at the dump area.

The productivity of a single dump truck (hauler) is governed by how many cycles per hour it can achieve - one cycle being a complete trip from load point to dump area to load point again. The number of cycles per hour is dependent upon the cycle time and so calculation of the cycle time is the starting point for productivity estimations. Therefore, if the four component times (spot, load travel and dump) bucket volume, buckets per load and number of trucks can be determined for a future operation, the productivity of that operation can be determined. This is the basis for the first simple model proposed, but before calculations based on data collected are carried out, the idea of *plant match* must be investigated.

The production rate of a multiple hauler earthmoving operation is dependent on the rate at which the excavator can win earth from the ground and load the dump trucks. The number of trucks used in an operation must be matched to the speed of the excavator if the efficiency of the operation is to be maintained. If too many trucks are deployed then efficiency will be lost due to queueing trucks. Too few and the loader will be idle as it waits for trucks to return. The ideal situation therefore is for the rates of the trucks and excavator to be such that as one truck leaves the loader another is just arriving. Such an operation is said to be perfectly matched and the state of any operation can be expressed using the dimensionless term *match factor*. This is a very important term and will be used extensively throughout this work. It was first proposed by the research department of the (then) Caterpillar Tractor Company (Morgan and Peterson, 1968) and is dependent on the quantities and cycle times of the plant used:

$$\text{Match Factor} = \frac{\text{No. of Haulers x Loader Cycle Time}}{\text{No. of Loaders x Hauler Cycle Time}} \tag{3.1}$$

Op. No.	Date	Resources				Haul	Load	Cycle	Dump	Wait	Travel	Vol. per	Vol per	Target
		Loader	No.	Haul.	No.	Length	Time	Time	Time	Time	Time	Veh./Hr.	Hour	
274	11-Aug-92	245	2	D400/A35/A30	13	1860	01:56	17:23	(1:30)	02:10	11:47	41	530	375
274	01-Sep-92	245	1	D400	4	800	02:08	12:42	(1:30)	02:16	06:48	53	212	375
274	02-Sep-92	245	2	D400	6	3280	01:50	23:48	(1:30)	00:00	20:28	30	182	375
274	02-Sep-92	245	2	A35	7	3280	01:45	22:49	(1:30)	00:24	19:10	29	199	375
274	02-Sep-92	245	2	A30	2	3280	01:12	21:34	(1:30)	00:18	18:34	26	52	375
274	02-Sep-92	245	2	D400	6	3380	02:28	25:33	(1:30)	00:37	20:58	31	183	375
274	02-Sep-92	245	2	A30/A35	6,2	3380	02:04	26:18	(1:30)	00:12	22:32	25	202	375
204/304	17-Jul-92	235	1	D400/R35	2,3	240	02:13	15:30	(1:30)	09:23	02:24	42	210	381
207/307	17-Jun-92	245	2	A25	4	2150	(1:10)	16:14	(1:30)	00:00	13:34	33	130	385
207/307	17-Jun-92	245	2	D400	5	2150	(1:55)	16:46	(1:30)	00:00	13:21	52	257	385
207/307	17-Jun-92	245	2	R35	2	2150	(1:40)	22:49	(1:30)	00:00	19:39	35	70	385
207/307	18-Jun-92	245	2	D400	7	2230	01:56	17:19	(1:30)	00:48	13:05	48	338	385
207/307	18-Jun-92	245	2	A25	3	2230	01:11	16:57	(1:30)	00:59	13:17	33	96	385
207/307	18-Jun-92	245	2	R35	2	2230	01:41	22:42	(1:30)	00:39	18:52	32	65	385
207/307	22-Jun-92	245	2	vol.cat.euc	12	2350	02:05	18:22	(1:30)	01:46	13:01	41	497	385
207/307	27-Jul-92	235	1	D400/300/A30	3	1450	02:38	16:54	(1:30)	00:00	12:46	37	111	200
233/333	26-Jun-92	245	1	A25	3	400	01:11	13:33	(1:30)	07:35	03:17	36	108	202
233/333	26-Jun-92	245	1	D400	3	400	02:03	12:54	(1:30)	05:40	03:41	61	183	202
234/334	24-Aug-92	235	1	A25/A30/D300	4,1,2	1150	01:43	16:48	(1:30)	03:19	10:16	27	192	188
246/346	13-May-92	245	1	R35	3	340	02:00	07:55	(1:30)	00:52	03:33	95	284	367
246/346	14-May-92	245	1	R35	3	860	02:10	13:13	(1:30)	00:00	09:33	55	166	367
246/346	15-May-92	245	1	R35	3	670	01:43	12:21	(1:30)	00:00	09:08	50	151	367
246/346	18-May-92	245	1	R35	5	1300	02:34	17:24	(1:30)	00:00	13:20	45	223	367
251/351	16-Jun-92	245	1	R35	3	120	01:48	07:25	(1:30)	01:37	02:30	95	286	365
251/351	16-Jun-92	245	1	R35	3	280	02:13	08:45	(1:30)	00:38	04:24	96	288	365
252/352	29-Jun-92	245	1	D400	5	1750	01:51	13:55	(1:30)	00:18	10:16	56	280	399
252/352	15-Jul-92	245	2	D400	12	1640	02:01	16:46	(1:30)	03:24	09:51	47	563	399
252/352	23-Jul-92	245	2	D400	10	1810	02:20	16:36	(1:30)	01:38	11:08	51	506	500
252/352	24-Jul-92	245	2	D400	9	1650	02:20	14:56	(1:30)	01:56	09:10	54	482	500
281/381	25-Jun-92	245	1	D400	6	2750	01:57	23:37	(1:30)	02:31	17:39	33	198	398
281/381	30-Jun-92	245	2	D400/300	9,2	2850	02:21	20:00	(1:30)	00:21	15:48	34	371	398
281/381	27-Jul-92	245	1	D400	5	2650	02:03	18:01	(1:30)	00:00	14:28	41	206	500
281/381	26-Aug-92	245	2	D400	5	3620	02:13	27:00	(1:30)	00:07	23:10	28	141	
281/381	26-Aug-92	245	2	A35	7	3620	02:10	30:51	(1:30)	00:00	26:51	22		

Table 3.1 Summary of Data Collected on M1/A1

The plant cycle times must not include any idle times, that is they must be the time the plant is actually working. An operation with match factor (MF) of one is perfectly matched. If MF is greater than one the operation has too many trucks; if MF < 1 then the operation is under resourced. The maximum productivity from an operation occurs when the match factor is one. If more trucks are then deployed the productivity cannot increase unless the rate at which the excavator works is increased. Figure 3.4 indicates how the productivity of an operation will increase as the match factor increases.

3.7 Productivity Estimation: Model 1

An initial model of the earthmoving process can now be proposed, which is deliberately very simple. The modelling process is iterative and no modeller should assume that the first attempt at analysing a system in this way will provide the optimum answer; this model will be developed by refining the assumptions and relationships until the output it provides represents that of the real system it is supposed to replace to a required degree of accuracy. The simplicity of this model means that its solution is an analytical one that has been performed on a spreadsheet. Before the method of solution is described, the assumptions used need to be stated.

3.7.1 Assumptions used in model 1

i. All time components (spot, bucket swing, travel and dump) are constant for the operation in consideration.

ii. The time to load a truck is the product of the number of buckets per load less one and the bucket swing time. This allows for the assumption that the first bucketful of material is deposited at the moment the load time starts.

iii. The bucket volume and the number of buckets per load are constant.

iv. All time components have been calculated by some other means before being input to the model.

v. All plant is assumed to work constantly. Therefore the production rate the model calculates will be invalid if the trucks in the actual system breakdown for example. This can be allowed for by reducing the number of working minutes per hour. If 50 minutes per hour is used this assumes that the plant is stopped for 10 minutes in every hour and is equivalent to an efficiency of 50/60 = 83%.

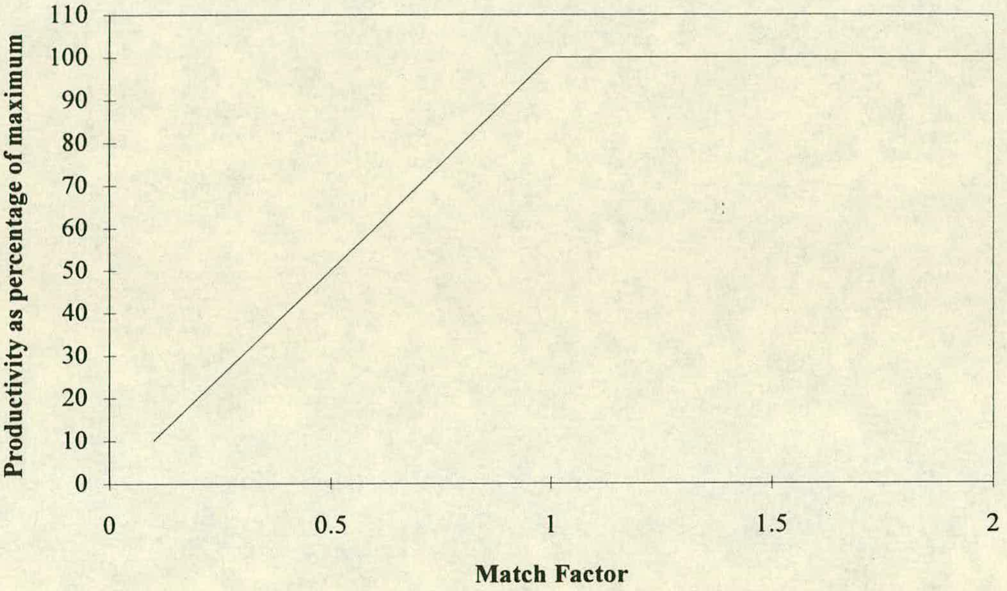


Figure 3.4 The effect of plant match on operation productivity

3.7.2 Determination of time components

This is an area of great importance in the calculation of the overall productivity of an earthmoving operation. If the bucket swing time, for example, cannot be reliably determined then the productivity calculated is questionable. A full description of the work involved in estimating the time components is beyond the scope of this thesis; indeed it is the subject of two additional research projects that have been undertaken in parallel with this one. However a brief description of the factors influencing the time components is given here.

i. *Spot Time*. The time taken for a truck to get from the queue into position ready for loading depends on the type of truck (size, rigid or articulated), the amount of area for manoeuvring and the condition of the ground surface. Therefore a small articulated dumptruck (say Volvo BM A25) manoeuvring in a rock cutting for a three lane motorway will have a much lower spot time than a rigid dumptruck spotting in a soft clay cutting for a single carriageway relief road. There is no analytical method available for determination of spot time and values taken from experience must be used.

ii. *Bucket Swing Time.* The factors affecting this time component are: the type of material, the excavator used, the depth of cut, the relative heights of the excavator and the swing angle. Therefore a low powered excavator (for example a Cat 225) excavating with a cut depth of 0.8m in wet clay with a swing angle of 150 degrees will have a much longer swing time than a high powered excavator cutting a 3m depth of soft, dry chalk with a swing angle of 30 degrees. (Approximately 30 seconds to 10 seconds). Again, like spot time, there is no established analytical method of calculating swing time and values taken from experience will have to be used.

iii. *Travel Time.* Specifically, this should be two separate times: haul time and return time but because these times could not be recorded separately, a collective travel time has been used. This time is the most complex to calculate as there are many different factors that will influence it. Firstly, the type of truck used and its payload will affect the speed. The physical nature of the haul road has a large influence; the gradient, number and angle of any bends and obstructions such as plant crossings and Bailey bridges are all important. However, the factor that is most difficult to determine is the rolling resistance of the haul road. This one factor is the subject of a whole research project. It is the resistance to motion due to the vehicle/terrain interaction and is affected by the type of plant, soil conditions, haul road conditions (i.e. rut depth), driver ability and tyre pressure. Computer packages are available that can calculate the travel time for a particular operation (notably Accelerator and Caterpillar's Vehsim). These packages work by inputting the vehicle specification (taken from manufacturers' handbooks), the payload and the haul details, i.e. length, gradient and rolling resistance.

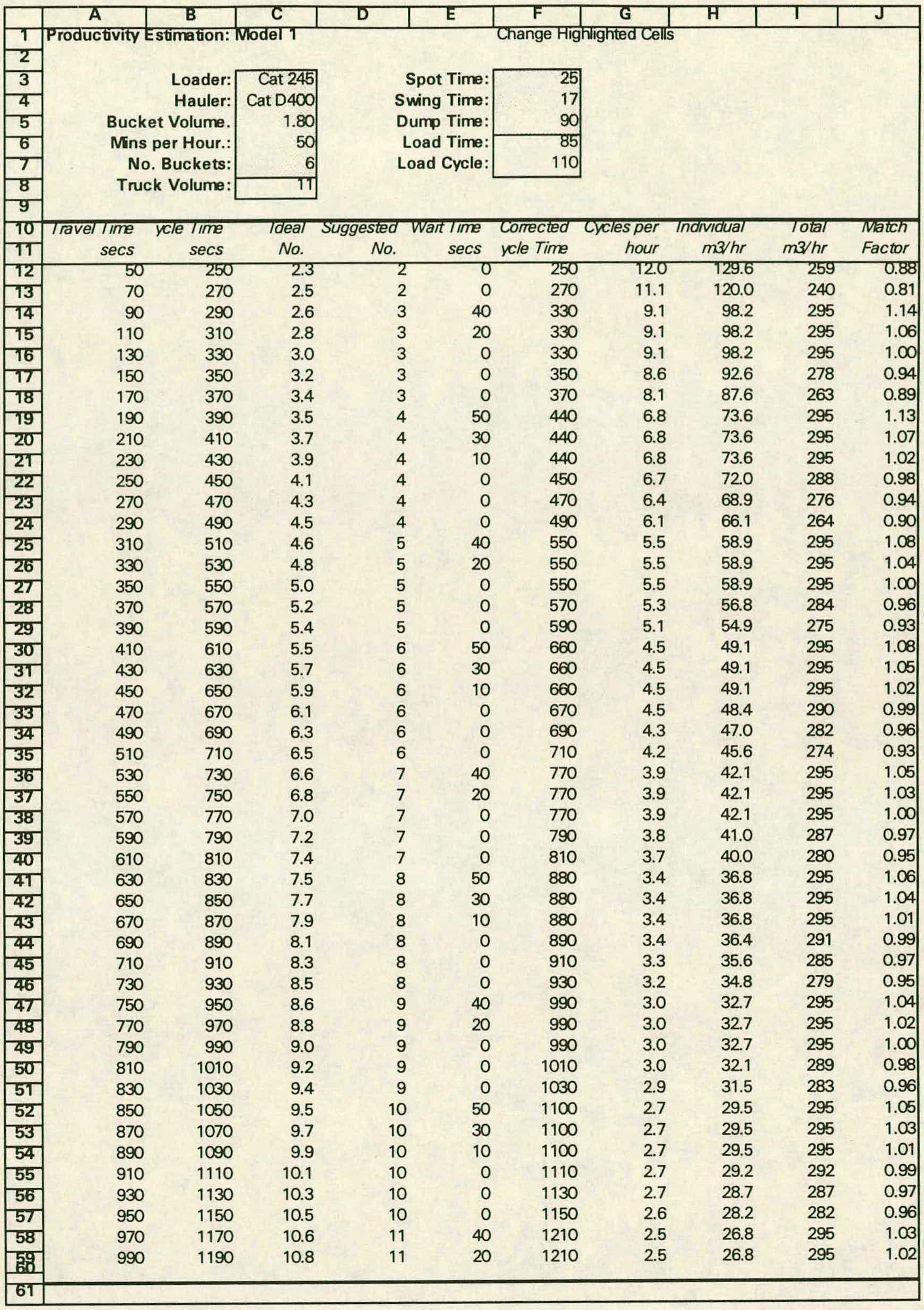
iv. *Dump Time.* This is the time component for which there is little data as the dump area could not be observed for most of the time. It is affected by the type of material, the size of the dumptruck and the area available for manoeuvring. A value of 90 seconds is usually taken which allows for manoeuvring of the truck as it enters the dump area as well as the actual tipping operation.

3.7.3 Solution to Model 1

This model has been set up on a spreadsheet so that the output of the system can be seen quickly to changes in the initial conditions of the model. Once the component times have been calculated, they can be entered into the spreadsheet model. Figure 3.5 shows the spreadsheet set up with the following initial conditions:

Bucket Volume:	1.80 m^3
Minutes per hour	50
Buckets per load	6
Swing time	17 secs
Dump time	90 secs
Spot time	25 secs

The minutes per hour can be set to a figure less than 60 to allow for any stoppages that the model does not already allow for, such as breakdowns. The travel time that has been calculated for the truck is read off column A. The spreadsheet model then calculates the cycle time, the number of trucks that should be used, adjusts the cycle time for any waiting of the trucks and finally calculates the productivity and match factor of the operation. To outline the assumptions and equations used, the following is an example from line 15 of the spreadsheet in figure 3.5



	A	B	C	D	E	F	G	H	I	J
1	Productivity Estimation: Model 1					Change Highlighted Cells				
2										
3		Loader:	Cat 245		Spot Time:	25				
4		Hauler:	Cat D400		Swing Time:	17				
5		Bucket Volume.	1.80		Dump Time:	90				
6		Mins per Hour.:	50		Load Time:	85				
7		No. Buckets:	6		Load Cycle:	110				
8		Truck Volume:	11							
9										
10	Travel Time	ycle Time	Ideal	Suggested	Wait Time	Corrected	Cycles per	Individual	Total	Match
11	secs	secs	No.	No.	secs	ycle Time	hour	m3/hr	m3/hr	Factor
12	50	250	2.3	2	0	250	12.0	129.6	259	0.88
13	70	270	2.5	2	0	270	11.1	120.0	240	0.81
14	90	290	2.6	3	40	330	9.1	98.2	295	1.14
15	110	310	2.8	3	20	330	9.1	98.2	295	1.06
16	130	330	3.0	3	0	330	9.1	98.2	295	1.00
17	150	350	3.2	3	0	350	8.6	92.6	278	0.94
18	170	370	3.4	3	0	370	8.1	87.6	263	0.89
19	190	390	3.5	4	50	440	6.8	73.6	295	1.13
20	210	410	3.7	4	30	440	6.8	73.6	295	1.07
21	230	430	3.9	4	10	440	6.8	73.6	295	1.02
22	250	450	4.1	4	0	450	6.7	72.0	288	0.98
23	270	470	4.3	4	0	470	6.4	68.9	276	0.94
24	290	490	4.5	4	0	490	6.1	66.1	264	0.90
25	310	510	4.6	5	40	550	5.5	58.9	295	1.08
26	330	530	4.8	5	20	550	5.5	58.9	295	1.04
27	350	550	5.0	5	0	550	5.5	58.9	295	1.00
28	370	570	5.2	5	0	570	5.3	56.8	284	0.96
29	390	590	5.4	5	0	590	5.1	54.9	275	0.93
30	410	610	5.5	6	50	660	4.5	49.1	295	1.08
31	430	630	5.7	6	30	660	4.5	49.1	295	1.05
32	450	650	5.9	6	10	660	4.5	49.1	295	1.02
33	470	670	6.1	6	0	670	4.5	48.4	290	0.99
34	490	690	6.3	6	0	690	4.3	47.0	282	0.96
35	510	710	6.5	6	0	710	4.2	45.6	274	0.93
36	530	730	6.6	7	40	770	3.9	42.1	295	1.05
37	550	750	6.8	7	20	770	3.9	42.1	295	1.03
38	570	770	7.0	7	0	770	3.9	42.1	295	1.00
39	590	790	7.2	7	0	790	3.8	41.0	287	0.97
40	610	810	7.4	7	0	810	3.7	40.0	280	0.95
41	630	830	7.5	8	50	880	3.4	36.8	295	1.06
42	650	850	7.7	8	30	880	3.4	36.8	295	1.04
43	670	870	7.9	8	10	880	3.4	36.8	295	1.01
44	690	890	8.1	8	0	890	3.4	36.4	291	0.99
45	710	910	8.3	8	0	910	3.3	35.6	285	0.97
46	730	930	8.5	8	0	930	3.2	34.8	279	0.95
47	750	950	8.6	9	40	990	3.0	32.7	295	1.04
48	770	970	8.8	9	20	990	3.0	32.7	295	1.02
49	790	990	9.0	9	0	990	3.0	32.7	295	1.00
50	810	1010	9.2	9	0	1010	3.0	32.1	289	0.98
51	830	1030	9.4	9	0	1030	2.9	31.5	283	0.96
52	850	1050	9.5	10	50	1100	2.7	29.5	295	1.05
53	870	1070	9.7	10	30	1100	2.7	29.5	295	1.03
54	890	1090	9.9	10	10	1100	2.7	29.5	295	1.01
55	910	1110	10.1	10	0	1110	2.7	29.2	292	0.99
56	930	1130	10.3	10	0	1130	2.7	28.7	287	0.97
57	950	1150	10.5	10	0	1150	2.6	28.2	282	0.96
58	970	1170	10.6	11	40	1210	2.5	26.8	295	1.03
59	990	1190	10.8	11	20	1210	2.5	26.8	295	1.02
60										
61										

Figure 3.5 Spreadsheet model for productivity estimation.

i. Column B. Truck cycle time (excluding idle time)

Cycle Time = Spot + Load + Travel + Dump
= 310 seconds

(The load time is calculated by multiplying the swing time by the number of buckets per load less one. This is because a full bucket load will be waiting and deposited as soon as the truck is in position. In this case, 6 bucket loads are deposited, but the excavator swings only 5 times.)

ii. Column C. The ideal number of trucks for the cycle time. This occurs when match factor is one. Substituting MF = 1 in equation 3.1 gives:

Ideal no. of trucks = truck cycle time / loader cycle time
= 310 / 110
= 2.8 trucks

iii. Column D. The spreadsheet suggests a number of trucks by rounding the ideal number to the nearest whole number (3 trucks).

iv. Column E. If the number of trucks suggested is higher than the ideal then the operation will be over resourced and the trucks will have to wait for a certain amount of time in the queue. This is calculated by

Truck wait time = (load cycle x no. trucks) - truck cycle
= (110 x 3) - 310
= 20 seconds

v. Column F. Corrected cycle time, including wait time.

Corrected cycle time = 310 + 20
= 330 seconds

vi. Column G. Cycles per hour (in this case, a 50 min hour is used)

cph = (50 x 60) / 330
= 9.1

vii. Column H. Individual truck production rate.

Individual prod. = Truck volume x cycles per hour
= Buckets per load x bucket volume x cph
= 6 x 1.80 x 9.1
= 98.3 m^3/ hr

viii. Total production rate = individual rate x no. of trucks

= 98.3 x 3

= 295 m^3/hr

ix. Match factor = (no. trucks x load cycle) / truck cycle time

= (3 x 110) / 310

= 1.06

These are the calculated results from one line of the spreadsheet, for a travel time of 110 seconds. Figure 3.6 is a graphical representation of this initial setup. The graph shows lines for productivity and match factor. As can be seen, for each number of trucks, the match factor of 1 equates with the maximum productivity; any further increases in match factor results in no further increase in productivity.

3.7.4 Results from Model 1

Figures 3.7 - 3.10 show various results from this model for the initial conditions shown in table 3.2. The initial conditions that are represented by these curves represent the wide range of conditions experienced on the sites that have been studied. The best production rate is possible with a Caterpillar 245 excavator working with initial conditions A in table 3.2 The production range with this setup is between 340 and 420 cubic metres per hour. This is because the large bucket on the Cat 245 can be moved very quickly; the swing angle is small (30 degrees) the cut depth is at the optimum for a Cat 245(3m) and the material it is moving is easy to excavate and falls out of the bucket without sticking.

This is in contrast to initial conditions B. Here the material is a wet clay that is difficult to excavate and dump because of the high adhesive forces the clay applies to the bucket. To exacerbate the loading procedure, the swing angle is large (90 degrees) - this may be due to the cut area being very narrow - and the loading area is badly rutted which will cause long spotting times (30 seconds). The production range for these conditions is only between 145 and 175 cubic metres per hour.

These results show how quickly and easily this initial model can provide estimates for productivity, providing estimates for the component times have been calculated. However, before the model can be used in a real

	A	B	C	D
Excavator	Cat 245	Cat 235	Cat 245	Cat 235
Dumptruck	Cat D400	Volvo BM A25	Volvo BM A35	Volvo BM A35
Cut Depth	3m	1m	2m	1.5m
Material Type	Soft Rock	Wet Clay	Gravelly Clay	Dry Clay
Swing Angle (degrees)	30	90	45	45
Load Area	Large, Smooth	Rutted	Large	Small
Dump Area	Very Large	Large	Large	Large
Bucket Volume (m3)	2.0	1.5	2.0	1.5
Buckets per Load	7	6	6	8
Spot Time (seconds)	15	30	17	22
Swing Time (seconds)	14	25	18	20
Dump Time (seconds)	75	90	90	90

Table 3.2 Initial Conditions for Figures 3.7 - 3.10

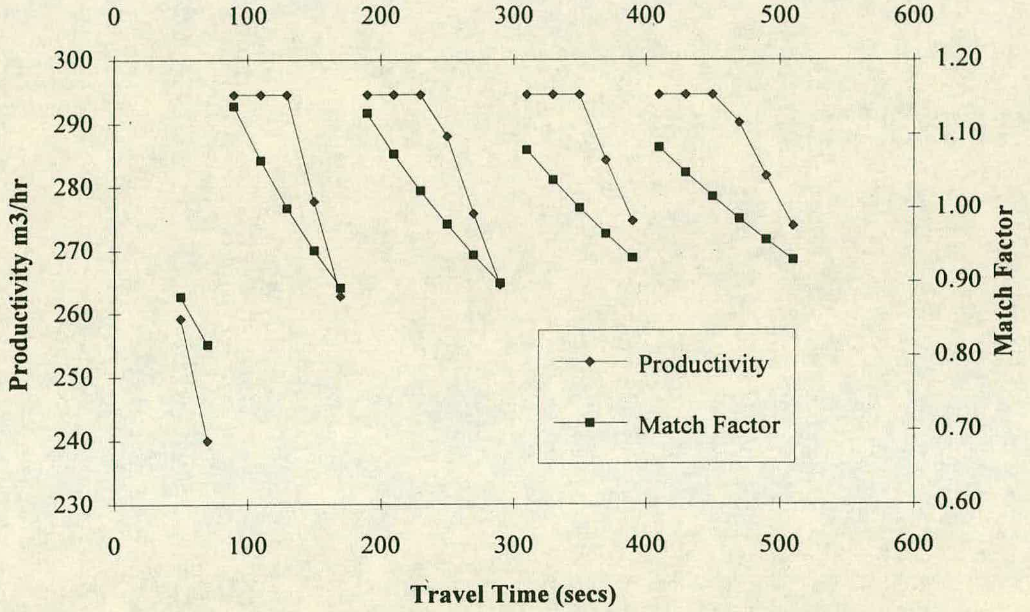


Figure 3.6 Graphical Representation of model 1

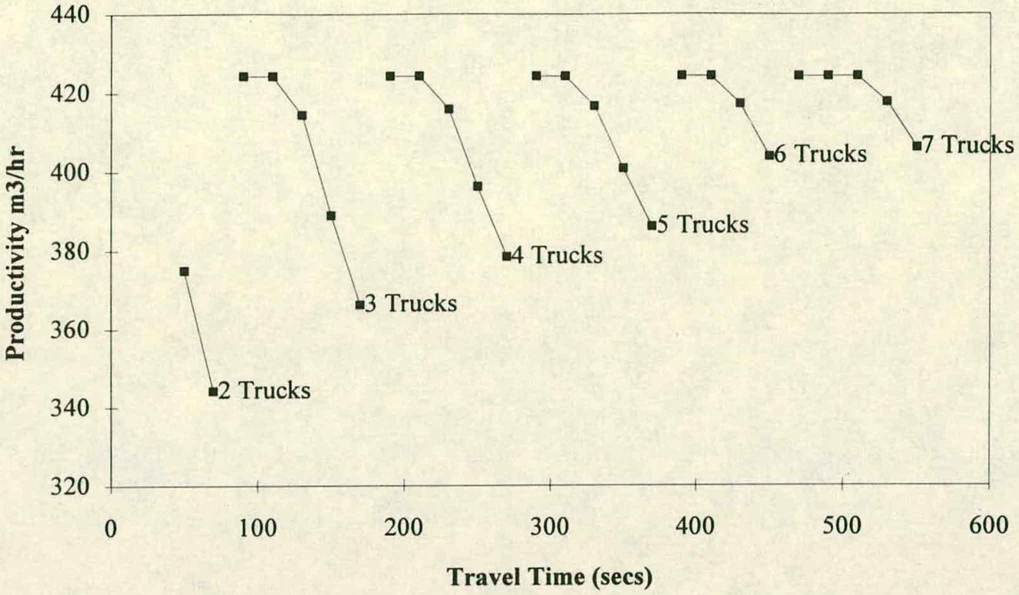


Figure 3.7 Productivity vs. Travel Time for initial conditions A

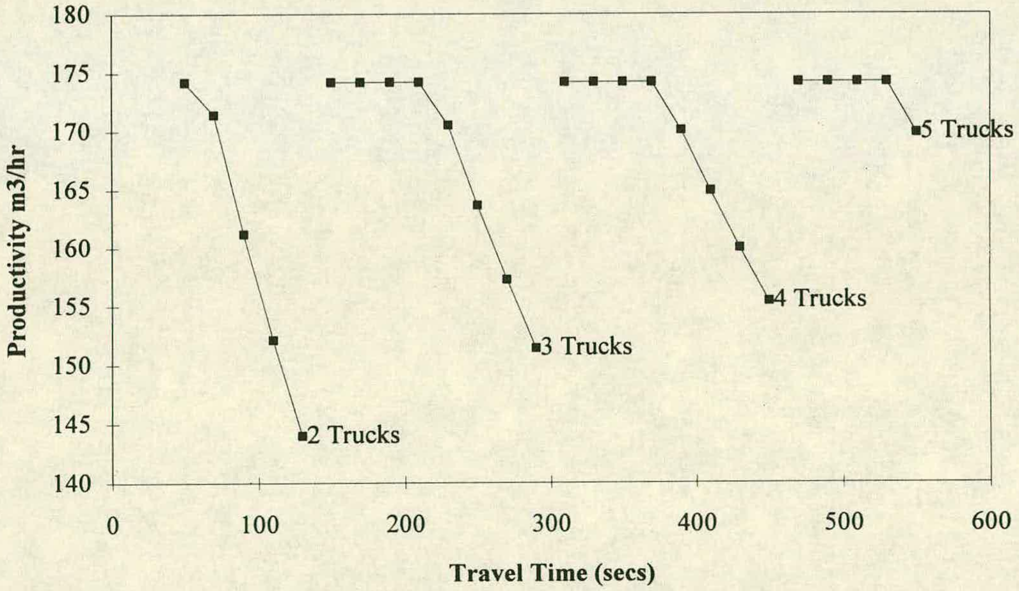


Figure 3.8 Productivity vs. Travel Time for initial conditions B

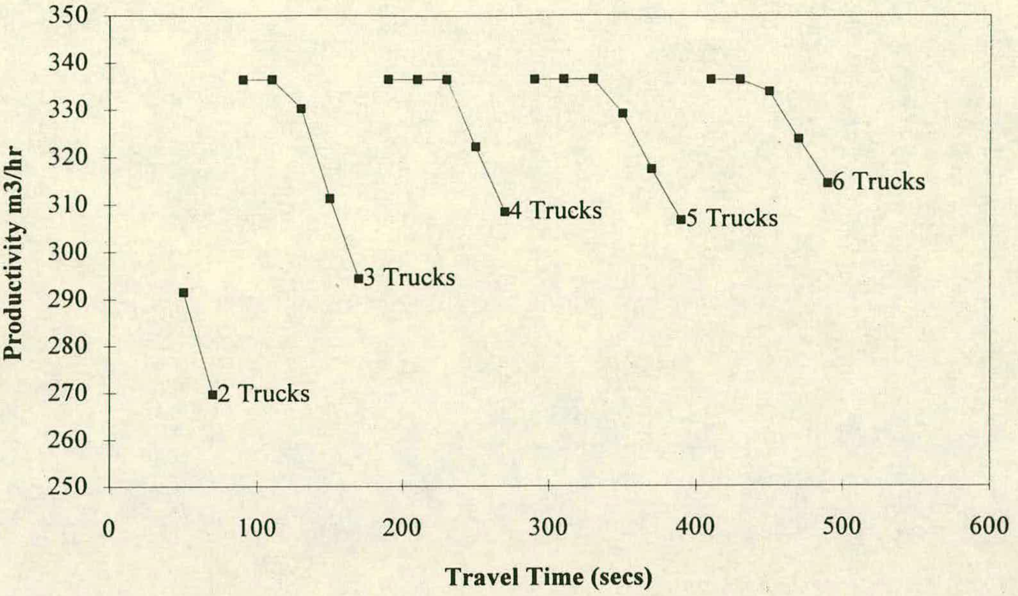


Figure 3.9 Productivity vs. Travel Time for initial conditions C

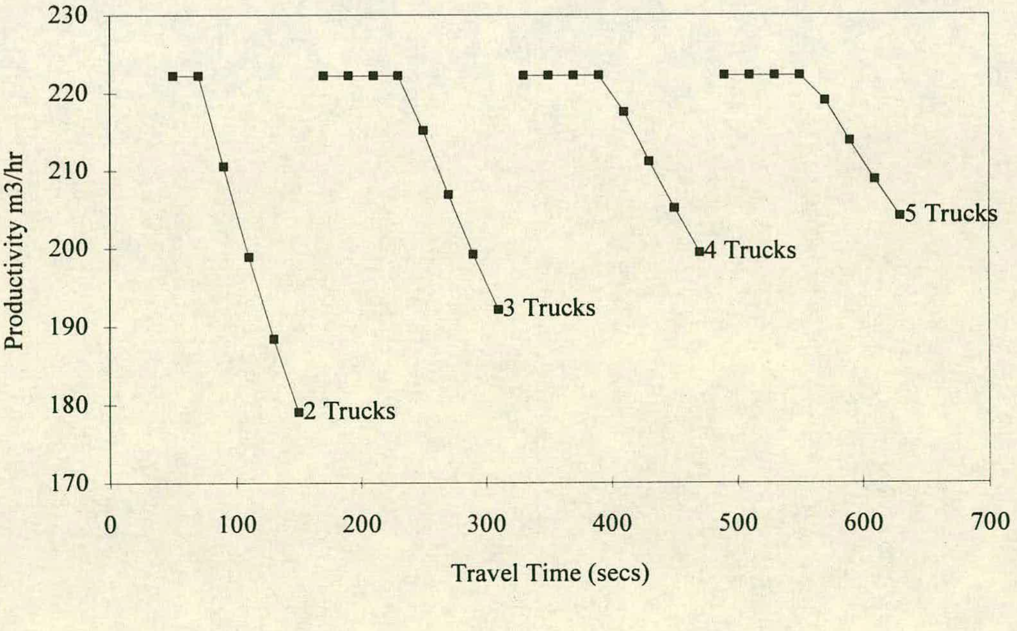


Figure 3.10 Productivity vs. Travel Time for initial conditions D

situation, the validity of the model must be determined i.e. whether or not the model accurately represents the actual system it is supposed to replace.

3.7.5 Validation of Model 1

The full techniques available for validating mathematical models will not be applied to this simple, analytical model. (Further work will be discussed when the second model in this thesis is studied. See section 6). At this stage, the output from the model will be directly compared with that from operations observed on the M1/A1 site. If the results do not match, the reasons for this will need to be investigated so that the assumptions used can be refined and further models developed.

Table 3.3 has the results from 13 separate observed operations. The bucket volume, buckets per load, component times and plant numbers from the actual operation were used as input to the spreadsheet model. The actual production is compared with the calculated and in all but 1 instance it has been over-estimated by the model. The largest difference is 12.3% with an average over-estimation of 4.6%. For a production rate of 250 m^3/hr this is only approximately 12 m^3/hr over-estimated but there is no way of knowing by how much the actual over-estimation is.

To conclude, this initial model does not estimate the productivity with as much accuracy as is perhaps required. The next stage therefore is to isolate the reasons for the errors and change any assumptions accordingly.

3.8 Invalidities of Model 1

In section 3.7.1 the assumptions upon which the first model is based were defined. This model, however, has been unable to consistently estimate productivities and indeed the results it gives are mostly over-estimates - which could prove dangerous in a real situation. Of the five assumptions made it is perhaps easy to see which one is invalid: the assumption that all the times in one operation are constant. If an earthmoving operation is observed it is very soon apparent that the plant cycle times are not constant. The number of buckets per load have been observed to be fairly constant, especially if all the haulers are the same model, and the bucket volume is also constant if the same volume per load is achieved. Whether or not this happens in an operation is quite noticeable: an under loaded dumptruck will have a smaller heap on the back.

				Actual Component Times (seconds)				Match	Production			
Loaders	Haulers	Bucket Vol.	Buckets per Load	Swing	Dump	Spot	Travel	Factor	Actual	Calculated	Difference	Act./Calc
2	15	1.90	7	18.1	90	37	1107	0.78	440	446	1.36%	0.99
1	4	1.95	5	19.8	90	32	191	1.11	265	260	-1.89%	1.02
1	3	2.04	6	16.7	90	40	254	0.80	224	237	5.80%	0.95
1	2	1.85	6	19.4	90	41	178	0.65	158	164	3.80%	0.96
1	4	2.18	5	19.8	90	29	129	1.39	274	301	9.85%	0.91
2	11	2.07	6	24.1	90	35	505	1.16	442	477	7.92%	0.93
2	11	2.02	6	18.6	90	42	676	0.83	418	442	5.74%	0.95
2	12	2.01	6	23.7	90	40	630	1.09	427	457	7.03%	0.93
1	6	2.19	5	15.4	90	45	509	0.96	271	277	2.21%	0.98
1	10	1.95	6	16.3	90	27	878	0.98	315	324	2.86%	0.97
2	13	2.04	6	16.4	90	34	916	0.67	423	428	1.18%	0.99
2	14	2.04	7	21.9	90	41	1272	0.79	384	389	1.30%	0.99
1	3	1.95	6	29.8	90	43	175	1.25	163	183	12.27%	0.89

Table 3.3 Comparison of observed results with calculated results from model 1

Therefore, the assumption that the plant cycle times are constant must be redefined and the way in which the variations of these times affect the overall production must be investigated.

3.8.1 Loss of production through variable cycle times: bunching

If the dumptrucks in an earthmoving operation all spotted, loaded, travelled and dumped at a constant rate then the time between trucks arriving at the queue for the loader (or the *interarrival* time) would be constant. The productive output would also be constant and, as has been shown, easy to calculate. In a real situation the plant will always work at different rates for a number of reasons:

- The dumptrucks may be of a different type or model or even of a different age. Small changes in engine specification will cause the trucks to run at different speeds.
- The drivers of the trucks will have varying ability and motivation and thus some will drive faster than others.
- If a haul road crosses a public highway then the trucks will have to wait at any traffic lights installed for varying amounts of time, depending on the traffic on the roads.
- The rate at which the trucks are loaded will also change. If the ground becomes harder, or if the cut has to be trimmed, at formation or at a batter for example, then the load times will increase thus altering the truck cycle time.

This inherent variance in the cycle times will cause the interarrival time to be variable also; trucks will arrive at random with long times with no arrival and then a few units arriving within a few seconds. Thus, the consequence of the cycle time variability is known as *bunching*, and the amount by which this effect reduces the production by can be called the *bunching factor*. The line on the graph in figure 3.4 can now be modified to allow for this extra factor. The bunching factor can be defined as:

$$\text{Bunching Factor} = \frac{\text{Actual Production}}{\text{Production after Mismatch}} \tag{3.2}$$

where production after mismatch is that which would occur if the cycle times were constant. The last column in table 3.3 shows the bunching factors for the different operations. The second line has a bunching factor of greater than one; this is not possible and is probably due to an error in taking the data on that operation. Figure 3.11 is an update of figure 3.4 where each point on the graph represents a single observed earthmoving operation.

From figure 3.11, the following observations can be made:

- For low match factors (0.6 and below) the effect of bunching is negligible. The reasons for this are easy to explain qualitatively: an operation with such a low match factor is very under resourced. The variability of cycle times on such an operation will not have a large effect because when a truck arrives at the loader there is unlikely to be a queue and it can be loaded immediately.
- The rest of the points have two distinct areas. Up to match factors of approximately 1.2, the points follow a general envelope with the largest loss through bunching at match factors of 1. Again this can be explained qualitatively. At perfect match a hauler will arrive at the loader just as another has completed the loading process. If this arriving hauler were to be a few seconds late then the loader would become idle thus reducing efficiency. Succeeding haulers would then have to wait to be loaded decreasing efficiency further. In other words a perfectly matched operation is very sensitive to changes in cycle time whereas an under resourced operation will have an idle loader to act as a buffer to changes in arrival time. The converse should be true for over resourced operations: if trucks are waiting in a queue then the loader will always have a supply of trucks and production rate can be maintained if not cost effectively. Irregularly arriving trucks will have little effect on the loader's operation. From figure 3.11 this can be seen to be true for match factors up to 1.2.
- Although there is little data above match factor 1.2 it can be seen that there is no general trend or even envelope for the points. From the argument above, the points should approach the 100% production line but many are at 90% or lower. The reason for this is harder to explain. The reduction in production must be due to the fact that the loader cannot maintain its output because it is idle for some of the time but why is this when there is a large over capacity of trucks? One reason may be because the number of trucks is simply too great for the haul route to cope with: the trucks are simply causing a kind of traffic jam. Certainly, on narrow haul routes, it has been observed that trucks must slow down or even stop

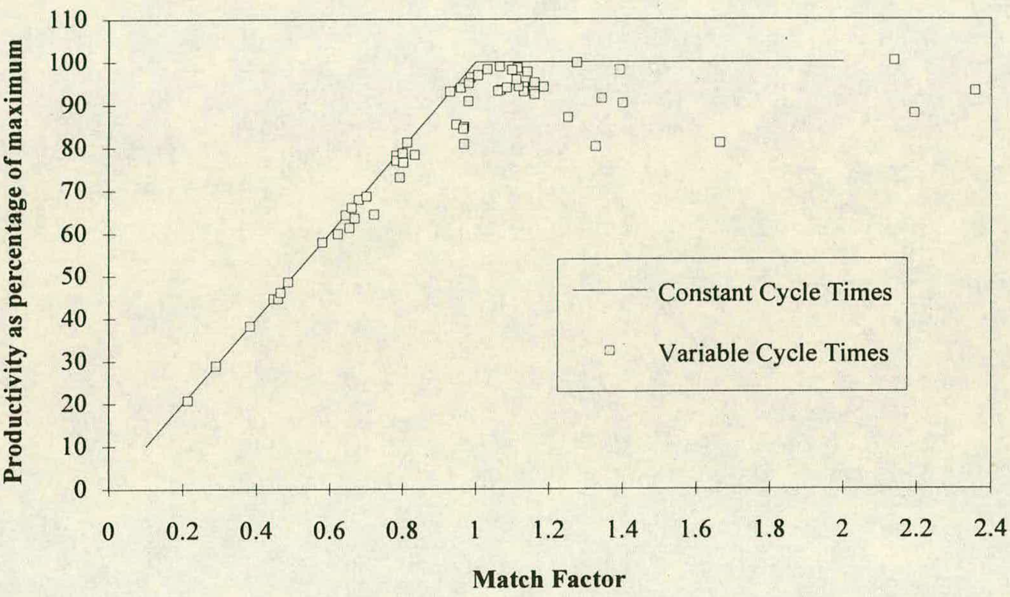


Figure 3.11 The effect of plant match and cycle time variability on operation productivity

when passing a truck coming in the other direction; as the number of trucks increases, the instances when this occurs will increase.

3.8.2 Model 1 redefined

The way in which the earthmoving system works has been shown to be subtly different from that first proposed in section 3.4. Not only is the production from an operation more difficult to predict but the time the trucks wait in the queue is also now random. The greater the queue time, the lower the utilisation and this is a factor that would be useful to know if the efficiency of an operation has to be controlled. The amount of time the loader is idle is also important and will vary; if the bunching is very bad, a loader may be idle for some time before a large number of trucks arrive at once. The first model therefore needs to be redefined and more data gathered to develop and test this new model. The following points will have to be taken into account:

- The distributions of the various time components will have to be determined. Operations will ideally be watched for longer than has previously been the case as component time standard deviations as well as mean times need to be determined. The more cycles observed, the greater the accuracy of the standard deviation.
- The future model will have to be validated and so the extra factors such as plant utilisation times will have to be determined for real operations.

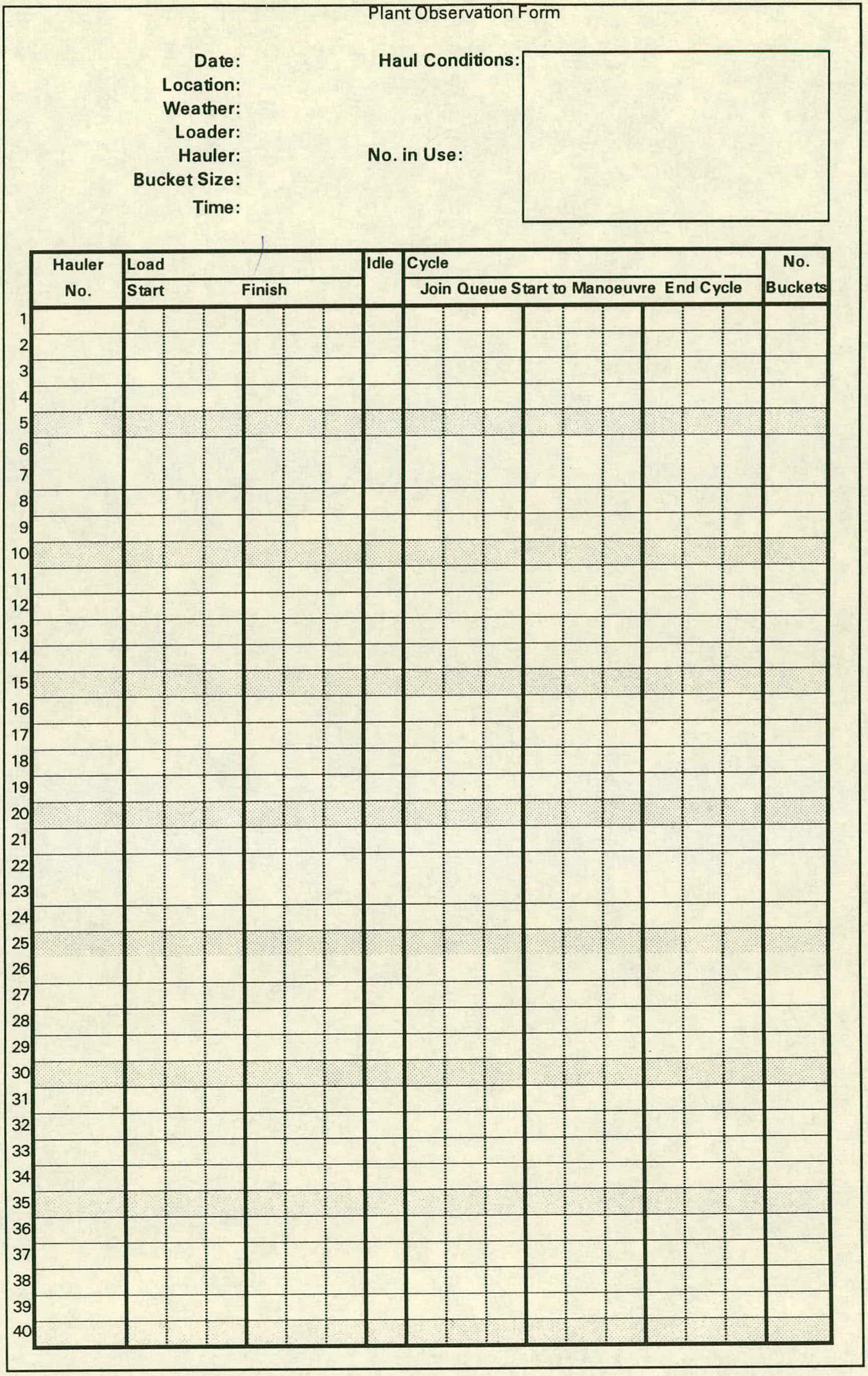
The existing method study has been revised to enable loader idle time to be recorded and other components such as spotting time to be taken with more accuracy. The revised data collection form can be seen in figure 3.12.

3.9 Conclusions

The conclusions that can be drawn from this chapter are:

1. The plasticity index of the haul road soil can affect the productivity of an earthmoving operation due to its susceptibility to moisture. The undrained shear strength of a soil with low plasticity (10% or lower) is quickly reduced with small increases in moisture content.

2. The earthmoving system can be studied as a queueing system. The hauling units (dumptrucks or scrapers) are the customers, the loading units (excavators or pushers) are the servers, the arrival, or more specifically the return of the hauling units at the queue becomes the input process and the service or loading cycle is



Plant Observation Form

Date:
Location:
Weather:
Loader:
Hauler:
Bucket Size:
Time:

Haul Conditions:

No. in Use:

	Hauler No.	Load		Idle	Cycle			No. Buckets
		Start	Finish		Join Queue	Start to Manoeuvre	End Cycle	
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								

Figure 3.12 Updated Plant Observation Form

the output process. The queue discipline of this queueing system is first come first served (FCFS).

3. The match factor of an operation (as defined by equation 3.1) indicates the efficiency of an operation with respect to its plant match. A match factor of one indicates perfect match with an over-resourced operation (i.e. too many trucks) being indicated by a match factor greater than one. Conversely, a match factor less than one indicates an under-resourced operation.

4. A simple spreadsheet model can be defined which predictively estimates the productivity of an earthmoving operation. This model assumes that the component times of the earthmoving cycle are constant. It thus provides a calculation for productivity that is consistently over-estimated.

5. An appraisal of the assumptions made in formulating the first model, and the results obtained from it indicate that the variation, or randomness, of the cycle times must be taken into account when estimating the productivity.

6. This randomness is one of the causes of bunching of the hauling plant. Bunching is when a number of haulers travel close together thus preventing the loader from having a constant supply of customers to service. The lower utilisation of the loader brings about a reduction of the productivity of the operation that can be quantified with the bunching factor (equation 3.2). This bunching factor has been shown, through a study of actual operations, to vary with the match factor of the operation: greatest bunching occurs at match factors at or around one.

7. A match factor of 1.2 can be considered, by investigation of actual operations, to be the upper limit of efficiency for an earthmoving operation. Above MF = 1.2, productivities have little relationship with the match factor which indicates that other factors, such as over-crowding of the haul road, come into play.

Chapter 4

Development of the Simulation Model

Simulation and more particularly, discrete-event simulation, are introduced as a means of estimating the productivity of earthmoving operations whilst taking account of the inherent variability of earthmoving plant cycle times. The independence of the observed, raw data is established and a random variate generator is developed to provide the stochastic data to drive the simulation model.

4.1 Introduction

Chapter 3 discussed a model of the earthmoving system based on the means of the components of the cycle times but, as was seen in section 3.8, these component times are never fixed at one value but are random. A redefined earthmoving model will therefore attempt to recreate the randomness of the system and provide a more accurate estimate of the output of an actual system. Such a model can be analysed using two operational research techniques, namely queuing theory and simulation.

Queuing theory can provide estimates for various output characteristics of the queue; for example, the number of customers (trucks) in the queue, the expected length of time the customer spends in the queue, the probability distribution of the number of customers in the queue and the amount of time the server (loader) spends idle. Whilst these values are useful to the earthmoving engineer, especially utilisation factors, the productivity of the system is of major interest and this is difficult to estimate using queuing theory. Furthermore, most queuing theory solutions require estimates of the service rate (μ), or the average number of service completion's per unit time. For a fully resourced system, the service rate is merely the inverse of the service time but the server in an under resourced system will spend an unknown amount of time idle. The service rate in this instance cannot be reliably determined. These are two reasons why queuing theory is difficult to apply to the earthmoving system and although various researchers have found solutions to provide estimates for loader utilisation (Carmichael, 1987a, 1987b; Donjerkovich, 1983), which may be useful to contractors, the rest of the queuing theory literature related to construction is almost entirely dominated by one person: David G. Carmichael of the University of Western Australia. (1986, 1987a,b,c and 1990.) This may reflect the very esoteric and 'inaccessible' nature of queuing theory, as opposed to simulation, which is far easier to comprehend.

Simulation, is a very versatile method for estimating the output of a system when all other methods have been found to be unsuitable. It has therefore become thought of as a 'method of last resort' but in recent years, with the advances in computing power, simulation has become an extremely common analysis technique used extensively in the areas of production and industrial engineering. The flexibility of the technique means that a simulation model can be repeatedly analysed for different initial conditions. Therefore simulation can be used to experiment with the real system; such techniques have provided valuable insights to the earthmoving system and will be discussed in chapter 7.

4.1.1 Developments in simulation techniques in the Construction Industry

Simulation applied to construction operations is not new. Much work has been done in this area since the seventies, the majority of research being carried out in the United States at institutions such as the Construction Institute at Stanford University and at The University of Illinois.

Early work was started by Fondahl (1960) at Stanford University using photographic techniques to analyse the effectiveness of construction operations and further developed by Teicholz (1963) and Gaarslev (1969). In the seventies Halpin developed CYCLONE which was later extended at Stanford University (Halpin 1977). This program has instigated most of the work in simulation techniques in the last two decades, with most of the research in the United States using CYCLONE or one of its many offsprings. These modifications have included INSIGHT (Kalk and Douglas 1980; Paulson et al. 1983 and 1985). Bernold developed a version called UM-CYCLONE (University of Michigan - CYCLONE) as part of his PhD research and used the program to aid in earthworks quantities and resource allocations (Bernold 1986) as an alternative to the mass haul and linear programming methods discussed in chapter 2 (2.3). The latest version of CYCLONE, MicroCYCLONE (Halpin 1990) is available on a PC rather than the minicomputer versions of the 1980s.

Clemmens and Willenbrock (1978), at Pennsylvania State University, developed a computer simulation program independent of the work going on at Stanford specifically for the estimation of scraper earthmoving operations. *Scrapesim* used data from three construction sites to set up probability distributions for the various parts of a scraper cycle. However, the authors commented, surprisingly, that they considered the haul road condition had no significant effect on the scraper travel time and that the capacity of the scraper and the size of the load obtained were insignificant and could not in any case be accurately predicted for estimation purposes. On the variability of travel times, the authors "discovered that those scrapers with twin engines generally travelled at a faster rate than those with a single engine." Not surprisingly perhaps, this work does not seen to have been developed further and few references to this work exist.

More recently, a very prolific researcher into computer simulation applications in the construction industry is Simaan AbouRizk at the University of Alberta in Edmonton. Starting with research with Halpin at

Purdue University, variance reduction techniques were developed for simulation programs (mainly CYCLONE and MicroCYCLONE) to reduce the number of simulation runs required (and hence run time) without sacrificing the level of confidence of the output (AbouRizk et al. 1990). This of course is not as necessary any more as the speed of computer simulation programs has increased along with computer development in the last five years. Computer simulation run time is no longer a limiting factor in the development of simulation models. Further work was done validating the CYCLONE program using queuing theory (AbouRizk et al. 1991a) and developing techniques to fit beta distributions to activity times in simulation models (AbouRizk et al. 1991b and AbouRizk and Halpin 1992). With Gonzalez-Quevedo et al. (1993), MicroCYCLONE was compared with SLAM II, a mainly manufacturing industry based simulation package. Similar results were established thus validating MicroCYCLONE due to SLAM II's established and tested reputation. AbouRizk and Wales (1993) used the simulation techniques in a rare excursion from the more usual earthmoving applications in developing a model to estimate the impact of adverse weather on a construction programme. The two most recent papers from this author have reviewed the most recent work in simulation (mainly MicroCYCLONE) and the fitting of beta distributions to sample data (AbouRizk and Shi, 1994 and AbouRizk et al. 1994 respectively).

Other workers in this field include Amr Oloufa, currently at Pennsylvania State University. Three papers (Oloufa and Crandall 1992 and Oloufa 1993a and b) cover very similar ground by first reviewing the history of simulation in construction and then developing object - oriented techniques based on CYCLONE and developed into a new package called MODSIM. The object oriented view that a system is composed of interacting physical objects intends to 'bridge the gap' between system and model.

Riggs and Han (1991) and Touran (1992) have developed the simulation technique further by simplifying the interface between computer and user. Both sets of researchers have independently used an expert system base to generate computer code, Riggs and Han into CYCLONE, Touran into SLAM II.

Finally, one of the most recent papers on simulation in the construction industry (Farid and Koning, 1994) outlined the use of MicroCYCLONE to verify (but not validate, see chapter 6) a queuing theory program (FLEET) developed to model multi-loader truck systems in 1988

(Karshenas, 1989a and b and Karshenas and Farid, 1988). Both queuing theory model and MicroCYCLONE were set up using the same assumptions about time distributions (exponential, see 4.3) and that operations were carried out at the steady-state (see 5.2.1). Not surprisingly therefore, very similar results were obtained thus indicating that the queuing theory model is verified (i.e. producing the correct results for the model) but in no way validated (i.e. estimating the correct results for the real life situation). The authors actually conclude that "the exponential distribution is not appropriate for modelling activity durations of construction operations" and also that "a more realistic model should include start-up, operating and shut-down transitions" i.e. the actual system does not operate at the steady-state.

4.2 Simulation: Application to the Earthmoving System

The simulation model of the earthmoving system has already been defined in terms of mean component times. The assumptions of the model given in section 3.7.1 all hold true for the simulation model with the exception of assumption i: all time components in the simulation model are represented by probability distributions, that is the input to the system is random. There are two immediate aspects that need to be considered before simulation can be applied to the earthmoving system: what type of simulation technique is to be used and what form will the input probability distributions take. Selection of the input probability distribution will be discussed in section 4.3. This section will determine the type of simulation model to be studied and the aspects of the method chosen to analyse this model.

4.2.1 Classification of the simulation model of the earthmoving system

There are various types of simulation technique such as continuous, combined discrete-continuous and Monte Carlo simulation. To determine the type of technique to be used the following characteristics need to be considered to classify the simulation model:

i. *Dynamic or Static Model.* The earthmoving simulation model is dynamic because it represents a system that evolves over time. A static model would be one in which time plays no role; such systems can be analysed using Monte Carlo simulation techniques. (Note that some authors consider Monte Carlo simulation to be any that involves the use of random numbers, which would include the simulation of the earthmoving system. This work will use that definition of Monte Carlo simulation that restricts the term to those simulations that analyse a system at one particular time only.)

ii. *Deterministic or Stochastic Model.* Model one, proposed in chapter 3, is a deterministic model because it contains no random elements. The exact output of this model could be determined once the initial conditions had been set. The second model has random input and is therefore known as stochastic. The output from a stochastic model is also random and therefore can be considered only as estimates.

iii. *Discrete or Continuous Model.* The model of the earthmoving system is discrete because the state* of the system changes at instantaneous points in time. For example, the state of the earthmoving system would change with the arrival of a truck and remain so until the loader has loaded all trucks in the queue. A continuous model would represent a system in which the state of the system changes continuously with respect to time. To emphasise the difference between the two types of system, work by Braun, 1975 and Gordon, 1978 can be consulted. Law and Kelton (1991, p. 110) use these works to describe continuous simulation with the following example:

> Consider a *predator-prey* situation that has two populations that interact with each other: the predator depends upon the prey for its food and existence. The state of the two populations at a particular time can be determined using differential equations. Let $x(t)$ and $y(t)$ denote the numbers of prey and predators in the system at time = t.
>
> In the absence of predators, the prey has a natural growth rate of $rx(t)$ and in the absence of prey, the predators have a negative growth rate of $-sy(t)$. With interaction between the two populations, the death rate of the prey and the growth rate of the predators can be assumed to be proportional to the product of the two population's $x(t)y(t)$. Therefore, the overall growth rates, dx/dt and dy/dt are given by:

---

* The *state* of the system is defined as 'the collection of variables necessary to describe the status of the system at any given time.' (Winston, 1987 p. 934). These variables, or *state variables*, are in this case the number of trucks in the queue, the number of busy loaders, the departure time of the next truck and the arrival time of the next truck. These variables will describe the state at the present time and in the future. In a system, an object of interest is an *entity* and any properties of the entities are called *attributes* which define the system state. In this system, the entities are the trucks (customers) and loaders (servers). The loader will have the attributes *busy* or *idle* and also the number of customers in the queue, while the trucks attributes are *arrival time* and *departure time*.

$$\frac{dx}{dt} = rx(t) - ax(t)y(t) \tag{4.1}$$

$$\frac{dy}{dt} = -sx(t) + bx(t)y(t) \tag{4.2}$$

The numerical solution to these differential equations can be found using a computer package such as *Maple* (see Braun 1975).

The above example is completely deterministic but continuous systems need not be so restricted. The constants in equations 4.1 and 4.2 could be stochastic quantities that change their values with time.

The model of the earthmoving system is therefore dynamic, stochastic and discrete. The type of simulation technique used to analyse such a model is called *Discrete-Event Simulation.*

4.2.2 Discrete-Event Simulation

This technique is used when the system evolves over time and the state of the system changes instantaneously at discrete points in time. (Law and Kelton, 1991 and Banks and Carson, 1984 provided much of the background theory to this type of simulation for the developments presented here.) When the state of the system changes, an *event* is said to have occurred. The earthmoving system has events of two types: arrivals and departures. An arrival occurs when a truck returns and joins the queue. For example, if the loader is idle and a truck arrives, the state of the system changes from idle to busy with one truck in service. If a truck arrives at an empty queue but with the loader busy, then the state of the system is busy with one truck in service and one waiting. When a truck has been loaded and starts to haul to the load area a departure event has occurred. When this event occurs the state of the system will change by reducing the number of trucks queuing by one. If there are no trucks waiting the system is idle.

4.2.3 Time advance mechanisms and the event list

An important concept in simulation is that of the *simulation clock,* that is a variable in a simulation exercise that gives the simulated time. The units of the simulation clock are the same as those of the input parameters. For example, if the cycle component times are expressed in seconds then the simulation clock would also be represented in seconds. This variable is advanced depending on the length of time to complete a certain phase of the simulation and there are two mechanisms that can be used:

i. *Next event time advance.* The simulation clock is initialised at zero and then advanced to the time of the occurrence of the next event. The state of the system is updated depending on the type of event and the times and types of future events are also determined. The clock is then advanced to the time of the next event and the process is repeated until some condition has been reached, for example when the simulation clock has reached a certain value or a pre-set number of cycles have been completed.

ii. *Fixed increment time advance.* This method advances the simulation clock by a pre-determined increment. After every update of the clock, any events that occurred during the interval are said to have occurred at the end of the interval. The system state is updated accordingly and the time advanced again by the same increment. This method will increase errors due to changing the event times if they occur within an interval. Reducing the size of the increment will reduce errors but increase execution time. For these reasons, this type of time advance mechanism is rarely used and will not be used here.

Both of the above methods need to have some way of determining the type and time of occurrence of the next event. This is done by having an *event list* that is a schedule of the type and time of events that are yet to occur. When an event has been processed it is removed from the list. Note that for some simulations, including those of the earthmoving system, the time of future events cannot be determined until some other event has occurred. For example, the time of arrival of a truck can only be calculated when the time of its previous departure is known. In such cases, the event list is continuously updated.

The method of discrete-event simulation is best explained by use of an example. The data in the following is taken from an actual observed operation. In a real simulation the times of the events will be determined by generation of random component times.

4.2.4 Example of discrete-event simulation

Table 4.1 has the cycle component times for four cycles each of 4 trucks. The loader has a bucket with a heaped, bank* volume of 1.8 m^3. Eight buckets of material are loaded per truck giving a total load of 14.4 m^3 per cycle.

Initialisation. The simulation clock is initialised at $t = 0$ and the status of the system is initialised with the server idle, 4 trucks in the queue and next arrival (from the event list) at $t = 0$. The event list is actually initialised with 4 arrivals, one for each truck, at $t = 0$.

Simulation. The next (first) event is read from the event list and processed. Because 4 events happen at $t = 0$, the events are processed in truck number order. Truck 1 has a spot time of 30 seconds and its load time is 143 seconds giving a departure time at $t = 0 + 30 + 143 = 173$ seconds. This departure event is added to the bottom of the event list. Trucks 2 to 4 are processed in the same way with departure events scheduled at times $t = 358$, 554 and 776 seconds respectively. The next nine events occur at the following times:

$t = 173$. Truck 1 departs, its next arrival event is at $t = 173 + 205$ (travel) + 63 (dump) = 441 seconds.

$t = 358$. Truck 2 departs. Next arrival event for truck two is at $t = 682$ seconds.

t=441 Truck 1 arrives, completing its first cycle. Cycle time is therefore 441 seconds. This truck has to wait in the queue until truck 4 departs at $t = 776$. A departure time is scheduled for $776 + 18 + 133 = 927$ seconds.

t=554 Truck 3 departs, with next arrival scheduled at $t = 837$.

t=682 Truck 2 arrives, waits till 927 (departure of truck 1), next departure at $t = 1105$.

t=776 Truck 4 departs, next arrival at $t = 1076$.

t=837 Truck 3 arrives, waits till 1105, next departure at $t = 1256$.

t=927 Truck 1 departs, next arrival at $t = 1223$.

t=1076 Truck 4 arrives, waits till 1256, next departure at $t = 1412$.

---

* 'Bank' volume is the volume in the unexcavated state as opposed to loose volume, after excavation.

Truck	Spot Time	Load Time	Buckets / Load	Volume (m3)	Travel Time	Dump Time
1	30	143	8	14.4	205	63
2	34	151	8	14.4	250	74
3	68	128	8	14.4	217	66
4	87	135	8	14.4	231	69
1	18	133	8	14.4	208	88
2	38	140	8	14.4	237	59
3	21	130	8	14.4	220	63
4	26	130	8	14.4	257	71
1	25	159	8	14.4	208	48
2	28	148	8	14.4	283	50
3	38	130	8	14.4	183	67
4	20	135	8	14.4	271	76
1	48	123	8	14.4	212	43
2	63	137	8	14.4	246	52
3	81	126	8	14.4	225	58
4	17	160	8	14.4	231	47

Table 4.1 Component times, in seconds, for example in 4.2.4

This continues with the event list being updated as events happen. The cycle time of each truck is recorded as a cycle is completed so that the number of cycles per hour can be calculated and from that a productivity. If the simulation is continued by hand as above, truck 4 will arrive after its third cycle at time $t = 2442$ (or 40 minutes and 42 seconds). 12 cycles have been completed at 14.4 m^3 per cycle giving a total of 172.8 m^3 of material moved. This equates to a production rate of $172.8 / (2442/60^2) = 254.7$ m^3/hour. Other performance measures can be calculated if necessary. For example, the utilisation of the trucks can be calculated by summing the amount of time the trucks spend idle (in this case a total of 3841 seconds) and dividing by the total machine running time. Truck utilisation is therefore $3841 / (4 \times 2442) = 0.39$ or 39%. Note that in this case loader utilisation is 100% as it is always in a busy state.

4.3 Selection of probability distributions

The example in the previous section used component times taken from an actual operation but for a simulation to perform the task it is designed for, these times must be random samples from some probability distribution (or *probability density function, pdf*) that agrees with the distribution that occurs in real life. The samples from such probability distributions are called *random variates*. The algorithm that produces the samples is a *random variate generator*. These will be discussed in section 4.4 but before that, the type of probability distribution to be used must be selected.

There are two approaches that can be used in selecting a probability distribution. Firstly an empirical distribution can be formulated which would be unique to an observed distribution. Such a distribution is a piecemeal mathematical representation of the *actual* data observed. The second approach is to fit a theoretical distribution to the data. There are many theoretical distributions such as normal, lognormal, exponential and Erlang to name a few. This second approach is preferable for a number of reasons. In the case of the earthmoving system, there are four time components that need to be represented; each situation will require four different empirical distributions. A theoretical distribution that is much more compact should be able to be adjusted to represent different situations thus avoiding the need to store large quantities of standard empirical distributions. For the earthmoving system it is therefore ideal if a theoretical distribution can be found which represents all the time components. To do this the form of the observed data must be studied.

4.3.1 Form of observed data

The component times taken from observed earthmoving operations can be shown graphicaly as cumulative distribution plots (or *cumulative distribution function, cdf*). Theoretical distributions can then be visually compared with these plots to see if a good 'fit' is obtained. If a theoretical distribution is found it must be assumed that this distribution is valid for all data ranges outside those observed. Figures 4.1 to 4.6 show cdfs for time components taken from the M1/A1 and M3 sites.

i. *Spot Time component.* Figures 4.1 and 4.2 show the spot time cdf for both the Cat D400 and Volvo A35 articulated dump trucks. They both have similar shapes: the flat top portion of the curve indicating that the spot time can be quite long in comparison to the average while the lower part of the curve indicates that there is a sharply defined lower bound of about 15 seconds. The average times are 25.2 seconds for the Cat D400 (based on 217 observations) and 30.6 seconds for the Volvo A35 (based on 312 observations).

ii. *Bucket Swing Time component.* Figures 4.3 and 4.4 show the bucket swing time (BST) cdfs for two different situations. Figure 4.3 is the cdf for a Cat 245B backhoe excavator in clay while figure 4.4 is for a Cat 245D excavating in chalk. The clay excavated is that on the M1/A1, which has been described in section 3.2.2. Both distributions have similar shapes with the higher 25% of observed times occurring within approximately the upper 60% of the time range. The curves are smoother than those for the spot times due to the number of observations. The BST for the Cat 245B in clay has an average of 19.4 seconds based on 985 observations, the Cat 245D in chalk has an average of 14.6 seconds based on 1082 observations. It is interesting to note that Lewis and Schexnayder (1986) report their studies of the Caterpillar 245 excavator and obtained very similar bucket swing times as those obtained here. Both works, however, indicate that the set-up of the excavator and the working conditions heavily influence the speed at which it works. For example, figure 4.4 shows a faster BST for two reasons: the chalk is a much drier, more friable material than clay and hence is quicker to

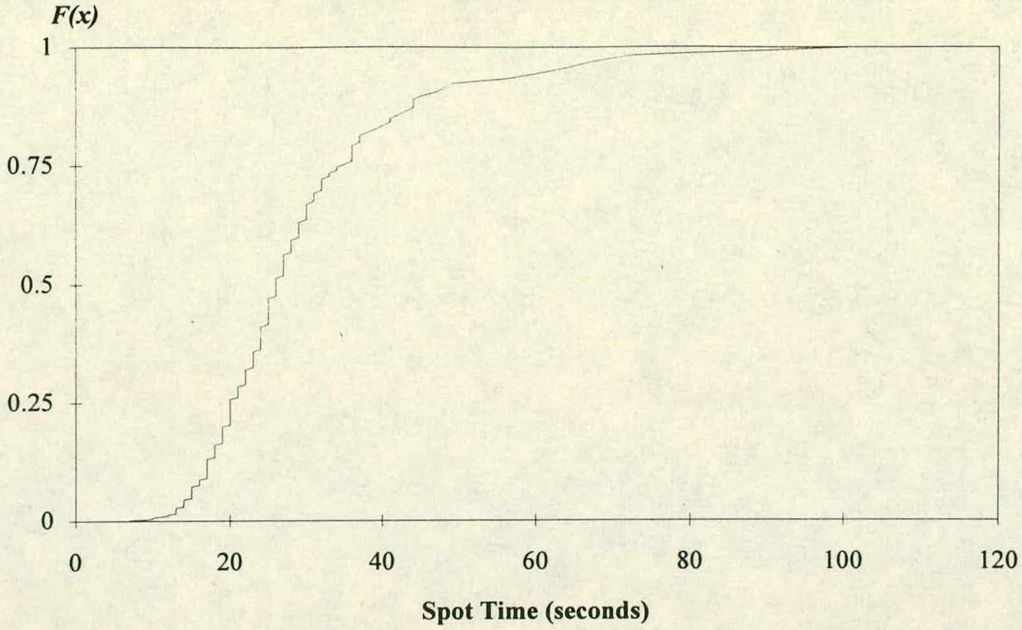


Figure 4.1 Cumulative Distribution Function for Spot Time for Cat D400D. (Observed Data).

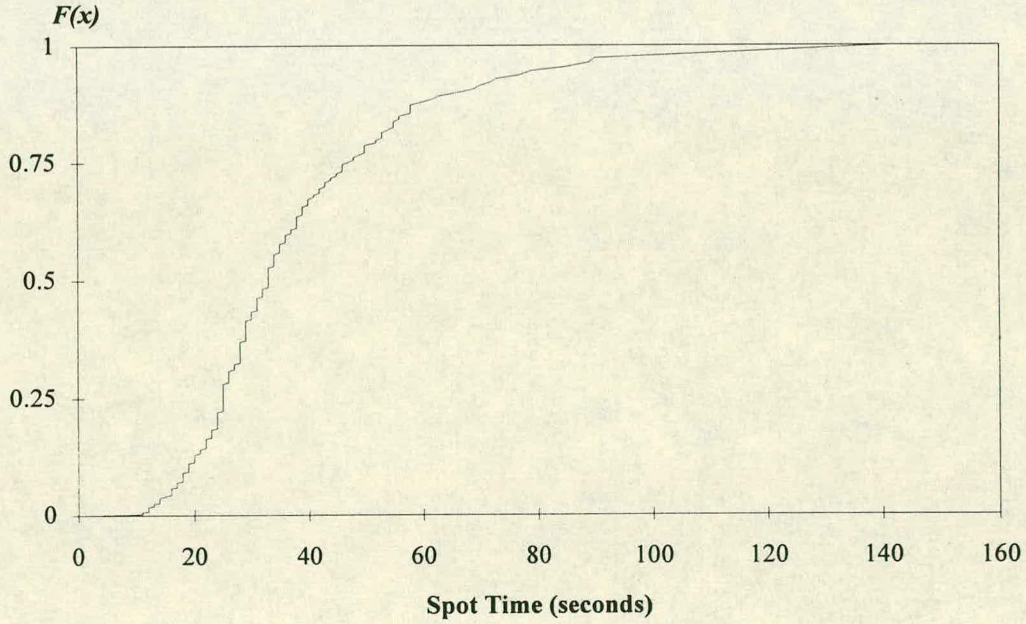


Figure 4.2 Cumulative Distribution Function for Spot Time for Volvo A35. (Observed Data)

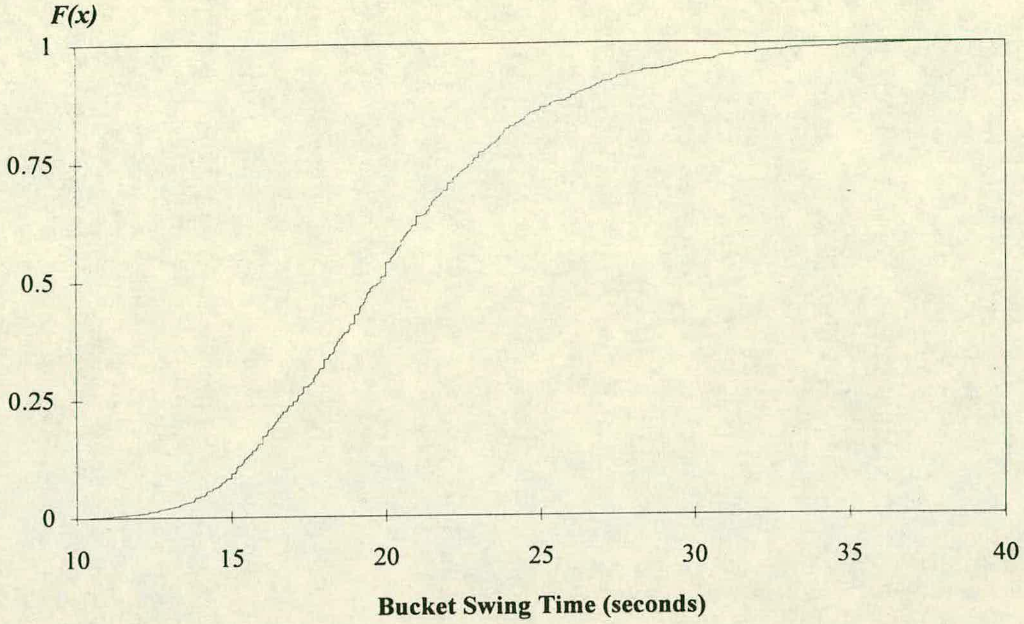


Figure 4.3 Cumulative Distribution Function for Bucket Swing Time for Cat 245B in Clay (Observed Data).

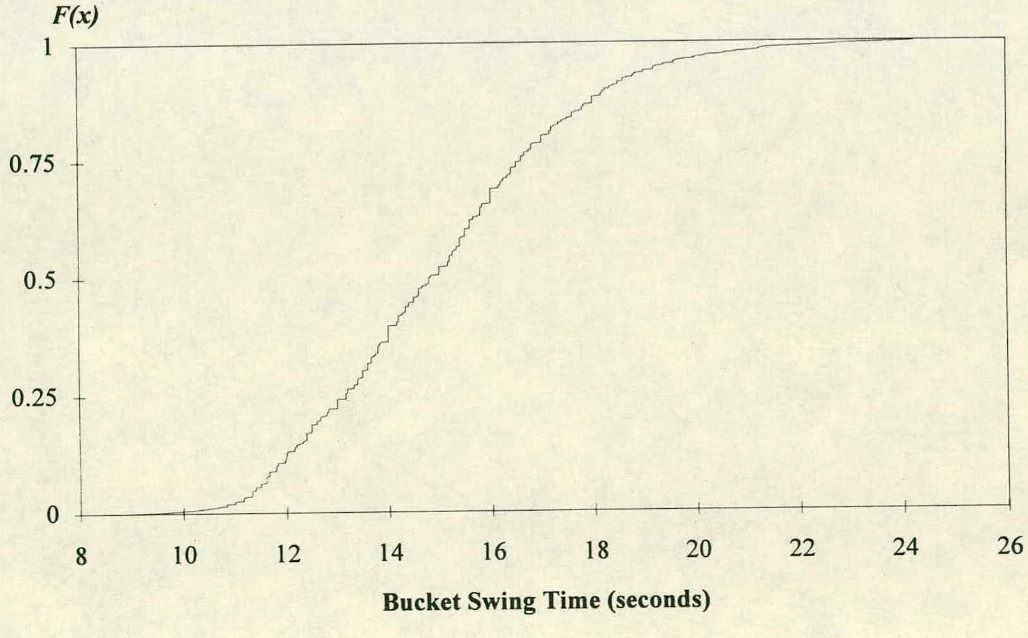


Figure 4.4 Cumulative Distribution Function for Bucket Swing Time for Cat 245D in Chalk. (Observed Data).

excavate and dump than the more cohesive and adhesive clay of figure 4.3.

iii. *Dump Time component.* The cdf for this component is shown in figure 4.5 and it is somewhat different from those for the spot and bucket swing times in that the 's' curve is the other way around. The lower part of this curve is flatter than the upper part indicating that a disproportionate number of observations (25%) occur in the first 50% of the time range. The curve is not very smooth due to the small number of observations (126) and has an average value of 61.5 seconds.

iv. *Travel Time component.* Figure 4.6 shows 6 cdfs for different haul lengths. Unfortunately, the curves are very irregular due to the small number of observations (they must be taken from single operations with total observation time being between 1 and 2 hours). The short observation time also means that these curves show what is probably only a small part of the overall distribution for a truck's travel time. However, the curves do show that the travel times conform to a general 's' shape, and also the range of the curves tends to increase as the haul length increases. This can be expected; small changes in truck speed will result in larger differences in travel time for longer hauls.

Table 4.2 has a summary of the curves shown in figures 4.1 to 4.6.

4.3.2 Independence of observed data

The previous section showed the form of the observed data but before this data can be used to determine theoretical distributions, it must be established that the data is not dependent in any way. If data is dependent, that is if the value of one observation is linked to the value of preceding observations, then a random process such as simulation cannot recreate this dependence. If the data is independent then the stochastic process in simulation can be used safely to replace it. Winston (1987) has expressed independence in terms of the probabilities of random variables. Two random variables x and y are independent if and only if for any two sets A and B:

$$P(x \in A \text{ and } y \in B) = P(x \in A)P(y \in B) \tag{4.3}$$

This definition shows that if x and y are independent the value of y does not change the probability of any event involving x. To determine the dependence

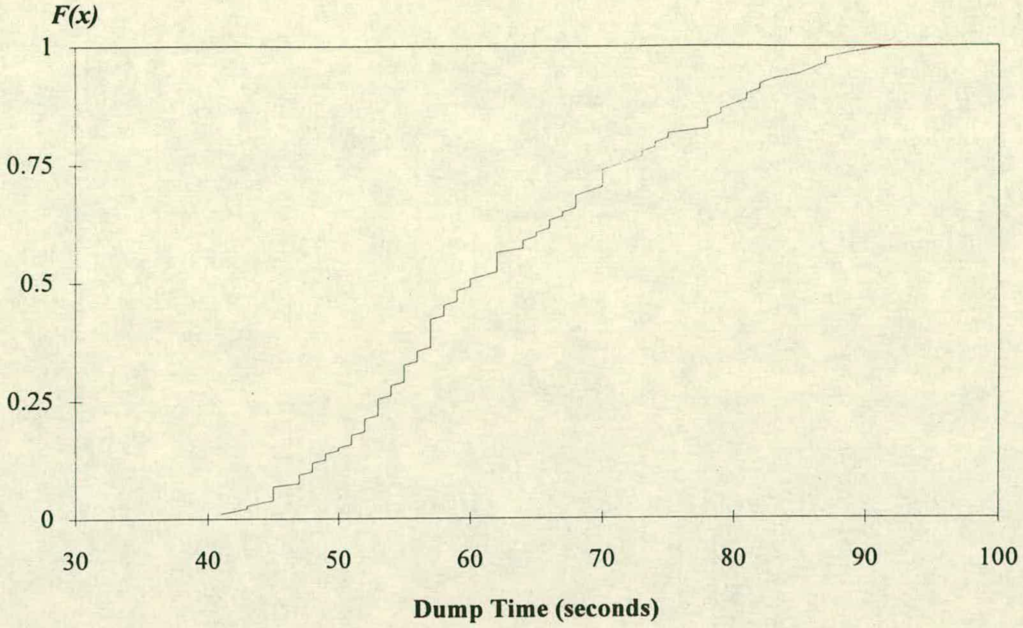


Figure 4.5 Cumulative Distribution Function for Dump Time for all vehicles. (Observed Data).

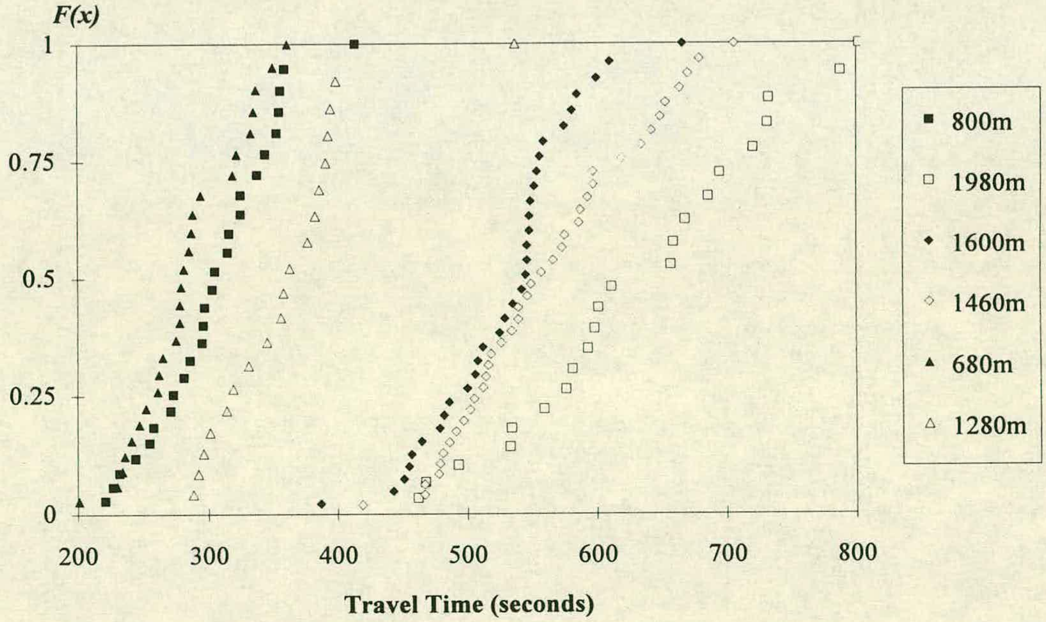


Figure 4.6 Cumulative Distribution Function for Travel Time for all vehicles on chalk haul roads. (Observed Data).

Component	Figure	Sample Mean	Sample Varience	Coefficient of Variation	No. of Points	Lower Limit	Upper Limit	Erlang Parameters	
								k	R
Spot Time (Cat D400)	4.1	25.16	127.92	0.45	217	9	101	5	0.1967
Spot Time (Volvo A35)	4.2	30.64	264.39	0.53	312	4	141	4	0.1159
BST (Cat 245B in Clay)	4.3	19.42	18.15	0.22	985	10	40	21	1.0700
BST (Cat 245D in Chalk)	4.4	14.62	5.95	0.17	1082	9	24	36	2.4571
Dump Time	4.5	61.47	124.54	0.18	126	41	92	30	0.4936
Travel Time	4.6								
680m		281.96	1647.80	0.14	26	201	361	48	0.1711
800m		301.69	2232.54	0.16	26	221	414	41	0.1351
1280m		359.16	3326.81	0.16	19	288	377	39	0.1080
1460m		556.72	4851.21	0.13	39	419	706	64	0.1148
1600m		526.73	3130.46	0.11	33	387	666	89	0.1683
1980m		625.00	9499.14	0.16	22	462	802	41	0.0658

Table 4.2 Summary statistics for figures 4.1 to 4.6

of one set of variables on another there are two statistical calculations that can be done.

i. *Covariance.* For two sets of random variables, x and y the covariance is a measure of the linear association between x and y. It is expressed as the average of the product of the deviations of the random variables from their respective means. If μ_x and μ_y are the mean values of x and y respectively then the covariance of x and y is defined as

$$\mathrm{Cov}(x,y) = E[(x-\mu_x)(y-\mu_y)] \tag{4.4}$$

where $E(x-\mu_x)$ is the expected deviation of x from its mean. If $\mathrm{Cov}(x,y) = 1$ then x and y are totally dependent. Conversely, if $\mathrm{Cov}(x,y) = 0$ then the random variables are independent.

ii. *Correlation.* Covariance is not a dimensionless term and therefore can make interpretation difficult. For example, if x and y are measured in seconds then $\mathrm{Cov}(x,y)$ is in units of seconds squared. Therefore, correlation can be used to establish dependence between variables; for two random variables x and y, the correlation is the covariance divided by the product of the standard deviations:

$$\mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y} = \rho_{x,y} \tag{4.5}$$

Correlation can be used to determine whether two sets of data are independent of one another, but it can also be used to determine if values in one data set are independent of others in the same set by comparing the whole set with subsets. This can be done graphically in two ways: by plotting correlation plots and scatter diagrams. To explain, the spot time data for the Volvo A35 shall be used as an example.

i. *Correlation Plot.* The spot times for the Volvo A35 can be expressed as $x_1, x_2, x_3, \dots x_n$ listed in order of collection. A sample correlation, ρ'_j, can be estimated as the true correlation between two observations that are j observations apart. To form a correlation plot, ρ'_j is calculated for $j = 1, 2, 3, \dots m$ where m is a positive integer. Figure 4.7 shows the correlation plot for values of ρ'_j up to $j = 20$. The correlations have been calculated using a spreadsheet and each calculation is based on the comparison between 292 pairs of random spot times. As can be seen from figure 4.7,

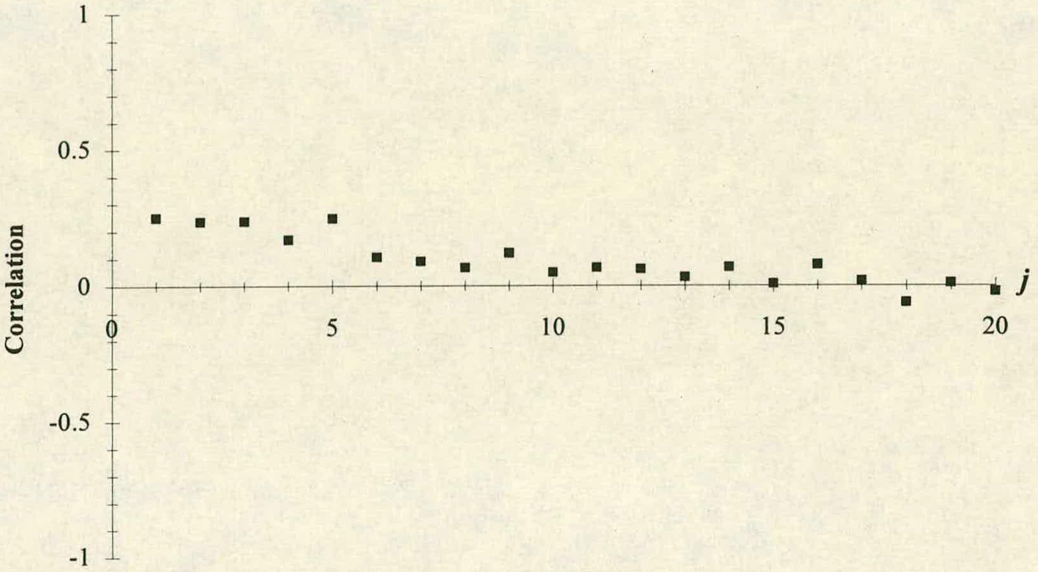


Figure 4.7 Correlation Plot for Volvo A35 Spot Times.

the sample correlations are close to zero but not quite. There is also a trend as j gets larger for ρ'_j to get smaller which is an indication that the data set is slightly correlated in some way and not wholly independent. There are two possible reasons for this. Firstly, ρ'_j is an estimator for the true correlation and as such is a random variable. The second reason is more instinctive. There may be some dependence between spot times, especially when j is small; the longer a truck takes to spot the higher the probability that another truck will join the queue. If the queue is large then there is less space available for manoeuvring and hence spot times will be longer. Also, spot time is partly dependent on the size of the manoeuvring area. This area will not change rapidly which will result in slight positive correlation. The correlation will decrease as j gets larger because an observed spot time will have less effect on one that is, say, 20 observations later.

ii. *Scatter Diagrams*. These are a simple way to establish the dependence between values in a set of data. Again, if the spot time observations are expressed as $x_1, x_2, x_3, \dots x_n$ then the scatter diagram is a plot of (x_i, x_{i+1}) for $i = 1, 2, 3, \dots, n-1$. If the points were positively correlated then they would lie along a line with a positive slope. Figure 4.8 shows the scatter diagram for the A35 spot times for (x_i, x_{i+1}). Figure 4.9 is another scatter diagram for (x_i, x_{i+20}) and it can be seen from both plots that the data is very randomly distributed indicating that the data has little dependence on each other.

Figures 4.10 through to 4.13 show the correlation plots for the other time components shown in figures 4.1 to 4.5. These are discussed below.

i. *Cat 245D Bucket Swing Times.* This actually shows a fairly high level of correlation especially at low values of j. Even at high values of j the correlation is between 0.1 and 0.2. Although this may upset the need for total independence for the observed data, reasons for it can be given in quantitative terms. At all values of j a positive correlation is shown which means that if an observed data point has a high value then subsequent observations will also have high values. Figure 4.10 implies that this is what happens when a Cat 245D operates in chalk. When the observations were taken (on the M3), the excavator was usually set up in the same position for prolonged periods - the cut depth was uniform as was the swing angle and relative positions of the loader and trucks. Chalk also tends to be a uniform material so if the excavator had a bucket swing time

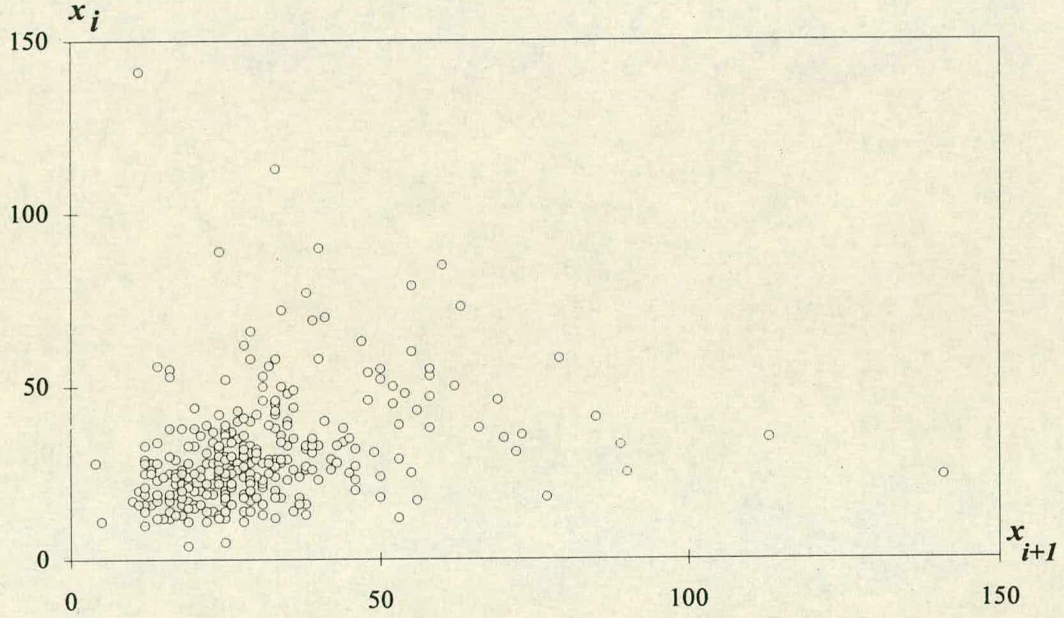


Figure 4.8 Scatter Diagram for (x_i, x_{i+1}). Volvo A35 Spot Times.

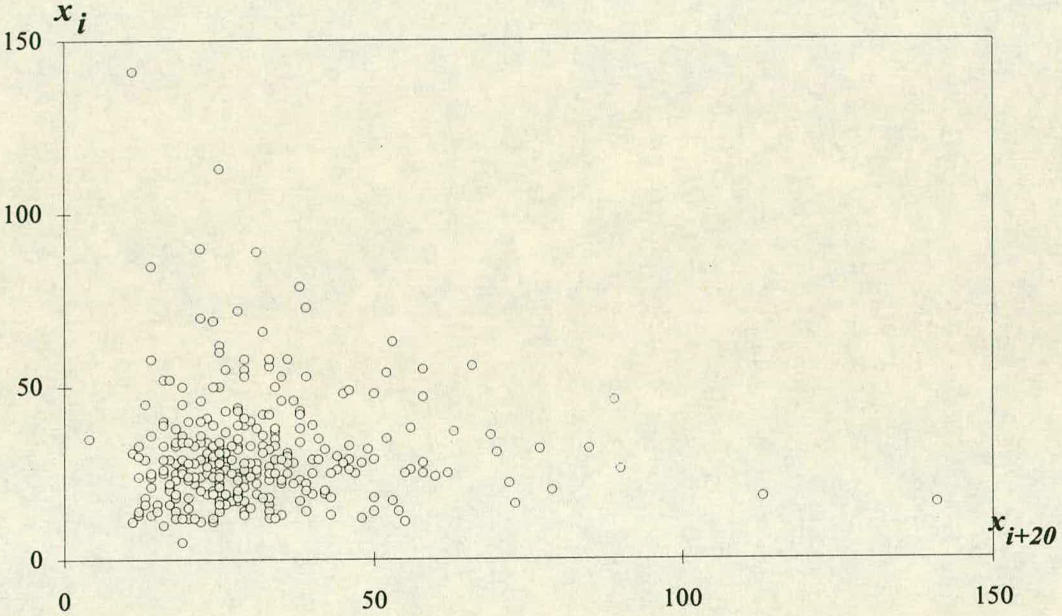


Figure 4.9 Scatter Diagram for (x_i, x_{i+20}). Volvo A35 Spot Times.

of, say, 15 seconds then there would be a high probability that the next bucket swing time would be around 15 seconds. It is not until j is greater than about 30 that low correlations can be computed. (ρ'_j was calculated for j = 30, 40, 50 and 60 and values of 0.1299, 0.1077, 0.0737 and -0.0872 were obtained respectively.)

ii. *Cat 245B Bucket Swing Times in Clay.* Compared with figure 4.10, figure 4.11 has quite low correlations. This can be explained in the same way as for the observations in chalk. The observations for the 245B were taken on the M1/A1 where the cuts were much smaller than those on the chalk cutting of the M3; consequently the loading conditions (i.e. cut depth, swing angle etc.) changed much more rapidly with resulting lower correlations. To compare, the sample correlation in the clay cutting is approximately 0.1 at j = 7 whereas for a similar sample correlation in the chalk cutting, j is between 30 and 40. It can be concluded, therefore, that if the loading conditions remain constant for prolonged periods, the observed data cannot be considered independent.

iii. *Cat D400 Spot Times.* Figure 4.12 shows the correlation plot for the Cat D400 spot times that can be compared with that of the Volvo A35 in figure 4.7. At first the conclusion can be reached that observations on the D400 are much more independent than those of the A35 but the way in which the data was taken must be considered. The A35 dumptruck was much more common than the D400 and consequently observations of D400 spot times did not necessarily occur concurrently. The j values in figure 4.12 are therefore not strictly true. It would be difficult to obtain, with the data available, true values of the sample correlation for the spot times.

iv. *Dump Times.* Figure 4.13 shows the correlation plot for this component. This graph indicates that the observed dump times are independent with the highest correlation value being approximately 0.18 and the remainder at or below 0.1.

Figure 4.14 shows scatter plots for (x_i, x_{i+1}) to complement figures 4.7 to 4.13. All show random scattering but a very slight trend can be detected in figure 4.14c, the scatter diagram for Cat 245D bucket swing times.

It has been shown that the observed data is not totally independent. It is also probably true that totally independe of, say, 15 seconds then there would be a high probability that the next bucket swing time would be around 15 seconds. It is not until j is greater than about 30 that low correlations can be computed. (ρ'_j was calculated for $j = 30$, 40, 50 and 60 and values of 0.1299, 0.1077, 0.0737 and -0.0872 were obtained respectively.)

ii. *Cat 245B Bucket Swing Times in Clay.* Compared with figure 4.10, figure 4.11 has quite low correlations. This can be explained in the same way as for the observations in chalk. The observations for the 245B were taken on the M1/A1 where the cuts were much smaller than those on the chalk cutting of the M3; consequently the loading conditions (i.e. cut depth, swing angle etc.) changed much more rapidly with resulting lower correlations. To compare, the sample correlation in the clay cutting is approximately 0.1 at $j = 7$ whereas for a similar sample correlation in the chalk cutting, j is between 30 and 40. It can be concluded, therefore, that if the loading conditions remain constant for prolonged periods, the observed data cannot be considered independent.

iii. *Cat D400 Spot Times.* Figure 4.12 shows the correlation plot for the Cat D400 spot times that can be compared with that of the Volvo A35 in figure 4.7. At first the conclusion can be reached that observations on the D400 are much more independent than those of the A35 but the way in which the data was taken must be considered. The A35 dumptruck was much more common than the D400 and consequently observations of D400 spot times did not necessarily occur concurrently. The j values in figure 4.12 are therefore not strictly true. It would be difficult to obtain, with the data available, true values of the sample correlation for the spot times.

iv. *Dump Times.* Figure 4.13 shows the correlation plot for this component. This graph indicates that the observed dump times are independent with the highest correlation value being approximately 0.18 and the remainder at or below 0.1.

Figure 4.14 shows scatter plots for (x_i, x_{i+1}) to complement figures 4.7 to 4.13. All show random scattering but a very slight trend can be detected in figure 4.14c, the scatter diagram for Cat 245D bucket swing times.

It has been shown that the observed data is not totally independent. It is also probably true that totally independent data could not be obtained as it seems that the earthmoving system is not as random as was at first thought. It

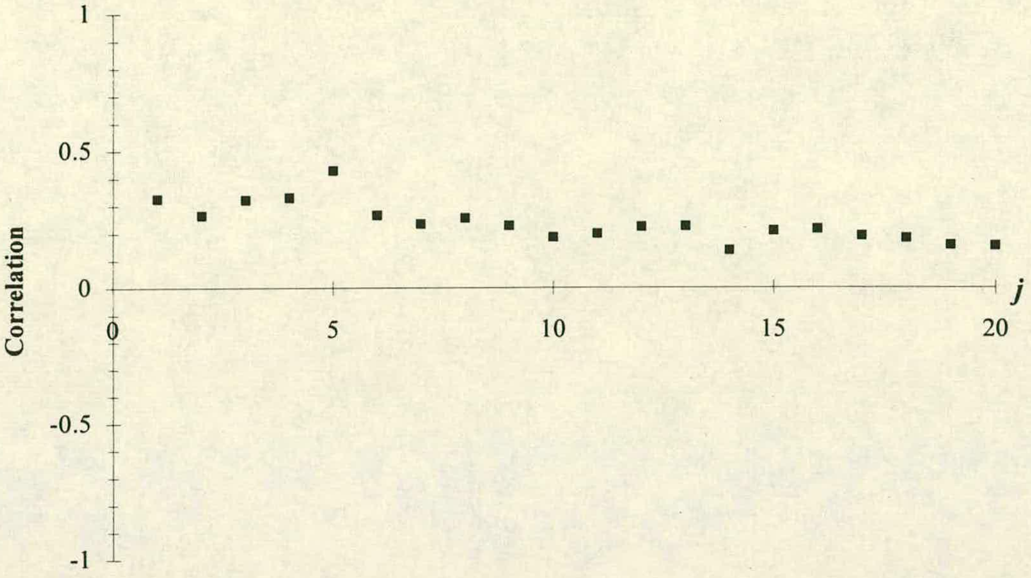


Figure 4.10 Correlation Plot for Cat 245D Bucket Swing Times in Chalk

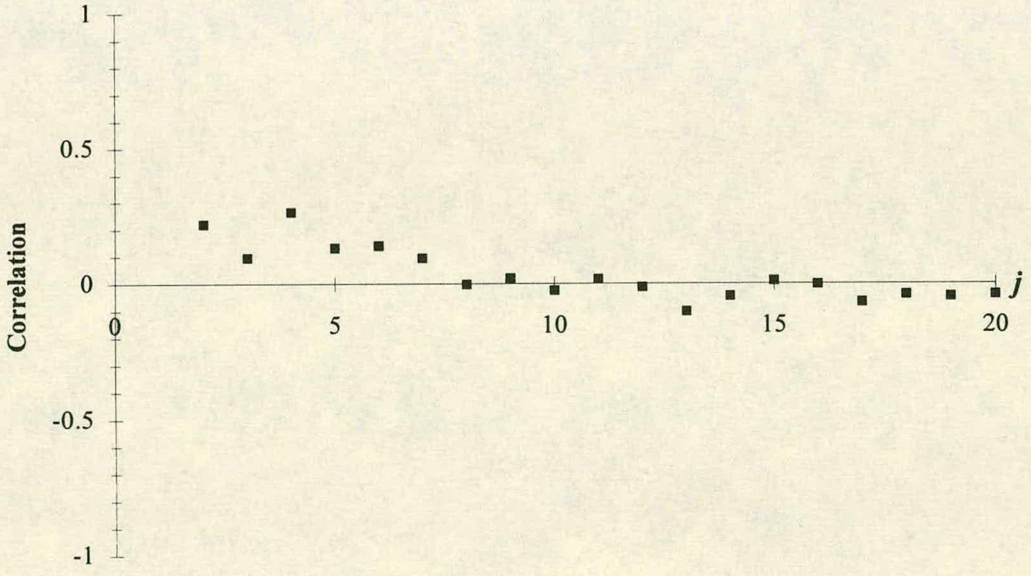


Figure 4.11 Correlation Plot for Cat 245B Bucket Swing Times in Clay

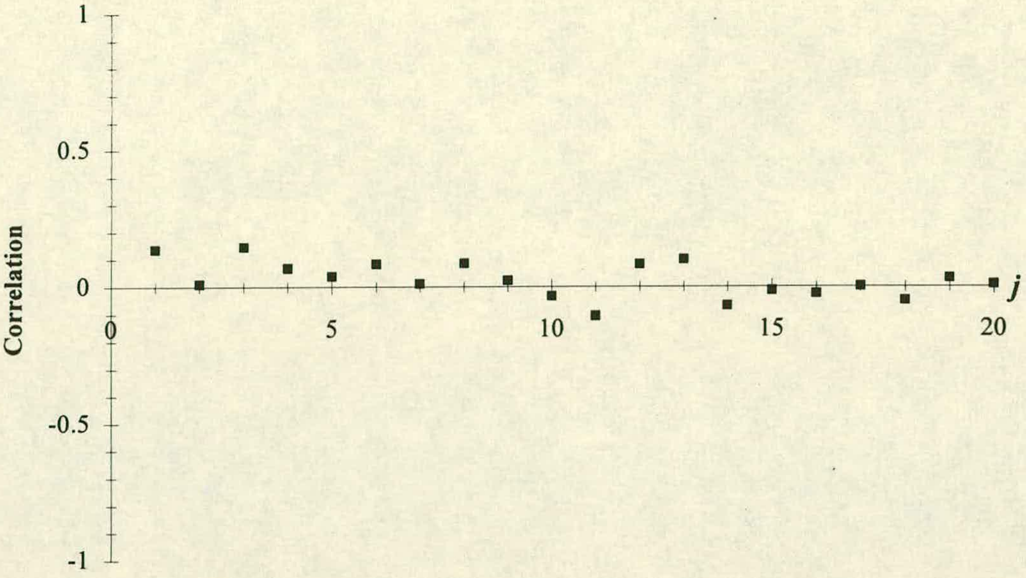


Figure 4.12 Correlation Plot for Cat D400 Spot Times.

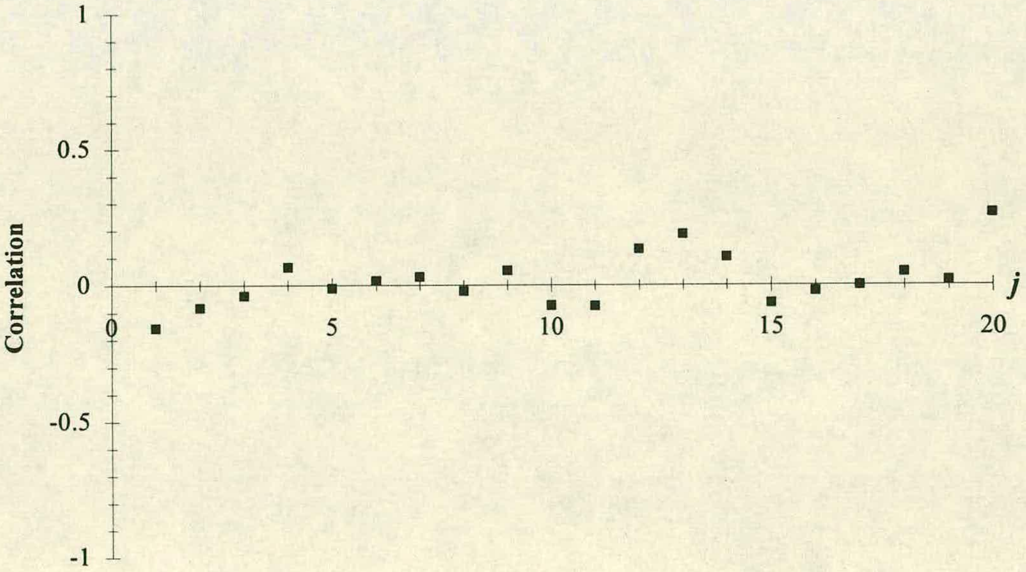


Figure 4.13 Correlation Plot for Dump Times, all vehicles.

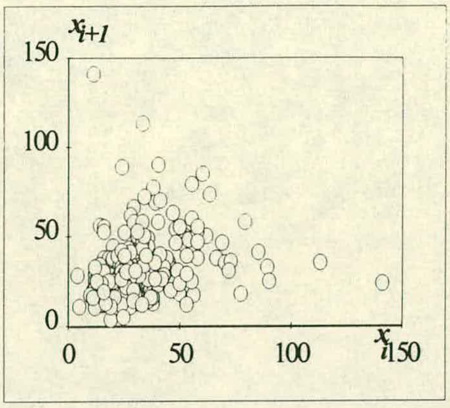


Fig 4.14a A35 Spot Times

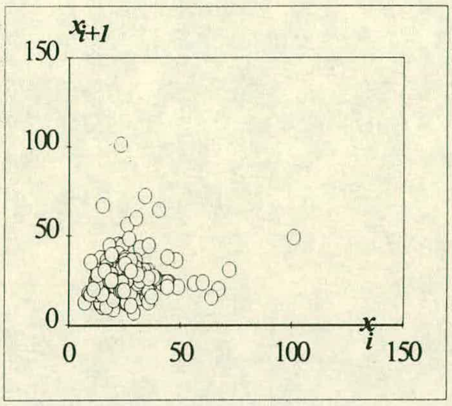


Fig 4.14b D400 Spot Times

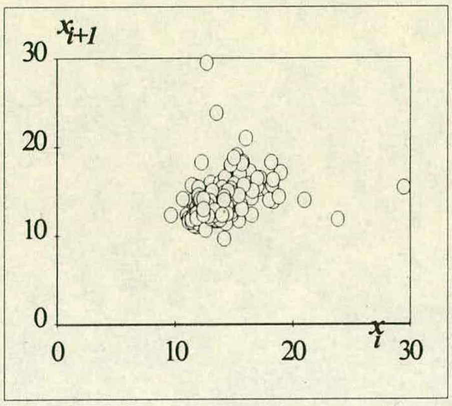


Fig 4.14c 245D Bucket Swing Times in Chalk

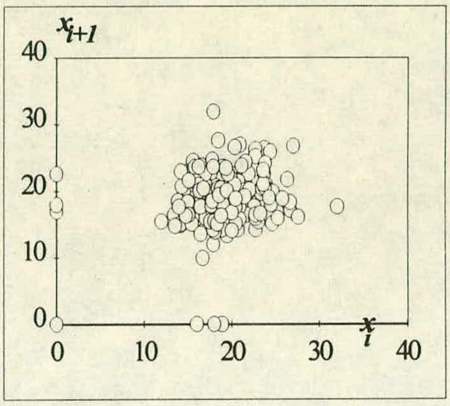


Fig4.14d 245B Bucket Swing Times in Clay

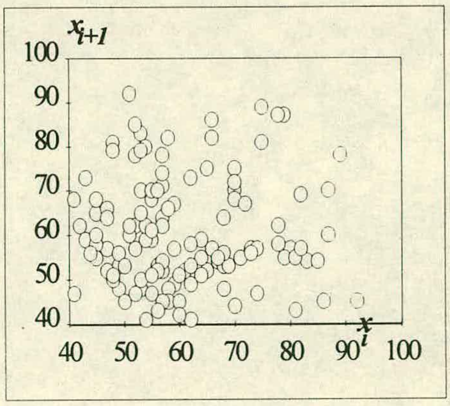


Fig 4.14e Dump Times for all vehicles

could be concluded at this stage that simulation is therefore an unsuitable method to use but, if actual operations are not as random as their simulated counterparts then conservative estimates of production would be obtained. It remains to be seen how close simulated output data is to the real data and it must be remembered that it is only a model that is being worked with.

4.3.3 Theoretical Probability Distributions

Probability distributions can now be selected to fit, as well as possible, the observed cdfs in figures 4.1 to 4.6. There are many distributions available, many of which can be discounted immediately. Law and Kelton (1991), Hahn and Shapiro (1967) and Lawless (1982) provide reviews of most of the continuous probability distributions that can be used in simulation. Some of these are discussed below.

i. *Uniform.* This is the distribution for data that varies uniformly between two values *a* and *b*. The U(0,1) distribution is the basis for generation of random values of other distributions. (See 4.4).

ii. *Exponential.* This is the pdf that is used to represent the interarrival times of customers to a system, usually from an infinite source. It is not applicable to any of the component times in the earthmoving system.

iii. *Gamma.* Usually used to represent the time taken to complete some task and is therefore applicable to the earthmoving system. It is represented by two parameters: α, a shape parameter and β, a scale parameter. For positive integer values of α the gamma distribution is known as the *Erlang* distribution.

iv. *Normal.* This distribution is easy to recreate as only a sample mean and sample variance are required to represent it. However, this simplicity means that it cannot be 'fine tuned' by use of upper and lower limits and the shape of the distribution is symmetrical; hence it is perhaps unsuitable for the asymmetrical distributions shown in figures 4.1 to 4.6.

Other distributions, such as *beta, lognormal* and *Pearson* are available. These have similar shapes to the gamma distributions but have certain characteristics that are useful in certain situations. In deciding which distribution to use, reference has been made to the work undertaken by Carmichael (1986, 1987a) who is one of the only authors to have studied the queueing nature of earthmoving operations. He has made extensive use of the Erlang distribution that can be altered to fit to an almost infinite number of shapes. The observed cdfs will therefore be initially compared with Erlang distributions.

4.3.4 Erlang Distribution

The Erlang distribution is named after a Dutch telephone engineer who developed the distribution in the study of a telephone queueing network. It is a continuous random variable (*T*) with a density function *f*(*t*) defined by two parameters, R and k, the scale (or Rate) parameter and the shape parameter. (k must be a positive integer). If R and k are known then the probability density function is given by:

$$f(t) = \frac{R(Rt)^{k-1} e^{-Rt}}{(k-1)!} \quad (t \geq 0) \tag{4.6}$$

The mean and the variance of this distribution are given by:

$$\mu = \hat{\mu} = E(\mathrm{T}) = \frac{k}{R} \tag{4.7}$$

$$\sigma^2 = \hat{s} = \mathrm{var(T)} = \frac{k}{R^2} \tag{4.8}$$

The actual mean and variance of the distribution can be estimated by the sample mean and variance taken from the observed site data. Therefore, to represent those distributions shown in figures 4.1 to 4.6, µ\$ and s\$ taken from the actual sample of times can be used to provide estimates for the distribution parameters. The cumulative distribution function for the Erlang distribution does not exist in a closed form so it cannot be plotted directly. Therefore a different approach is needed if the observed and theoretical distributions are to be compared. One way is to generate numbers from the distribution for the values of k and R in question and build up the cdf from them in the same way that they were done for the observed data. To do this, a random variate generator is required and although these will not be discussed until section 4.4 this has been done for the observed distributions shown in table 4.2. Table 4.2 also includes the Erlang distribution parameters.

4.3.5 Comparison of Observed and Theoretical Distributions

Visual fits between the observed and theoretical (generated) distributions can be made with figures 4.15 to 4.20. In some cases these fits are quite poor but others, (that of bucket swing times in chalk for example, figure 4.18) the fit is very good. A visual inspection of the curves cannot reliably indicate just how good the fit is and therefore, other methods must be employed.

i. *Probability Plots.* If comparisons had to be made with a straight line, this task would be somewhat easier. This is possible using probability plots (Law and Kelton (1991), Barnett (1975) and Hahn and Shapiro (1967)). Two methods will be used here which work by normalising the plots of the CDFs by using *quantile-quantile* (*Q-Q*) or *probability-probability* (*P-P*) plots. Referring to figure 4.21, for each quantile q on the ordinate, there is a value t for both the observed and the theoretical distributions. If these two values of t are plotted for all quantiles then the result will be a straight line through the origin with a slope of one if the fit is perfect. The amount of deviation from this line will indicate how good the fit is. The *P-P* plot is constructed in the same way but this time for the theoretical probability verses the observed probability at each value of t. To do this, $F(t)$ needs to be calculated for values of t for each distribution; unfortunately, $F(t)$ does not exist in a closed form for the Erlang distribution and so cannot be calculated directly for values of t. Alternatively, the actual plots of the cumulative distribution functions can be used to provide values of the quantiles and probabilities for the two curves being compared. This has been done and figures 4.22 to 4.26 show the *Q-Q* and *P-P* plots for the pairs of distributions shown in figures 4.15 to 4.19. The probability plots for the spot times clearly indicate that the fit is not good, particularly in the middle of the range. The best fits occur with the bucket swing times; the plots in figures 4.24 and 4.25 agree very well with the straight line. Figure 4.26 shows the plots for the dump time and while the plots tend to deviate from the straight line, the trend is there and discrepancies will be due, in part, to the small amount of dump time data available. Probability plots are a heuristic method of comparing two distributions but they still do not indicate formally whether or not a proposed theoretical distribution should be used to replace random observations. To do this, statistical hypothesis tests should be used if possible to assess whether the values $t_1, t_2, \ldots, t_i$ are a sample from some particular distribution.

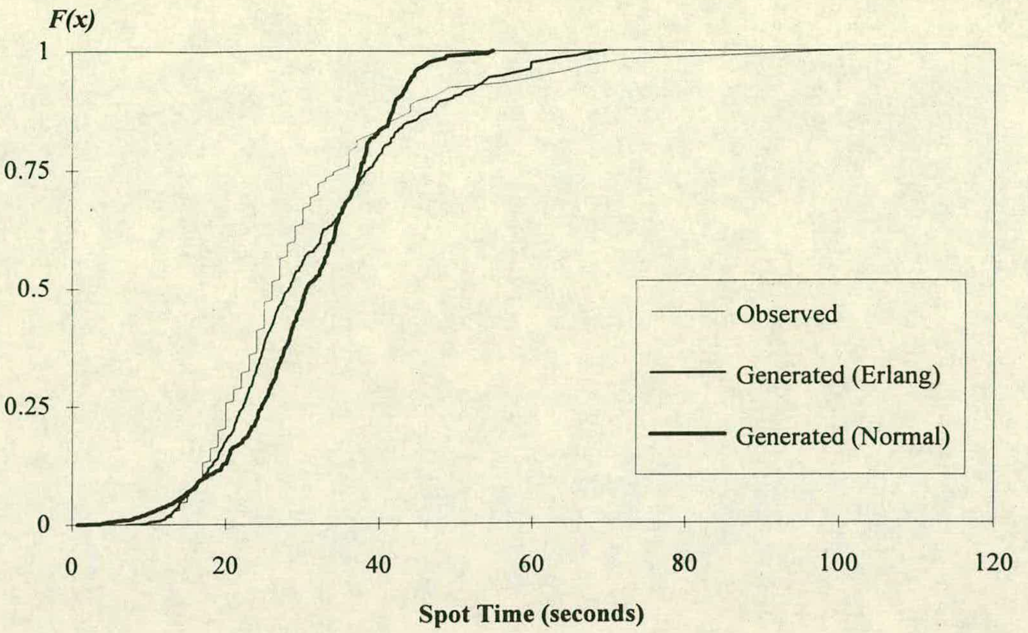


Figure 4.15 Cumulative Distribution Function for Spot Time for Cat D400D.

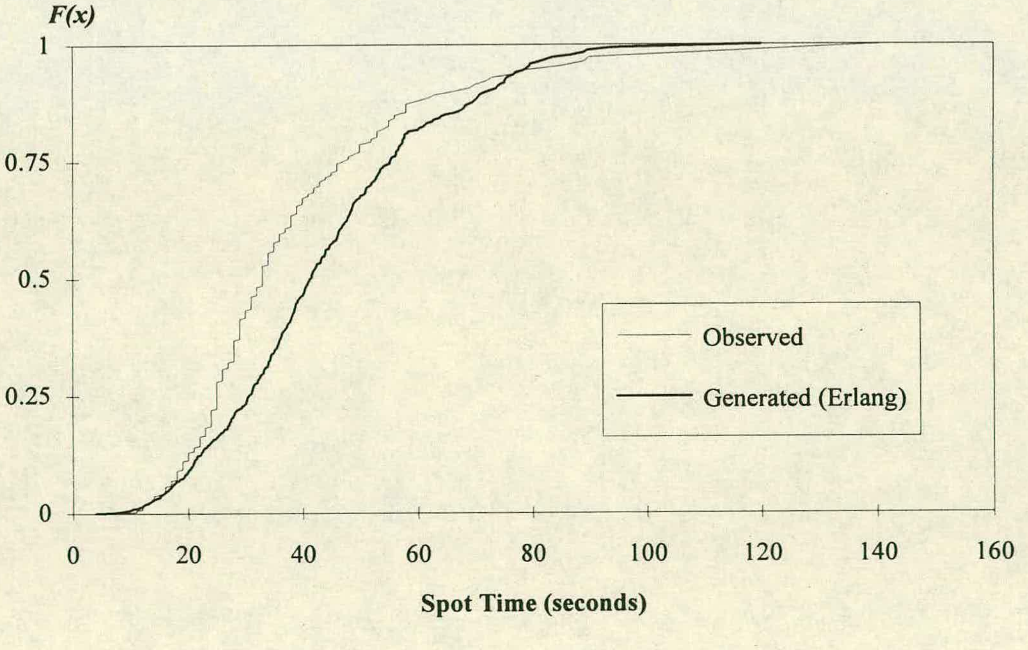


Figure 4.16 Cumulative Distribution Function Spot Time for Volvo A35.

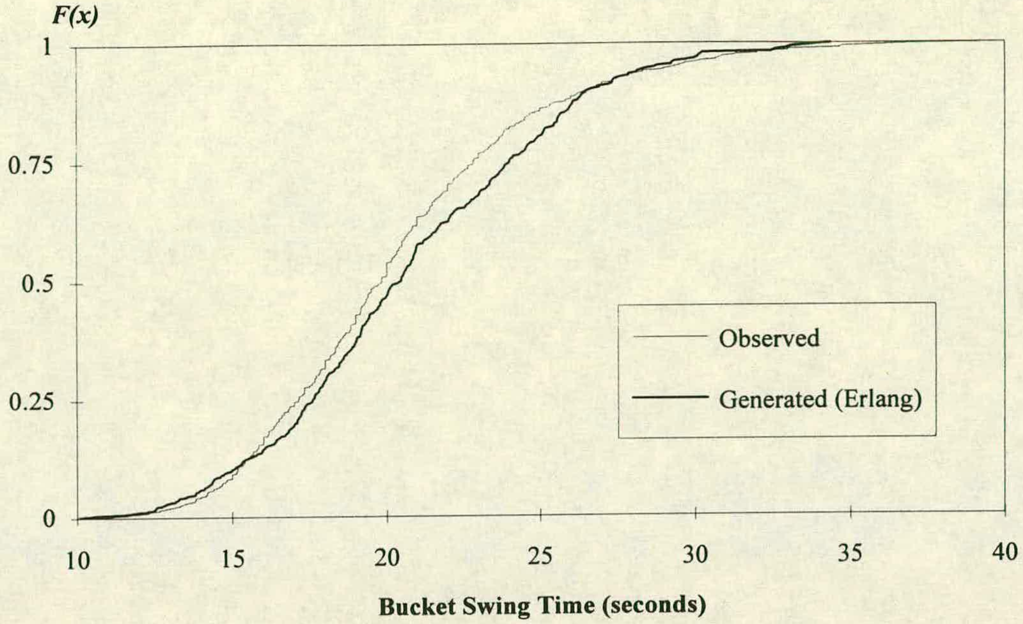


Figure 4.17 Cumulative Distribution Function for Bucket Swing Time for Cat 245B in Clay.

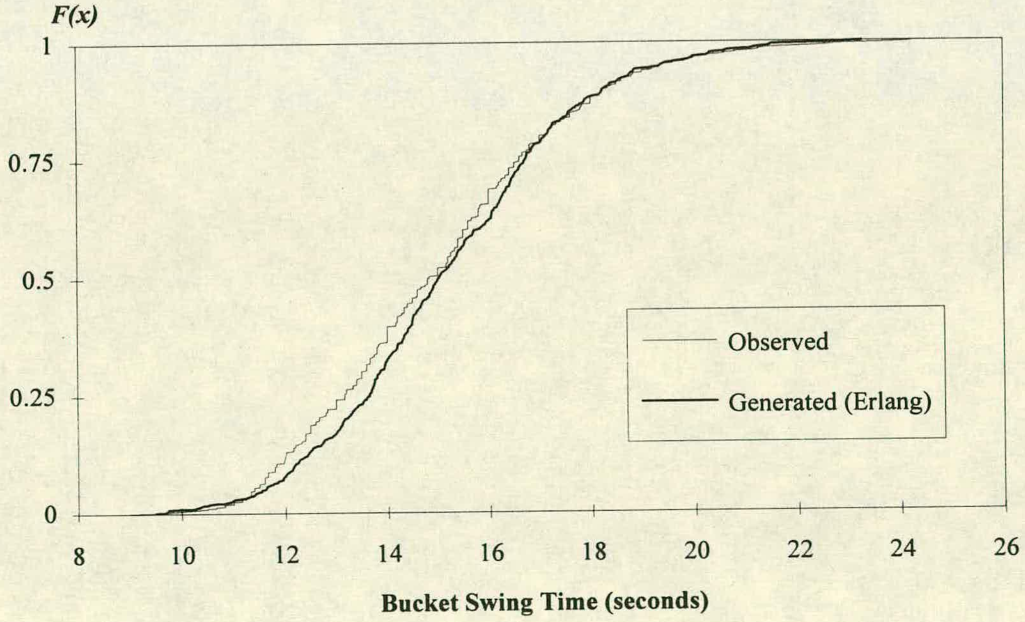


Figure 4.18 Cumulative Distribution Function for Bucket Swing Time for Cat 245D in Chalk.

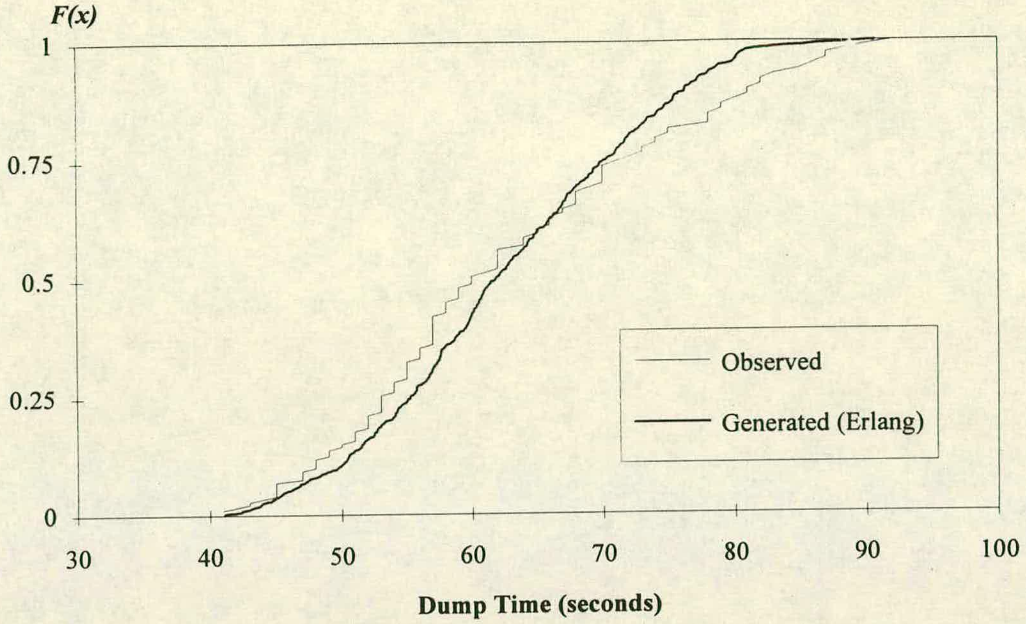


Figure 4.19 Cumulative Distribution Function for Dump Time all vehicles.

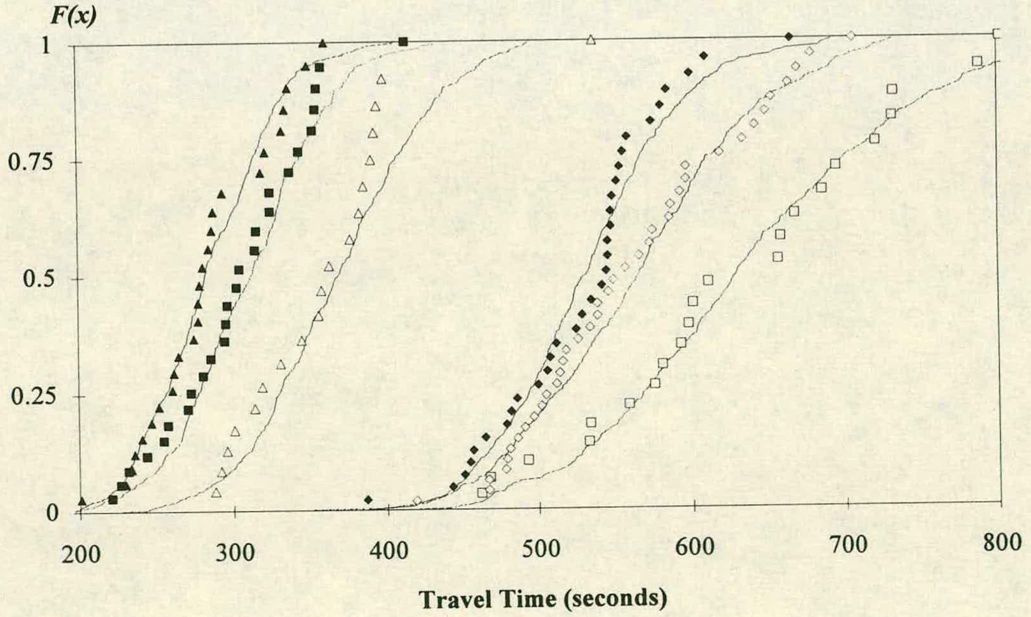


Figure 4.20 Cumulative Distribution Function for Travel Time for all vehicles on chalk haul roads.

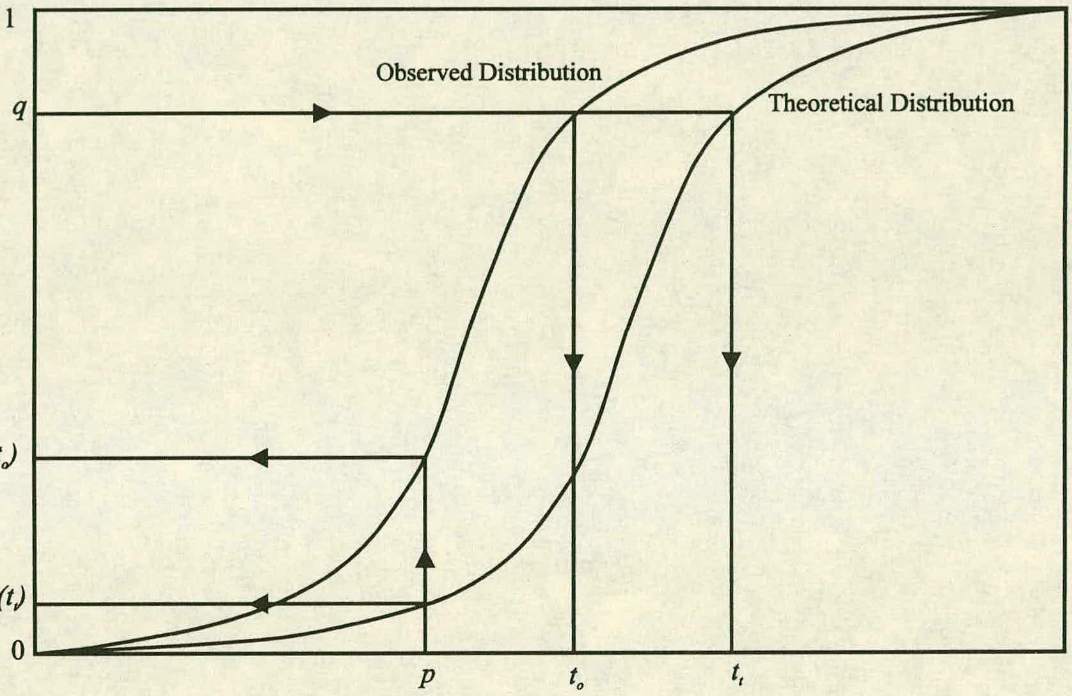


Figure 4.21. *Q-Q* and *P-P* plots

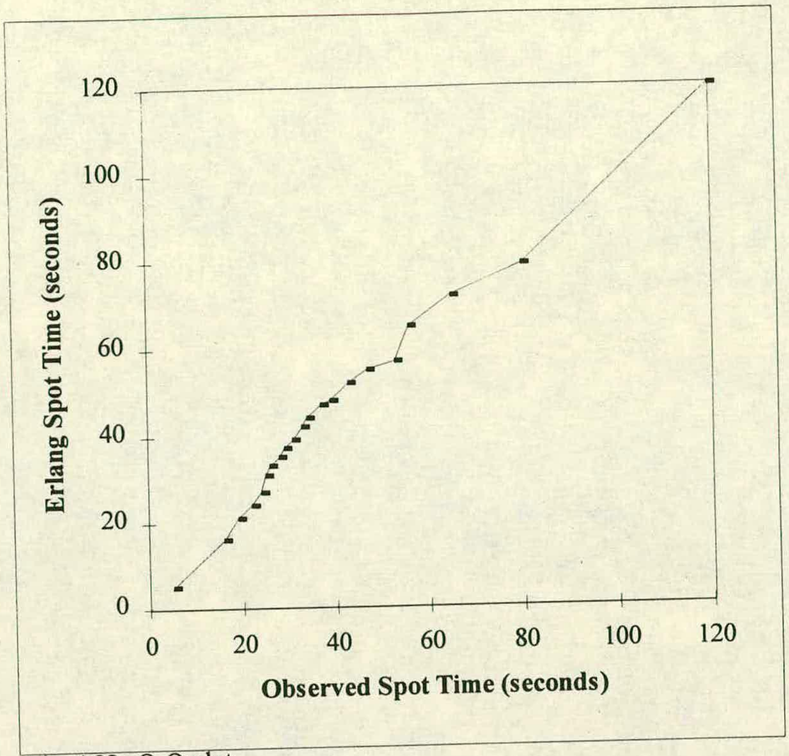


Fig. 4.22a Q-Q plot

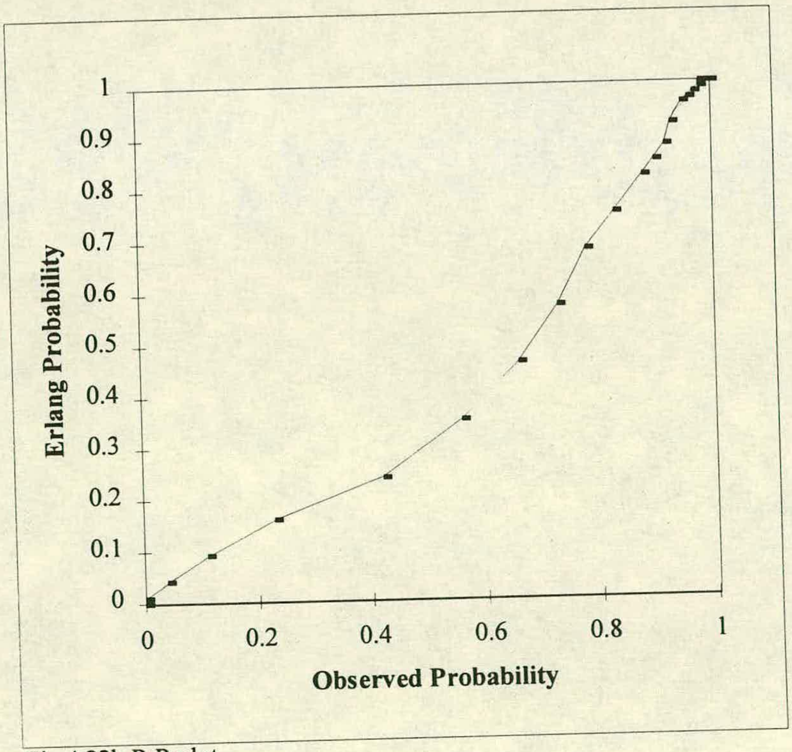


Fig 4.22b P-P plot

Figure 4.22 Q-Q and P-P plots for spot time for Cat D400D

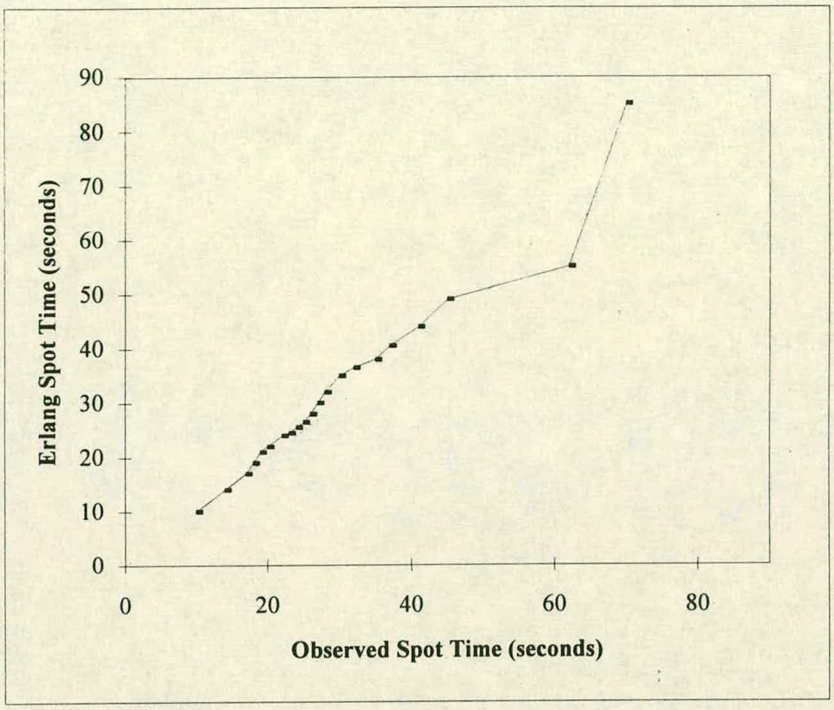


Fig. 4.23a Q-Q plot

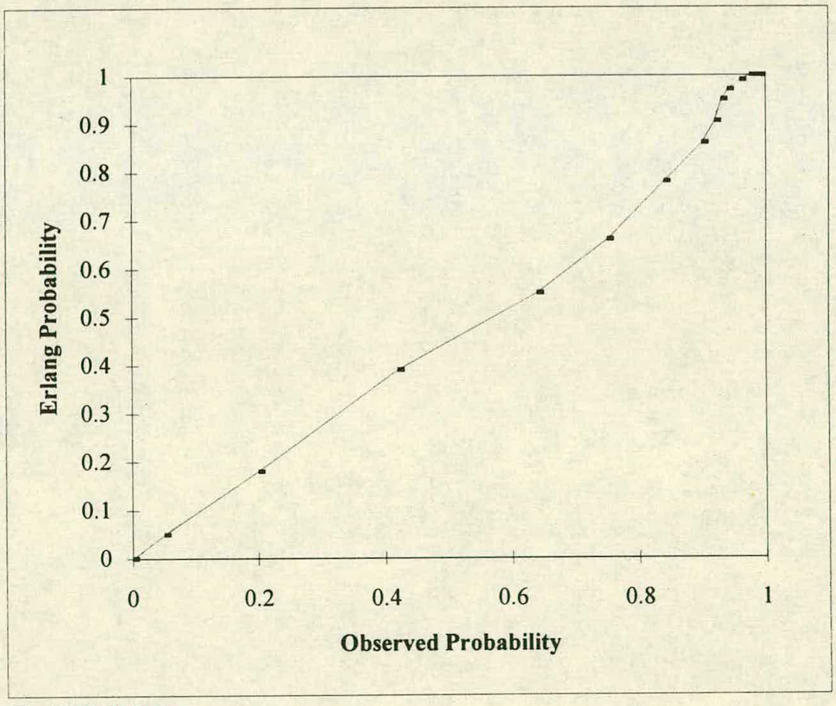


Fig 4.23b P-P plot

Figure 4.23 Q-Q and P-P plots for spot time for Volvo A35

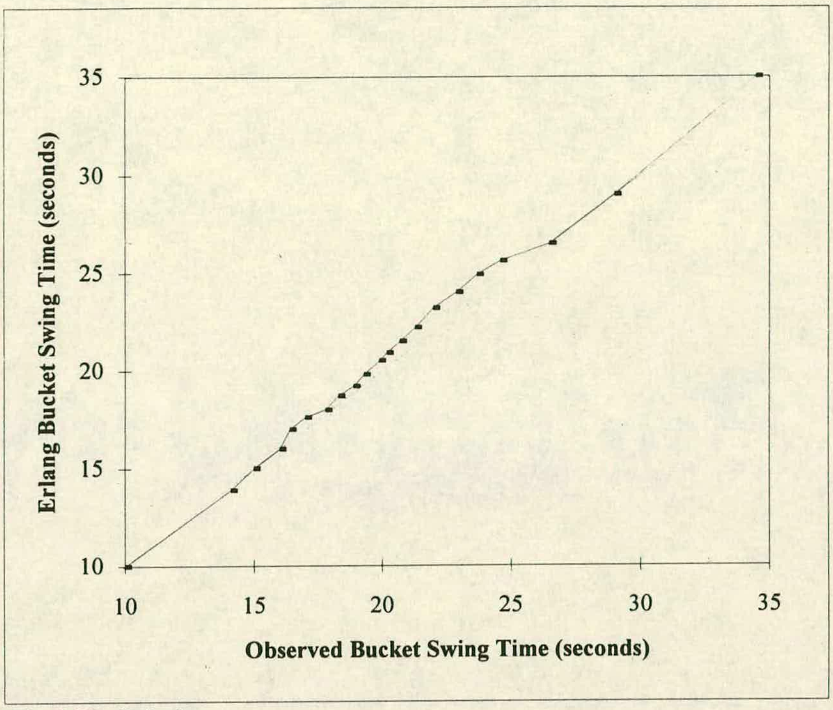


Fig. 4.24a Q-Q plot

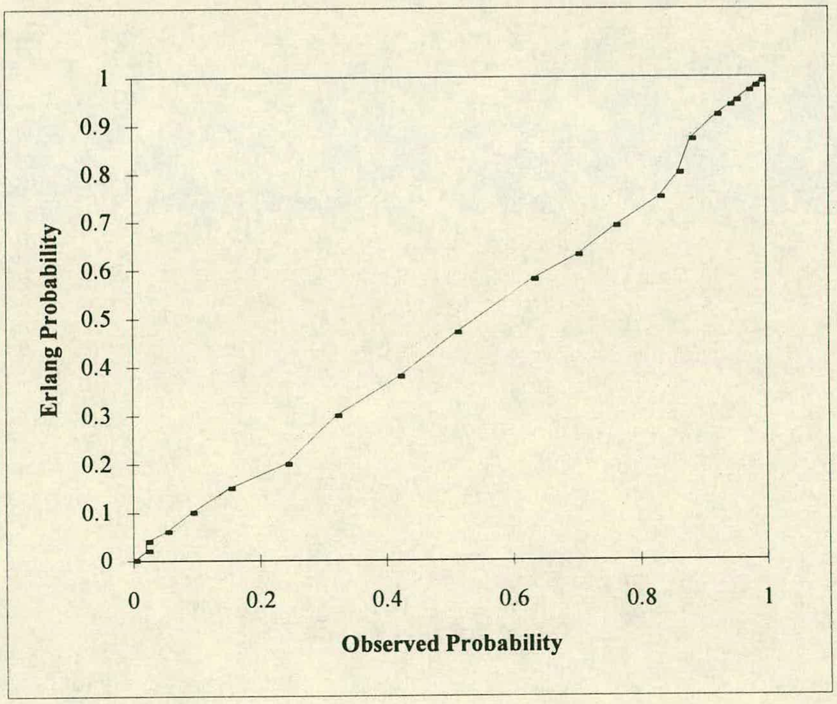


Fig 4.24b P-P plot

Figure 4.24 Q-Q and P-P plots for bucket swing time for Cat 245B in Clay

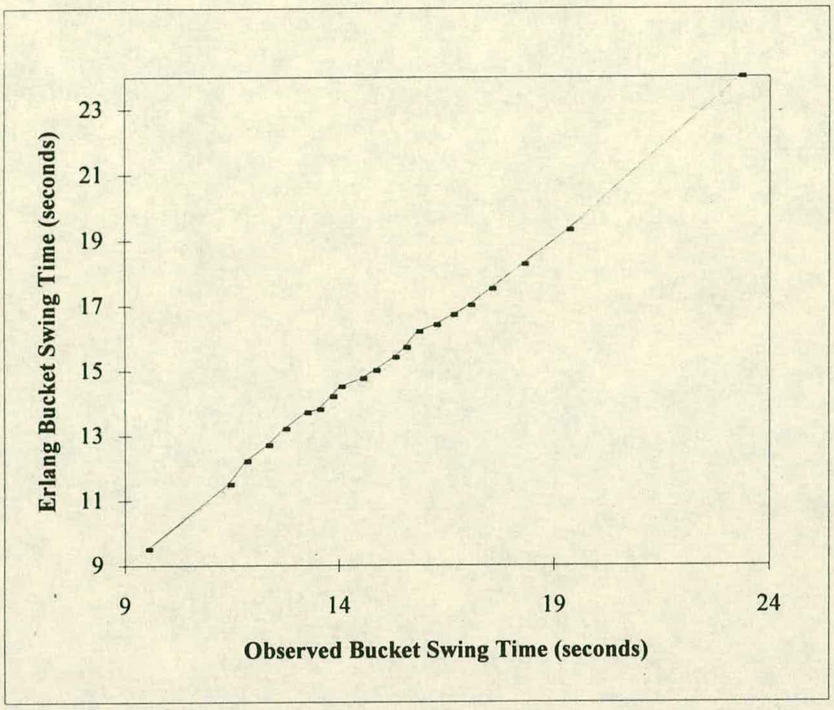


Fig. 4.25a Q-Q plot

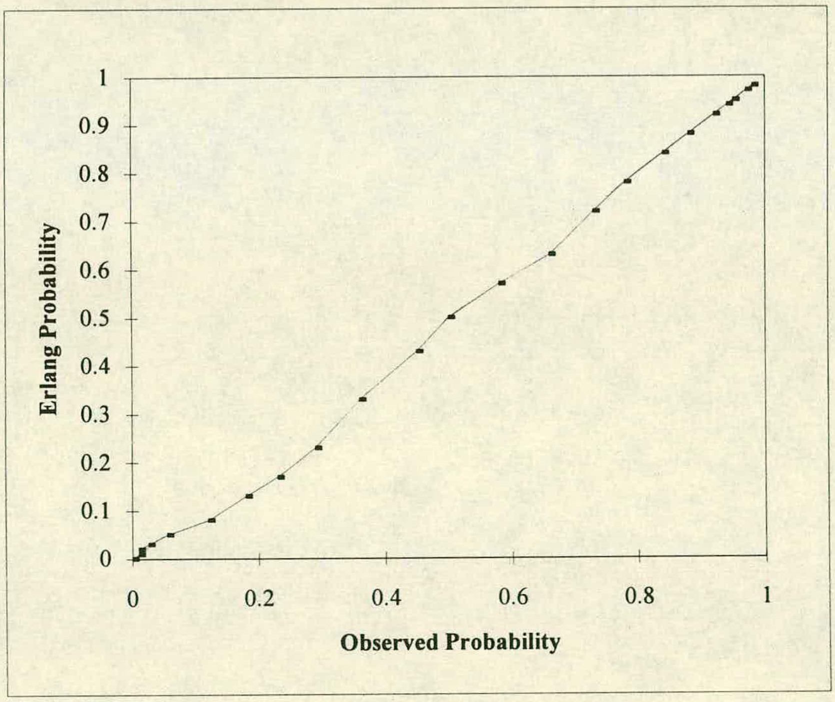


Fig 4.25b P-P plot

Figure 4.25 Q-Q and P-P plots for bucket swing time for Cat 245D in chalk

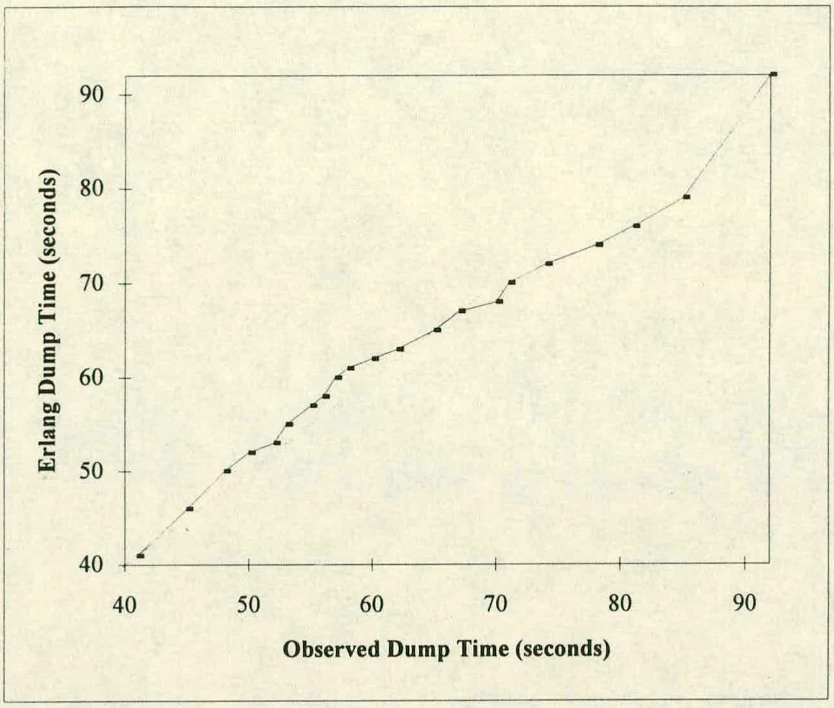


Fig. 4.26a Q-Q plot

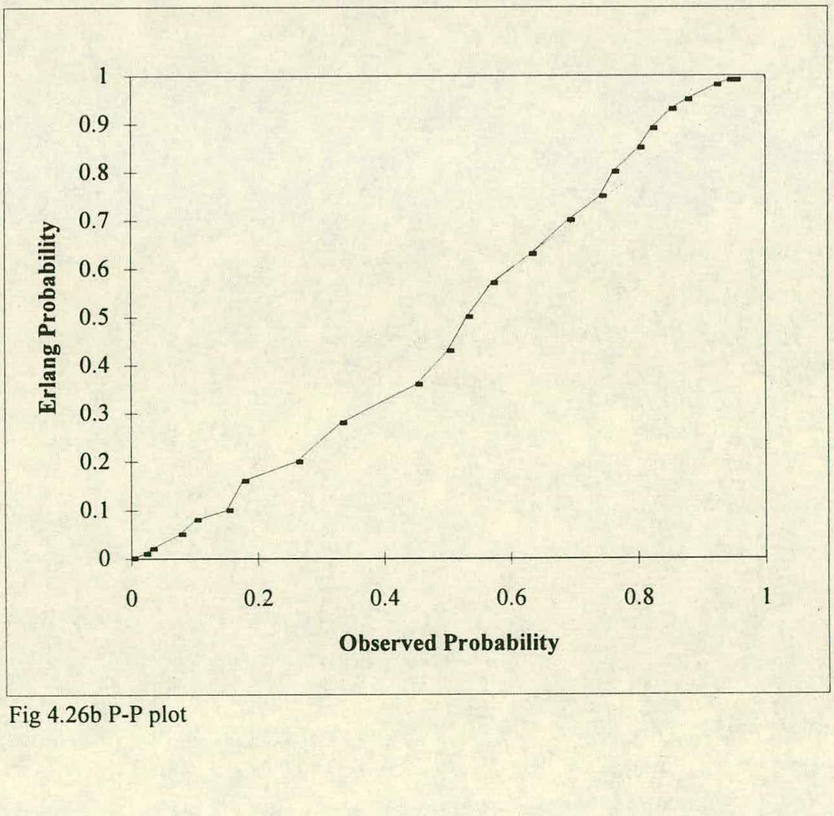


Fig 4.26b P-P plot

Figure 4.26 Q-Q and P-P plots for dump time for all vehicles.

ii. *Chi-Square Goodness-of-Fit Test.* This is perhaps the oldest goodness-of-fit test that was first proposed in 1900 by Pearson. The distribution to be tested is split into k intervals and the number of observations falling into each interval is recorded; N_j is the number of observations in the jth interval. The next stage is to compute the proportion, p_j of observations that would fall into each interval of the fitted, theoretical distribution. For an observed distribution containing n observations, the test statistic is:

$$\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \tag{4.9}$$

since np_j is the *expected* number of observations that should fall in the jth interval. For a perfect fit, χ^2 would be zero and hence we reject the hypothesis that the observed distribution is a sample from the theoretical distribution if χ^2 is too large. Unfortunately, to calculate the proportion of expected results in each interval we need to invert the cumulative distribution function, *F(x)* that, in the case of the Erlang distribution, does not exist in a closed form. Numerical approximations are available for the calculation of *F(x)* (for example, Bhattacharjee, 1970) but it is felt that the amount of computation time required does not justify the results that would be obtained. It can be seen from the plots in figures 4.22 to 4.26 that the Erlang distribution does not give an ideal fit but this distribution is very simple and easy to work with. For an initial simulation model it is ideal. Other goodness-of-fit tests available such as the Kolmogorov-Smirnov tests (Law and Kelton, 1991) which requires the parameters from the assumed distribution to be known (i.e. not estimated from the data) as well as requiring the distribution to exist in a closed form.

4.4 Random Numbers and Random Variate Generators.

At this stage, the method of simulation has been investigated and the form of the distribution of the time components has been selected. The next stage in the development of this simulation model is to form a method of generating the numbers that agree to the distribution chosen (the Erlang distribution in this case). This is a two stage process. Firstly, a series of numbers must be generated which are IID (independent and identically distributed) from the uniform distribution within the range (0,1). The second stage is to generate random *variates* from this uniform distribution using some algorithm. These variates are samples from the chosen distribution that can then be used directly in the simulation.

4.4.1 Random Number Generators

The process of generating random numbers has a long and interesting history but it shall not be dwelt upon here except to point out the features that are important in a generator. A random number generator can either be physical generator, for example ERNIE, the Electronic Random Number Indicator Equipment that uses a pulsating vacuum tube to give random numbers at the rate of 50 per second (Thompson, 1959), or an arithmetic generator. With the power of modern computers physical generators are rarely used any more as arithmetic generators have been developed that are, whilst not exactly random in the sense that the numbers are unpredictable, very good at producing numbers that agree to the IID U(0,1) distribution. An arithmetic generator works, in the simplest sense, by applying some mathematical function to an initial number. The result of this calculation becomes the input to the next calculation and so on to produce the numbers required. The fact that the numbers are predictable if the initial value is known is actually an advantage in stochastic simulation, as shall be seen below. Because there are so many different generators available it is important to choose the right one; Law and Kelton (1991) proposed the following properties for generators to be used in simulation studies:

i. Numbers produced by the generator should be uncorrelated and appear to be distributed uniformly on (0,1).
ii. The generator should be fast and not require large amounts of storage.
iii. The numbers produced should be able to be reproduced for two reasons. Firstly, it makes debugging and verification of the computer code used easier but more importantly, the same simulation can be repeated if so desired or, the same random numbers can be used in comparing different systems to provide more precise comparisons.
iv. Ideally, the generator should be able to provide separate 'streams' of random numbers which are separate subsegments of the string of numbers produced. A stream can then be allocated to the generation of different variates (for example, one stream for spot time and another for travel time). Whilst this feature is ideal, it is not provided for in the generators that have been used in this study and so a separate generator would have

to be obtained or coded. It was felt that this was unnecessary in the initial modelling.

The random generators used are those provided (i.e. the 'RAND' functions in the spreadsheet and language compiler used) and as such comply with the second and third points above. The first and most important aspect of a generator is its ability to produce the IID numbers required and this ability must be determined before any results of the simulations are used. This has been done (see Appendix 1) and it can be said with a high degree of confidence that the generators used are satisfactory.

4.4.2 Random Variate Generators

The previous section outlined what is the basic ingredient of all stochastic processes, the generation of an IID U(0,1) distribution. These numbers are then input to an algorithm that produces numbers that agree to the distribution desired. This algorithm is known as a random variate generator and there are many that can be used depending on the type of distribution to be modelled and the degree of exactness required. Some generators are also more efficient than others and this is an important feature in large scale simulation studies. The algorithm chosen for the study described here was one of the first to be investigated and works extremely well; it was decided therefore that a broad study of the various types was not warranted. There is however, a large range of published work concerning the generation of random variates notably Dagpugnar (1988) and Devroye (1986).

The simplest generator is the *inverse-transform* method that works by first generating a number distributed as U(0,1) and then returning the variate as the inverse of the distribution function *F(x)* of the distribution required. This method will not work for distributions that do not have a distribution function that exists in a closed form. Therefore, other methods have to be looked at.

i. *Convolution Method.* This method is applicable to those distributions that can be expressed as the sum of a series of other random variables that may be easier to generate. This is the case for the Erlang distribution. An Erlang distribution with mean μ and shape factor k is actually the sum of k exponential random variables with a mean of μ/k. (Therefore, an Erlang distribution with $k = 1$ is the exponential distribution). Exponential random variables can easily be generated by the inverse-transform method and so an Erlang variate can easily be 'built' up. For an exponential variable with mean μ/k, the probability density function is:

$$F(x) = 1 - e^{-kx/\mu} \tag{4.10}$$

To generate an exponential variate with the inverse-transform method, a number $u = \mathrm{U}(0,1)$ is generated and set as $F(x)$. Rearranging (4.10):

$$1 - u = e^{-kx/\mu} \tag{4.11}$$

and taking natural logarithms of both sides:

$$\ln(1-u) = \frac{-kx}{\mu} \tag{4.12}$$

For a random $u = \mathrm{U}(0,1)$, u-1 is also random. Therefore:

$$x = \frac{-\mu}{k}\ln(u) \tag{4.13}$$

Therefore, for an Erlang distribution with shape factor k, the variate X is calculated using:

$$X = \sum_{i=1}^{k} \frac{-\mu}{k}\ln(u_i) = -\frac{-\mu}{k}\ln\left(\prod_{i=1}^{k} u_i\right) \tag{4.14}$$

where $u_1, u_2, \ldots, u_m$ are IID U(0,1) random variates.

ii. *Acceptance-Rejection Method (ARM).* This method can be used for those distributions that are defined over a finite range. Although this is not technically the case with the Erlang distribution, finite limits must be set for the range over which the actual distribution is expected to fall. If this range is that of the observed distribution then inaccuracies may occur as will be seen below. Equation 4.6 defined the probability function for the Erlang distribution:

$$f(t) = \frac{R(Rt)^{k-1}e^{-Rt}}{(k-1)!} \quad (t \geq 0) \tag{4.6}$$

This is now used in the following method to generate variates:

Step 1. A constant M is calculated such that M is the largest value of (4.6) over the interval [a,b].

Step 2. Two U(0,1) numbers are generated, r_1 and r_2.

Step 3. Calculate $t^* = a + (b - a)r_1$.

Step 4. Evaluate $f(t^*)$.

Step 5. If r_2 is less than or equal to $f(t^*)/M$ then accept t^* as a random variate. Otherwise, reject and go back to step 2.

The above are the two methods that have been investigated to generate Erlang random variates and each have advantages over the other. Figure 4.27 shows a comparison of the cdfs generated by the two curves for D400 spot times. They are very similar except for the tails; the reason for this is that the curves generated by ARM are controlled to some extent by the upper and lower bounds. One could argue that this curve is therefore not a true Erlang distribution and that the one generated by convolution should be used; however, if the limits are chosen carefully, that is, allowing for the small chance of actual times falling outside the range of those times observed, then the ARM will be a suitable algorithm to use. Figure 4.28 shows how the curves differ if the wrong limits are chosen. This is for dump times and the ARM uses an upper limit of 92 and a lower limit of 41. The convolution method actually generates numbers in the range (29, 105) so causing the difference in the curves. In this case the convolution generated curve is probably the best fit but unfortunately this took much longer to generate than the one in figure 4.27; the dump times have a shape factor of $k = 30$ and so 30 random numbers on U(0,1) must be generated for every one variate. The method is fine for distributions with low shape factors (the distribution in figure 4.27 has $k = 5$) but for some of the travel time distributions, k can get very large. For example, referring to table 4.2, the observed distribution for a haul length of 1600 m has a shape factor of 89. Generating such a distribution by the convolution method would be extremely time consuming. Therefore, although the ARM is not ideal it is the algorithm that will be used in the simulation studies.

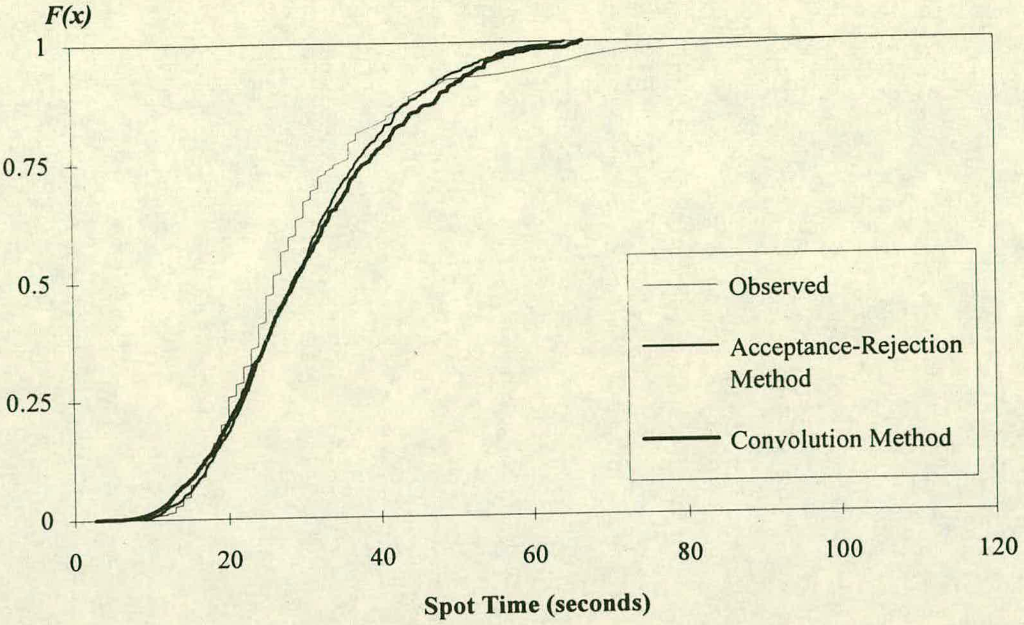


Figure 4.27 Comparison of Acceptance-Rejection and Convolution methods of generating Random Variates.

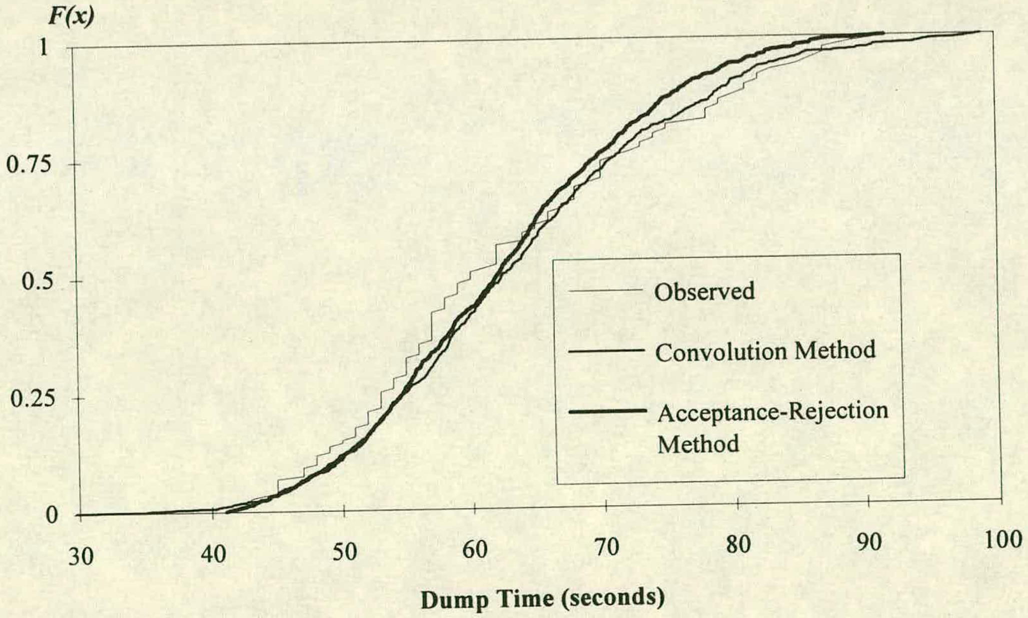


Figure 4.28 Comparison of Acceptance-Rejection and Convolution methods for generating dump times.

4.5 Conclusions

This chapter has been devoted to the development of a model of the earthmoving system to be analysed using simulation techniques. The next chapter will deal with two methods of carrying out the simulations but before that, the conclusions obtained from this chapter will be outlined. Words in italics are important keywords and were discussed in more detail in the body of the chapter.

1. The model of the earthmoving system is *dynamic* (i.e. it models the system as it evolves), *stochastic* (i.e. random as opposed to deterministic; model one in chapter 3 is deterministic) and *discrete*, that is the *state variables* (those which describe the state of the system; see section 4.2.2) change at discrete points in time.
2. A dynamic, stochastic and discrete model can be analysed using *discrete-event simulation*. Such a technique works by advancing from event to event, updating statistics about the system until a pre-set stop condition has been met. The events used in the earthmoving system are *arrivals* at the queue and *departures* from the loader. The time advance mechanism used is *next-event time advance* where the time and type of the next event is selected from an *event list* and processed accordingly.
3. The observed data taken from the sites investigated was looked at to determine if there is any *dependence* of the data. If a stochastic analysis of the data is to be carried out then this method would not be completely valid if the data were not completely random. *Correlation* was used via *correlation plots* and *scatter diagrams* and some degree of dependence was found, especially in the bucket swing times. Therefore, independent input data is another assumption of the model.
4. Probability distributions were determined for the observed data. *Probability plots* showed that the *Erlang* distribution provided a reasonable fit for all time components and is ideal in that the parameters for these distributions are easy to determine. Other distributions could have been investigated (such as lognormal or Weibull) but the parameters for these distributions would have to be determined using the more complicated method of maximum likelihood estimation.
5. Two types of *random variate generator* (i.e. an algorithm producing numbers corresponding to a desired distribution) were investigated and it was found that the better of the two was the *acceptance-rejection method* as opposed to the *convolution* method. A pre-requisite of a good variate generator is a *random number generator* that returns numbers that are independent and identically distributed on the uniform distribution over the range (0,1). The chi-squared

goodness-of-fit test was used to determine whether this property existed in the generator that came with the computer used and the result was positive. (See Appendix 1).

6. The discrete-event simulations of the model can be carried out using both a spreadsheet package and a dedicated program.

Chapter 5

Simulation of the Earthmoving System

Two methods for carrying out discrete-event simulations of the type described in chapter 4 are discussed, namely by computer spreadsheet and by dedicated computer program. The transient and steady-state behaviour of the simulation output is investigated and the use of 'Production Charts' for estimation purposes is outlined.

5.1 Introduction

This chapter can be split into two parts. Firstly, the methods used to carry out the simulations will be discussed. The first method, using a spreadsheet, is impractical from a user's standpoint but it indicates much about how the process works - especially in terms of *transient* and *steady-state* output (see section 5.5). The second method is by use of a specially written program in the language C that provides a much quicker and easier route to simulation results; this is essential if the model is to be validated and experimented with.

The second part of this chapter investigates the transient and steady-state behaviour of the simulation output in detail - procedures are used to determine the transient length of the output, in effect, how long an earthmoving operation will take before full production is achieved.

5.2 Simulation via a Computer Spreadsheet

The method of discrete-event simulation as described in section 4.2.2, is best applied using a dedicated computer program. To do this requires knowledge not only of the simulation process but of the language to be used for the computer code. For these reasons it was decided to carry out initial simulations using a spreadsheet package. The simulations were anticipated to be cumbersome to carry out and to take a long time (the random variates would have to be generated beforehand and pasted into the spreadsheet) but it was felt that this method would be, and indeed turned out to be, very constructive. To explain how to spreadsheet simulation works, an example shall be used. The input data for this example is taken from an actual operation.

5.2.1 Case study example

The operation from which the data for the case study comes from was a single loader operation on the M1/A1 contract. The excavator, (a Cat 245B) was working in an approximate 1.5m depth bench but with a fairly tight truck manoeuvring area. The haul length was 800m with an average travel time of 311 seconds. Table 5.1 shows the observed means and variances of the time component distributions together with estimates of the parameters of the Erlang distribution that will be used to represent the actual distributions. Note that the load cycle comprised an average of 8 buckets with a total volume loaded into the D400 dumptrucks of approximately 15 bank cubic metres. Therefore the load cycle time is equal to (8-1)x18 seconds - an average of 126

	Observed Average (secs)	Observed Varience	Erlang Parameters		Bounds	
			k	R	Lower	Upper
Spot Time	37	235	5	0.1462	20	100
Bucket Swing Time	18	20.3	16	0.8889	10	45
Dump Time	61	107.5	35	0.5672	30	100
Travel Time	311	2476	39	0.1256	220	450

Notes:

Load Time based on 8 Buckets. Therefore average load time = 7 x 18 = 126 seconds
Haul length = 800m
Hauler units: 3 no. Cat D400 articulated dumptrucks
Loader unit: 1 no. Cat 245 backhoe excavator

Table 5.1 Parameters to be used in case study example, (section 5.2.1) seconds. For every load time variate required, 7 separate bucket swing time variates must be generated and summed.

Two hundred separate variates were generated for the four time components. This was done using the acceptance-rejection algorithm (section 4.4.2 ii) and can be easily (but slowly) done using a separate spreadsheet. The problem with using a spreadsheet to carry out an ARM calculation is that if the generation is rejected, it must be removed from the spreadsheet by hand after the sheet has finished calculating. A more complex spreadsheet package (such as *Microsoft Excel*) can be used to build macros for such a calculation but a far more efficient method is to write a dedicated computer program to do the task. A file handling facility can be incorporated to allow the generated variates to be stored and then opened as a spreadsheet file.* This has been done (see Appendix 2) and it is this program on which the variates were generated for this example. The 800 variates were pasted into the spreadsheet that calculates almost instantaneously the equivalent of over 10 hours of operation time. Part of this spreadsheet can be seen in table 5.2.

The initial conditions of the simulation are very important in that they affect how the operation proceeds in the first few cycles. In this example (and indeed in all simulated operations that will be carried out) the initial conditions are as follows. The loader is initially idle and the number of trucks in the simulation all start in the queue at time = 0. Therefore all trucks have an initial arrival time of 0; consequently, the status of the loader changes instantaneously from idle to busy at t = 0. The statistics are all set to zero - for example, the number of cycles completed and the total idle time of both the trucks and loaders. Alternative initial conditions could be that the trucks do not all start at the queue at time = 0, but each truck could arrive at the queue at some random time after t = 0.

Simulation by this method is not strictly discrete-event. D-E progresses through a simulation run from event to event calculating the statistics (such as production and utilisation) as the events occur. This would be difficult to carry out on a spreadsheet as calculations are calculated instantaneously (or as near as the speed of the computer allows). Therefore, the spreadsheet makes the assumption that all trucks arrive at the queue in the same order and each row

---

* At this stage, the question 'why was a dedicated simulation program not written in the same way as the variate generator was?' The answer to this is that the spreadsheet was ideal in learning how simulation itself worked. Writing a simulation program from scratch would have been far more difficult if the spreadsheet was not used.

Simulation Sheet for 3 Trucks

Haul Length = 800m Loader : Cat 245 Hauler : Cat D400

All times in seconds unless otherwise stated. Erlang random variates shown shaded.

Truck No.	No. Cycles	Arival Time	Cycle Time	Queue Time	Spot Time	Load Time	Departure Time	Loader Status	Loader Idle Time	Dump Time	Travel Time	Next Arrival	Cycles per Hour	Cum. Time (hh:mm:ss)	Cumulative Production (m3)	Prod. m3/hr
1	0	0	\\\	0	20	150	171	busy	0	50	411	631	0.00	00:00:00	0	0.00
2	0	0	\\\	171	35	142	348	busy	0	68	297	713	0.00	00:00:00	0	0.00
3	0	0	\\\	348	34	148	530	idle	101	76	390	995	0.00	00:00:00	0	0.00
1	1	631	631	0	44	161	836	busy	0	78	297	1212	5.70	00:10:31	15	85.57
2	2	713	713	123	27	157	1020	busy	0	47	346	1413	10.10	00:11:53	30	151.50
3	3	995	995	25	32	154	1206	idle	5	54	292	1553	10.85	00:16:35	45	162.76
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1	183	37097	674	2	74	147	37320	busy	0	65	440	37825	17.76	10:18:17	2745	266.38
2	184	37315	631	6	44	147	37511	busy	0	56	282	37849	17.75	10:21:55	2760	266.28
3	185	37427	523	84	28	153	37692	idle	133	50	305	38048	17.79	10:23:47	2775	266.92
1	186	37825	728	0	31	143	38000	busy	0	60	358	38418	17.70	10:30:25	2790	265.54
2	187	37849	535	150	45	155	38199	busy	0	59	332	38590	17.79	10:30:49	2805	266.79
3	188	38048	621	151	42	139	38380	idle	38	48	305	38733	17.79	10:34:08	2820	266.82
1	189	38418	592	0	23	139	38580	idle	9	45	271	38896	17.71	10:40:18	2835	265.66
2	190	38590	740	0	35	156	38781	busy	0	42	287	39111	17.72	10:43:10	2850	265.87
3	191	38733	685	48	56	128	38964	busy	0	59	325	39349	17.75	10:45:33	2865	266.29

Table 5.2 Part of simulation sheet for case study example

of the spreadsheet represents one cycle. Each column of the sheet is explained below:

- *Truck no.* Each truck allocated a number and this is included to aid ease of reading. This sheet is for 3 trucks only and so if other resource levels were to be simulated then separate sheets would have to be used.
- *No. Cycles.* Increased by one every time a cycle has been completed.
- *Arrival Time.* The time at which the truck arrives at the queue. For trucks after their first cycle this is the time of their previous departure plus the variates for travel time and dump time.
- *Cycle time.* Calculated by subtracting the previous arrival time from the current arrival time. For truck number one, this is 631 - 0 = 631 seconds.
- *Queue time.* The time the trucks spend idle, waiting to be loaded. It is calculated by the truck's arrival time from the departure time of the previous truck. If this is less than zero then there is no queue time.
- *Spot and load times.* These are the random variates generated externally and pasted into the spreadsheet. The times are in seconds.
- *Departure time.* Simply the arrival time plus the queue, spot and load times.
- *Loader status.* Either busy or idle. If the time of departure of the last truck is later than the time of arrival of the next then the loader is busy. It is idle at all other times.
- *Loader idle time.* Time spent idle, if any, by the loader. It is the time of the arrival of the next truck less the time of departure of the previous truck.
- *Dump and Travel Times.* Random variates for the dump and travel components.
- *Next arrival.* Departure time plus the variates for dump and travel times.
- *Cycles per hour.* Total number of cycles completed divided by the arrival time of the current truck.
- *Cumulative time.* Taken as the start of each cycle, i.e. arrival time.
- *Cumulative production.* Total volume moved so far. It is the number of cycles completed multiplied by the volume carried per cycle by each truck (i.e. 15m^3.)
- *Production.* Cumulative production divided by cumulative time. It is therefore a running average of the productivity up to the current time.

The utilisation of the trucks and the loader is also calculated (although not shown in table 5.2). The idle time for each type of plant is cumulated and the utilisation of the loaders or haulers (η_l or η_h), as a percentage, is calculated by:

$$\eta_{l,h} = 100\left(1-\left(\frac{t_i}{nt}\right)\right) \tag{5.1}$$

where t_i is the cumulative time spent idle, t is the total machine time and n is the number of items of plant. Figure 5.1 plots the cumulative production and utilisation percentages for 10 hours of simulated run time. This operation is almost perfectly matched (MF = 0.97) and so the plant utilisation factors are very high. Note however that they are not at 100% and this is due to the bunching effect on the trucks. Hence it can already be seen that simulation provides a more realistic estimate.

The graph clearly shows the two distinct parts to the simulation output. There is a transient stage and a steady-state output. Strictly, steady state is only reached as t - ∞. There is a finite point, however, when the outputs at successive times are approximately equal. Note that this does not mean that individual runs (or *replications*) will produce the same values at steady state, but rather they will have the same distribution. This transient or warm-up period is very important in a system such as this because maximum production is not possible until steady state has been reached. The rate of convergence to steady state depends on the initial conditions of the simulation and, furthermore, every time the simulation is restarted with the same initial conditions, the same transition will occur. Therefore, if an earthmoving operation is stopped before steady state production is reached and re-started with the same initial conditions, the operation can never be at maximum efficiency. Section 5.5 will investigate the length of this transient period and its consequences on actual earthmoving operations. It is noteworthy that if the model described in chapter 3 were used to estimate production for this operation, the figure of 303 m^3/hr would be obtained. This first simulation obtained a steady state production of approximately 266 m^3/hr that compares with an observed production rate of 262 m^3/hr.

5.2.2 Utilisation and Plant Match

Figures 5.2 to 5.5 show how the output changes for different plant match factors. (See section 3.6 for a definition of match factor.) The closer the match factor is to one, the better the utilisation's of the plant. Figure 5.2 shows the output for a match factor of 0.61. The truck utilisation in this case is very high

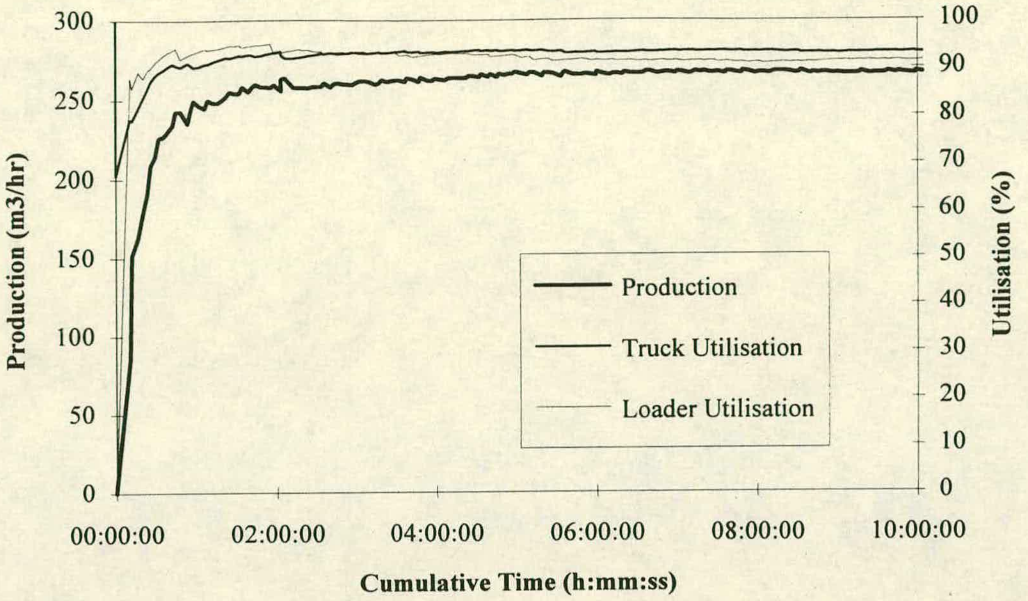


Figure 5.1 Graphical plot of simulation results for case study example (section 5.2.1). Match Factor = 0.97.

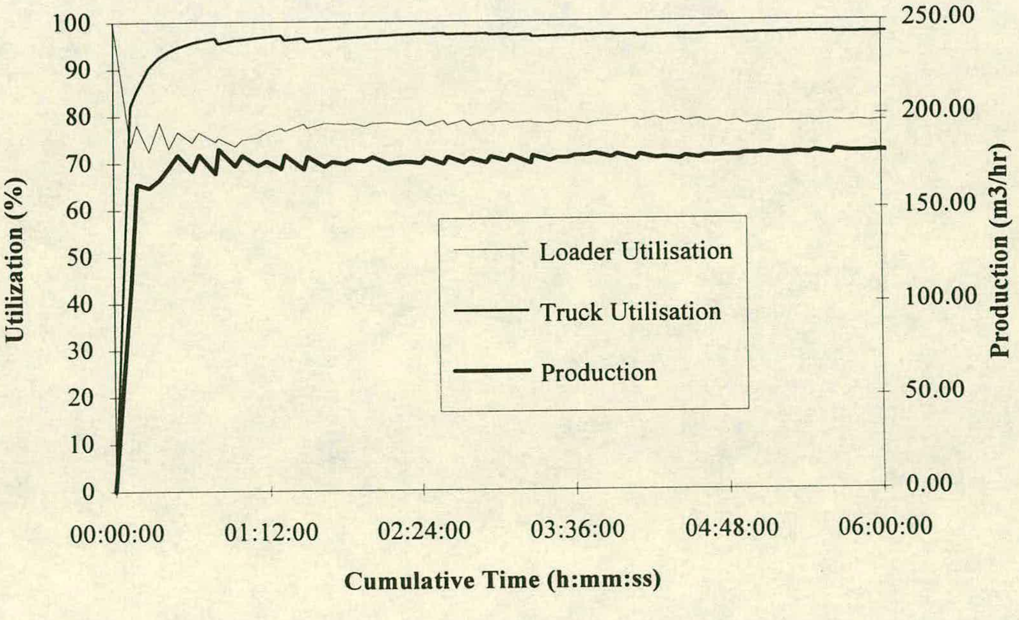


Figure 5.2 Simulation output for match factor of 0.61

and the loader spends approximately 25 % of the time idle. This inefficiency therefore results in low productivity (180 m^3/hr at steady state). As the match factor increases, plant utilisation's approach (but will never reach) 100%, and as MF moves greater than one, loader utilisation will remain high as truck utilisation falls lower. The production reaches its maximum when loader utilisation is 100% and remains at that level; production cannot be increased by adding more trucks unless the loading conditions are improved - this is an important concept in earthmoving. The point at which this maximum production is reached will depend on the variability of the plant cycle times but will be approximately at MF = 1.2. Table 5.3 summarises the results shown in figures' 5.1 to 5.5.

5.2.3 Disadvantages with spreadsheet simulation

The spreadsheet simulation has been very useful in understanding how to apply the simulation technique to the earthmoving model. It indicated also that there is a warm up period within which the output of the system is not at its maximum. However, the method is very cumbersome and far from ideal from a users' standpoint as each simulation requires that the variates are generated first (and the number of variates for the total run time required is a matter of trial and error) and are then pasted into the spreadsheet. Altogether, a simulation run of the equivalent of, say, 10 hours of earthmoving, takes about 10 minutes to carry out. This is very impractical if an experimental analysis is required which needs as many replications for different situations as possible. A major disadvantage is that separate spreadsheets are needed for each resource level; this means that comparing between different numbers of trucks is a difficult procedure. It is anticipated that this will be one of the main uses of the simulation model. Therefore, the decision was made to develop a computer program that carries out a discrete-event simulation of the earthmoving system.

5.3 Dedicated computer simulation program

The most important decision to be made when developing a simulation program is which programming language is to be used. This is made more complicated by the fact that there are languages available for simulation programs only as well as the common general purpose languages such as FORTRAN, Pascal or C. The emergence of

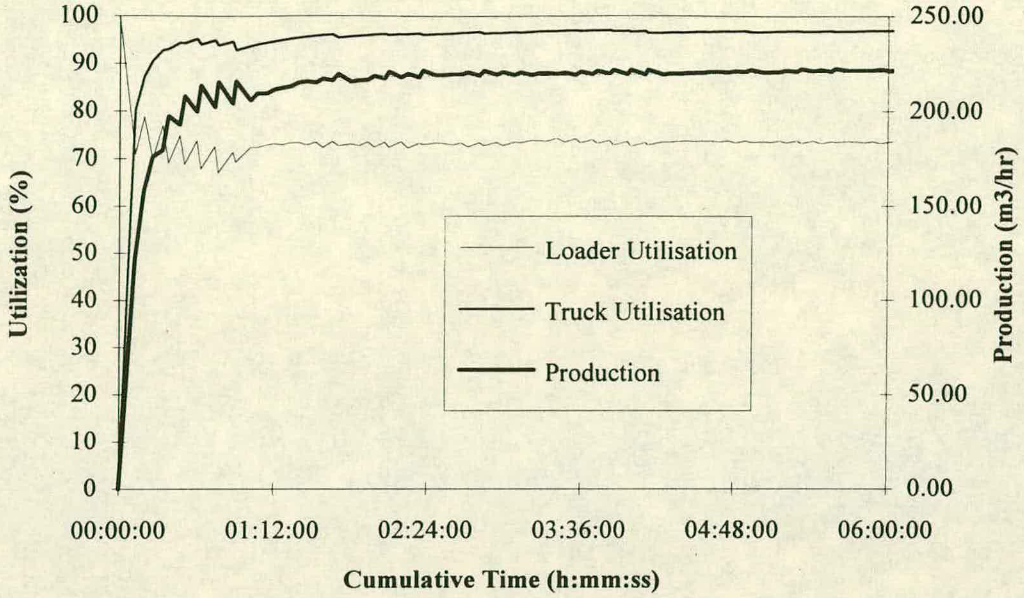


Figure 5.3 Simulation output for match factor of 0.76

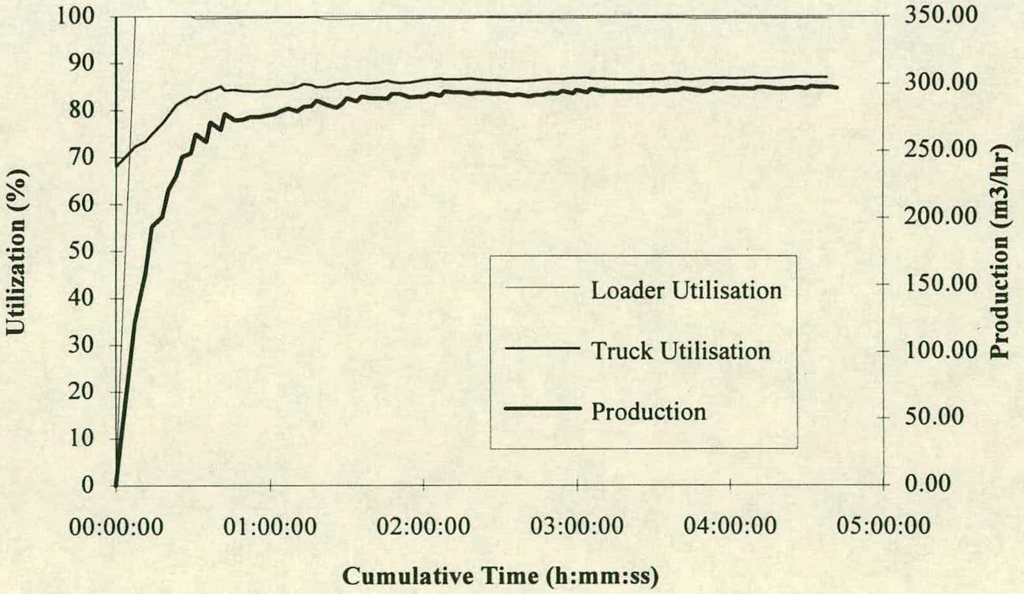


Figure 5.4 Simulation output for match factor of 1.15

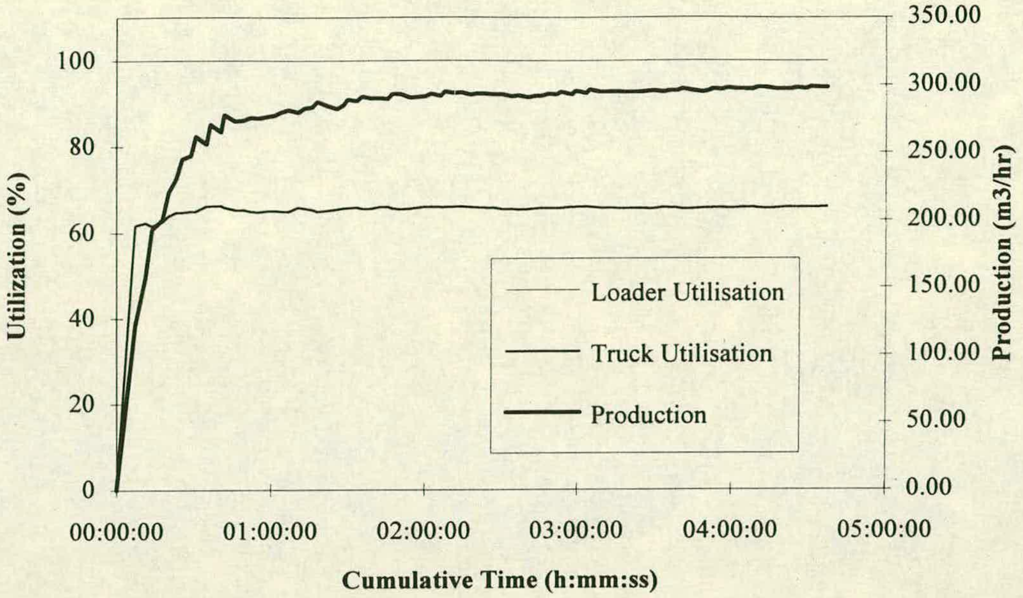


Figure 5.5 Simulation output for match factor of 1.52

		Utilisation (steady state) %		Steady-state
Match Factor	**No. Trucks**	**Loader**	**Hauler**	**Production m3/hr**
0.61	2	78	98	180
0.76	2	73	97	225
0.97	3	92	94	270
1.15	3	99	87	295
1.52	4	100	65	300

Table 5.3 Summarry of results for figures 5.1 to 5.5

		Low Bunching	High Bunching
Load Time			
	μ	18 seconds	
	k	16	
	R	0.889	
	a	10 seconds	
	b	45 seconds	
Spot Time			
	μ	37 seconds	
	k	5	
	R	0.1462	
	a	20 seconds	
	b	100 seconds	
Dump Time			
	μ	61 seconds	
	k	35	
	R	0.5672	
	a	30 seconds	
	b	100 seconds	
Travel Time			
	μ	126-1073 s	132-1093
	k	40	5
	R	0.322-0.0421	0.0402-0.0046
	a	62-543 s	31-272 s
	b	248-2173 s	497-4346 s
No. Trucks		4	
Passes per Load		5	
Bucket Volume		1.8 bcm	

Table 5.4 Parameters for figure 5.10

languages for simulation only occurred because of the common elements in simulation studies. All simulations for example, require the generation of random U(0,1) numbers and random variates and tasks such as advancing the simulation clock and updating lists of records are important in most programs. These languages provide the framework for a simulation model but they do have some disadvantages to overcome. The biggest difficulty is the availability of a simulation language. The four most widely used simulation languages are GPSS, SIMAN, SIMSCRIPT II.5 and SLAM II (a review of the features is given by Law and Haider, 1989) but they are much less common than general purpose languages that are installed on most university PC's. Most modelers will have at least some knowledge of a general purpose language but will have little idea how to program in a simulation language and it is perhaps for this reason why the general purpose language C was chosen. Although in the long run it may have been advantageous to develop a simulation program with a language specially developed for the task, the time taken to acquire the software, learn the language and, when written, verify and debug the code would have been prohibitive. At least with C there are many other users who can help in times of difficulty (including the Edinburgh University Computer Service, who provide a help service. This would not be available for GPSS for example.) Law and Kelton (1991) give a very good reason for using a general purpose language especially for those who have never simulated before:

> "By learning to simulate in a general purpose language, in which one must pay attention to every detail, there will be a greater understanding of how simulations actually operate, and thus less chance of conceptual errors if a switch is later made to a high level simulation language."

The program (called Trucksim) is explained in detail in appendix 4 and the next section will briefly describe the program by way of flowcharts.

5.3.1 Trucksim organisation

The trucksim program is a discrete-event simulation program (see section 4.2.2) capable of simulating a single-server (loader) earthmoving system with up to 15 trucks that are assumed to be of the same make and model, i.e. truck characteristics are the same for all trucks. The main simulation part of the program is very straightforward with the bulk of the program code being input/output controls. The random variate generator incorporated into the program is essentially the same as the one shown in appendix 2. When the program is first started, the user has to configure the program before a simulation can be run. This requires the file directory to be specified, the time limit (in seconds) to be set, the number of trucks to be used, the number of buckets per load and the volume of the loader bucket in bank cubic metres. When configuration is complete, the Erlang density function parameters can be loaded in from a previously written file. Each file contains 6 parameters for each of the four time components and these can be viewed, and changed, from within the program. When the program has the parameters the user requires, simulations can be run. This can be done either singly, with the results shown on the screen or multiply where the results for several replications are saved to file.

Figure 5.6 is the main flowchart for the simulation. The first task is to initialise the simulation (i.e. set the state variables to their initial values and zero the statistical counters). The program then loops continuously through the simulation (the *simulation loop)* until the pre-set time limit has been reached. On each pass of the loop the type of the next event is determined (by comparison of the times of the first arrival and departure events in the event list) and the appropriate event routine is invoked. The statistical counters are then updated (such as the simulation clock and the counters for the result averages: average truck delay time, utilisations, production, etc.) and, if the time limit has not yet been reached, the loop is started again. Note that the simulation clock is updated only when an event happens. This means that the simulation will finish when the last event has been processed and the clock updated for the last time. The finish time will probably not be the pre-set time but some short time after the stop time - for this reason averages are worked out on the final simulation clock time and not the pre-set time.

Figure 5.7 is the flowchart for the arrival event. The first task is to ensure that the truck that has just arrived cannot have an arrival event occurring before the next departure. This is done by setting the time of its next arrival to 1×10^{31} seconds. The state of the loader is then checked and if it is busy, the truck joins the queue, the queue length counter is increased by one and the time at which the truck starts its delay is recorded. The program is then returned to the simulation loop. If the loader is idle, then its status is changed to busy and a departure event is scheduled for the truck by generating a spot time and a load time and adding this to the current simulation clock time. The time the loader has spent idle is then computed and the program returned to the simulation loop.

Figure 5.8 is the flowchart for the departure event routine. The departing truck is scheduled an arrival time (adding generated travel times and

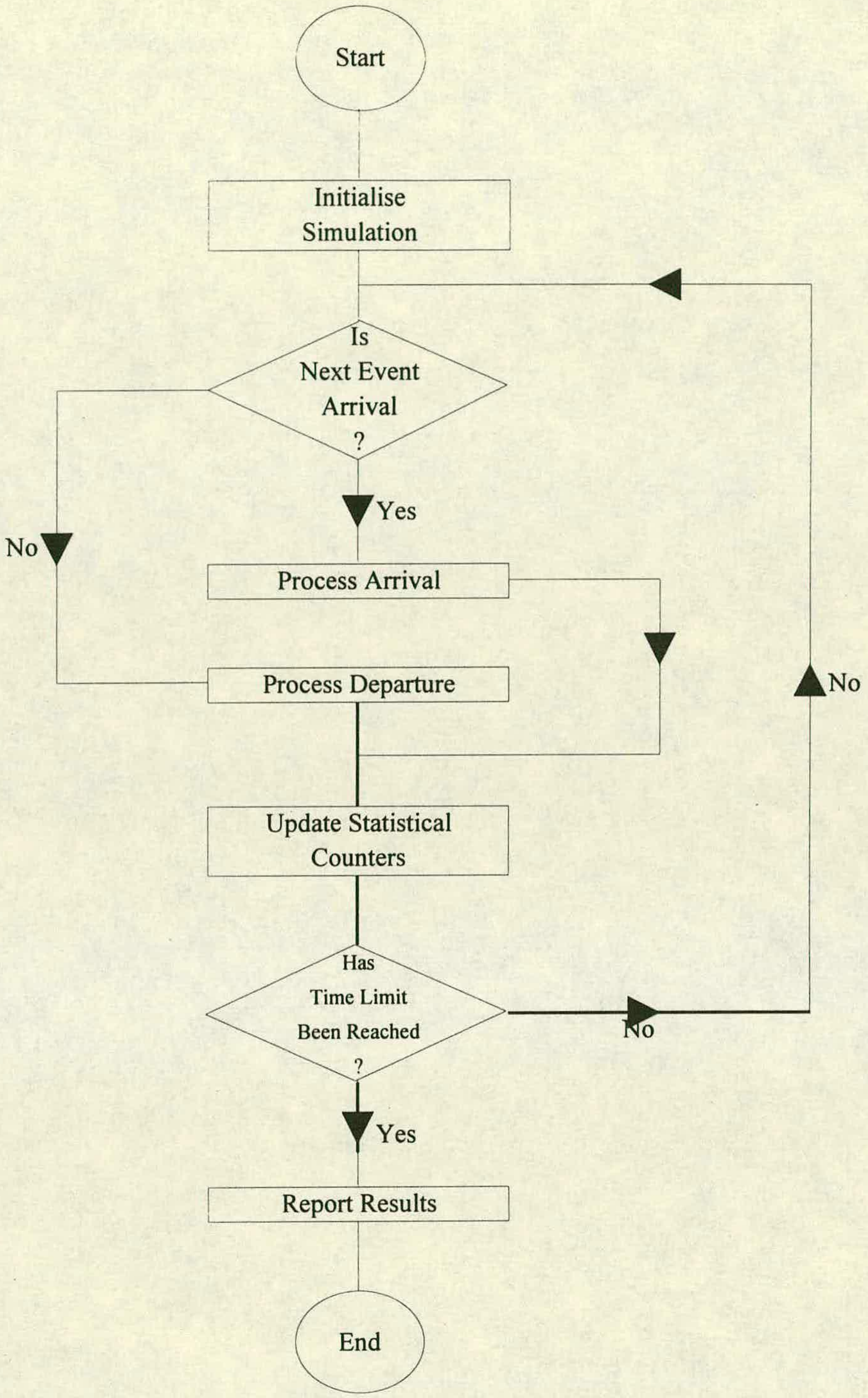


Figure 5.6 Main Simulation Flowchart

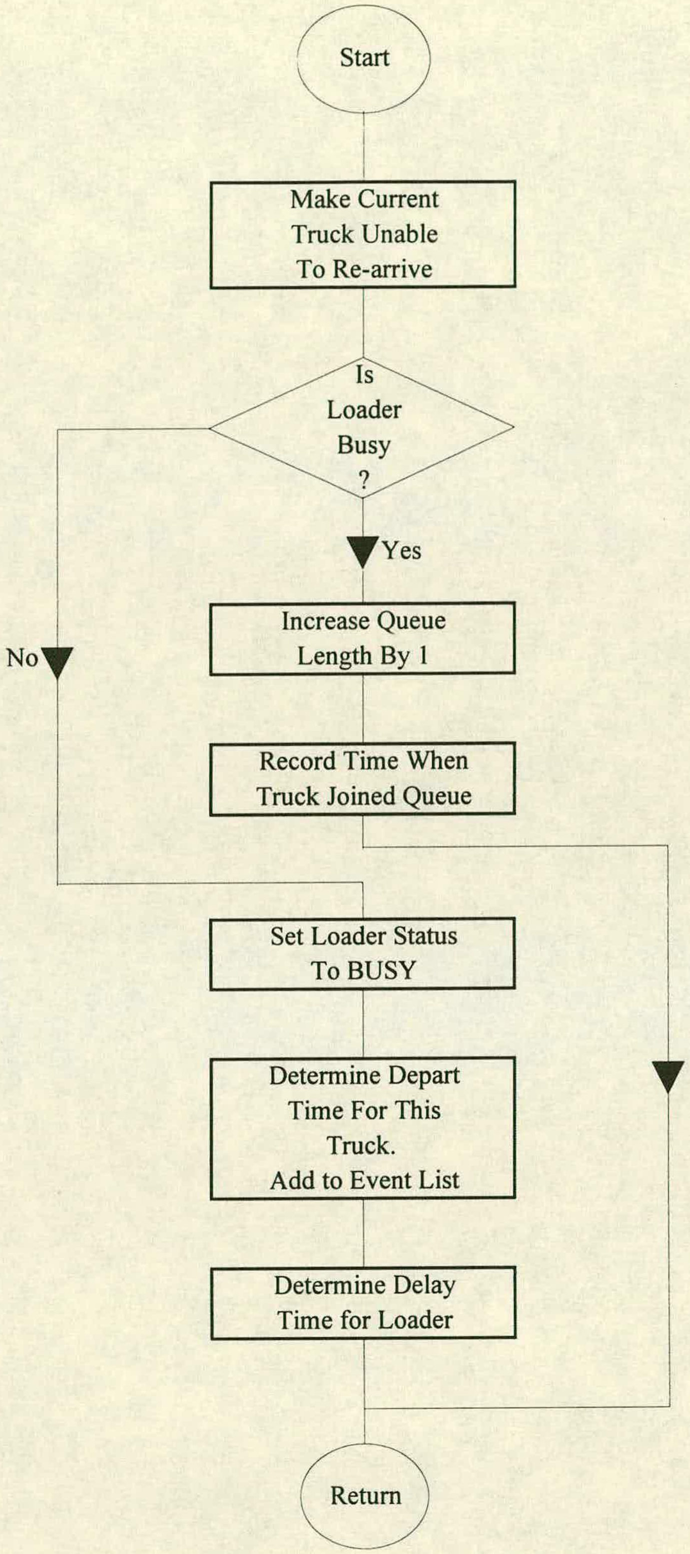


Figure 5.7 Arrival Event Flowchart

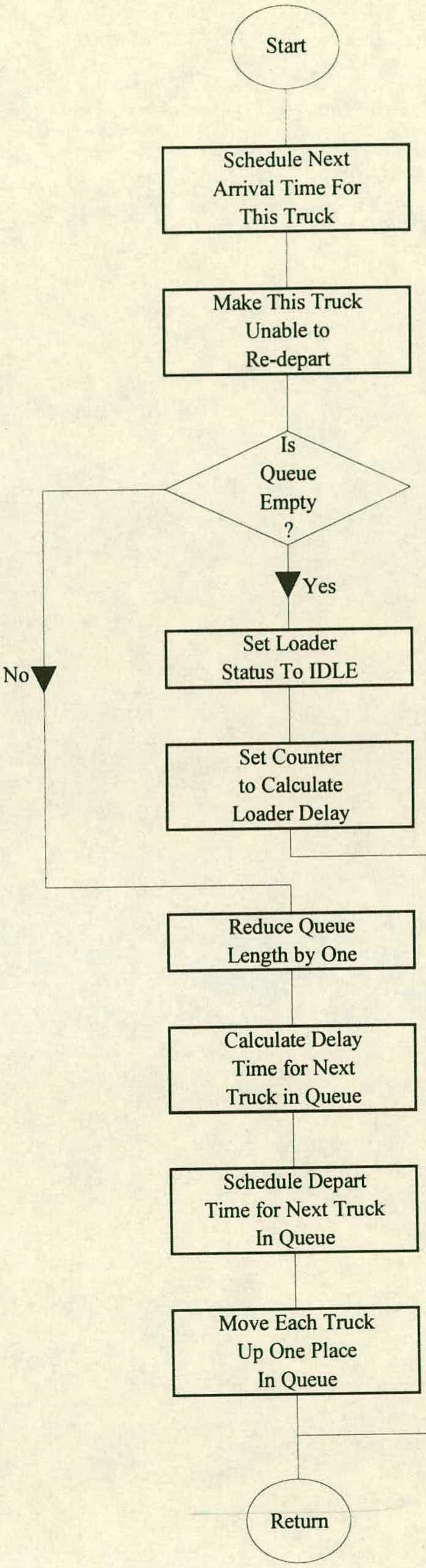


Figure 5.8 Departure Event Flowchart

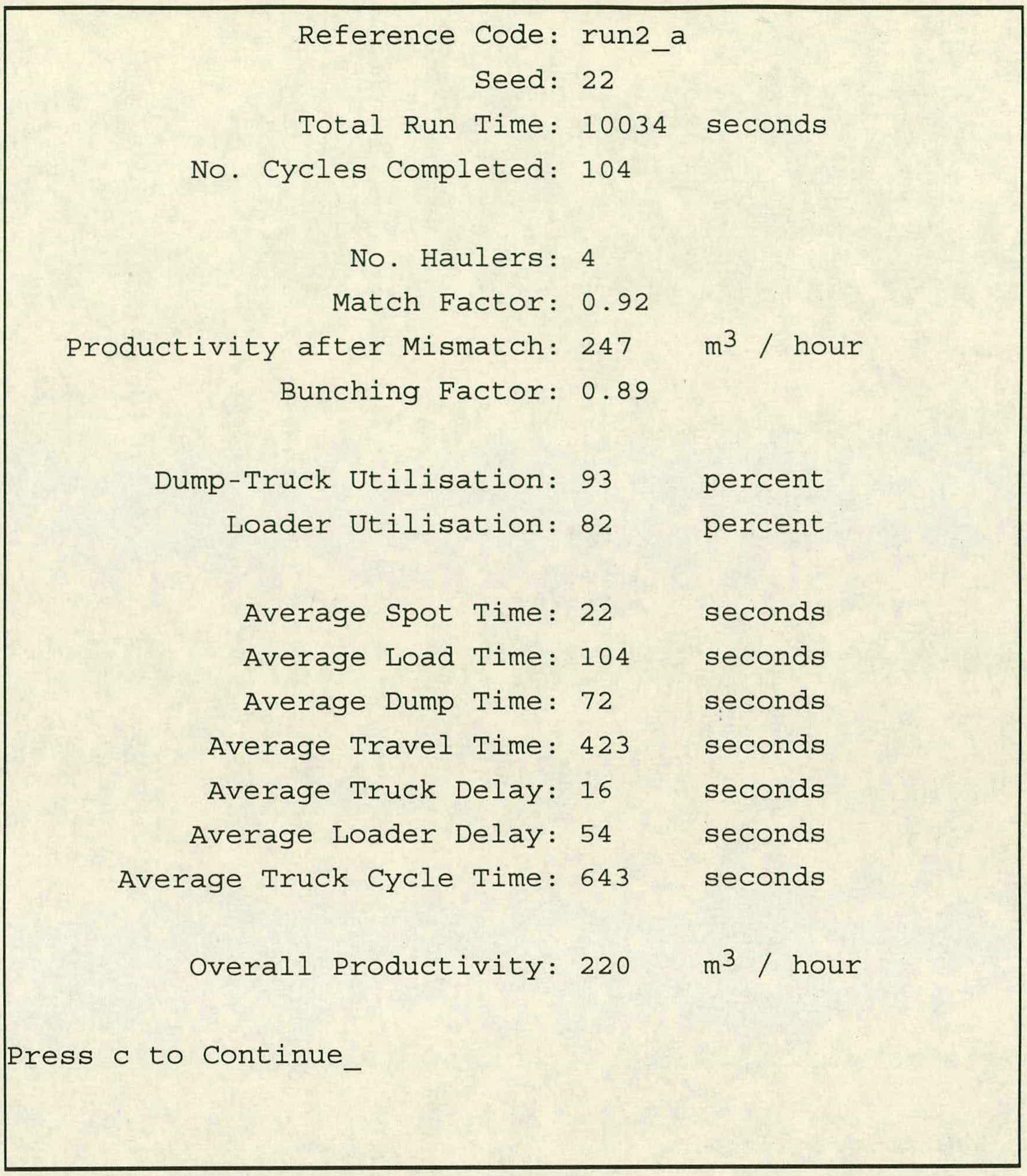
dump times to the current simulation clock time) and is forced to arrive before it can depart again by making its next departure time very large. The state of the system is checked and if the departing truck leaves the queue empty then the loader status is switched to idle and the loader idle counter is started. (This counter is stopped at the next arrival event.) The program is then returned to the simulation loop. If the queue is not empty then a truck is drawn from it, the queue length counter reduced by one and the delay time of this truck is computed (the arrival event starts a delay counter for every truck joining the queue). This truck is then scheduled with a departure event (using the variate generator) and the position of each truck in the queue is moved up by one. The program is then returned to the simulation loop.

When the simulation is complete, the results can be viewed (for single simulations) and, if necessary, the program configuration and Erlang parameters changed for new simulations.

5.3.2 Trucksim results

Figure 5.9 is a sample results sheet from the Trucksim program.

- *Reference Code.* A code entered by the user before simulation is started so particular results can be identified.
- *Seed.* This is an integer entered by the user before the simulation is started to provide a random seed for the internal random number generator. If the same results need to be provided again, the same random seed can be used.
- *Total Run Time.* The time of the simulation clock when the simulation was stopped. This will usually be a short time after the pre-set time limit.
- *No. Cycles Completed.* Number of cycles completed by all trucks (not individually) when the simulation was stopped.
- *No. Haulers.* The number of trucks (entered by the user at configuration).
- *Match Factor.* Indicates the amount of plant mis-match (see section 3.6)
- *Productivity after Mismatch.* If the system was not a stochastic process then no bunching would occur and this figure would be the productivity achievable with the plant that has been used in the simulation. This productivity is used in calculation of the bunching factor.
- *Bunching Factor.* Overall productivity (see below) divided by the productivity after mismatch. Therefore a bunching factor of 1 indicates zero bunching and a bunching factor of 0.7 can be considered very bad



```
                 Reference Code: run2_a
                           Seed: 22
                 Total Run Time: 10034  seconds
           No. Cycles Completed: 104

                    No. Haulers: 4
                   Match Factor: 0.92
    Productivity after Mismatch: 247    m^3 / hour
                Bunching Factor: 0.89

         Dump-Truck Utilisation: 93     percent
             Loader Utilisation: 82     percent

              Average Spot Time: 22     seconds
              Average Load Time: 104    seconds
              Average Dump Time: 72     seconds
            Average Travel Time: 423    seconds
            Average Truck Delay: 16     seconds
           Average Loader Delay: 54     seconds
       Average Truck Cycle Time: 643    seconds

           Overall Productivity: 220    m^3 / hour

Press c to Continue_
```

Figure 5.9 Sample Results from Trucksim

bunching. See section 3.8.1 for a more detailed explanation of this factor (which was the reason simulation was necessary).

- *Plant Utilisations.* Calculated for trucks and loader using equation 5.1. The dumptruck utilisation is the average utilisation for a single unit.
- *Average Component Time.* These averages are very important in that the random variate generator may not produce a series of numbers with the mean that was specified, especially for short simulations, simply because the numbers are random. These averages are used to calculate the match factor and are used in the experimental analyses (chapter 7).
- *Average plant delays.* The average delay per cycle of both individual trucks and of the loader. Calculated by summing the total delays and dividing by the number of cycles.
- *Average Truck Cycle Time.* This is calculated using a total cycle time counter and dividing the final total by the number of cycles completed. It is assumed that this provides a better estimate of the overall cycle time than summing the individual averages for the time components and the truck delay time.
- *Overall Productivity.* This can be calculated in several ways, for example by first calculating the number of cycles per hour. The best estimate is assumed to be by multiplying the number of cycles completed by the volume of each truck and dividing by the total run time.

5.4 Use of the Trucksim program

The uses of the simulation program are far reaching. At this stage, before any are investigated, it is perhaps prudent to review the procedure that most modelers use in a simulation study. This procedure is used by many authors (for example Law and McComas 1989, 1990; Shannon, 1975 and Gordon, 1978) and is as follows:

i. Define the problem. This was a long procedure that started with the simple model shown in chapter 3.
ii. Collect data and develop the model. The data collected on the M1\A1 contract was used initially in the first model but when the simulation model was developed (chapter 4) the data from both this site and the M3 contract was used extensively as the basis for the model.
iii. First validation. The simulation model was tested using the spreadsheet simulation reviewed earlier in this chapter. This was validated very briefly using a

case example and so the decision was made to write a dedicated computer program.

iv. Write computer code and verify. This has been done. Verification is sometimes confused with validation but they are quite distinct. Computer code is verified when all bugs have been detected and removed and the computerised model runs as intended. There are formal methods for verification, such as constructing a *trace* whereby the variables used are printed on the screen whenever they are updated and what is called a *structured walk through* where each line of code is inspected to determine that it does what is intended. The compiler used to write the code, *Microsoft Quick C*, (Microsoft Corp. 1990) contained an interactive debugger that provided a trace function; this was used extensively when the program was written as C is a language that will compile even if the code is incorrect thus hiding any possible errors. Fortunately, the simulation was kept very uncomplicated; no major problems were encountered and the program is now considered verified. It may however not be validated - this is when the program produces results that indicate the model is correct.

v. Validation. This is major step and is the subject of chapter 6. Validation must be carried out if the simulation is intended to be used for major decision making activities. If the program is verified then it may produce results that appear to be correct but are actually wrong. If this is the case, and these results are used, then the whole process of simulation will have been wasted.

vi. Design experiments, make runs and analyse the output data. This is the subject of chapter 7. Experiments are used to determine which input factors affect the output the most. It can also be described as sensitivity analysis whereby the sensitivity of the output to different inputs is determined.

vii. Document, present and implement results. Part of this process is contained in this thesis but to ensure that results are used correctly the implementation must be an ongoing process. It cannot be assumed that a simulation study is finished when the final report is presented.

The next logical step in this study is therefore validation of the model. Before that is done, a few of the results and uses of the program will be described. Analysis of the output data for determination of confidence limits and steady state output will be described in section 5.5 and 5.6.

5.4.1 Bunching and Match Factor

If a deterministic approach is used to estimate the productivity of an earthmoving operation then the match factor will influence the output in a way such as shown in figure 3.4. In chapter 3, however, this deterministic approach was shown to be unsuitable when the results from various actual operations were analysed, as seen in figure 3.11. The earthmoving system was therefore proposed to be stochastic and solved by use of simulation. If simulation results are plotted in a similar way to figure 3.4 then, if the assumption that earthmoving is stochastic and that simulation is the best analysis method available, these results should be more similar to those in figure 3.11.

Figure 5.10 shows the results of simulation runs producing various match factors. The vertical axis is the productivity of the operation as a percentage of the maximum productivity - calculated by multiplying the match and bunch factors together. (For match factors greater than one, a value of 1 is used - productivity cannot be greater than 100%.) The match factor in this study was varied by changing the travel time but keeping the spot, load and dump time components constant as well as keeping bucket volume, passes per load and number of trucks constant. Two curves were produced, high and low bunching, by using two sets of travel time with different distribution 'families'. Remember from section 4.3.4 that the Erlang distribution has two parameters, R (rate) and k (shape). These two parameters influence the shape of the distribution curve with $k = 1$ being an exponential distribution and for $k >$ (approximately) 20, the curve becomes more and more normal. (Winston, 1987, p. 872). If k becomes very large then the curve becomes a random variable with zero variance, i.e. a constant time. The deterministic line in figure 3.4 can therefore be thought of as the result of a stochastic analysis with very large shape factors*. If the stochastic nature causes bunching then the assumption can be made that the smaller the shape factor, the greater the bunching - this is how the two curves were produced in figure 5.10. The low bunching curve was produced using variable travel times but with k constant at 40. The high bunching curve has a constant k value of 5. Table 5.4 has the parameter values used in this exercise. At each different value of travel time, 5

---

* This would actually be difficult to prove. The generation of Erlang variates require the factorial of the shape factor to be calculated. On all but the more powerful computers, the largest factorial that can be calculated is $170! = 7.257 \times 10^{306}$. Even if such numbers could be calculated, the generation would still be difficult. The convolution method would require k numbers for each variate and the ARM would take a long time to accept a constant number within a range. Figure 5.11 shows an Erlang cdf for a shape factor of 150 which is still far from being constant.

replications were run to give a better estimate of the match and bunching factors.

It can be seen that the two different shape factors do indeed produce different degrees of bunching. The two curves can be considered the extremes for vehicle bunching with the space in between an envelope of possible bunching factors. Indeed, if an operation showed such high bunching as the upper limit in figure 5.10 then it would be an unnecessarily badly set up operation. The travel time variability's on such an operation would be caused by using different types of dumptruck with the haul road providing many obstacles with consequent long delays.

5.4.2 Production Charts

The trucksim program is ultimately intended to be used as an estimation tool for tendering purposes and as a site control method once operations have commenced. The program is, however, difficult to use repeatedly - parameters for the Erlang distributions need to be calculated and various combinations of truck types and numbers may need to be considered before the optimum solution is found for a particular operation. In a construction situation time is a very precious commodity and to carry out a long iteration using Trucksim directly would be very impractical. Even at a tender stage, the contractor has little time to investigate the site and produce an estimate and programme for the work.

Production charts are therefore a solution to this problem. If a series of charts for different site and vehicle conditions could be produced then determining a solution for an earthmoving operation would be a far easier task. Obviously there are an infinite number of possible combinations of spot time, bucket size, passes per load, dump time, etc. but if the most common situations could be determined then charts for these could be drawn up. Another approach is to work backwards. The earthmoving contractor could determine the maximum time that is available to carry out the works and use this as a target. Production charts could be drawn up which are lower bound solutions of productivity and if these targets are not reached on site then the contractor knows he has to change the operation conditions to increase the productivity.

As an example, the charts drawn up for an actual road contract shall be explained. The Northern Cross Route in Dublin, Republic of Ireland is a dual 2 lane motorway to relieve the traffic from the centre of the capital. The ground

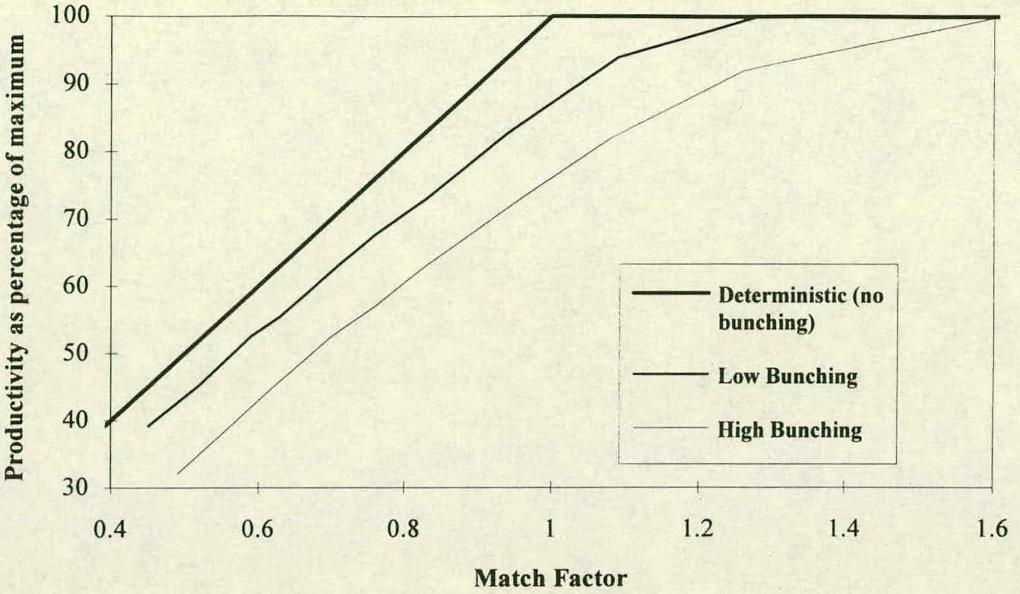


Figure 5.10 The effect of plant match on operation productivity

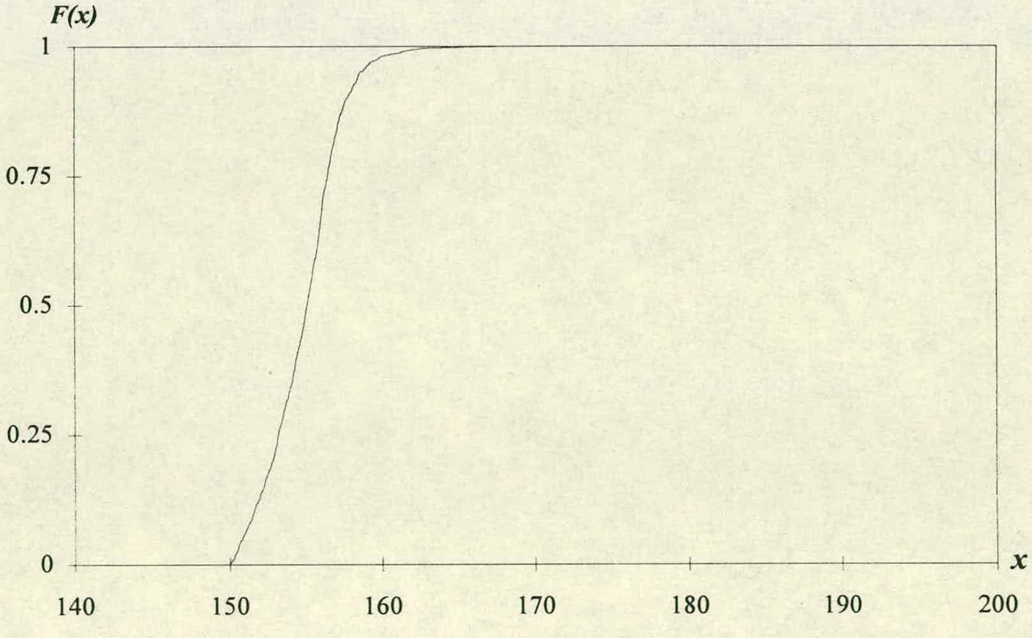


Figure 5.11 Erlang cumulative distribution function with shape factor of 150

conditions through the 10 km contract are fairly constant; the haul roads consist of a reasonably strong boulder clay with rolling resistances ranging from 4% to 12%. The plant proposed to be used on this contract were Caterpillar D400 dumptrucks coupled with Caterpillar 245 backhoe excavators. The size of the job and the similarity of the conditions with those experienced on the M1/A1 contract indicated that the loading conditions would be similar to those shown in table 5.5. The two columns are for different types of cut. Acceptable cut, where the material can be expected to be stronger, more consistent and in larger quantities than that in unacceptable cuts, should be easier to excavate. The poorer quality unacceptable material is likely to be found in smaller 'bands' than the acceptable material with a consequent smaller bank cubic volume per bucket and longer load time.

Using the basic simulation parameters shown in table 5.5, simulations were carried out for various combinations of travel time and truck numbers. For example, in acceptable cut, an operation with a travel time of 800 seconds and 7 dumptrucks could expect to realise 232 bank cubic metres every 50 minutes.* The charts shown in figures 5.12 and 5.13 are the production curves for the two types of cut. As travel time increases the number of trucks must also increase to keep the efficiency of an operation high and so each chart has a number of lines for different numbers of trucks. These are plotted using 10 simulation replications for each travel time used. In addition, for each truck number there will be a travel time where the match factor is equal to one and these are also plotted to give an indication as to the most efficient resource level to choose. As can be seen, the acceptable cut is expected to produce more material than the unacceptable cut but unless the actual operation works at exactly those rates shown in table 5.5 the actual productions shown in the charts are unlikely to be achieved. The charts are therefore intended to be used as targets as follows. Firstly, at the tender stage, the contract was split into separate operations with programmed durations to fit into the overall contract programme. The charts are then used to estimate the production and resources required for each operation keeping the intended durations in mind. These are then the targets for the operations. If, when works start, the required

---

* Earthmoving estimation has always been done using a basic hour of 50 minutes. Although simulation should be able to eliminate such rules of thumb it must be remembered that, (a) the simulation has not yet been validated and (b) the simulation model does not allow for stoppages such as breakdowns and management delays. Therefore an assumption is made that 10 minutes in every 60 are unproductive.

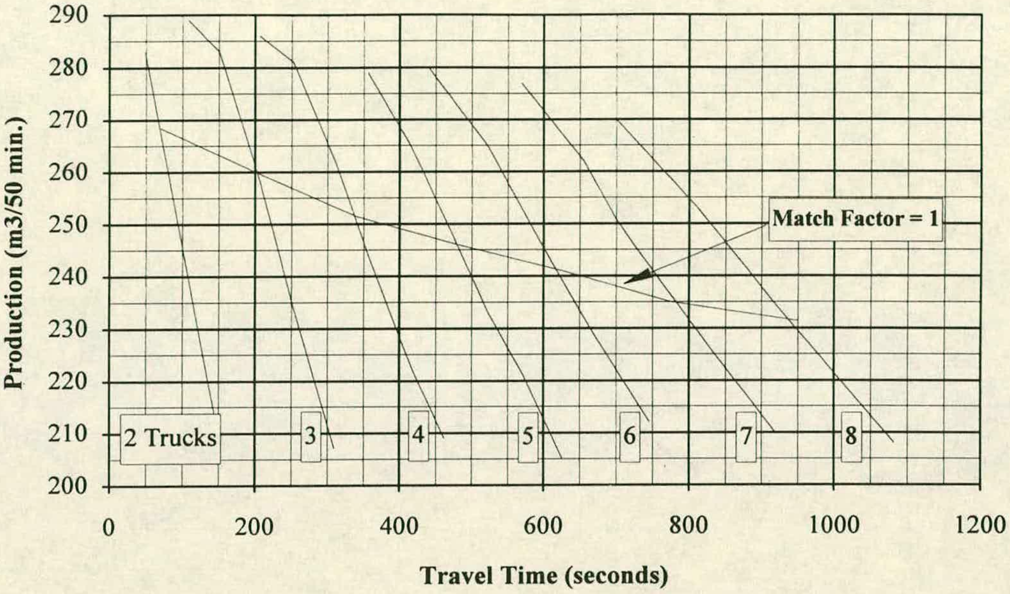


Figure 5.12 Production chart for acceptable cut

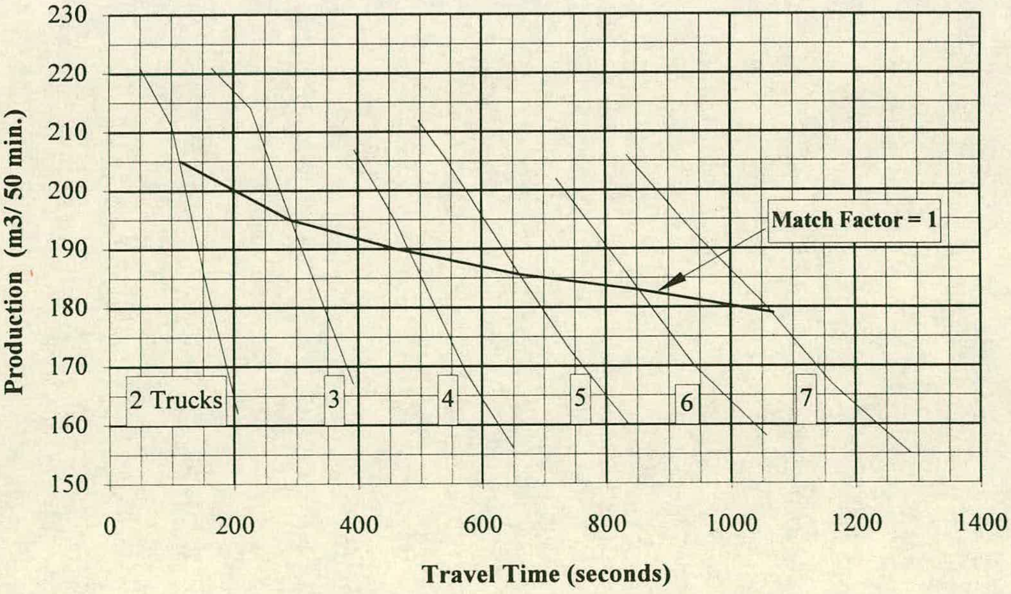


Figure 5.13 Production chart for unacceptable cut

production rate is not achieved then the management can take more effective action to resolve the problem.

5.5 Transient behaviour of simulation output

The plots in figures 5.1 to 5.5 show that the simulation output has a possible transient or warm-up period before the maximum is achieved at steady-state. This is very important in earthmoving where shifts may be of different lengths that may not be long enough to overcome the transient period. It is also important that any productions estimated by simulation take the time length of the simulation into account. For example, if a replication has a time length of 10,000 seconds (2 hours 46 minutes) then the production estimate given at the end of the simulation is not applicable if shifts last only, say, 2 hours or if breakdowns occur which effectively cut a long shift into several smaller shifts.

This apparent warm-up period is also a reflection of the way in which production is measured. In 5.2.1, production is taken as cumulative production divided by cumulative time. However, cumulative production is defined as the number of cycles completed multiplied by the volume carried per cycle: therefore, production is measured only when a truck has dumped its load, returned to the queue and the cycle has been completed. Alternative ways of measuring production, such as after each bucket has been excavated, or at the end of each load, or even when the load is actually dumped, will affect the apparent warm-up period due to the differences in 'time-lag' between when the production is actually achieved (i.e. the moment it leaves the ground) to when it is measured. The following results may not be applicable if a different method of measuring production is used.

It is possible to estimate the time length l of the warm up period. One of the methods available is a graphical procedure due to Welch (1983).

5.5.1 Welch's graphical procedure to determine transient length l.

If the plots in figures 5.1 to 5.5 are studied it can be seen that they fluctuate rapidly, especially during the transient. Welch's method smoothes out the curve by making n independent replications and following the procedure below:

i. n replications are made of the simulation (where $n \geq 5$) each of length m. Y_{ji} is the ith observation from the jth replication. $i = 1,2,3, \ldots ,m$ and $j = 1,2,3, \ldots,n$.

ii. Let $\overline{Y}_i = \sum_{j=1}^{n} Y_{ji} / n$ for i = 1,2, ...,m. This averaged process has means of $E(\overline{Y}_i) = E(Y_i)$ and variances $\mathrm{Var}(\overline{Y}_i) = \mathrm{Var}(Y_i)/n$. Thus, the averaged process has the same mean transient curve as the original process but its plot has only $1/n$ th the variance.

iii. A moving average is then defined to further smooth out the high frequency oscillations. This is defined as:

$$\overline{Y}_i(w) = \frac{\sum_{s=-w}^{w} \overline{Y}_{i+s}}{2w+1} \quad \text{for } i = w+1, \ldots, m-w \tag{5.2}$$

iv. Plot $\overline{Y}_i(w)$ for i = 1,2,..., m-w and choose l to be the value of i where $\overline{Y}_1(w), \overline{Y}_2, \ldots$ appear to have converged.

5.5.2 Determination of transient length using Trucksim.

The graphical method described above has been used to determine the transient length for four different earthmoving situations simulated using Trucksim. It is important to know not only the likely value of the transient length but whether or not its value is influenced by operating conditions. Two factors will be studied to determine their effect on the transient length: initial conditions and haul length. To investigate the effect of more factors (such as load time and passes per load) would have been a lengthy process (for every transient length that is determined, at least 110 replications are needed. Therefore for four factors with two levels set for each factor, 2^4 x 110 = 1760 replications would be needed) and, as will be seen from the results, this would probably be unnecessary anyway. The initial conditions set in the Trucksim program are that each truck starts empty and waiting in the queue at time = 0. It must therefore take a few cycles before the arrival time of the trucks becomes random and this will have an effect on the production for these first few cycles. As an alternative to these initial conditions, the program was altered so that the time of the first arrival of each truck is some random exponential time after the arrival of the previous truck. Transient lengths were determined for two haul lengths (800m and 1800m) for each set of initial conditions and the results from these can be seen in figures 5.14 to 5.17.

Each figure has plots of moving averages for 5 replications at values of i=1000,2000,...,22,000 seconds with w=1 and w=2; and a percentage of

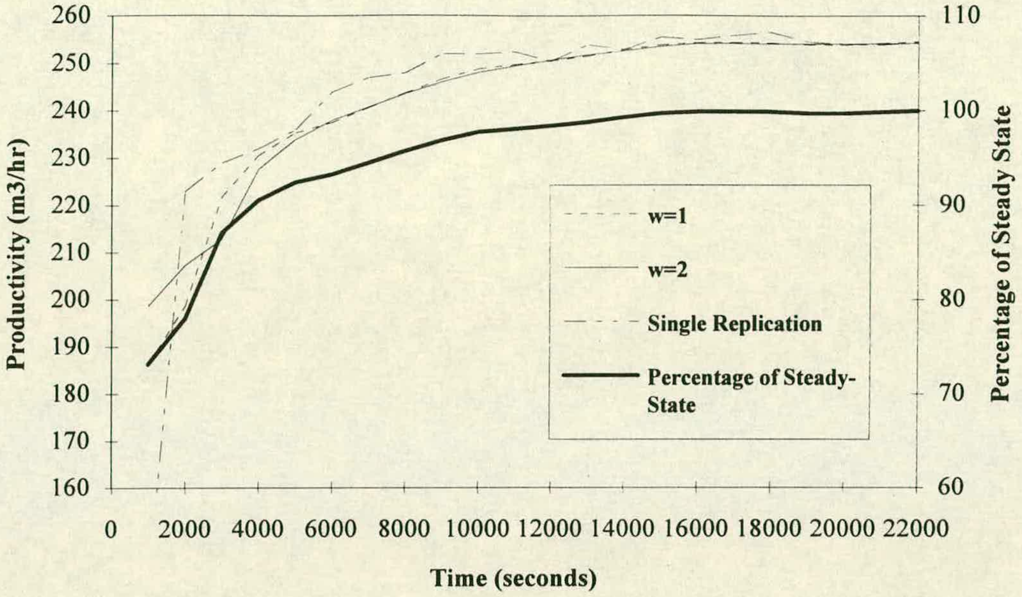


Figure 5.14 Moving averages for simulation output

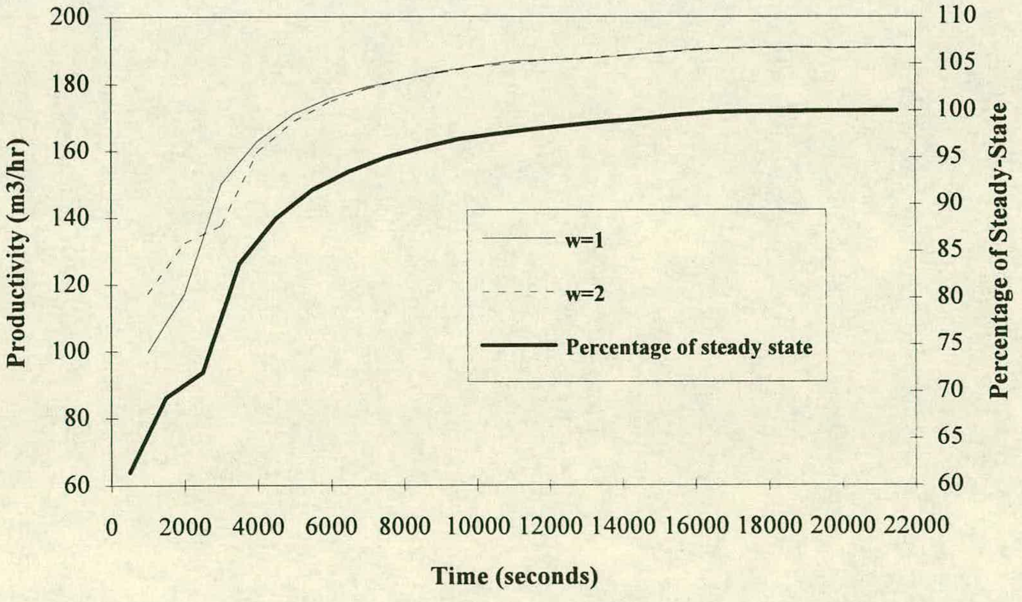


Figure 5.15 Moving averages for simulation output (haul length = 1800m)

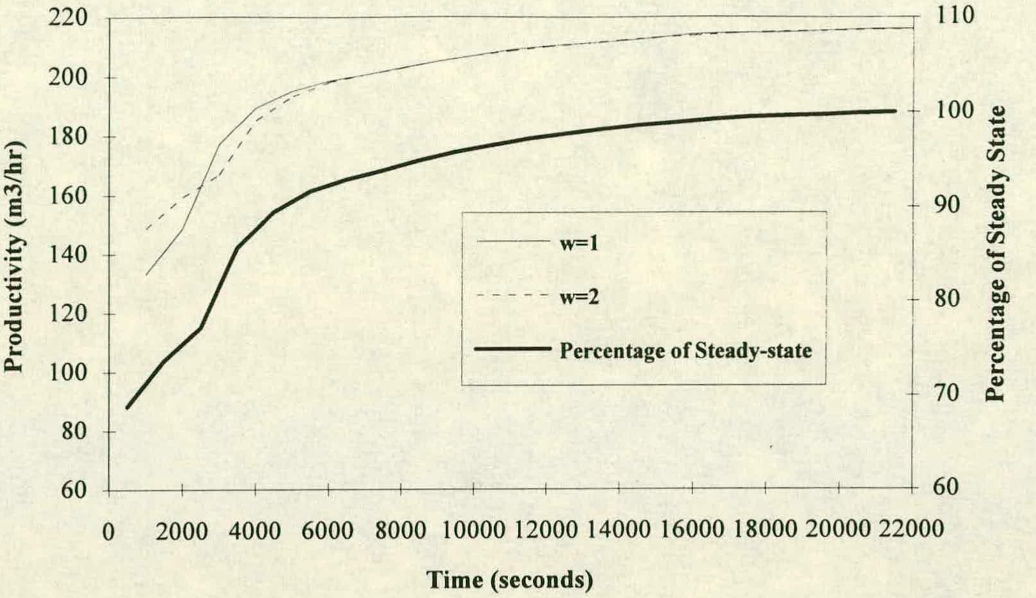


Figure 5.16 Moving averages for simulation output (haul length = 800m, exponential initial arrival times).

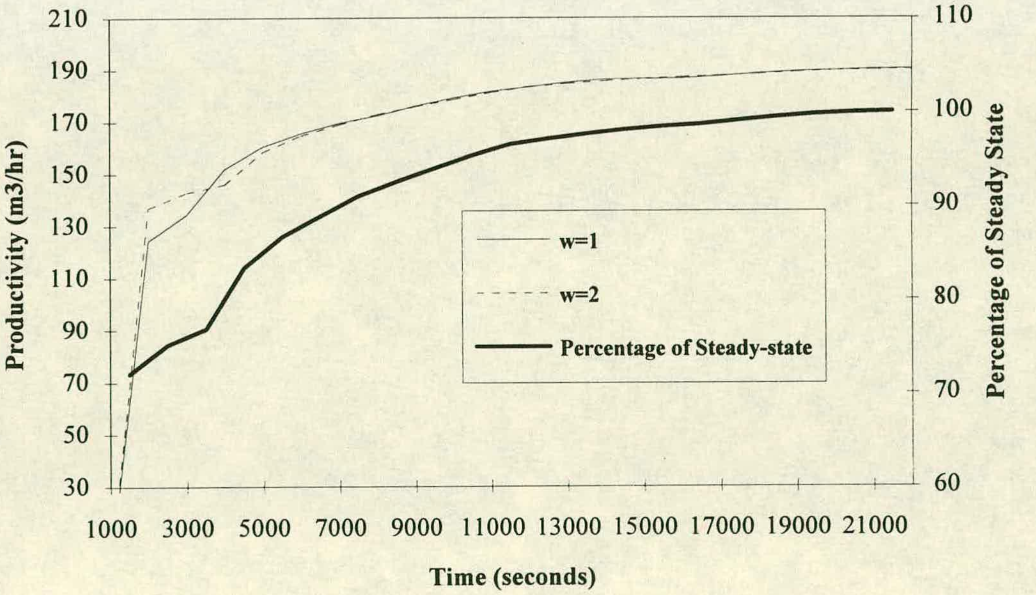


Figure 5.17 Moving averages for simulation output (haul length = 1800m, exponential initial arrival times).

steady-state line. This is calculated by taking the production at each value of *i* as a percentage of the production at *i*=22,000 seconds. (It is assumed that the production has reached steady state at this time.) This extra plot allows the results from different situations to be compared as the steady-state production may differ. Figure 5.14 also includes the output for a single replication so that the smoothing effect of Welch's method can be seen. Table 5.6 shows how the moving averages are calculated. The values in this table are taken from figure 5.14 (800m haul with zero arrival times) and it can be seen that, for example, the moving average for *i*=20,000 seconds is the average of the averaged replications for *i*=18,000 to *i*=22,000 seconds. The results from the four situations are summarised in figure 5.18 and table 5.7.

The broken lines in figure 5.18 represent the output for the zero initial arrival times; the solid lines represent the exponential initial arrival times and it can be clearly seen that the alternative initial conditions do not have the desired effect of reducing the transient time. The best situation occurs when the initial arrival times are zero and the haul length is short: in this case the length of the transient period is approximately 16,000 seconds. This contrasts with a long haul length with exponential initial arrival where *l* is approximately 19,500 seconds or nearly one hour later. Why is this? If the results from simulations with the alternative initial conditions are looked at, it is seen that the utilisation of the loader is lower than if a queue full of empty trucks is waiting, especially in the first few cycles. Maximum output from an operation is dependent on the loader and so the conclusion can be made that an initial queue of trucks speeds up the process from start to steady-state production. It is unclear, however, why the length of the haul should affect the transition length.

What is clear, however, is that even in the best situation, an earthmoving operation will rarely run for nearly 4½ hours without stoppages of some kind. The output of an earthmoving operation can therefore be considered transient and the fact that steady-state will probably not be reached must be taken into account when using Trucksim for estimation purposes. Table 5.7 shows the production as a percentage of the steady-state production at various times in the simulation run. The following procedure can be used when estimates of production are made:

i. Simulation runs are always made at one particular time length, say 10,000 seconds (2 hours 47 minutes).

	Acceptable Cut	Unacceptable Cut
Bucket Swing Time	21 secs	24 secs
Passes per Load	6	7
Load Time	105 secs	144 secs
Spot Time	26 secs	26 secs
Dump Time	60 secs	60 secs
Bucket Volume (bcm)	2	1.72
Volume per Cycle (bcm)	12	12

Table 5.5 Operation conditions for production conditions in figs. 5.12 & 5.13

	Time (seconds)									
Replication	1000	2000	3000	4000	...	18000	19000	20000	21000	22000
1	138	220	218	229	...	260	252	253	253	253
2	175	204	218	239	...	250	255	255	253	253
3	188	202	229	228	...	257	253	251	257	259
3	138	223	229	232	...	253	255	254	254	255
4	171	192	219	240	...	251	251	256	252	254
5	174	209	229	232	...	255	258	253	256	257
Averaged	164.00	208.33	223.67	233.33	...	254.27	253.93	253.70	254.15	255.10
Moving average (w=1)	186.17	198.67	221.78	230.56	...	254.39	253.97	253.93	254.32	254.63
Moving average (w=2)	198.67	207.33	212.80	227.60	...	254.21	254.21	254.23	254.22	254.32

Table 5.6 Moving averages for simulation output (figure 5.14)

				Production as percentage of steady-state at time =				
Figure	**Haul Length**	**Intialisation**	***l***	**1hr 7mins**	**1 hr 40 mins**	**2 hrs 47 mins**	**4 hrs 10 mins**	**5 hrs 30 mins**
5.14	800	zero	4h 27m	91	93	98	100	100
5.15	1800	zero	4h 43m	84	92	97	99	100
5.16	800	exponential	5h 30m	86	92	96	99	100
5.17	1800	exponential	5h 53m	77	87	94	98	100

Notes:

'Zero' initialisation: all trucks start in the queue at time = 0

'Exponential' initialisation: trucks start in queue at a random exponential time after time = 0

Steady-state production taken as being at time = 22000 seconds or 6 hours and 7 minutes

Table 5.7 Summary of results for transient behavior of simulation output study

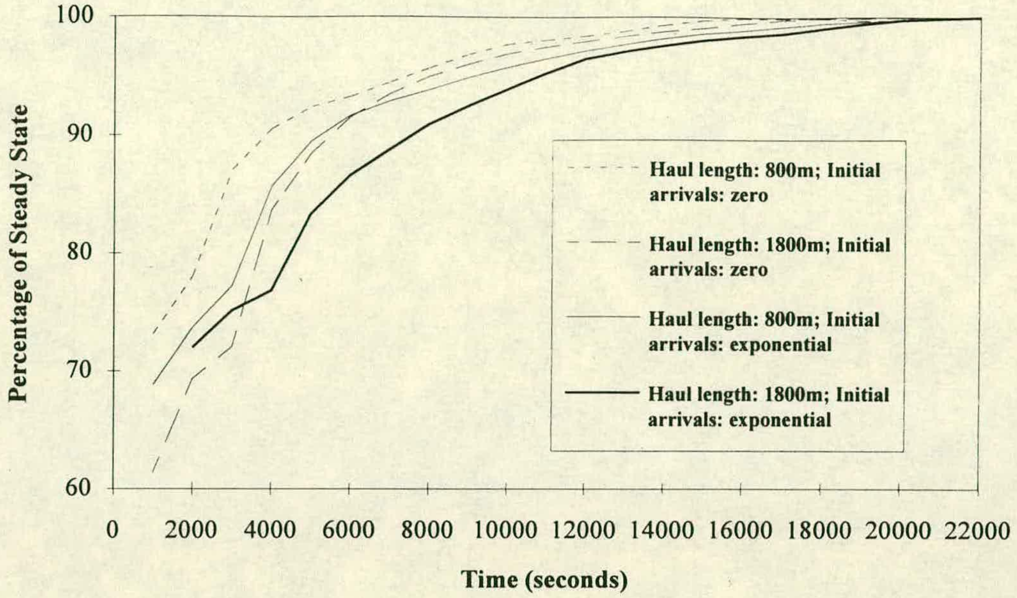


Figure 5.18 Percentages of steady-state for figures 5.14 to 5.17

ii. This production is then reduced by a factor dependent on the maximum probable duration of an earthmoving shift.

iii. For example, say an estimate is required for an operation with a short haul length with zero initial arrival times. Simulation runs are made with a time limit of 10,000 seconds. This production is 98% of the possible steady-state production (from table 5.7) or:

$$P_{10,000} = 0.98 P_{ss} \tag{5.3}$$

If the maximum probable duration is set at 1 hour and 40 minutes (6000 seconds) then the production from such an operation will be:

$$\begin{aligned} P_{6,000} &= 0.93 P_{ss} \\ &= 0.93 \frac{P_{10,000}}{0.98} \end{aligned}$$

$$P_{6,000} = 0.95 P_{10,000} \tag{5.4}$$

The factor can be calculated using table 5.7 as a guide for all combinations of conditions and shift length. Keeping the simulation run length constant for all replications means that estimates of probable shift length can be made at a later stage or amended on site. It would be bad practice to simulate using a short run length and then extrapolate for longer shifts due to the higher variances of the replication results at lower times (this can be clearly seen from table 5.6).

5.6 Conclusions

The conclusions that can be drawn from chapter 5 are:

1. Discrete-event simulation can be carried out using a spreadsheet package. Although this indicates some interesting features in the simulation output, it has a few disadvantages. Firstly, the spreadsheet is cumbersome to use: variates need to be generated first and then pasted into the sheet before calculation can take place. This means that a spreadsheet simulation takes a long time to run: one hour of actual time takes about a minute to simulate which is too long for the purposes required. Another disadvantage is that separate spreadsheets are required for different resource levels.

2. For a dedicated program, three methods can be used: simulators, simulation languages or general purpose languages. It was decided that for the purposes of this work, the general purpose language C should be used. This was mainly for reasons of cost and that more assistance would be available. It is also a general belief that general purpose languages help the modeller understand the subject better.
3. The program written provides 8 types of output (such as productivity, utilisation and match factor) from 7 types of input (component time parameters, number of trucks and excavator parameters). The program cannot simulate for more than one excavator.
4. Trucksim provides estimates for bunching which could not be provided by the deterministic model outlined in chapter 3. This should provide more accurate estimates of the operation productivity.
5. Trucksim can be used to draw up production charts to be used at an estimating and construction stage. These will provide a quicker and easier way of estimating or checking earthmoving production.
6. Simulation output (and hence by inference actual earthmoving output) has a transient (or warm-up) and a steady-state phase.
7. Smallest transient length occurs for short haul lengths and when all trucks start in the queue at the start of the simulation run. Longest transient lengths occur when the haul length is long and the trucks start with random arrival times.
8. The transient length has been shown to be not less than 4½ hours. It can be assumed therefore that an earthmoving operation will not continue, without stoppages, until steady-state production is reached. Estimates from Trucksim therefore need to be factored according to how far along the transient curve the operation is expected to go.

This chapter has examined the two different methods of carrying out simulations on an earthmoving system and ways in which the Trucksim program can be used in practice. The importance of allowing for the transient behaviour of this systems output has been outlined and transient lengths have been calculated for various operating conditions. The next step in the simulation procedure is to validate the model and the program used. Results from the simulation package cannot be used with confidence until it has been shown that simulation model accurately represents the system it has been designed to replace.

Chapter 6

Validation of the Simulation Model

If a model is to give acceptable results which match those found in real life, its validity must be established. This chapter discusses various aspects of validation and outlines the steps which have been taken to determined whether or not the simulation model is valid.

6.1 Introduction

Validation is perhaps the single most important step in a simulation study. Without some degree of validation, results from simulation runs are worthless and cannot be used with any confidence when applied to a real world situation. For this reason, the subject of model validation occurs regularly in simulation and model literature with philosophical discussions of the subject appearing at least as far back as 1935 (Robbins, 1935). Many authors have summed up the importance of this subject (for example, Fishman and Kiviat, 1968; Emshoff and Sisson, 1970 and Quade, 1980) with one of the most pertinent from Sprowls, 1964. When referring to computer models that had not been validated, Sprowls commented that:

> "I am prepared to look at each of them as an interesting isolated case which can be described to me but from which I shall draw no conclusions."

The process of validating (or attempting to validate) the model described in this thesis is no less important than for any other model. An earthmoving contractor attempting to estimate the production and resources for a multi-million pound contract will not even consider the results from the Trucksim program unless he is convinced that these results are *credible* and that there is no chance that the model will cause expensive assumptions and errors to be made. Many models suffer from the disadvantage that they represent a system that is not real, that is, in the sense that actual data cannot be retrieved, for example a model of an economical or political policy or of a weapons system for some future war. Establishing the credibility of the earthmoving system should therefore be a relatively easy task because there is a large amount of historical data that can be compared to the simulated data.

The first part of this chapter will briefly discuss the different aspects of validation and associated processes. These include evaluation, credibility, documentation, verification and various aspects of validation itself. The second part will outline the steps taken and results gained in a validation study of the earthmoving system described in this thesis.

6.2 Terminology used in the assessment of simulation models.

The literature in which references are made to the subject of validation is very wide ranging and as such a jargon of related terms has been developed over the years. Some of these terms, such as verification and validation, have become confused (indeed, Naylor and Finger have two versions of the same article (1967 and 1971) which use the term 'verification' in one and 'validation' in the other to mean the same thing). This section intends to clarify the whole subject area of validation and bring together some of the ideas that have been proposed.

6.2.1 Evaluation

Model evaluation is a process generally restricted to large scale models but its ideas can be used for smaller models to assess the successfulness of a simulation study. Gass and Thompson (1980) give the basic details of the evaluation process developed by the U.S. General Accounting Office. The intention of model evaluation is to "[set down] guidelines ... for a team which evaluates the model and assesses its results by using an established set of criteria to accumulate evidence regarding the credibility and applicability of the model." They go on to say that "evaluation seeks to improve the model's usefulness by identifying its strengths, weaknesses and appropriate uses as explicitly as possible." For a large scale model, a team of assessors, who could be either connected or independent from the system under study, would follow these guidelines and give a written report to the client detailing their opinions on how successful the simulation study is. The following 5 stage approach could be used:

i. Documentation
ii. Validity
iii. Computer model verification
iv. Maintainability
v. Useability

For smaller scale models, the full approach would be both unnecessary and costly but it complements the procedure detailed in section 5.4 on the simulation procedure.

6.2.2 Credibility

If the ultimate aim of a simulation study is to implement more efficient, cost effective and predictive methods into an industrial situation then only when the credibility of the model, or its ability to demonstrate improved efficiency and predictivity, has been established will it be accepted by the management of the particular system. Therefore even if the modeller himself is happy that the model is as valid as it can be he must convince the management of this. The modeller must present the results of the validation study in such a way that the ability of the model to carry out its aims is in no doubt. One possible way to do this is to present the management of the system two sets of results, one being historical data from the actual system and the other being simulated results - if the simulated results can be detected from the other then credibility is in doubt. This test is also known as the *Turing test* after A.M. Turing, one of the developers of the first digital computers at the University of Manchester. Another method available that is becoming increasingly popular is to use animation in presenting the simulation's results. This technique has the advantage in that the ‘black box’ idea of a computer simulation model (whereby data is input at one end and results come out of the other) is reduced but animation also has at least two pitfalls. Firstly, it can be too successful in that what is seen on the screen is truly believed to be what will actually happen; animation has the ability to reduce any questioning of the model. Secondly, for animation to work, the simulation runs must be slowed down to such a speed that the representations of the events can be seen. This is defeating one of the fundamental objects of simulation in that many runs can be done at high speed to evaluate different system configurations. Users of a simulation program may only be ultimately convinced of its credibility once it has been used to estimate a real system and it has been shown to produce ‘good’ results. This could, unfortunately, turn out to be a catch 22 situation. Credibility of industrial models is discussed by Carson (1986).

6.2.3 Documentation

This is the first heading in the approach suggested by Gass and Thompson (1980 and above). Documentation should:

- Outline claims for the model and detail evidence to support these claims.
- Describe the procedures taken in developing the model, why they were taken and how they were carried out.
- Be sufficient enough to allow independent analysts to replicate the model's results.

This thesis is providing all the documentation required for the simulation model although in far greater depth than is required from a user's point of view.

6.2.4 Certification

In certain instances, a model developer may guarantee that “the model yields outputs or results that are suitably accurate for a particular application.” Such a model can be considered certified but these are understandably rare.

Because a model cannot be 100 percent valid (see next section) no guarantees of its results can be made unless large restrictions are imposed.

6.2.5 Validation

There are many aspects to this procedure of testing the agreement between the model and the real world system and they will be reviewed here in three areas: validation levels, the philosophy of validation and the methods of validation. They will all be discussed with respect to the earthmoving system model.

6.2.5.1 Validation levels

Within the model as a whole, validation can be viewed on various different levels. Three of these levels are discussed here.

i. *Replicative validity.* This level of validation is achieved if the data from the model matches data already acquired from the real system. It is the easiest form of validity to achieve because it is not based upon occurrences that have not yet happened.

ii. *Predictive validity.* If a model can match data with the real system for operations that have not yet occurred then it is predictively valid. (Zeigler, 1976.) It is much more difficult to achieve than replicative validity, especially with the earthmoving model being described here, because the model is based only on historical data. The earthmoving model requires estimates of the component times to be entered in before simulation takes place. Predictive validity is therefore unlikely to be achieved.

iii. *Structural validity.* Such a model is not only predictivily valid but carries out the simulation in exactly the way in which the real system works. At this stage, the earthmoving model can never be structurally valid because many assumptions are made which are simplifications of the real system. For example, the model assumes that for every truck that is loaded there are the same number of bucket passes by the excavator, with the same amount of material in the bucket each pass. It was not felt necessary for the model to represent the real system so accurately at the development stage and this does not prevent the model from achieving replicative validity.

6.2.5.2 The philosophy of validation

Computer modelling is not restricted to industrial arenas and a large area of work has been with economic and political systems. From this field of research have come many workers who have contributed to the general philosophy behind the validation of a model of a system. Perhaps the first area of discussion by many of these authors is the acceptance that full validation of a model is a near impossibility. If a model were proved to be completely validated then it would be able to predict future performance with total accuracy, which is both unlikely and unnecessary. The analyst requires the simulation to provide estimates of the performance of the real system within the bounds of the resources available and so any model used can only be accepted for what it is - a simplification of the actual system without necessarily having a 'one-to-one' agreement between elements in the real system and the model. What validation intends to show is that the model predicts certain areas of the performance of the real system to some desired level of accuracy. The level of validation, and hence accuracy, will depend upon the importance of the results and the time and money available to produce the model. As noted by Law and Kelton (1991), the most valid model is not necessarily the most cost-effective.

Naylor and Finger (1967 and 1971) conducted a philosophical investigation of validation. They proposed that validation could be viewed on three levels:

i. *Rationalism.* This is viewing a model as a series of basic assumptions that underlie the system of interest. If all the assumptions are stated then proponents of rational validation consider the model validated, presumably because all assumptions are so basic that they "do not need controlled experiments to establish their validity." (Robbins 1935, p.80) However, to identify all the basic assumptions behind a system would be a difficult task and some are surely not so obvious that they require no testing. For example, the assumption that component times come from an Erlang distribution could not be used with confidence without some investigation into the actual data. This position on validation may hold for theories of economic or political systems but other approaches must be considered for the explicit mathematical based system being investigated here.

ii. *Empiricism.* This is perhaps the complete opposite of rationalism in that the basis of validation begins with facts, not assumptions. (Blaug, 1962. pp. 612-613). An empirical approach will not allow any assumptions that cannot be validated independently and in this sense, the assumptions about and selection of the probability distributions (chapter 4) for this model have an empirical base. However, both empirical and rational approaches are based upon the validity of the assumptions behind the model. The third approach proposed by Naylor and Finger is based on the results of the model.

iii. *Positive Economics.* Milton Friedman (1953 pp.14) insists that the validation of a model rests only on its ability to accurately predict the variables of interest. It is a tempting argument because to ensure that an assumption agrees with reality is a difficult task whereas the testing of the results of a model to see if they agree with actual results is relatively simple. Surely if the simulated results can be shown to be an accurate representation of the real world then this should be all the validator is interested in.

Blaug (1962, pp.612-613), however, criticises Friedman on this stance as "he seems to be saying that it makes no difference whatever to what extent the assumptions falsify reality." If this approach were rigidly adhered to in the development of the earthmoving model then the correct solution would probably have taken longer to achieve - if the assumptions the model was built on were not tested as they were proposed then it would be unlikely that a final check on the validity of the results from the model would be positive.

After consideration of all three approaches, Naylor and Finger proposed their three-stage validation procedure that has been adopted by most modelers of both economic and scientific systems. It also encourages the idea that validation is an ongoing process throughout a model's development.

6.2.5.3 Multistage Validation

This procedure proposes that all the three aspects of validation considered above are necessary in solving the problem of validation but none of them on their own. The three stages are discussed below:

i. Formulate the hypotheses and postulates for the system under study using the available knowledge such as observations, relevant theory and even intuition. This stage is based on the rationalist approach put forward by Robbins (1935) and others. Naylor and Finger however, say that "having arrived at a set of basic postulates on which to build our simulation model, we are not willing to assume that these postulates are of such a nature as to require no further validation. Instead, we merely submit these postulates as tentative hypotheses about the behaviour of the system." Naylor and Finger leave the first stage there and go on to the second stage. No mention is given as to whether or not these postulates are 'reasonable' - this should be considered before any empirical tests are carried out.

Law and Kelton therefore proposed a widening of this first stage by considering the *face validity* of the model, that is, whether or not it appears, on the surface, to be reasonable to those who are knowledgeable about the system. This can be done by consultations with system experts, observations of the system, knowledge of existing theories and similar simulation models and, as suggested by Naylor and Finger, using one's experience and intuition. Of these, there was no existing experience of the system or similar simulation models to be studied but extensive observations of the earthmoving system were carried out, as described in chapter 3. By far the most useful exercise in the assessment of the earthmoving models face validity, however, was consultation with and presentation to experts in the earthmoving system. These are as follows:

- *Tarmac Construction Major Projects Division.* This division of one of the UK's largest civil engineering contractors carries out large road building and widening schemes. It is this company whose sites were investigated and detailed in chapter 3. Therefore, opportunities were taken to speak to those involved with the earthmoving aspects at all levels. At a site level, fitters, foremen, gangers and plant operators were all willing to give their opinions on the way in which earthmoving operations were carried out and the research that is being detailed here was being conducted. For example, the proposition that productivity could be increased considerably if trucks were filled to their maximum was not put forward simply because this would increase the occurrence of breakdowns with eventual long term durability problems. The theories and assumptions were also presented to some of the top management of this company (managing director, director of estimating, director of operations and director and managers of the earthmoving division). This provided very useful feedback on the reasonableness of the model. For example, the transient behaviour of the simulation output was presented and the suggestion that productivity increased as the operation progressed was given. This was suggested not to be the case - in the audience's experience, productivity reached a peak some time after the start of the operation and then tailed off. Reasons for this were proposed as being breakdowns and general fatigue in the operators. It could be, therefore, that the underlying distributions of the cycle's time components change as the operation progresses.

- *Blackwell Earthmoving Limited.* This company is one of the largest independent earthmoving contractors in the UK who carried out the earthworks on the Twyford Down deep chalk cutting on the M3 (see 3.2.3) Site staff were obviously very knowledgeable about earthmoving procedures and were very encouraging about the work that had been carried out by that stage. In particular, the terms match and bunching factor (3.6 and 3.8.1) were unknown to this company and were considered to be very useful.
- *Caterpillar Inc., Illinois, U.S.A.* A presentation similar but in more depth to the one given to the management of Tarmac Construction was given to the research staff of this high profile company at its corporate headquarters. Caterpillar are the world's largest manufacturers of earthmoving plant and as such were ideal for the testing of the face validity of the simulation model. The research staff at Caterpillar had done similar work in the 1960s and 1970s and so were in a position to comment on this work.
- *Department of Mechanical Engineering, University of Illinois U.S.A.* This department was undertaking simulation work on hydraulic excavators at the time of the visit to Caterpillar and so the opportunity was taken to give the same presentation to a selection of the staff and students. Again, the model was accepted as reasonable.
- *Accelerator Inc., Florida, U.S.A.* This company has produced one of the few vehicle simulation programs available in the world (one of the others coming from Caterpillar) which has been used extensively throughout this research to estimate travel times from the rolling and grade resistances of the haul roads. The work was presented and much useful feedback was given with the face validity of the model again being accepted.
- *Artix Ltd., Peterlee, England.* This company is the sole licensee to manufacture Caterpillar articulated dumptrucks that have been used on most of the sites that have been studied. The company has also used simulation software in the design and production of the plant and so

were in a good position to comment on the validity of the simulation model. After presentation of the work, the face validity was again accepted.

The face validity of the simulation model, being the first criteria of the Naylor and Finger/Law and Kelton multistage validation procedure, can therefore be accepted with some degree of confidence. This does not mean, however, that the model can be given credibility - the second and third stages have to be passed also.

ii. The second stage of the multistage procedure is to validate the assumptions upon which the model is based by subjecting them to empirical testing. This is of course the second aspect of validation (6.2.5.1) suggested by Naylor and Finger (1967 and 1971) and should be done at the time the assumptions are postulated. It is no less important than the other stages in the multistage procedure - "a model with untested, untestable or refuted assumptions is at least disturbing." (Van Horn, 1971). With the earthmoving model, this stage has already been undertaken for the most major assumptions - that the observations of the time components are independent and follow Erlang probability distributions. Empirical testing (correlation plots, scatter diagrams and probability plots) for these assumptions have already been outlined in chapter 4 and the validity of the assumptions has been validated to a certain extent. It was found that the assumptions were not 100 percent applicable but acceptable within the time and cost constraints of the project. It has been proposed by Law and Kelton (1991, p.310) that *sensitivity analysis* is an important tool that can be used in the second stage of the validation procedure. Sensitivity analysis is used to determine if the simulation output changes significantly with small changes in the input parameters. Whilst the importance of this procedure is not in doubt, it surely cannot be carried out until the simulation model has been set up on a computer ready for testing. The empirical testing of the assumptions should be done "during the initial stages of model development" (Law and Kelton, 1991, p.310) and so sensitivity analysis does not belong to the second stage but after the third stage (see below) when the model is up and running. For this reason, extensive experimental analysis has been undertaken to check the sensitivity of the model and will be detailed in chapter 7.

iii. The third stage of the multistage procedure is perhaps the most definitive test of the model's validity. It is based upon Friedman's assertion that the validity of a model is demonstrated if it has the ability to predict the behaviour of the real systems variables irrespective of whether or not the assumptions the model is based on are reasonable. The remainder of this chapter will concentrate upon the comparison of the simulation output data with the real-world observations. If the two sets of data agree then the validity of the simulation model can be accepted.

6.3 Comparison of simulation output with real-world observations

This activity is, fortunately, a relatively simple task if compared with the validation of, say, a military weapons system (where real-world data would be scarce). The earthmoving model is not intended to replace the existing system but to represent it as closely as possible so that estimations of its output can be made. Therefore, data from the existing system can be used to ascertain whether or not the model is replicativily valid.

Many authors have proposed that classical statistical tests can be used to determine the 'goodness-of-fit' between the model and real systems output (for example Naylor and Finger, 1967 and Van Horn, 1971). Such tests are analysis of variance (ANOVA) tests, Chi-square test, factor analysis, and Kolmogorov-Smirnov test. However, most of these tests are not applicable because simulation output data (for a single time series) is usually highly autocorrelated (that is the observations in the output process are correlated with each other. See 4.3.2 for further discussion.) The hypothesis test that the two systems are the same would also be inapplicable simply because the model is only intended to be a representation of the actual system. Other approaches need to used therefore to determine whether the differences between the two outputs are significantly different. Three of these approaches will be reviewed here.

6.3.1 Confidence Interval approach

At first glance, this procedure would seem to be appropriate due to the existence of actual data but it does in fact suffer from disadvantages that shall be seen. The basis for this test is to compute the difference between the observations from the system and model and construct confidence limits for this difference. If the limits contain zero then the observed difference is not *statistically significant* at the confidence level the limits are set at. Even if the confidence limits do not contain zero then this does not automatically invalidate the model - differences are said to be *practically significant* only if the magnitude of the difference is large enough to question the validity of the models predictive powers. Setting a limit for this magnitude is a subjective decision and will depend on the use of the model. If the model is to be used merely for comparison purposes then any differences could be allowed to be quite large. Law and Kelton (1991, pp.319-321) discuss this further.

Construction of confidence limits, however, requires a large number of observations from both system and model (the larger the number, the greater the confidence) for the performance parameter of interest. For example, suppose the validity of the model was to be tested by comparing the productivity estimated by the model with that for a particular earthmoving operation for which, say, two hours worth of observation was available. The simulation model could be replicated as many times as necessary to obtain *n* estimates of the productivity but the observations from the real system will provide only one value for the actual productivity. If observations were repeated then a second productivity value would be in applicable because the real system parameters have changed, even if slightly, from when the first observations were taken. For this reason, unfortunately, the confidence interval approach is not applicable.

6.3.2 Time Series approach: Spectral Analysis

Spectral analysis (Naylor, et al., 1969; Fishman and Kiviat, 1967 and Blackman and Tukey, 1958) is a method of comparing system outputs which highlights one of the inadequacies of the earthmoving simulation model. The method works by taking the time series outputs (that is, the output from the system over a period of time, such as used to determine the transient length in section 5.5) from the system and the model and computing their sample (or power) spectra. This is defined as the Fourier cosine transformation of the autocovarience of the time series. If the sample spectra for both system and model outputs are known then a confidence interval can be constructed to assess the degree of similarity of the two autocorrelation functions. Therefore, this method is primarily a tool to investigate the autocorrelation of output processes, which is not of interest in the validation of the earthmoving system.

There are other drawbacks. Firstly, the sample size must be fairly large to give a good spectra estimates. Blackman and Tukey have suggested that the sample size should ideally be greater than 100 but not less than 80; these sample sizes are simply not available for the existing system. The second drawback requires that the sample data is *covariance stationary*. Such data has its mean and variance stationary over time that is unlikely to be the case for the real system. This raises questions not previously considered about the validity of the assumptions behind the model because the input to the simulation model is covariance stationary. How this affects the estimating ability of the model is difficult to say - it would depend on how much the distributions of an actual earthmoving operations component times change over the period of operation, which would be difficult to assess. Spectral analysis is therefore an approach to output comparison that is inapplicable here.

6.3.3 Inspection approach

This is the simplest method of comparing the outputs of two or more systems that requires no statistical procedures. Performance parameters are calculated from both systems and compared. Because historical data from the actual system is used, this approach will determine the level of replicative validity of the model. As with the other approaches discussed above, this method has its drawbacks but these do not prevent the procedure being carried out. If the disadvantages are born in mind then the results should yield some indication of the replicative validity. Firstly, the data is compared with no statistical tests, which means the validity is based only a subjective comparison. Secondly, each performance parameter from the real system is a sample of size one. These values are random numbers and so provide a bad estimate of the actual performance of the operation. This is not a problem with the data from the model because the estimation can be improved with repeated replications.

Law and Kelton (1991) suggest a *correlated inspection* approach (see figure 6.1) which in its extreme form requires that the model is driven by exactly the same observations as the real system. This pre-supposes that such data from a historical system is available but it is difficult to argue whether or not this indicates the predictive powers of a model. Therefore, the approach that shall be used here is to drive the model with random numbers generated using the means and variances from actual operations to produce the Erlang parameters.

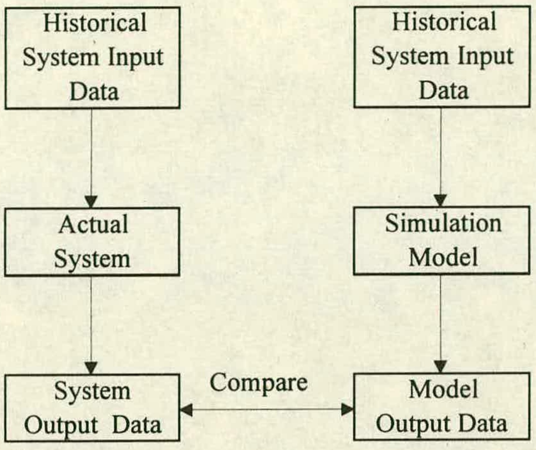


Figure 6.1 Correlated Inspection Approach to Model Validity Testing. (After Law and Kelton, 1991)

6.3.4 Correlated Inspection of Earthmoving Model Data

To attempt to satisfy the third stage of the multistage validation procedure, 12 earthmoving operations were picked, at random, from two different sites that were studied (see section 3.2). The sites were the M3 and the A52 and so they represent two extremes of road building earthmoving. The M3 consists of very large chalk cuts excavated using Caterpillar 245 backhoes and transported with Caterpillar D400 and Volvo A35 articulated dumptrucks. The average production rate on this site was approximately 300 bank cubic metres per hour, which compares with approximately 120 bank cubic metres per hour on the A52 contract. This second site used Caterpillar 235 excavators working in a sandy clay with Volvo A25 articulated dumptrucks.

Each operation was analysed to determine the mean and variance of the spot time, the load time and the travel time so that the Erlang distribution parameters could be calculated. The dump time component was assumed to have a mean of 90 seconds with a variance of 370 (standard deviation = 19s). Five performance parameters were chosen for comparison and these are match factor, bunching factor, truck utilisation, loader utilisation and productivity. Trucksim was then used to carry out simulation runs to provide estimates of the means of these 5 parameters using 5 replications for each operation to give a better estimate. It could be argued that since the actual operation can only give an estimate sample of size one then like is not being compared with like. However, if more than one replication is done then any differences due to the randomness of both outputs can be eliminated to a certain extent from the simulation output. The replications were stopped at time = 10,000 seconds and the averaged results were not adjusted for transient time (section 5.5.2) as the total shift time of the observed operations was unknown. It is felt, however, that the uncorrected results are good enough as they stand and any further adjustments would increase validity confidence. Table 6.1 shows the results of this validation study and it can be seen immediately that there is a strong similarity between the simulated and observed performance parameters. These results have been plotted against each other for clarity and can be seen in figures 6.2 to 6.6.

Figure 6.2 shows the observed and simulated match factors plotted against each other for the 12 simulated operations. Perfect match would result in a straight line through the origin and this line has been shown dotted for comparison. It can be seen that the simulation model is a very good estimator of match factor with 2 out of 12 simulations providing the exact value

Type	Match Factor	Bunching Factor	Truck Util.	Loader Util.	Prod.
A52 Contract					
Sim.	1	0.91	92	90	142
Obs.	1.05	1	96	86	142
diff.	-4.76%	-9.00%	-4.17%	4.65%	0.00%
Sim.	0.84	0.95	97	80	108
Obs.	0.84	0.94	95	76	121
diff.	0.00%	1.06%	2.11%	5.26%	-10.74%
Sim.	0.35	0.94	97	33	59
Obs.	0.31	1	100	33	66
diff.	12.90%	-6.00%	-3.00%	0.00%	-10.61%
Sim.	0.55	0.96	98	53	121
Obs.	0.49	1	100	44	115
diff.	12.24%	-4.00%	-2.00%	20.45%	5.22%
Sim.	1.38	1	75	100	117
Obs.	1.11	0.86	81	92	110
diff.	24.32%	16.28%	-7.41%	8.70%	6.36%
Sim.	1.13	0.98	88	97	115
Obs.	1.13	1	94	62	112
diff.	0.00%	-2.00%	-6.38%	56.45%	2.68%
M3 Contract					
Sim.	0.85	0.92	93	78	354
Obs.	0.86	0.92	92	77	343
diff.	-1.16%	0.00%	1.09%	1.30%	3.21%
Sim.	1.27	1	81	99	274
Obs.	1.19	0.99	91	97	288
diff.	6.72%	1.01%	-10.99%	2.06%	-4.86%
Sim.	0.6	0.92	94	55	202
Obs.	0.58	0.98	99	52	223
diff.	3.45%	-6.12%	-5.05%	5.77%	-9.42%
Sim.	0.96	0.9	91	86	386
Obs.	0.89	0.92	94	74	381
diff.	7.87%	-2.17%	-3.19%	16.22%	1.31%
Sim.	1.06	0.91	88	91	289
Obs.	1.04	0.93	95	93	299
diff.	1.92%	-2.15%	-7.37%	-2.15%	-3.34%
Sim.	0.78	0.9	92	70	272
Obs.	0.76	0.92	93	70	278
diff.	2.63%	-2.17%	-1.08%	0.00%	-2.16%

Table 6.1 Comparison of 12 observed and simulated operations

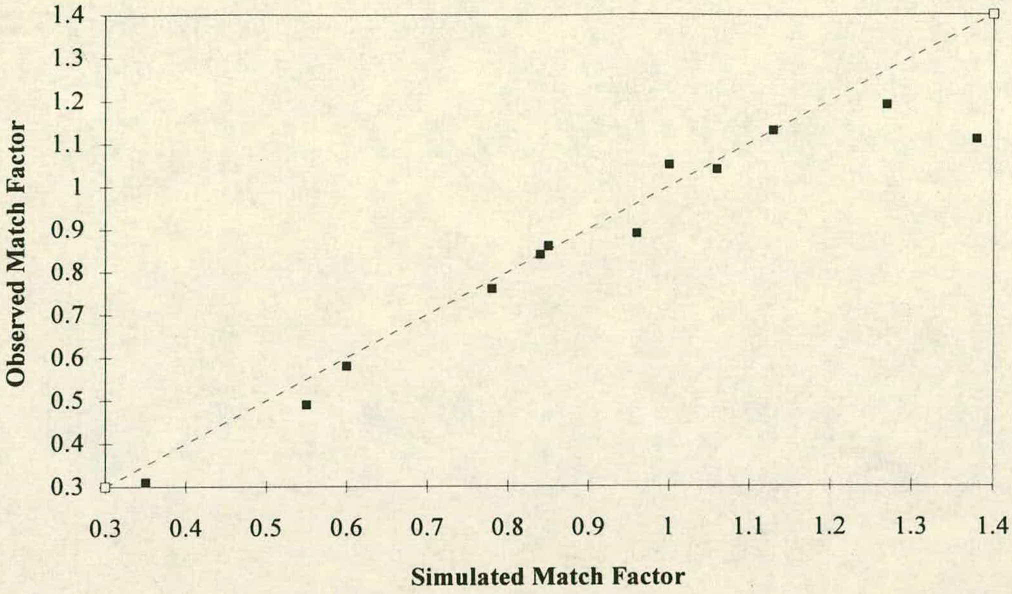


Figure 6.2a

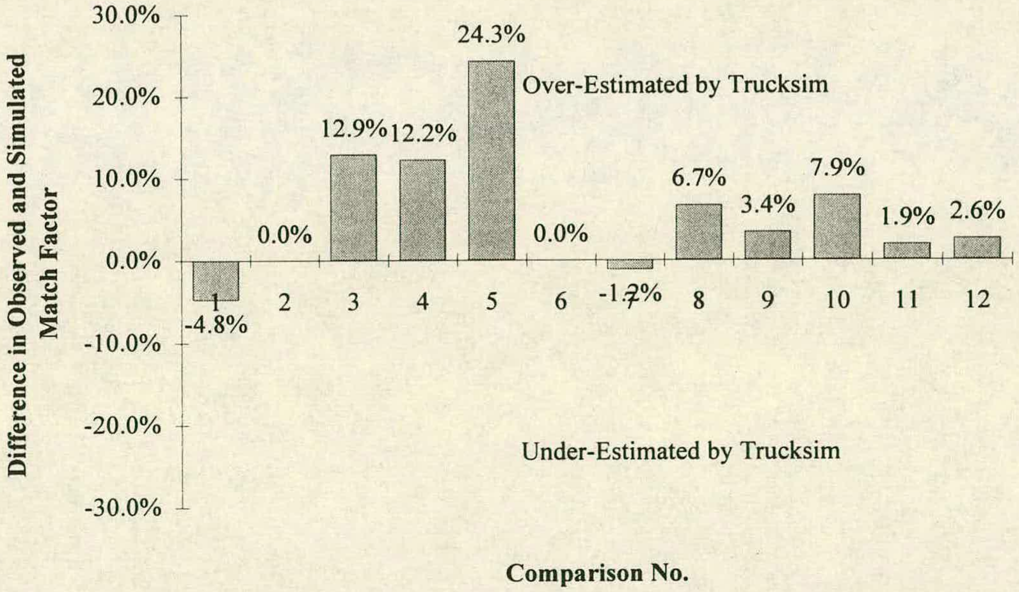


Figure 6.2b

Figure 6.2 Validation results for match factor obtained from the observations. It can also be seen from such plots whether or not the model has a tendency to over or under estimate. If the plotted points appear above the line then this indicates over estimation. It can be seen from figure 6.2 that Trucksim slightly over estimates match factor in general but not by a large amount. The largest match factor difference observed was 24.3%.

Figure 6.3 shows the same type of plot for bunching factor and as can be seen this performance parameter is poorly estimated by trucksim (although the method of plotting will amplify differences if the range of the results is low - in this case, the range is only 0.1). The large number of points above the line indicates that Trucksim tends to under estimate bunching factor although the largest difference being an over estimate of 16.3%. (This point has not been shown on the plot.) Many of the results showed an observed bunching factor of 1, which is zero bunching. It is felt, therefore, that the observation procedure is not particularly sensitive to different levels of bunching. It is unfortunate that this is the case as simulation is one of the only ways to calculate bunching in cyclic operations but it is also fortunate that this factor is under rather than over estimated.

Figure 6.4 shows the plot for the truck utilisation factor. Again, the plot amplifies differences (the largest difference is -11%) so it can be said that Trucksim is a fairly good estimator of this important performance parameter- with trucks costing up to £50 per hour it is important that the utilisation of the trucks is kept high. The plot shows that Trucksim tends to under estimate truck utilisation.

Figure 6.5 has the results for loader utilisation. Apart from one point (where the difference is an over estimate of 56.5%) there is strong similarity between the results. It can be seen that loader utilisations tend to be over estimated by Trucksim, which could be expected as truck utilisation is under estimated - it is natural that a low truck utilisation should increase loader utilisation.

Figure 6.6 shows the results for productivity, which is perhaps the most important performance parameter. High production in an earthmoving operation is paramount as this affects both the cost and duration of a project. If Trucksim could not estimate production with a fair degree of accuracy then the development of this simulation model could be considered unsuccessful. However, it can be see that there is no strong tendency either way to over or under estimation with the largest difference being 10.7% under estimation.

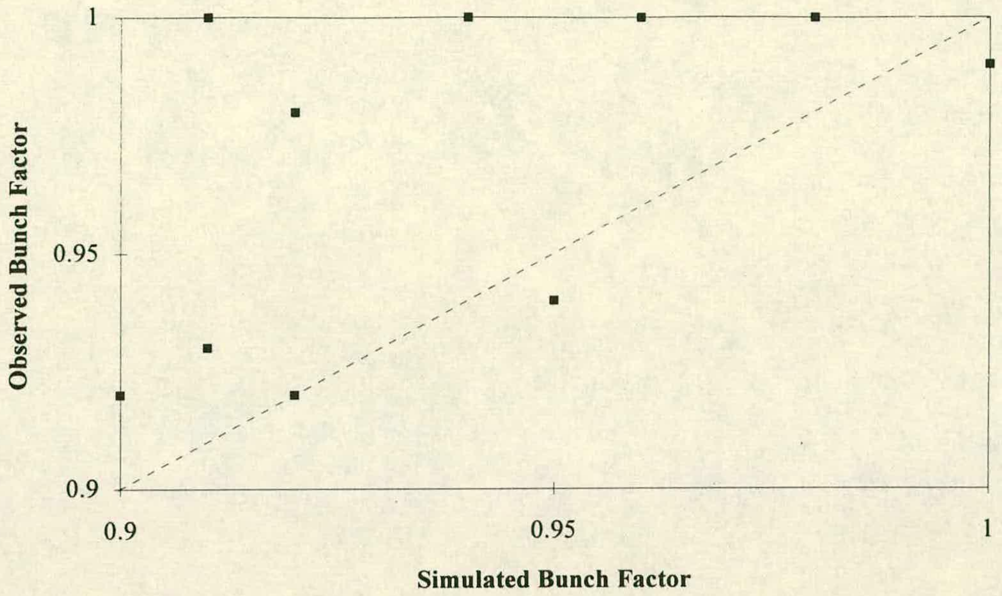


Figure 6.3a

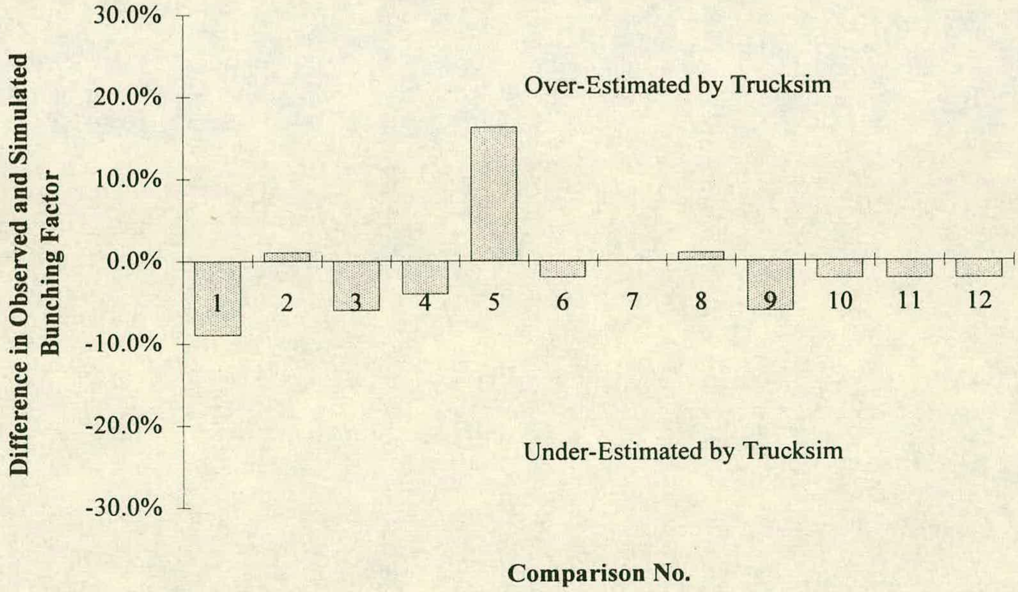


Figure 6.3b

Figure 6.3 Validation results for Bunch Factor

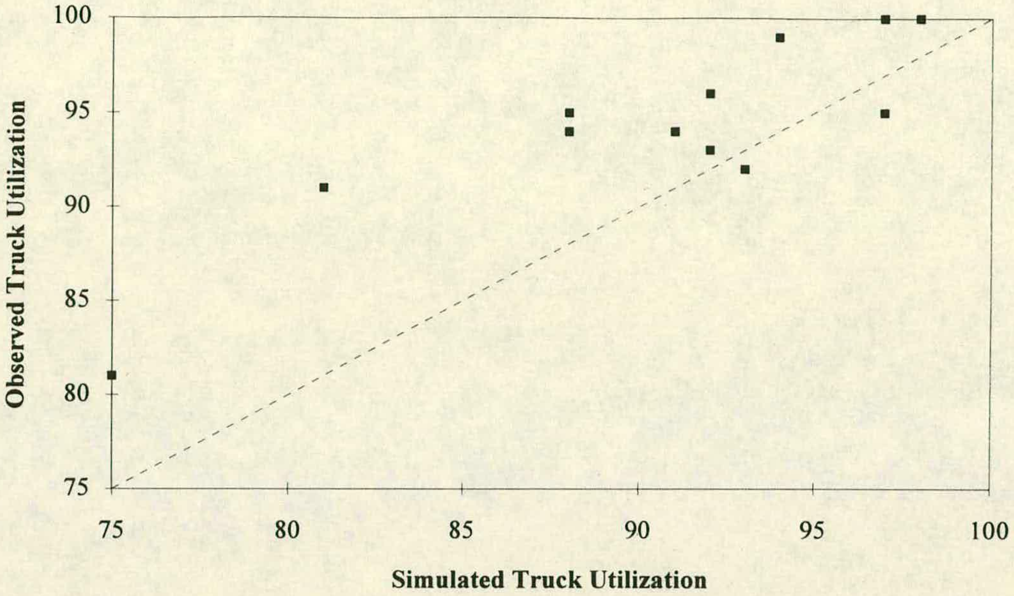


Figure 6.4a

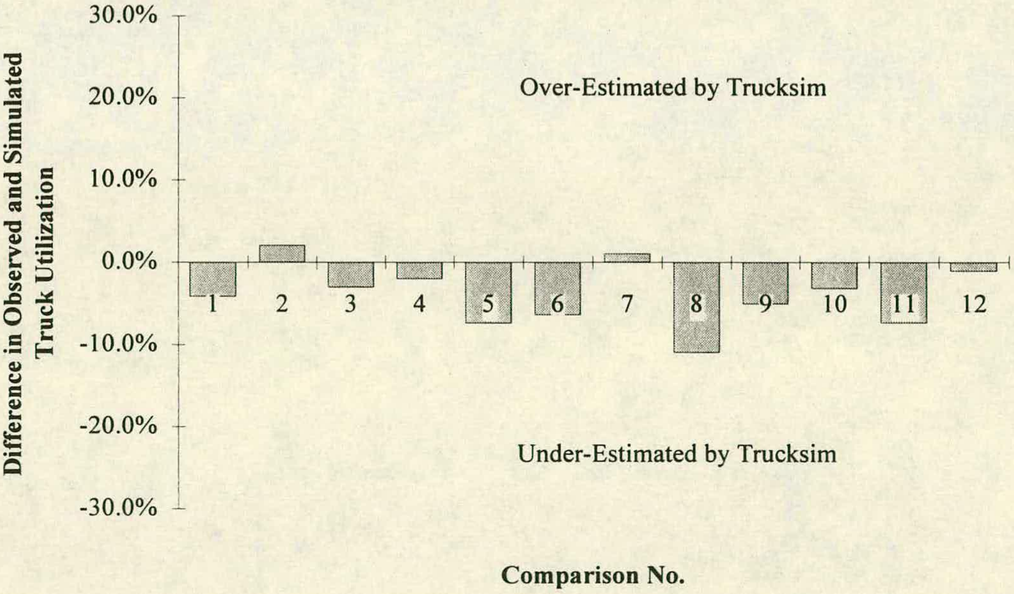


Figure 6.4b

Figure 6.4 Validation results for Truck Utilization

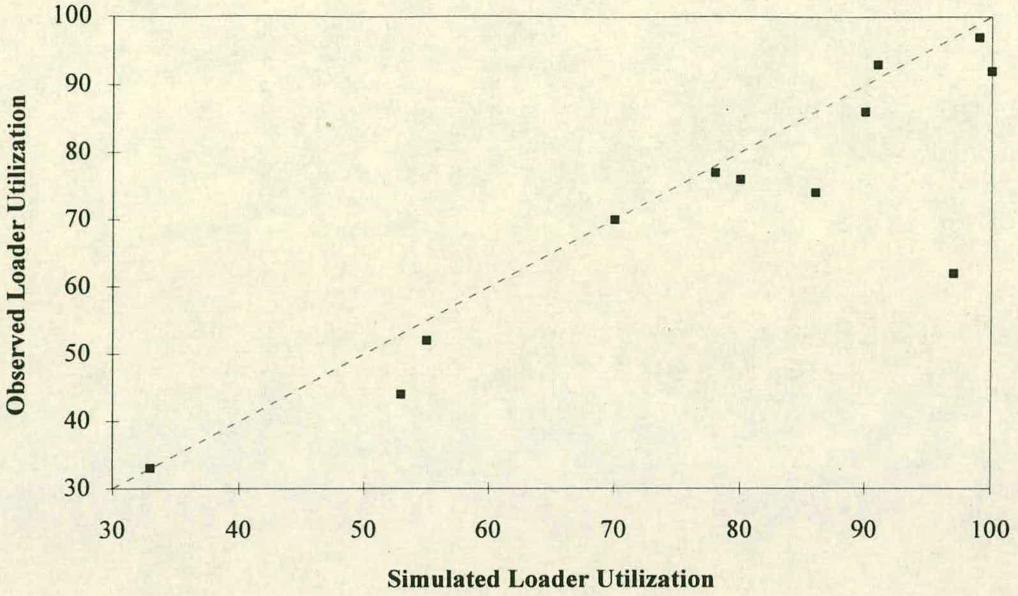


Figure 6.5a

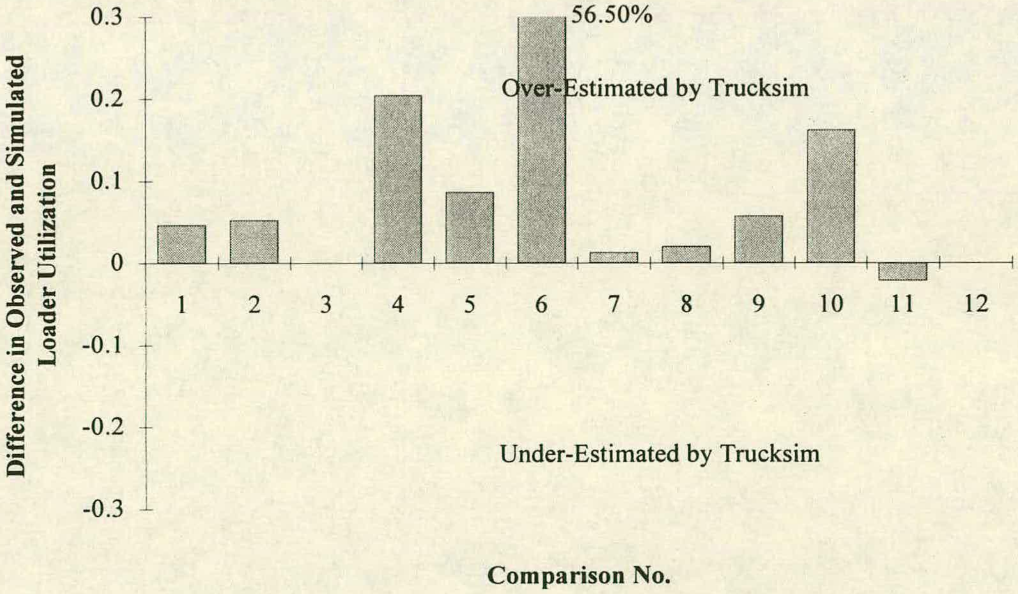


Figure 6.5b

Figure 6.5 Validation results for Loader Utilization

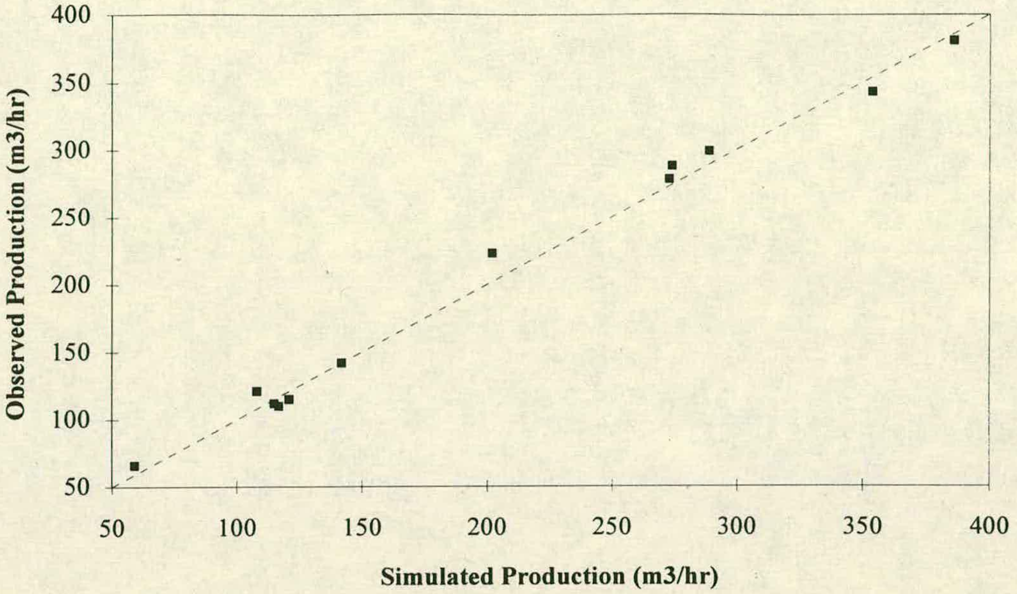


Figure 6.6a

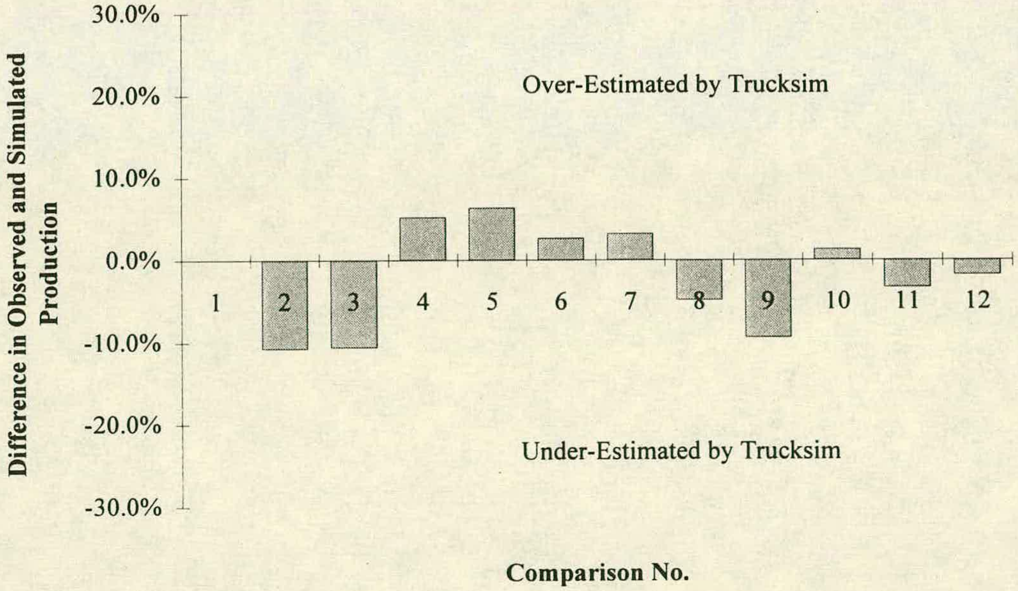


Figure 6.6b

Figure 6.6 Validation results for observed production

6.4 Conclusions

This chapter has first set out guidelines for validation that have been set down over the past 50 years or so by various different authors and, secondly, outlined how these guidelines have been used to attempt to validate the simulation model. The main question that needs to be answered at this stage is "is the model valid?." This has to be answered in several steps.

1. It is perhaps impossible to completely validate any model as a model, by definition, is only a representation of a real system. A model is therefore satisfactory if the discrepancies the validation procedure has highlighted are deemed to be acceptable. Secondly, there are at least three different levels of valid model (see section 6.2.5.1): structural, predictive and replicative validity. A model can still be acceptable even if all three levels are not achieved.

2. It can be said with certainty that the earthmoving model is not, and will never be, structurally valid as it stands at the moment. The model makes too many assumptions about the way the real system works for exact representation of simulated operations. This is however not a disadvantage as such a level of validation was never intended at the outset.

3. The earthmoving model can also not be considered predictivily valid. At this stage, the operations for which Trucksim has been used for estimation purposes have not been carried out and so this level of validation has yet to be shown. It is unlikely that full predictive validation will be achieved as this requires predictive abilities beyond the power of any simulation tool, such as determination of future ground conditions, weather and any other unexpected delays. It also requires the average time components of a future operation - the accuracy of these estimations can only be improved as a database of the way various vehicles behave is built up over a number of years.

4. The model can, however, be considered replicativily valid. This has been shown by the results in the previous section. Section 6.2.5.3 has also shown that high face validity has been achieved and chapter 4 has shown that the main assumptions behind the model are also valid.

Chapter 7

Experimental Analysis of the Earthmoving System via the Simulation Model

The simulation model, now validated, can be used to perform experiments which will yield results similar to that which would happen with the real system. This chapter outlines the factors and responses of the earthmoving system and two methods by which the experimental results can be shown

7.1 The simulation model as an experimental platform

The subject of experimental analysis is not common in civil engineering projects; such activities conjure up images of test tubes, Bunsen burners and laboratory mice. Indeed the theory of experimental design and analysis owes much to the fields of biology and chemistry via journals such as *Biometrics* and *Biometrika*. However, experimental procedures need not be restricted to the laboratory but to any system where the inputs can be varied. The simulation model described in this thesis can therefore act as surrogate to perform quasi-experimental analyses of the earthmoving system.

The real earthmoving system is extremely complex which is one of the reasons why it is so notoriously difficult to estimate its output. The model of this system has been developed to be much simpler but still requires many different inputs set at an almost infinite number of levels to obtain the ‘answer’ (or estimate of the answer) for a particular operation. This model has been shown to have an acceptable level of replicative validity and so can therefore be used with confidence to experiment, with the aim of determining which of the many inputs to the system the output is sensitive to. This type of investigation, sometimes called *sensitivity analysis*, is carried out to some degree in almost all simulation models, but it is felt that it is very important that an in depth analysis is carried out here for another reason to the usual one (that of determining which aspects of the system need to be modelled carefully). The model will not be able to provide totally accurate predictions of how an earthmoving operation will proceed and so site control will be necessary to reduce costs and improve efficiency. If it can be determined which of the factors which are input to the system (e.g. spot time) affect the output the most then site control can be concentrated on these. It may turn out that some factor which has previously been considered important is minor in comparison with another. It is felt that this approach is essential to an efficient and cost effective earthmoving project.

The literature on experimental design for computer simulation (for example Myers, Khuri and Carter, 1989; Box, Hunter and Hunter, 1978; Naylor, Burdick and Sasser, 1966; Hunter and Naylor, 1970; Kleijnen, 1977; Biles and Swain, 1980; and Montgomery, 1991) usually define a system mathematically by letting:

$$Y = f(k_1, k_2, k_3 \ldots k_n) \tag{7.1}$$

where Y is a *response* of the system to certain values of $k_1, k_2, k_3 \ldots k_n$ which are the *factors* or inputs to the system. In experimental design and analysis, the values of these factors and responses are known as *levels.* A system can have more than one response each of which requires a separate function. These functions can possibly be defined, over certain ranges of the factors, as an algebraic expression called a *metamodel* which is essentially a model of a model. Such expressions are probably very difficult to define for the earthmoving system but will be discussed at the end of this chapter. *Response surface methodology* (RSM) is one way to determine the form of these metamodels and the earthmoving model has been briefly investigated in terms of the RSM later on in this chapter. This procedure was found to be unsuitable due to the large number of factors in the earthmoving system. An alternative method is *factorial design* and this method has been used to provide the main conclusions of this chapter.

7.2 Factors and Responses of the Earthmoving System

To investigate the responses from the earthmoving system, it is important that all factors are considered. If some are ignored or simply not considered to be factors then it cannot be said for certainty that the level of a particular response is due wholly to the factors being considered. A good example of this occurred with the original, deterministic, model of the earthmoving system (chapter 3). Model one did not have plant variability as a factor and hence the production rate response from this system was inaccurate. If, after investigation, particular factors do not influence the response then these can be ignored (or at least kept at one level) from future calculations. However, if all factors which influence the earthmoving output were to be considered then the experimental analysis would be an extremely complex task; the fewer the number of factors the less experimental runs need to be carried out. There are at least 16 basic factors that determine how productive an operation is which must be reduced for an efficient investigation. Many authors have suggested factor screening strategies (for example, Plackett and Burman, 1946; Kleijnen, 1975; Mauro and Smith, 1982 and Mauro, 1984) whereby n factors can be investigated with as little as n-1 experiments. However, these methods are more suitable for systems with many factors (one hundred plus) and so a different approach has been used here which is more relevant to the way the earthmoving model has been set up.

Figure 3.1 shows the earthmoving system schematically. Table 7.1and figure 7.1 shows how this system can be split into first and second order factors for the individual sub-systems. The first order factors are combined to give responses which can be considered as second order factors. It is these second order factors, of which there are 7, that are the main factors of the simulation model. These factors contribute to the main responses from the whole system which are shown in the third column of table 7.1. It can be seen therefore how the simulation model is a simplification of the actual earthmoving system as it considers none of the first order factors. The four subsystems of the model and actual system correspond with the four time components; travel, spot, load and dump. There are also three other factors to be considered which belong to none of these sub-systems.

The travel system has as its first order factors all those which contribute to the speed of the truck over the haul road. It is easy to see when looked at in this way that the simulation model described in this thesis cannot possibly consider all factors; when using the Trucksim program, separate calculations are done using the travel sub-system's first order factors to obtain the responses, which are the travel time mean and variance. These two factors to the earthmoving model, can be calculated using programs such as Caterpillar's *Vehsim* and *Accelerator* from Accelerator, inc.; and also with a knowledge of how certain types of plant perform on different haul roads. Some of the first order factors are separate systems in their own right. For example, the truck specification factor can be considered an umbrella term for all those factors affecting the performance of the truck which are not separately listed, such as tyre size and pressure, engine size, transmission type, drive system, steering system, suspension and type of brakes. Also, most dumptrucks have retarders (engine braking systems) which cannot be controlled by the driver; they cut in when the downhill gradient reaches a certain level. All these separate factors in some way influence the speed at which the truck can travel. Factor 6, the number and type of obstructions, influences the overall speed and bunching of a truck. An automatic traffic light system on a plant crossing, for example, will slow a truck down more than a manned system.

The spot sub-system, which is the part of the truck cycle where the truck leaves the queue and positions itself ready for loading, is also dependent on the specification but the speed of the truck here is not as essential. The size and type of truck (whether articulated or rigid) affect the spot time more. The first order factors for this sub-system also include load area factors such as size for manoeuvring and ground

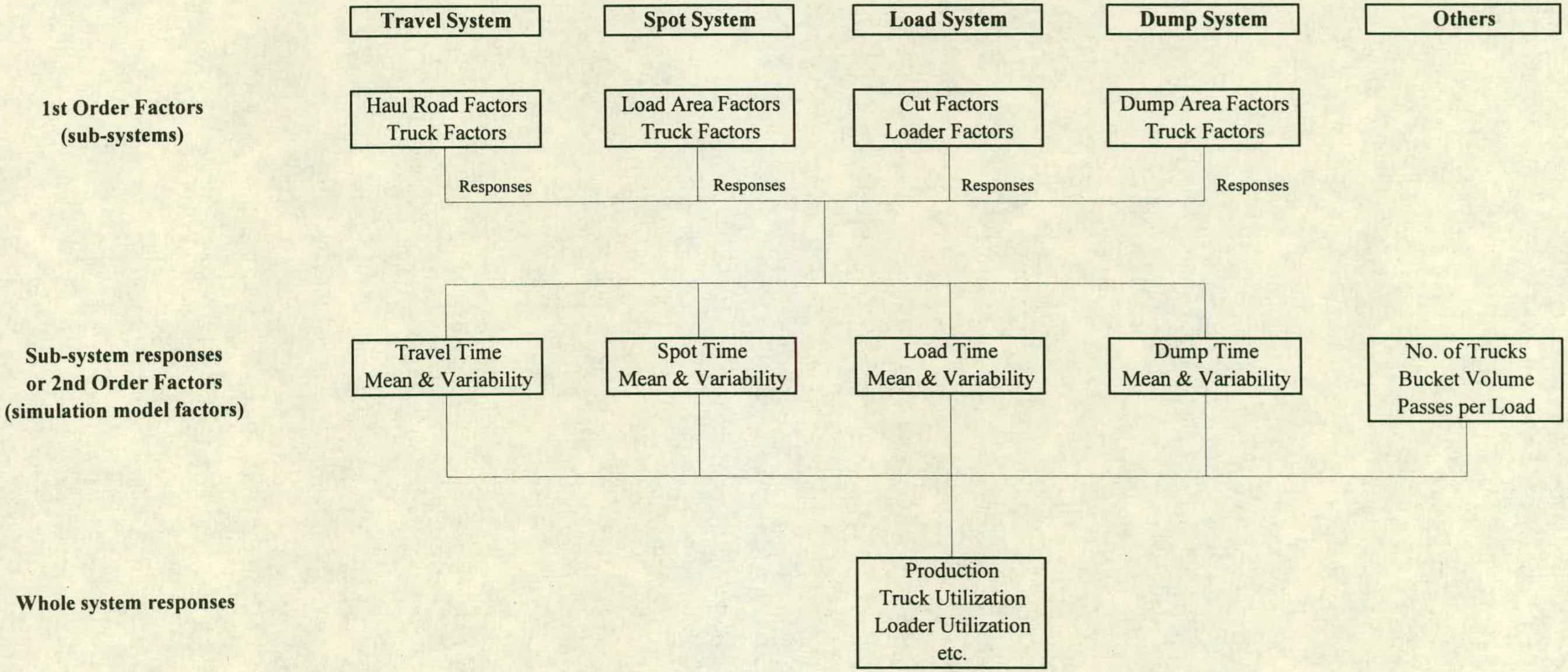


Figure 7.1 Schematic of earthmoving sub-system, factors and responses

1st Order Factors	2nd Order Factors	Responses
Travel System		1. Production (m3 per hour)
1. Haul Road Soil Parameters		2. Truck Utilisation (%)
2. Haul Road Gradient		3. Excavator Utilisation (%)
3. Haul Road Length		4. Queue Wait Time (seconds)
4. Truck Weight	1. Travel Time (mean)	5. Match Factor
5. Vehicle Specification:	2. Travel Time (variability)	6. Bunch Factor
Engine size		7. Cost per cubic metre (to excavate, load and haul)
Transmission Type		
Drive System		
Suspension System		
etc.		
6. Number & Type of Obstructions:		
Type of Plant Crossings		
Size of Bailey Bridges		
7. Driver Ability		
Spot System		
8. Load Area Size	3. Spot Time (mean)	
9. Size & Type of Trucks	4. Spot Time (variability)	
10. Driver Ability		
Load System		
11. Cut Depth		
12. Soil Parameters	5. Bucket Pass Time (mean)	
13. Type of Excavator	6. Bucket Pass Time (variability)	
14. Operator Ability		
Dump System		
15. Dump Area Size	7. Dump Time (mean)	
16. Driver Ability	8. Dump Time (variability)	
Other Factors	9. Number of Trucks	
	10.Bucket Volume	
	11. Bucket Passes per Load	

Table 7.1. Factors and Responses for the earthmoving System.

Factor	Level
Travel Time Variability	k = 40
Spot Time Mean	10 seconds
Spot Time Variability	k = 10
Bucket Swing Time Mean	16 seconds
BST variability	k = 10
Dump Time Mean	90 seconds
Dump Time Variability	k = 20
Passes per load	6
Bucket Volume	2 bank cubic metres
Number of Trucks	5

Table 7.2. Factor levels for response curve

conditions. The responses from this sub-system become the spot time mean and variance factors for the simulation model.

The load sub-system has many first order factors some of which are listed in table 7.1. The main influences on the responses of this sub-system are the type of material being excavated and the size of the area being cut. Like the truck specification, the loader will have many contributing factors which cannot be considered individually by the simulation model. For example, the type and size of the bucket, the boom length and the engine size.

The dump system is perhaps the simplest sub-system in the truck cycle. Contributing first order factors are the dump area factors (size and ground conditions for manoeuvrability) and truck factors which will be similar to those of the spot system.

In all the sub-systems, the one factor which probably has a greater influence than any other is the driver experience and skill. A good operator can make the difference between an inefficient operation and an very productive one and is therefore a very valuable commodity to the earthmoving contractor. Unfortunately, the driver ability is difficult to quantify: how can this factor be incorporated into the estimation of any of the second order factors in the same way the haul road gradient, for example, can? The best way to allow for driver ability in these calculations is to build up a database over a number of years on how a driver's experience (in number of years say) influences the speed the plant operates at. Other factors which have been classed as second order (i.e. they are input directly into the simulation model) are the number of trucks, bucket volume and passes per load. The last two of these could be considered first order factors for the load system, but they were considered important enough to stand alone.

This section has attempted to show the number of factors which influence the responses from an earthmoving system. Most of these factors are difficult to quantify, let alone experiment with, and so the term ‘second order factors ’ is used to indicate those which are factors to the simulation model. It is these factors which will be experimented.

7.3 Response Surface Designs

Response surface methodology as a means of determining optimum operating conditions was developed in the 1950s in the chemical industries. Much of the early work was pioneered by G.E.P.Box whose paper with K.B.Wilson entitled 'On the Experimental Attainment of Optimum Conditions' was one of the first to be published on the subject. Equation 7.1 set out the basics for RSM which is essentially a group of techniques used in the study of the relationships between the input factors and the output responses.

Of the eleven factors of the simulation model of the earthmoving system, suppose 10 of the factors are kept at a constant level and the eleventh is varied over a particular range of interest. If estimates of the response are calculated from the Trucksim program and are plotted against the varying factor, the resulting curve is a *response curve*. For example, table 7.2 shows the values of 10 factors and figure 7.2 shows the response curve obtained by varying travel time from 100 to 1200 seconds. The central portion of this graph is similar to the curves obtained for the production chart illustrated in figures 5.12 and 5.13; the extents of the response curve represent very inefficient production (i.e. over and under resourced) and so are not included in the production curves. The curve in figure 7.2 was obtained using 5 replications for each point and so a smooth line is obtained. Figures 7.3 and 7.4 show curves that are obtained with single replications for each point and the variability, or randomness, of the output can easily be seen. These two curves also include lines for the match factor response and represent a wider range of the varied factor. Each plot, however, suffers from the same problem as figure 7.2 in that they cannot indicate how the response behaves outside the range of interest; the difference in match factor between the two plots shows this. The two curves show how the response changes by varying two factors; travel time in figure 7.3 and bucket swing time in figure 7.4. As can be seen the responses behave in different ways: with travel time, the production reaches a maximum at MF=1.2 whereas if the bucket swing time is reduced in figure 7.4, the production does not reach a maximum but continues increasing. These two figures only show two individual 'snapshots' of the overall response function; they cannot show how the two factors work together.

This problem can be overcome if there are two main factors of interest. If we vary these two factors, keeping any other factors constant we can get a response surface i.e. a 3 dimensional representation of the factors in question for the particular range of interest. For three or more factors, a response surface cannot be graphically

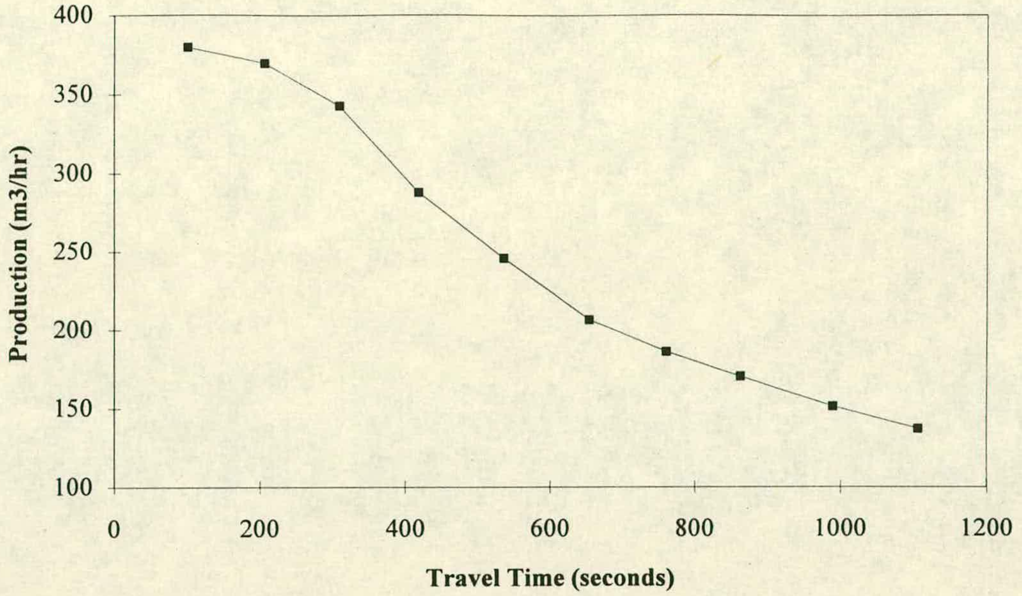


Figure 7.2 Production response curve

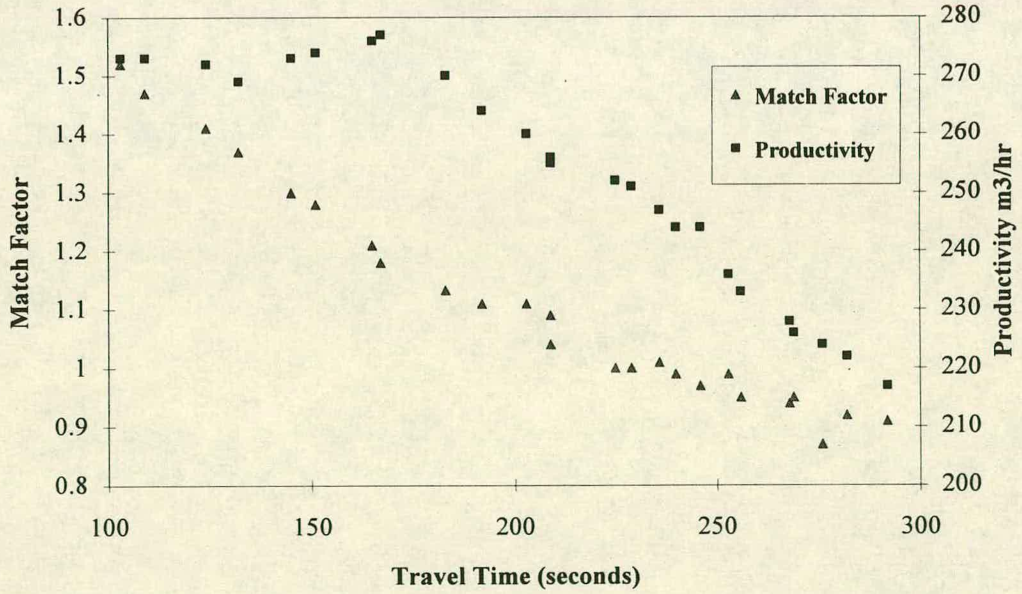


Figure 7.3 Response plots for match factor and production with varying travel times

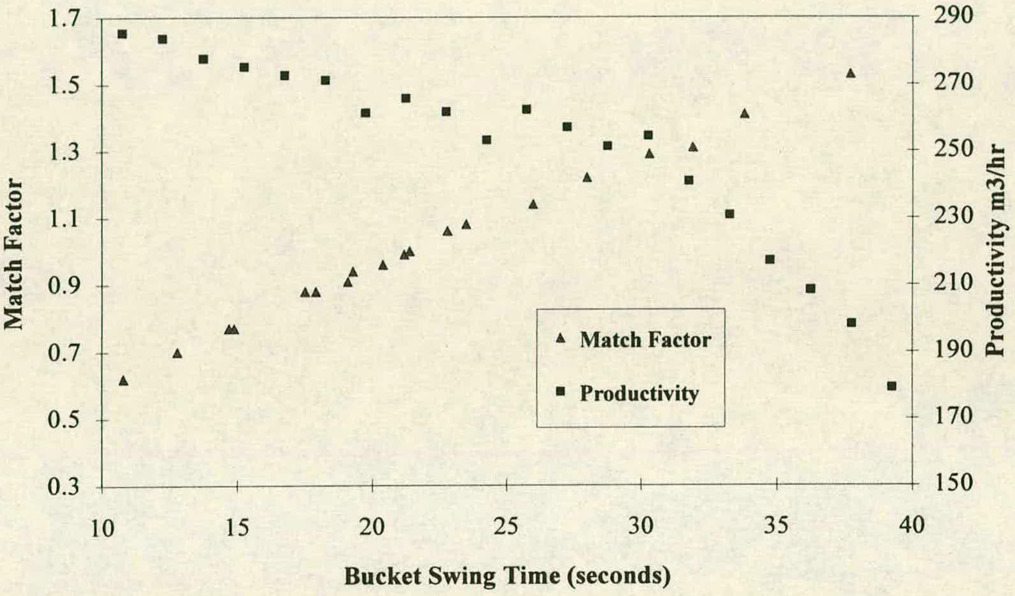


Figure 7.4 Response curves for match factor and production with varying bucket swing times

represented and the experimenter will have to be satisfied with sectional representations of the entire surface. Figure 7.5 shows a response surface obtained when spot time and travel time are varied. Spot time was varied from 10 to 22 seconds, travel time again varied from 100 to 1100 seconds and all other factors were kept at the same level as shown in table 7.2. The curve in figure 7.2 is therefore one single section through the surface shown in figure 7.5. From a brief inspection of figure 7.5 the simple conclusion that spot time has little effect on the production response could be drawn: At travel time = 100 seconds the response changes only 10% whilst at travel time = 1100 seconds the change in response is virtually nothing (0.7%). However, in this example, the travel times are set at much higher levels than the spot times and the increase in travel time is 1000% compared to only a 100% increase in spot time. This indicates that the response surface is little better than a response curve for this particular system: conclusions about how the levels of the factors *outside* the particular range studied will affect the response cannot be drawn.

Now consider figure 7.6. Here, travel time and spot time are varied again but travel time is varied from 50 seconds to 200 seconds. It can be clearly seen how spot time has a greater influence on the production response than in figure 7.5. However, what is unknown is how the levels at which the other factors have been set have affected the response and whether any *interactions* are present. For example, there may be an interaction between bucket swing time and travel time which influences the response. It is impossible to determine from these response surface designs whether such interactions exist and, if they do, whether they caused the difference in productions between figures 7.5 and 7.6.

7.4 Factorial Design

So far in this chapter it has been seen that the earthmoving system has at least 11 factors, all of which will have some effect on at least 7 responses. First attempts at response surface design have shown perhaps the most fundamental problem facing all experimentalists: which factors are the most important and over which ranges shall they be studied. Box, Hunter and Hunter (1979, p. 303) indicated a paradoxical situation: the best time to design an experiment is when it has been completed. Experimentation is therefore iterative; after each successive experiment, more will be known about the system and future experiments will be more useful. Factorial designs are useful in this respect in that they can give a 'broad picture' of the overall system and will help eliminate factors which have little effect on responses.

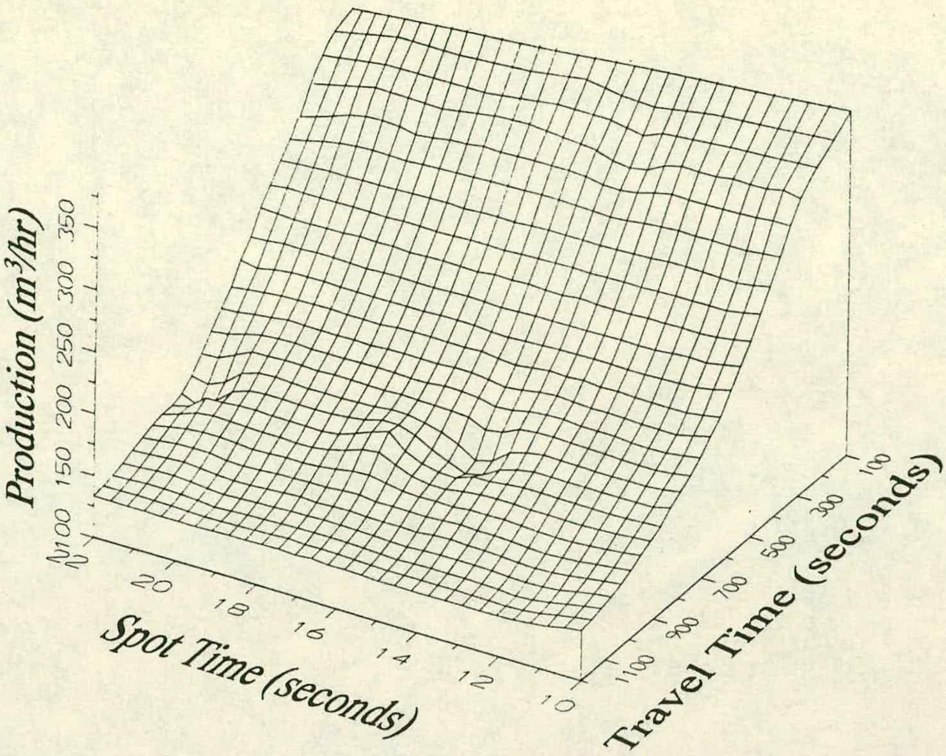


Figure 7.5. Response Surface for Spot Time and Travel Time

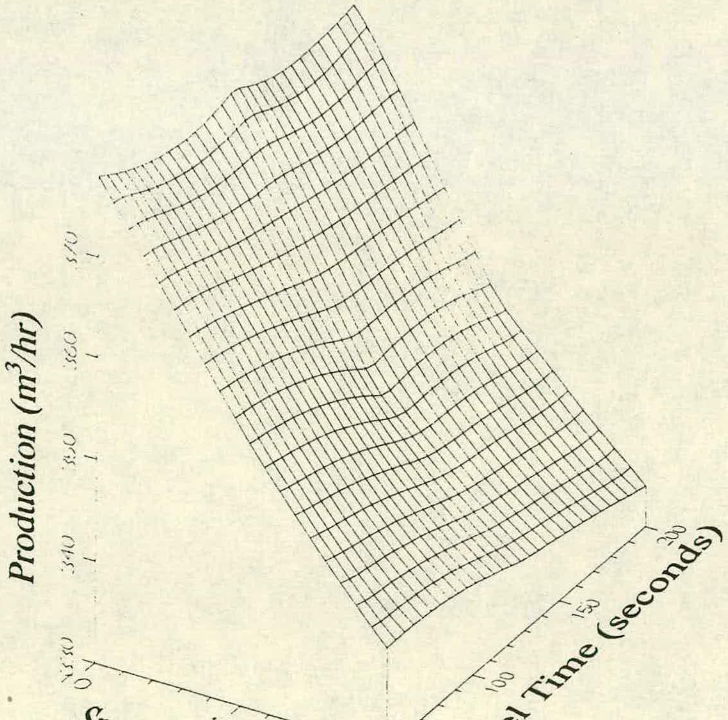


Figure 7.6. Response Surface for Spot Time and Travel Time (Reduced Travel Time)

In performing a factorial design, a set number of levels of each factor are chosen and experimental runs for all combinations of the factors at their different levels are carried out. For example, for a system with only 3 factors and 3 levels were chosen for each factor then $3^3 = 27$ experimental runs (or *design points*) would be required. Unfortunately, if 3 levels for each of the 11 factors in our model were chosen then $3^{11} = 177{,}147$ runs would have to be performed; this is clearly impractical. The simplest is 2 levels for each factor but even this requires 2048 runs which is also too many for an initial investigation. Therefore, for an initial investigation, the number of factors needs to be reduced even further than the first exercise (section 7.2) did and this can be done by grouping some of the factors.

The effective number of factors can be reduced by grouping four of the factors by making the assumption that pairs of factors (i.e. load and spot; dump and travel) have the same effect on the output. This is a fairly safe assumption because the simulation model itself pairs these time components. Referring to figure 3.1, the spot and total load times are collectively known as the 'service time'; the travel and dump times are collectively the 'back-cycle' times. [Carmichael (1987c) used queueing theory on such a refined model.] A further reduction can be made by eliminating the bucket volume from any designs. This factor clearly has an effect on the output but the assumption will be made that this is a linear relationship; in addition, any interactions between bucket volume and other factors have not been allowed for in the model. Quite simply, if the volume of the bucket doubles, the production will double.* The effective number of factors is now 6 and a two 2^6 factorial design requires 64 design points. This is a far more acceptable figure for an initial experiment.

7.4.1 2^6 Factorial Design: Initial Study

Table 7.3 shows the 2^6 experiment for the 6 remaining factors. The two levels for each factor are coded by a plus sign and a minus sign and in this case the minus sign is associated with the lower numerical value although in some cases, for example travel time mean, moving from the minus level to the plus

---

* In the real system, the size of the bucket will have some effect on the speed at which the loader can work, that is, if a large bucket is used then some of the gain in production due to a larger size will be offset by an increase in the load time.

Factor		-level	+level
1	Number of Trucks	2	6
2	Passes per Load	4	7
3	Bucket Swing Time (mean)	12	22
4	Bucket Swing Time (variability)	k = 20	k = 10
5	Travel Time (mean)	100	800
6	Travel Time (variabilty)	k = 70	k = 50

Table 7.3 Factor levels for 6 factor experiment

level will actually decrease some of the responses. It is a good convention to keep as it helps in the interpretation of results.

Runs were carried out using Trucksim at each of the 64 design points, with 11 replications for each point. Production and match factor responses were recorded for each replication and the average of the 11 replications for each design point are shown, along with the *design matrix* for the experiment in table 7.4.

The *main effect* of a factor is the average change in a response due to moving a factor from its minus level to its plus level, whilst all other factors are fixed. The average is taken for all combinations of the other factors in the design. The main effect on production of factor 1 (number of trucks) is therefore:

$$\begin{aligned} e_1 &= \frac{(P_2 - P_1) + (P_4 - P_3) + \ldots + (P_{64} - P_{63})}{32} \\ &= \frac{-P_1 + P_2 - P_3 + P_4 \ldots - P_{63} + P_{64}}{32} \\ &= \frac{-179.6 + 373.1 - 153.2 + 226.1 \ldots - 39.2 + 103.0}{32} \\ &= 82 \end{aligned}$$

Where P_n is the production response at design point *n*. By comparison between the above expressions and table 7.4 it can be seen that the effects for all factors can be computed by applying the relevant sign for the factor to the response at each design point, summing them, and dividing by 2^{k-1}, which in this case is 32. All main effects on both production and match factor for each factor are perhaps best calculated using a spreadsheet; 12 lengthy hand calculations would be prone to error.

As well as calculating the main effects for each factor, *interaction effects* between factors can also be determined in a similar way. (For full derivation of interaction effects see Box, Hunter and Hunter, 1978 or Law and Kelton, 1991). Interactions occur when the response due to one factor is dependent on the level of another factor. For example e_{25} is the interaction between factors 2 (passes per load) and 5 (travel time mean) and its value (if any) can be calculated by first adding another column to the design matrix headed 2x5. The signs for each design point in this column are simply the product of the signs in the factor 2 column and factor 5 column. This *2-factor* interaction is calculated in the same way as for main effects: summing the dot

Design Point	Factor Number						Production	Match Factor	Design Point	Factor Number						Production	Match Factor
	1	2	3	4	5	6				1	2	3	4	5	6		
1	-	-	-	-	-	-	179.6	0.49	33	-	-	-	-	-	+	176.3	0.49
2	+	-	-	-	-	-	373.1	1.50	34	+	-	-	-	-	+	372.4	1.47
3	-	+	-	-	-	-	153.2	0.70	35	-	+	-	-	-	+	150.8	0.70
4	+	+	-	-	-	-	226.1	2.18	36	+	+	-	-	-	+	225.7	2.16
5	-	-	+	-	-	-	156.9	0.67	37	-	-	+	-	-	+	154.2	0.66
6	+	-	+	-	-	-	242.0	2.08	38	+	-	+	-	-	+	242.3	2.04
7	-	+	+	-	-	-	122.7	0.95	39	-	+	+	-	-	+	120.9	0.95
8	+	+	+	-	-	-	135.7	3.16	40	+	+	+	-	-	+	135.5	3.08
9	-	-	-	+	-	-	179.8	0.49	41	-	-	-	+	-	+	173.1	0.49
10	+	-	-	+	-	-	375.9	1.49	42	+	-	-	+	-	+	361.1	1.50
11	-	+	-	+	-	-	153.6	0.78	43	-	+	-	+	-	+	148.6	0.71
12	+	+	-	+	-	-	220.0	2.22	44	+	+	-	+	-	+	218.5	2.19
13	-	-	+	+	-	-	155.4	0.69	45	-	-	+	+	-	+	152.2	0.67
14	+	-	+	+	-	-	233.5	2.14	46	+	-	+	+	-	+	233.6	2.09
15	-	+	+	+	-	-	119.6	0.98	47	-	+	+	+	-	+	118.6	0.96
16	+	+	+	+	-	-	130.9	3.18	48	+	+	+	+	-	+	131.6	3.07
17	-	-	-	-	+	-	44.2	0.13	49	-	-	-	-	+	+	43.5	0.13
18	+	-	-	-	+	-	124.2	0.40	50	+	-	-	-	+	+	122.8	0.41
19	-	+	-	-	+	-	42.0	0.21	51	-	+	-	-	+	+	42.5	0.21
20	+	+	-	-	+	-	116.3	0.64	52	+	+	-	-	+	+	115.6	0.64
21	-	-	+	-	+	-	42.1	0.20	53	-	-	+	-	+	+	42.8	0.20
22	+	-	+	-	+	-	119.1	0.60	54	+	-	+	-	+	+	116.9	0.60
23	-	+	+	-	+	-	39.6	0.33	55	-	+	+	-	+	+	39.5	0.33
24	+	+	+	-	+	-	107.0	0.99	56	+	+	+	-	+	+	104.2	1.01
25	-	-	-	+	+	-	44.0	0.13	57	-	-	-	+	+	+	43.6	0.13
26	+	-	-	+	+	-	123.3	0.41	58	+	-	-	+	+	+	121.6	0.41
27	-	+	-	+	+	-	42.4	0.22	59	-	+	-	+	+	+	41.0	0.21
28	+	+	-	+	+	-	117.3	0.66	60	+	+	-	+	+	+	114.7	0.65
29	-	-	+	+	+	-	42.6	0.20	61	-	-	+	+	+	+	42.9	0.21
30	+	-	+	+	+	-	119.5	0.62	62	+	-	+	+	+	+	117.2	0.63
31	-	+	+	+	+	-	39.7	0.34	63	-	+	+	+	+	+	39.2	0.34
32	+	+	+	+	+	-	105.3	1.05	64	+	+	+	+	+	+	103.0	1.04

Table 7.4. Design Matrix for Two Level Factorial Design

products of the 2x5 column and the response column and dividing again by 2^{k-1}. (2-factor interactions are symmetrical; i.e. $e_{25} = e_{52}$). Three- and higher-factor interactions are calculated in exactly the same way.

Figure 7.7 shows the main effects, two-factor and three-factor interactions calculated for this initial study. Four- and higher-factor interactions were negligible. It is important to *understand* that these are the results for the earthmoving system with the configuration shown in table 7.3. These effects and interactions may not be applicable for different levels of the factors. Each bar on the plot gives the average effect or interaction in moving from a minus to a plus level.

The largest main effect in this experiment is that on production by factor 5 which is the travel time; the average change in the production response when travel time is increased from 100 seconds to 800 seconds is a reduction in production of almost 120 m^3/hr. This factor also has a similar effect on match factor which is reduced by approximately 1.0 when the level moves from its minus to its plus level. However, this is not particularly surprising because the change in the levels is 700% and so a large effect should be expected. A similar effect would be expected with factor one, the number of trucks. In this case, the factor increases by 200% with a corresponding average increase in production of 98 m^3/hr. Factors 2 and 3 could therefore be much more significant because the increase in their levels is relatively much lower than the increases in number of trucks and travel time mean. Factor 2 (passes per load) has a factor level increase of 75% with a production response increase of 40 m^3/hr while factor 3 (bucket swing time mean) increases by 83% from its minus to its plus level giving a production response increase of 35 m^3/hr. It is interesting to note that the variability of both travel time and bucket swing time show virtually no main or interaction effects on either response.

Of the interactions, it can be seen that there is a strong production interaction between number of trucks and travel time mean; and between bucket swing time mean and travel time mean. This means that in this case the effect on production of travel time is highly dependent on the levels of the passes per load factor and the bucket swing time mean factor. This can be understood if the system is thought of intuitively: the number of passes per load and the bucket swing time influence the time it takes to load the truck (or the load cycle time) along with spot time, which is not being considered here. Therefore, if an increase in the travel time mean coincided with a reduction in

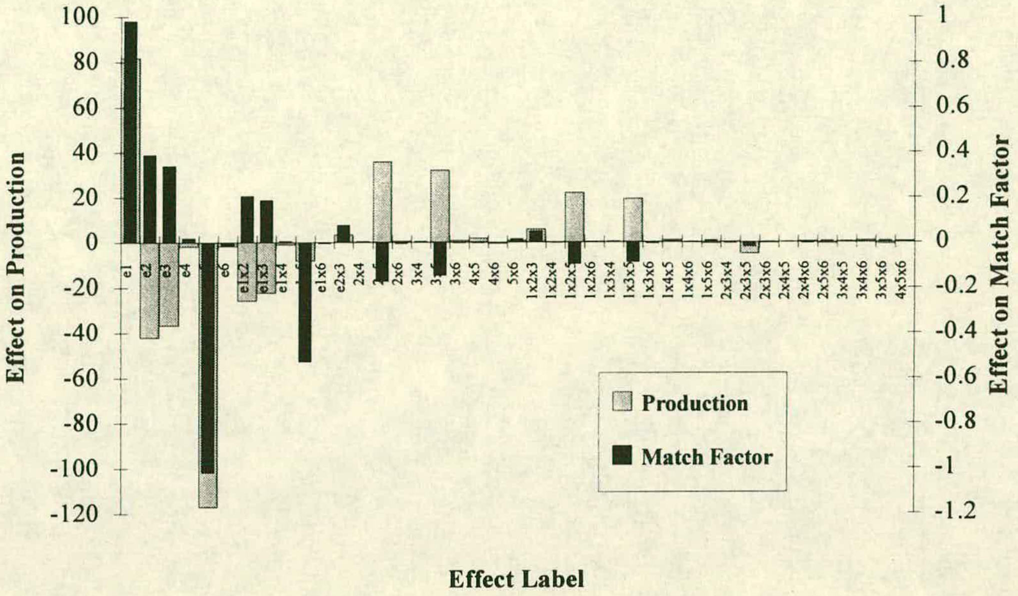


Figure 7.7 Main effects and 2/3 factor interactions for 6 factor experiment

the load cycle time, the reduction in production would not be as great than if the load swing time were to be kept at the same level. This interaction effect also works the other way. If the operation were to be improved by reducing the load cycle time then there must be no increase in the travel time otherwise the effort would be wasted. Load cycle time and travel time are both considered in the calculation of match factor (equation 3.1) and so it is not surprising that match factor response interactions occur between these factors: consider for example the three factor interactions e_{125} and e_{135}.

7.4.2 Testing the significance of main and interaction effects

It is difficult to say for some of the smaller effects whether they are real or just the result of random fluctuations or *noise* between run replications. To determine the significance of an effect, its variance must be estimated. There are several methods to achieve this; Law and Kelton (1991, p. 664) suggest replicating the design n times and obtaining n independent values for each effect. Confidence intervals for the expected effect can thus be obtained using the t distribution. If the confidence interval contains the value zero then the effect can be said to be statistically insignificant. In this case, with 11 separate replications for each design point such a method is a daunting task and therefore the method proposed by Box, Hunter and Hunter (1978, pp. 319 - 322) has been used. This method is based on the assumption that the variations between replications for each set of conditions can be used to estimate the standard deviation of a single response and hence the standard deviation of the effects. The assumption requires the variation between runs at one set of experimental conditions to be a reflection of the total variability affecting runs at other conditions; in this case the assumption will hold as each replication is carried out with a different set of random numbers.

If g design points are replicated with n_i replications made at the ith point giving an estimate s_i^2 of the variance σ^2 with $v_i = n_i - 1$ degrees of freedom then the overall estimate of run variance is given by

$$s^2 = \frac{v_1 s_1^2 + v_2 s_2^2 + \cdots + v_g s_g^2}{v_1 + v_2 + \cdots + v_g} \qquad (7.2)$$

In this example, $g = 64$ and $n_i = 11$, $v_i = 10$ for each design point. Therefore the above equation reduces to

$$s^2 = \frac{\sum_{i=1}^{g} s_i^2}{g} \tag{7.3}$$

s_i has been calculated for each design point which gives $s^2 = 0.082$ for match factor and $s^2 = 5.248$ for production. Each main effect and interaction is the difference between the averages of the responses when the factors are at the plus level and the averages of the responses at the minus level. Each average contains a total of 352 responses (i.e. (64/2)x11) which gives the variance of each effect to be:

$$V(effect) = (1/352 + 1/352)s^2 \tag{7.4}$$

giving *V(match factor)* = 4.65×10^{-4} and *V(production)* = 0.030. The estimate for standard error for each response is $\sqrt{V}$ giving *se(match factor)* = 0.022 and *se(production)* = 0.173. These errors are quite small, almost negligible, and is a reflection of the high number of replications carried out for each design point. We can conclude, therefore that effects and interactions shown in figure 7.7 that have values greater than these standard errors are real, or at least, *statistically* significant.

This initial factorial design has indicated that:

- The change in variability of the component times (factors 4 and 6) have very little effect on both the production and the match factor compared with the other factors. Furthermore, these factors do not significantly interact with any other factors. This does not mean that these factors do not influence the value of the response but rather that any change in these factors induces little change in the responses. This may be because levels of these factors do not change much in table 7.3 and, indeed, it has already been shown that cycle time variability has some effect on production (figure 5.10). However, it is probably quite safe, due to the lack of interactions, to keep these factors at a constant level in future experiments.
- There are a number of 2-factor interactions and two significant 3-factor interactions. The effects of factors 1, 2, 3 and 5 cannot be interpreted separately because of these interactions. In particular, the match factor is very sensitive to the levels at which the number of trucks and the travel time mean are set.

- Further study should be undertaken at different levels of the factors. These studies should be more realistic in that the level changes are appropriate to a real situation. For example, in a single, real earthmoving operation, a change of travel time from 100 to 800 seconds is unlikely.

7.4.3 2^4 Factorial Design: Case Study

The exercise that this design is based on was an actual earthmoving operation for the construction of a major motorway in southern England. The excavator used was a Caterpillar 245B backhoe excavator moving an average of 1.80 bank cubic metres of chalk every pass. The dumptrucks were Volvo BM A35 articulated dumptrucks carrying 6 passes i.e. 10.8 m^3 of chalk per load. Average spot time was 29 seconds, average dump time was 87 seconds. The remaining four factors were recorded and used as the minus level in a 2^4 factorial design. Table 7.5 shows the plus and minus levels used. The values used at the plus level were chosen to represent what would be an anticipated improvement in the overall efficiency of the operation. Firstly, the number of trucks was reduced by one at the plus level as the operation was initially over-resourced (indicated by an initial match factor of 1.23). A reduction by one truck should bring the match factor closer to one and hence give greater efficiency. The number of passes per load was increased by one at the plus level; the dumptrucks used were able to carry the extra load (which is only 1.8m^3, or approximately 3.2 tonnes). The load and travel times were both reduced by 17.5% to give the plus level. The design matrix is shown in table 7.6.

The design was replicated 11 times at each design point and responses were recorded for the match factor, production and also cost per cubic metre as this will give a better indication of the efficiency of the operation (costs were calculated on an arbitrary basis and reflect the differences in the plant prices and not their actual cost per hour). The average values of the responses are shown in table 7.6. Figures 7.8a, b and c show graphically the main effects and interactions for the responses. The overall standard error was also calculated, as for the initial study, and again, due to the high number of replications, these errors are very low. (*se(match factor)* =0.0038, *se(production)* =0.603). Several observations can be made from these results:

- The factor with the greatest effect is the loader pass time. In this example, a reduction from 17 seconds per pass to 14 seconds will, on average

Factor	- level	+ level
1. No. Trucks	7	6
2. Passes per Load	6	7
3. Load Time Mean	17	14
4. Travel Time Mean	492	405

Table 7.5 Factors for Four Factor Design

Design	Factor Number				Production	Match	Cost
Point	**1**	**2**	**3**	**4**	**(m3/hr)**	**Factor**	**(pence/m3)**
1	-	-	-	-	255.4	1.30	102
2	+	-	-	-	237.5	1.10	97
3	-	+	-	-	264.7	1.47	98
4	+	+	-	-	259.5	1.23	89
5	-	-	+	-	281.6	1.12	92
6	+	-	+	-	250.1	0.95	92
7	-	+	+	-	307.9	1.24	84
8	+	+	+	-	281.3	1.06	82
9	-	-	-	+	269.6	1.43	96
10	+	-	-	+	264.8	1.22	87
11	-	+	-	+	272.2	1.63	96
12	+	+	-	+	272.4	1.37	84
13	-	-	+	+	305.3	1.27	85
14	+	-	+	+	281.0	1.09	82
15	-	+	+	+	314.7	1.43	83
16	+	+	+	+	308.3	1.21	75

Table 7.6 Design Matrix for Four Factor Design

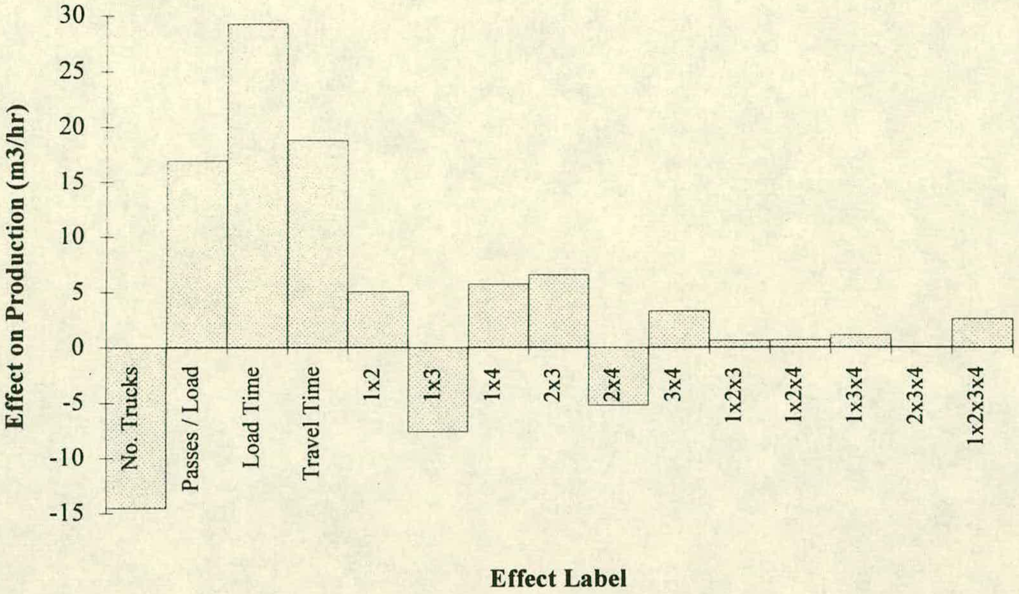


Figure 7.8a Main effects and 2,3,4 factor interactions on production for 4 factor design

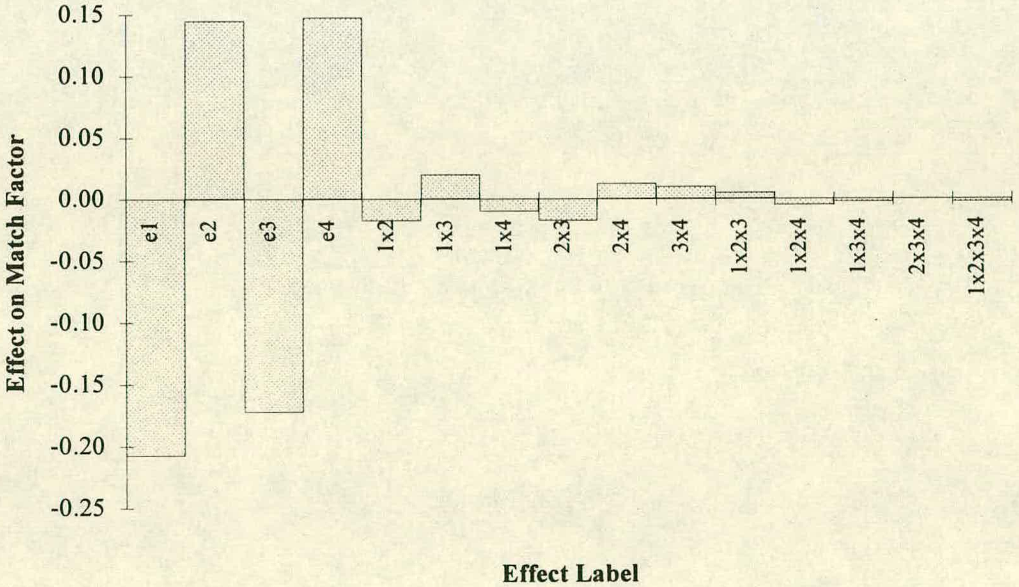


Figure 7.8b Main effects and 2,3,4 factor interactions on match factor for 4 factor design

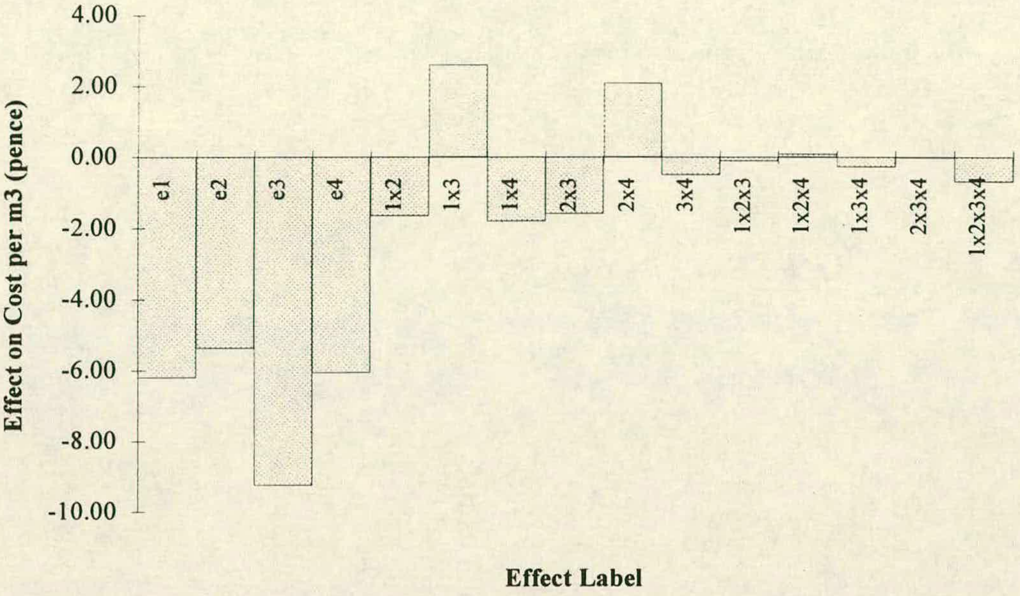


Figure 7.8c Main effects and 2,3,4 factor interactions on cost per m^3 for 4 factor design

decrease the cost per cubic metre by approximately 9 pence and increase the production by 29 cubic metres per hour. For an operation of, say 40,000 cubic metres. This translates into a time saving of 16 hours and a cost saving of £3,600. However, the presence of interactions means these figures are valid for this situation only.

- Of the interactions, three and four-factor interactions are virtually negligible but the two-factor interactions are significant. Surprisingly, although most factors have an improving effect on the responses, two interactions (number of trucks / load pass time and passes per load / travel time) have a negative effect which may seem at first to be difficult to explain. The reason can be seen from figure 7.8b, match factor effects. Factors 1 and 3 tend to reduce the match factor whilst factors 2 and 4 will increase the match factor. The most efficient operation will occur at a match factor of 1 so, for example, combining the effects of factors 1 and 3 will push the match factor to well below one and thus reducing efficiency. Factors 2 and 4 combined will force the match factor even higher than the initial value - again reducing efficiency.
- The effect of reducing the number of trucks by one did not bring about the anticipated increase in production. On its own, reducing the level will have reduced match factor to approximately one, the changes in the other factors did this also. In conclusion, the number of trucks should have been kept at seven although another experiment would have to be done to find out the exact effect. (If this factor were to be kept constant, a factorial design would be with effectively 3 factors giving $2^3 = 8$ design points).

7.5 Metamodel: model of the simulation model

A metamodel is a mathematical expression which represents another model. Equation 7.1 is the basic functional form of the metamodel and, if it could be shown to exist, there would be a separate expression for all required responses from the earthmoving system. Such a 'formula' would indeed be extremely valuable: the performance of an earthmoving operation could be determined very quickly on a hand-held calculator without the need to enter all the input into the Trucksim program and taking the average for a number of replications.

The most common form of a metamodel is a standard regression model: the simulation model is used to provide values of the responses of interest for various factor levels. These factor levels, or simulation input parameters, provide the independent variables for the regression analysis. There are various different regression models that can be determined for any system. The most simple would be a simple regression on two system factors for one response, for example an expression to approximate one of the response surfaces shown in figure 7.5 or 7.6. Alternatively, a multiple, polynomial regression could be determined for all the second order factors (figure 7.1) of the simulation model - recent statistical packages (for example, Minitab 1993) can formulate such regression equations in seconds. It can be anticipated at this stage, even before any regression calculations have been carried out that they will be of little use: firstly, like with response surfaces, a regression equation could not be used with confidence outside the range of the input variables used to construct it. This can of course be proved or disproved by comparison of additional results between the two models. Secondly, the presence of interactions between some of the factors indicates that the regression equation must be a polynomial of at least order 2; exactly what form the equation should take is difficult to determine because the extent of some of the interactions at all factor levels is still not known. This would require many factorial designs for a wide range of factor levels which is outside the scope of this thesis.

To attempt to quantify the usefulness of a metamodel of the simulation model, the statistical package Minitab (1993) has been used to provide regression equations for some of the experiments outlined in this chapter.

7.5.1 Regression analysis of response surface

Figure 7.5 is the response surface constructed from the simulation model with production as the response and spot time and travel time as the factors. This response surface can also be represented as a contour plot which is helpful in making comparisons between two different surfaces; the contour plot of this surface can be seen in Figure 7.9a. Using the data generated by the simulation model to construct this surface, Minitab has been used to determine a linear regression equation of the simulation model. This regression equation is therefore a metamodel of the simulation model. This linear equation is:

$$\text{Production} = 407 - 1.1(\text{Spot Time}) - 0.248(\text{Travel Time}) \qquad (\text{eqn.}7.5)$$

where production is given in bank cubic metres per hour and spot and travel times must be stated in seconds. It must be remembered that this is not a

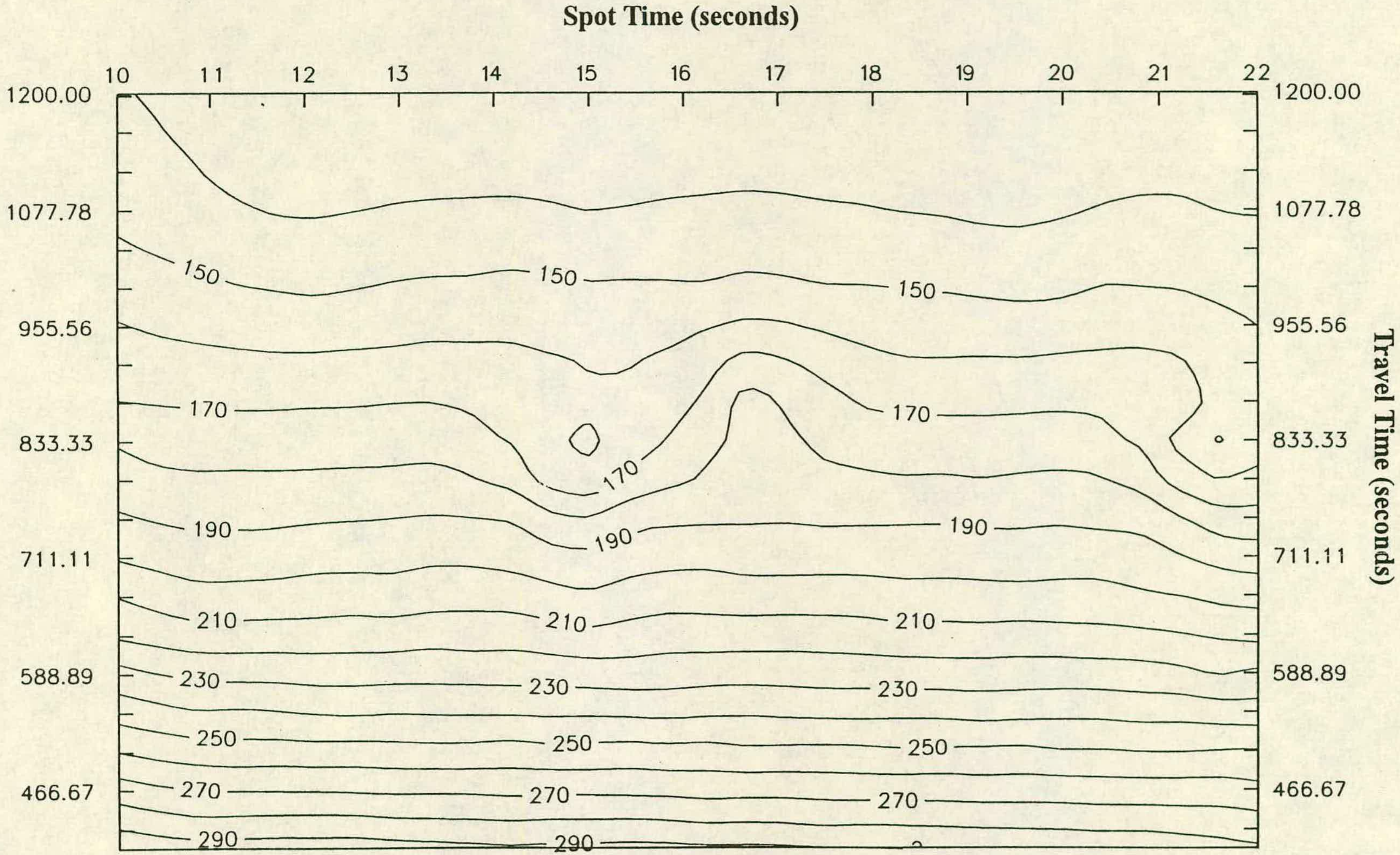


Figure 7.9a Production Response Surface for Simulation Metamodel

universal equation; it is only valid for the levels of the other factors for which the original simulation response surface was constructed (i.e. those values given in table 7.2). Equation 7.5 can be used to construct a contour plot of this linear metamodel by inserting values of spot and travel time over the range for which the simulation model response surface was constructed; this can be seen in figure 7.9b. The linear regression model can immediately be seen to be a poor representation of the simulation model by simply comparing figures 7.9a and b. In some instances, good estimates can be made but a bad estimate is just as likely to be given. The maximum differences between the simulation model and the linear regression model, over the range 10-22 seconds spot time and 100-1100 seconds travel time, are 23.01% under estimate and 25.12 over estimate. This is clearly an unsuitable estimation method as such percentage differences equate to an error of $60m^3/hr$ for an average production rate of $250m^3/hr$.

To improve this estimate, a second, quadratic regression equation was produced with Minitab. The equation is:

$$\text{Prod.}= 456 - 2.43ST - 0.415TT - 0.0115ST^2 - 0.000102TT^2 + 0.00289(ST\text{x}TT) \quad (7.6)$$

where ST is spot time and TT is travel time, both to be given in seconds. Figure 7.9c shows the response surface as a contour plot for the quadratic regression model and although not perfect, is a much better estimate for the simulation model. Both regression models do not indicate any 'bumps' in the response surface as shown in the simulation model and this is probably where the simulation model suffers from random fluctuation. However, this does not mean that the regression equations are better estimates because they have been formed using the same data as the simulation model response surface and hence will incorporate the same errors. Quantitivily, the quadratic regression model has maximum differences of 9.2% under estimate and 21.4% over estimate compared to the simulation model, which is better than the linear equation. The sum of the squares of the differences between a series of simulated estimates and regression estimates will give another indication of how good the regression models are. For 100 estimates over the range of spot time and travel time shown in figure 7.9, the sum of the squares of the differences was calculated to be 25,546 for the linear equation and 15,169 for the quadratic equation. Although this again shows that the quadratic equation

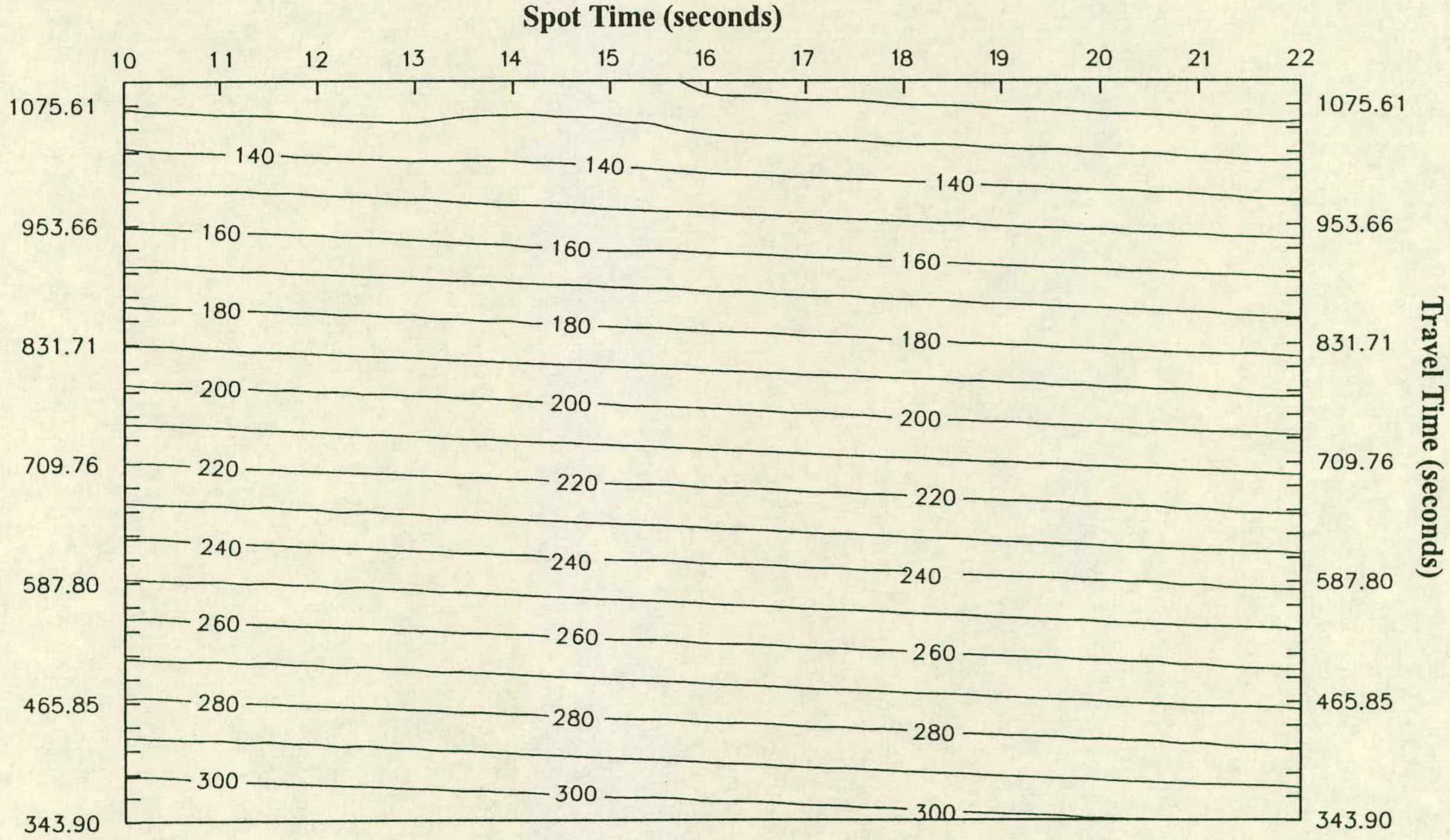


Figure 7.9b Production Response Surface for Linear Metamodel

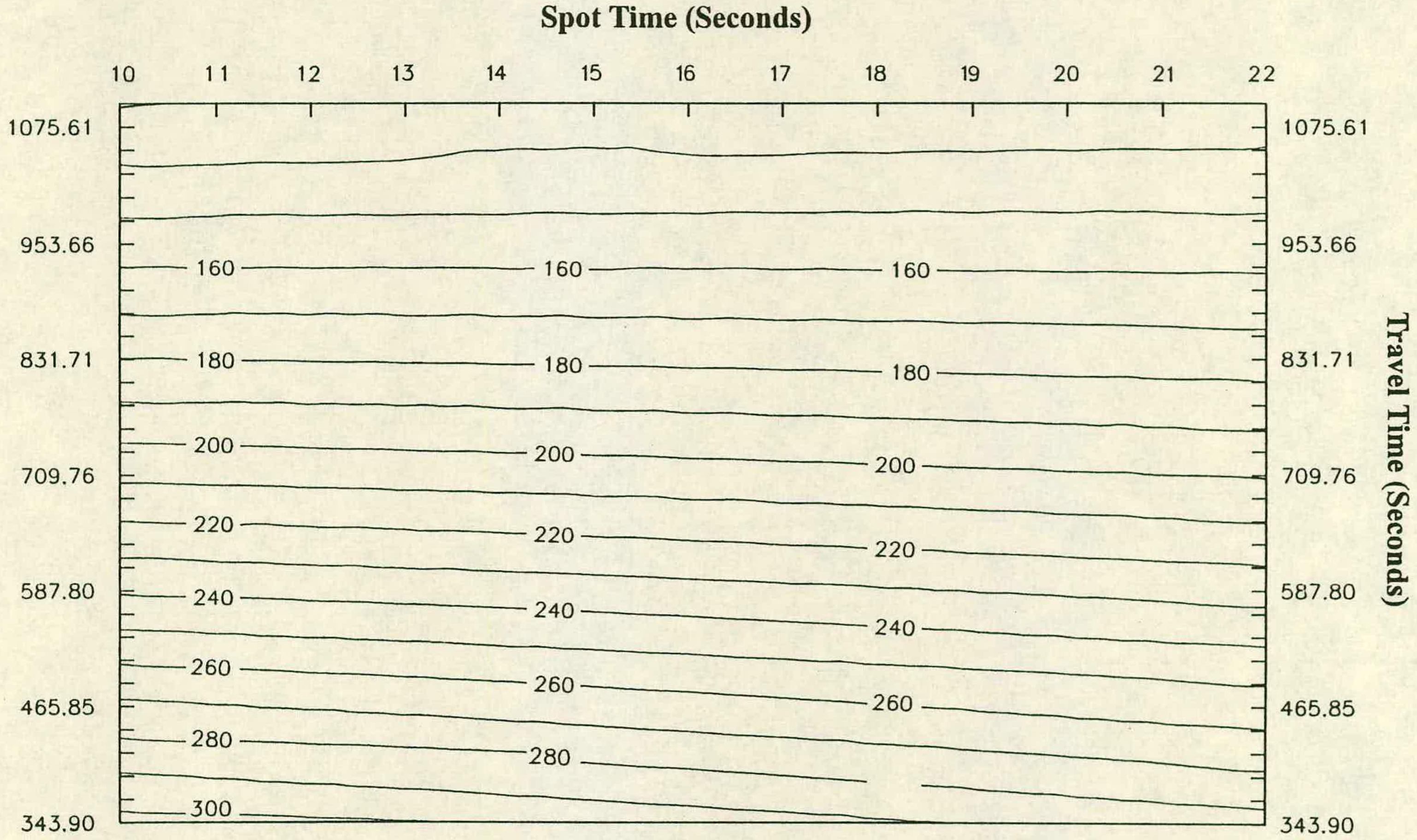


Figure 7.9c Production Response Surface for Quadratic Model

is a better estimator of the simulation output, the differences are probably far too high for the regression model to be used with confidence.

If both equations are used to estimate productivity for values of spot and travel time outside the range the simulated data was generated from then these estimates can also not be used with confidence. Table 7.7 shows a sample of some of the results of the simulation model compared with the estimates made with the regression equations. As can be seen, estimates made using values of spot and travel time that are within the range used to form the equations give a mixture of good and bad results. Once estimates are made without this range, however, very bad estimates are made. Clearly from table 7.7, the linear regression equation is useless and the quadratic equation fairs little better. The main question to be asked after looking at these results is "can these metamodels be used as a substitute for the simulation model?" Before this is answered, the results of another multiple regression analysis shall be looked at for more than 2 factors.

7.5.2 Multiple regression of the 2^6 factorial design

Section 7.4.1 discussed a 2^6 factorial experimental design of the simulation model where the 6 factors included were deemed to be the most important factors in the simulation system. These were: Number of trucks, passes per load, mean bucket swing time, bucket swing time variability (expressed as k), mean travel time and k for travel time. The levels of these factors were set as shown in table 7.3 and it is these levels that were used in the construction of a linear multiple regression equation. Table 7.4 is the design matrix for the 2^6 experiment including the response levels for each design point. The regression analysis was based on these 64 points on the simulation response surface; not surprisingly therefore, a very good fit was obtained (R^2=83.8% where R^2 is the square of the correlation coefficient discussed in 4.3.2; the correlation between the fitted line and the observed data is therefore 0.92 which is very high). This high correlation does not necessarily mean that the equation provides good estimates of the simulation output as shall be seen. The equation is:

$$P=262+20.5n-14.1ps-3.69bstm+0.207bstv-0.168ttm+0.094ttv \qquad (7.7)$$

Table 7.8 compares simulation results with the linear regression results. As with the linear regression of the spot and travel time response surface, fairly

		Simulation	Regression Model Productivity		Differences (%)	
Spot Time	Travel Time	Productivity	Linear	Quadratic	Linear	Quadratic
Within range of simulation model						
10	101	380.0	370.7	392.1	-2.5%	3.1%
14	1112	137.6	116.2	129.2	-15.5%	-6.5%
14	425	283.7	286.3	279.2	0.9%	-1.6%
15	807	152.4	190.7	183.7	25.1%	17.0%
15	426	283.9	284.7	277.0	0.3%	-2.5%
16	1125	137.4	110.1	128.6	-19.9%	-6.8%
22	1123	135.7	104.5	131.2	-23.0%	-3.4%
Outside range of simulation model						
22	1240	122.5	75.0	118.4	-38.8%	-3.5%
28	2101	74.8	-144.4	126.1	-293.1%	40.7%
35	1611	93.2	-31.1	116.1	-133.4%	19.7%
45	1879	86.1	-108.5	148.1	-226.0%	41.9%
52	1485	99.3	-18.5	130.4	-118.6%	23.8%

Table 7.7 Comparison between simulation model and regression model results

good estimates are provided within the range for which the simulation results were provided but outside this range the regression equation cannot be used with confidence.

7.6 Implications for Earthmoving Practice

This chapter attempted to determine which factors, and what levels of these factors, affect the output responses of an earthmoving model the most. It can be seen that the output of the model is sensitive to six factors: number of trucks, passes per load, load pass time, spot time, travel time and dump time. Four of these factors can be split into two groups of two, that is the effect on the response of dump time is the same as that for travel time; spot time has the same effect as load pass time. The following is a discussion on the factors and how they can be controlled in a site situation.

- **Number of trucks.** This factor is perhaps the one which is easiest to control by the contractor. The last experiment in this chapter has shown that the correct number of trucks is essential for maximum efficiency. Experienced contractors will be able to instinctively assess a live operation and adjust the number of trucks. However, for operations yet to be carried out and for long haul operations with many trucks this is not as easy: the impact on the output response to changes in fleet numbers should be understood.
- **Passes per load.** Like the number of trucks, this is easily controlled and it has been shown that despite an increase in load time, an extra bucketful per load is advantageous. There are two points that need to be considered. An extra bucket should never cause the truck payload to exceed its limit. Apart from long term damage and safety considerations, some materials, especially wet clays, will take longer to dump if the truck is overloaded. Observations made on real sites has shown that if an operation is over-resourced, loader operators will tend to under-load a truck to reduce the queue length. If this is happening, the number of trucks should be reduced and the trucks filled to their maximum limit.
- **Spot and load pass time.** These factors both have the same effect on the output. For the operation studied with the 2^4 design, load pass time was seen to have the greatest effect on cost and production and this will generally hold true for operations with short hauls. (On longer hauls, the travel time will become the dominant factor.) Therefore, effort should be made by the contractor to reduce spot and load times, although comments made in section 7.4.1, that the load cycle and the travel cycle times interact, should be born in mind. Spot time can be controlled by having a large, clear manoeuvring area and will reduce with good

	Trucks	**Passes**	**BSTmean**	**BSTk**	**TTmean**	**TTk**	**Production (metamodel)**	**Production (sim. model)**	**%diff**
Results from 6 factor design (within range)									
1	2	4	12	20	100	70	196.24	179.6	9.3%
2	6	4	12	20	100	70	278.24	373.1	-25.4%
3	2	7	12	20	100	70	153.94	153.2	0.5%
11	2	7	12	10	100	70	151.87	153.6	-1.1%
12	6	7	12	10	100	70	233.87	220	6.3%
13	2	4	22	10	100	70	157.27	155.4	1.2%
14	6	4	22	10	100	70	239.27	233.5	2.5%
15	2	7	22	10	100	70	114.97	119.6	-3.9%
16	6	7	22	10	100	70	196.97	130.9	50.5%
17	2	4	12	20	800	70	78.64	44.2	77.9%
27	2	7	12	10	800	70	34.27	42.4	-19.2%
28	6	7	12	10	800	70	116.27	117.3	-0.9%
29	2	4	22	10	800	70	39.67	42.6	-6.9%
30	6	4	22	10	800	70	121.67	119.5	1.8%
31	2	7	22	10	800	70	-2.63	39.7	-106.6%
32	6	7	22	10	800	70	79.37	105.3	-24.6%
33	2	4	12	20	100	50	194.36	176.3	10.2%
34	6	4	12	20	100	50	276.36	372.4	-25.8%
35	2	7	12	20	100	50	152.06	150.8	0.8%
36	6	7	12	20	100	50	234.06	225.7	3.7%
37	2	4	22	20	100	50	157.46	154.2	2.1%
50	6	4	12	20	800	50	158.76	122.8	29.3%
51	2	7	12	20	800	50	34.46	42.5	-18.9%
52	6	7	12	20	800	50	116.46	115.6	0.7%
53	2	4	22	20	800	50	39.86	42.8	-6.9%
60	6	7	12	10	800	50	114.39	114.7	-0.3%
61	2	4	22	10	800	50	37.79	42.9	-11.9%
62	6	4	22	10	800	50	119.79	117.2	2.2%
63	2	7	22	10	800	50	-4.51	39.2	-111.5%
64	6	7	22	10	800	50	77.49	103	-24.8%
Results from 4 factor design (outside range)									
1	7	6	17	20	492	50	184.354	255.4	-27.8%
2	6	6	17	20	492	50	163.854	237.5	-31.0%
3	7	7	17	20	492	50	170.254	264.7	-35.7%
8	6	7	14	20	492	50	160.824	281.3	-42.8%
9	7	6	17	20	405	50	198.97	269.6	-26.2%
10	6	6	17	20	405	50	178.47	264.8	-32.6%
11	7	7	17	20	405	50	184.87	272.2	-32.1%
12	6	7	17	20	405	50	164.37	272.4	-39.7%
15	7	7	14	20	405	50	195.94	314.7	-37.7%
16	6	7	14	20	405	50	175.44	308.3	-43.1%

Table 7.8 Comparison of sample of results between simulation model and second regression model

truck discipline i.e. queuing as close to the loader as possible and moving into place as soon as the previous truck has departed. For operations with large numbers of trucks, it is worth having a ganger in charge of 'directing' the plant into place. Load time is governed by a number of factors, some more controllable than others. For example, the contractor has little control over the type of material but the excavating operation should be set up to minimise swing angle and maximise cutting depth.

- **Travel and dump times.** Especially on long hauls, the travel time is a major factor in the earthmoving system and is influenced itself by many factors. Unless the haul road is purpose built, the plant will have to run over the actual cuttings and embankments that make up the works and little can be done to change the soil properties. However, careful maintenance of haul roads and an awareness of how rain will affect the strength of the soil is needed to maximise the running speed. Other factors influencing the travel time are obstructions, such as bailey bridges and plant crossings, haul road gradient and truck specification.

The presence of interactions between the factors, which can only be detected by certain experimental designs, have indicated the following:

- Care should be taken when trying to improve an operation that changes made do not have a detrimental effect on the output. Figure 7.8 has shown that certain combinations of level changes will, at best, leave the output unchanged. The largest interaction is between number of trucks and load pass time; essentially, there is no point in reducing load time if there are not enough trucks to satisfy the loader.

Finally, the formulation of metamodels was investigated briefly using the method of regression analysis. The existence of simple, solvable equations would obviously be of extreme importance in the estimation of earthmoving problems, but it was easily anticipated after getting so far in this thesis that earthmoving is not such a simple system and that such equations were not going to be possible. Two separate parts of the whole production response surface of the simulation model were looked at and regression equations provided by Minitab and these provided the following points:

- The regression equations constructed were in one instance quadratic and in the other two linear (equations 7.5 to 7.7). The linear equations were found to be a bad estimator of production for the range of the factor levels that were used to form them and this could be anticipated by simply looking at the response surfaces in figures 7.5 and 7.6. The quadratic equation was the better estimator of production but even in this case, certain points of the response surface could not be reliably calculated by this metamodel. This could, in part, be because the simulation model itself provided inaccurate, fluctuating results (as can be seen by the unevenness of the surface in figure 7.5) and the regression metamodel has smoothed out these inaccuracies. One way to test this would be to compare the metamodel with actual results in a similar way to what was done for the simulation model in chapter 6, but as the results from the metamodel were poorer than those from the simulation model, it is felt that this would be a pointless task. Another reason for the poor estimation could be that the response surface is not actually quadratic but perhaps some higher degree polynomial or logarithmic equation. The investigation into such equations could be useful if it were not for the second point.

- The regression equations, both linear and quadratic, were totally unsuitable for estimating production on points of the response surface outside the range for which the equations were formed. If it could be anticipated that earthmoving operations would always fall into one area of the response surface then regression analysis could have some part to play in the estimation and planning of such work. Unfortunately, the earthmoving system rarely works that way with operational conditions changing on an daily and even hourly basis. It is for this reason that simulation methods were proposed in the first place and so the conclusion has to be made that regression analysis plays little part in this research project. It could form part of future work as the field of metamodelling is very active at the present time.

7.7 Conclusions

1. Earthmoving operational output is most sensitive to the following factors: number of trucks, passes per load, load pass time, spot time, travel time, and dump time.
2. The correct number of trucks matched to the loader is essential for maximum efficiency of an earthmoving operation. For live operations, the plant match can be easily controlled by an experienced contractor but this may be a difficult task for operations which are only in the planning stage.
3. Despite the increase in load time, extra bucketfuls per load are advantageous. This must never overload the truck for safety and plant longevity.

4. Spot and load time both have the same effect on the output. The 2^4 design has shown that the load pass time has the greatest effect on cost and production. As haul length increases, travel time will become the dominant factor. In both cases, keeping the component times to a minimum is essential if maximum production is to be achieved.
5. The presence of interactions between some of the factors indicates that any changes made to an operation in attempting to improve the output should not have a detrimental effect. For example, if load time is reduced then production will not increase if the operation is already under-resourced. Such a decrease in load time should coincide with an increase in the number of trucks.
6. Interactions also mean that the response function cannot be represented by a simple linear function. This indicates firstly that simulation is the best way to investigate the earthmoving system and, secondly, that the effects of the factors cannot be interpreted outside the experimental range.

Chapter 8

Conclusions

Individual conclusions are given at the end of chapters 3 to 7 but this final chapter brings together general and specific conclusions about the work presented in this thesis as a whole. It is important also that research such as this does not simply end when the thesis is complete, and so the further work section outlines suggestions to not only continue this work but ensure that the research carried out so far is implemented successfully.

8.1 General Conclusions

1. Cost savings are undoubtedly achievable if a more systematic approach to the estimation of earthmoving productivity and resource allocation is used. Methods such as simulation models have been studied and implemented in the United States for the past two decades and have shown, along with the work in this thesis, that the earthmoving system can be represented mathematically and give accurate estimations of how operations will perform. The current state of the road building industry, in the UK in particular, is in a state of decline and contractors have to take on board new ways of thinking about how traditional construction methods have been undertaken in the past if they are to survive.
2. Chapter two outlined that the dumptruck / excavator combination is more cost effective than the scraper / pusher method of earthmoving (in the UK and Europe: lower fuel prices in the United States and the middle east keep the fuel hungry scrapers a viable alternative). Whilst the scraper is by no means obsolete (it is extremely productive at very short hauls for example) the versatility of the dumptruck, and in particular the articulated dumptruck, has been the reason that the rest of the thesis has concentrated on this type of plant.
3. Various methods can be used to estimate the productivity of earthmoving operations: deterministicly or stochasticly; spreadsheet or computer program; pre-written simulator or 'bespoke' program. Chapters three and four indicated that the deterministic method was not accurate enough and that a stochastic simulation could be used. The choice of writing a program specifically rather than using an off the shelf package was a simple one: much more can be learnt about the whole process being modelled if it has to be translated completely into code rather than have a program carry out the task.
4. Any program written can and must be verified (ensured that it performs the task it has been written to do) but it must also be validated - that is, the model must give results that are very close to those that would be obtained in real life. It is also important to realise that true, 100% validation is a near impossibility - the model is after all only intended to be an approximation of the real system - but the model output must be shown to resemble the real output as close as is practically possible. The model outlined in this thesis has been shown to be replicativily valid (reproduces fairly accurately past events) but not particularly predictivily valid. This of course is a much harder task as *all* factors about an earthmoving operation would need to be known in advance.
5. The model is also a good platform upon which experiments can be performed. The program can be used to see how small adjustments to the setup of an

operation will affect the output which, if they were to be done with the real system, would be prohibitively expensive. This could have far reaching possibilities: it could help in deciding whether the plant that has been used for years is the ideal combination. Are bigger trucks better than smaller ones? In certain situations would a very large excavator be more cost effective than two smaller ones? Every earthmoving operation is unique, and so there will never be a single answer to these questions. But it has been shown that a computer can put managers in much better positions to determine how to go about the next project.

8.2 Specific Conclusions

1. The earthmoving system is effectively a queueing system: the haulers are the customers, the loaders are the servers, the arrival, or more specifically the return of the hauling units at the queue becomes the input process and the service or loading cycle is the output process. The queue discipline of this queueing system is first come first served.
2. The match factor of an operation (as defined by equation 3.1) indicates the efficiency of an operation with respect to its match between trucks and excavators. A match factor of one indicates perfect match, with an over-resourced operation (i.e. too many trucks) being indicated by a match factor greater than one. Conversely, a match factor less than one indicates an under-resourced operation.
3. An appraisal of the assumptions made in formulating the first model (chapter 3, using a spreadsheet), and the results obtained from it indicate that the variation, or randomness, of the earthmoving cycle times must be taken into account when estimating the productivity.
4. This randomness is one of the causes of bunching of the hauling plant. Bunching occurs when the variability of the plant operating speeds means the trucks travel in groups, following the slowest truck thus preventing the loader from having a constant supply of customers to service. The lower utilisation of the loader brings about a reduction of the productivity of the operation that can be quantified with the bunching factor (equation 3.2). This bunching factor has been shown, through a study of actual operations, to vary with the match factor of the operation: greatest bunching occurs at match factors at or around one.

5. The model of the earthmoving system is dynamic (modeling the system as it evolves), stochastic (rather than deterministic) and discrete, that is the state variables change at discrete points in time.
6. A dynamic, stochastic and discrete model can be analysed using discrete-event simulation. Such a technique works by advancing from event to event, updating statistics about the system until a pre-set stop condition has been met. The events used in the earthmoving system are arrivals at the queue and departures from the loader. The time advance mechanism used is next-event time advance where the time and type of the next event is selected from an event list and processed accordingly.
7. Probability distributions were determined for the observed data. Probability plots showed that the Erlang distribution provided a reasonable fit for all time components and is ideal in that the parameters for these distributions are easy to determine.
8. The program written (Trucksim) provides 8 types of output (such as productivity, utilisation and match factor) from 7 types of input (component time parameters, number of trucks and excavator parameters). The program cannot simulate for more than one excavator.
9. Trucksim provides estimates for bunching which could not be provided by the deterministic model outlined in chapter 3. This should provide more accurate estimates of the operation productivity.
10. It is perhaps impossible to completely validate any model. A model is therefore satisfactory if the discrepancies the validation procedure has highlighted are deemed to be acceptable. Secondly, there are at least three different levels of valid model: structural, predictive and replicative validity. A model can still be acceptable even if all three levels are not achieved.
11. The earthmoving model can also not be considered predictivily valid. It is unlikely that full predictive validation will be achieved as this requires predictive abilities beyond the power of any simulation tool, such as determination of future ground conditions, weather and any other unexpected delays. It also requires the average time components of a future operation - the accuracy of these estimations can only be improved as a database of the way various vehicles behave is built up over a number of years.
12. The model can, however, be considered replicativily valid. This has been shown by the results in chapter 6, which has also shown that high face validity has been achieved. Chapter 4 has shown that the main assumptions behind the model are also valid.

13. Earthmoving operational output is most sensitive to the following factors: number of trucks, passes per load, load pass time, spot time, travel time, and dump time.
14. The presence of interactions between some of the factors indicates that any changes made to an operation in attempting to improve the output should not have a detrimental effect. For example, if load time is reduced then production will not increase if the operation is already under-resourced. Such a decrease in load time should coincide with an increase in the number of trucks.
15. Interactions also mean that the response function cannot be represented by a simple linear function. This indicates firstly that simulation is the best way to investigate the earthmoving system and, secondly, that the effects of the factors cannot be interpreted outside the experimental range.

8.3 Future Work

1. If the work presented in this thesis is to be successful then it has to be accepted as a credible means of estimating earthmoving; an alternative to the traditionally used tables and rules of thumb. To do this the work has to be implemented on live sites, in addition to those used to gather data used in this thesis. The implementation process must also not be rigid: the findings and processes of this research will change and evolve as the hugely complicated business of earthmoving is explored and the researchers must be prepared to accept that their own findings may be impractical in the same way that traditional methods have been shown to be. At the time of writing, this implementation procedure is already underway on two of the UK's largest road building contracts: the M65 Blackburn Southern Bypass and the Derby Southern Bypass.
2. The work on simulation presented here has, in some ways, been very inward looking and has not taken on board other work that has been done in this field, mainly in the United States. A possible area of future work would be to compare the simulation model and program developed here with those that have already been written, such as CYCLONE for construction operations and also those written for industrial applications, such as SLAM II and SLAMSYSTEM.
3. Probability distributions other than the Erlang distribution used here should be investigated. Much work has been done using the beta distribution and comparisons between the two would indicate which provided the best estimation of the real system.
4. The Trucksim program as it stands can only simulate for single loader operations. Work could be done to improve the program by making it able to

simulate for multi loader operations. General improvements, such as upgrading to the Windows environment, would increase the useability of the program.

5. The quality of the output from the model is only as good as the data that has been inputted, and in this case it is a simple fact that the more data the better. Therefore, future work could concentrate on more efficient and less time consuming methods of recording plant speeds. One such way could be to use video and image analysis: operations could be recorded and a computer used to recognise individual trucks and the points in time when the different cycle components start and finish.
6. Data collection on a broader range of sites, with different plant and different soil conditions would also increase the quality of the database. Currently, only a few types of plant (mainly Caterpillar and Volvo) and soils (silty clays and chalk) have been investigated, but there are many other plant manufacturers (Komatsu in particular) and, for example, no investigation has been done into earthmoving in sand and gravels.
7. Different earthmoving subcontractors should be investigated. Methods of working will vary throughout the industry and some may reveal ways of earthmoving not yet considered. More likely, some will respond better to new ways of working.
8. Finally, there is no reason why the methods outlined in this thesis cannot be used elsewhere in the construction business. For example, structural work has many cyclic operations such as concreting and reuse of formwork. If it is shown that earthmoving benefits from different ways of thinking then so too will the other parts that make up a whole construction project.

References

AbouRizk, S.M.; A.A.Gonzalez-Quevedo and D.W.Halpin (1990) Application of Variance Reduction Techniques in Construction Simulation. *Microcomputers in Civil Engineering, 5.* 299-306.

AbouRizk, S.M.; D.W.Halpin and S.L.Hill, (1991a) Measuring Productivity and Validating Microcomputer Simulation. *Microcomputers in Civil Engineering, 6.* 205-215.

AbouRizk, S.M; D.W.Halpin and J.R.Wilson (1991b) Visual interactive fitting of beta distributions. *Journal of Construction Engineering and Management, ASCE 117(4).* 589-605.

AbouRizk, S.M and D.W.Halpin (1992) Modelling Input Data for Construction Simulation. *Proceedings of the 8th Conference on Computing in Civil Engineering, ASCE.*

AbouRizk, S.M and R.J.Wales. (1993) Incorporating weather effects in project simulation. *Proceedings of the 5th International Conference on Computing in Civil and Building Engineering, ASCE.* Anaheim, Calif.

AbouRizk, S.M and J.Shi (1994) Automated Construction-Simulation Optimization. *Journal of Construction Engineering and Management ASCE, 120(2)* 374-385.

AbouRizk, S.M; D.W.Halpin and J.R.Wilson (1994) Fitting beta distributions based on sample data. *Journal of Construction Engineering and Management ASCE, 120(2)* 288-305.

Accelerator, inc. (1987) *Accelerator Handbook.* Accelerator, inc. Fort Myers, Fla.

Arrowsmith, E.J. (1978). Roadwork fills - a material engineer's viewpoint. *Proceedings of The Clay Fills Conference.* ICE, London. 25-36.

Ashworth, R. (1966). *Highway Engineering.* Heinemann, London.

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

Barker, A. (1988). *Contract Journal.* 23 June. 21-34.

Barnett, V. (1975). Probability Plotting Methods and Order Statistics. *Applied Statistics, 24.* 95-108.

Bernold, L.E. (1986) Low level artificial intelligence and computer simulation to plan and control earthmoving operations. *Earthmoving and Heavy Equipment, Proceedings of the Conference, ASCE.* 156-165

Bhattacharjee, G.P. (1970) Algorithm AS32: The Incomplete Gamma Integral. *Applied Statistics, 19.* 285-287.

Biles,W.E. and J.J.Swain (1980). *Optimization and Industrial Experimentation.* John Wiley, New York.

Blackman, R.B. and J.W.Tukey, (1958). *The measurement of Power Spectra.* Dover Pub. Inc. New York.

Blaug, M. (1962) *Economic Theory in Retrospect.* Richard D. Irwin. Homewood, Ill.

Box, G. E. P.; W.G. Hunter and J.S. Hunter. (1978) *Statistics for Experimenters.* John Wiley, New York.

Box, G.E.P and K.B.Wilson. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B, 13.* 1-45.

Braun, M. (1975) *Differential Equations and Their Applications.* Springer-Verlag, New York, NY.

BS 1377 (1990). British Standards Institution, London.

BS 3138 (1969) Glossary of Terms used in Work Study. British Standards Institution, London.

Byles, R. (1984). *New Civil Engineer.* 22 November. 44-49.

Byles, R. (1976). Mud Sweat and Tyres. *New Civil Engineer.* 29 April. 5-47.

Byles, R. (1988). *New Civil Engineer.* 24 November. 36-47.

Carmichael, D.G. (1986). Erlang Loading Models in Earthmoving. *Civil Engineering Systems, 3*: 118 - 124.

Carmichael, D.G. (1987a). *Engineering Queues in Construction and Mining.* Ellis Horwood. Chichester.

Carmichael, D.G. (1987b). On the equivalence of the (E_h/M/c) and (M/M/c) finite source queues. *Civil Engineering Systems, 4.* June 1987. 87-93.

Carmichael, D.G. (1987c). A Refined Queuing Model for Earthmoving Operations. *Civil Engineering Systems, 4*: 153 - 159.

Carmichael, D.G. (1990). Heterogeneity in deterministic finite source queues. *Civil Engineering Systems, 7(1).* 11-19.

Carson, J.S. (1986) Convincing Users of Model's Validity is Challenging Aspect of Modeler's Job. *Industrial Engineering, 18.* June 1986. 74-85.

Caterpillar. (1987). *Vehsim Handbook.* Caterpillar, inc. Peoria, Ill.

Char, B.W et. al. (1992). *Maple V Reference Manual,* Springer-Verlag, New York.

Clemmens, J.P. and J.H.Willenbrock. (1978) The SCRAPESIM Computer Simulation. *The Journal of the Construction Division, ASCE.* CO4, Dec. 1978. 419-435.

Dagpugnar, J. (1988). *Principles of Random Variate Generation.* Clarendon Press, Oxford.

Dennehy, J.P. (1978). The remoulded undrained shear strength of cohesive soils and its influence on the stability of embankment fill. *Proceedings of The Clay Fills Conference.* ICE, London. 87-94.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* Springer-Verlag, New York, NY.

Donjerkovitch, D. (1983) Application of queuing theory to earthmoving and quarrying operations. *B.E. Report, University of Western Australia.*

Drake, J. et. al. (1969). *Motorways,* Faber and Faber.

Easa, S.M. (1988). Earthwork Allocations with Linear Unit Costs. *Journal of Construction Engineering and Management, ASCE 114(4).* 641-653.

Easa, S.M. (1987). Earthwork Allocations with Nonconstant Unit Costs. *Journal of Construction Engineering and Management, ASCE 113(1).* 34-50.

Emshoff, J.R. and R.L. Sisson, (1970). *Design and Use of Computer Simulation Models.* Macmillan. New York.

Farid, F and T.LKoning (1994) Simulation verifies queuing program for selecting loader truck fleets. *Journal of Construction Engineering and Management ASCE, 120(2)* 386-404.

Fishman, G. and P.J. Kiviat. (1967). The Analysis of Simulation-Generated Time Series. *Management Science, 13(7).* 525-557.

Fondahl, J.W. (1960) Photographic analysis for construction operations. *Journal of the Construction Division, ASCE. 86(2).*

Friedman, M. (1953) *Essays in Positive Economics.* University of Chicago Press. Chicago, Ill.

Gaarslev, A. (1969) Stochastic models to estimate the production of material handling systems in the construction industry. *Technical Report no. 111.* Construction Institute, Stanford University, Stanford, California.

Gass, S.I. and B.W. Thompson. (1980) Guidelines for Model Evaluation: An abridged version of the U.S. General Accounting Office Exposure Draft. *Operations Research, 28.* 431-439

Gonzalez-Quevedo, A; S.M.AbouRizk; D.T. Iseley and D.W.Halpin (1993) Comparison of two simulation methodologies in construction. *Journal of Construction Engineering and Management, ASCE, 119(3)* 573-589.

Gordon, G. (1978) *System Simulation, 2nd ed.* Prentice-Hall, Englewood Cliffs, NJ.

Hahn, G.J. and S.S. Shapiro (1967). *Statistical Models in Engineering.* John Wiley, New York, NY.

Halpin, D. W. (1977) CYCLONE: Method for modelling of job site processes. *Journal of the Construction Division, ASCE. 103,* 489-499.

Halpin, D.W. (1990) *MicroCYCLONE user's manual.* Division of Construction Engineering and Management, Purdue University, West Lafayette, Ind.

Handa, V.K. and R.M. Barcia. (1986). Construction Production Planning. *Journal of Construction Engineering and Management, ASCE. 112(2).* 163-177.

Hunter, J.S. and T.H.Naylor (1970). Experimental Designs for Computer Simulation Experiments. *Management Science 16(7)*: 422 - 433.

Institution of Civil Engineers (1991) *Inadequate Site Investigation.* Thomas Telford, London.

Institution of Civil Engineers (1991). *The ICE Conditions of Contract, 6th ed.* Thomas Telford, London.

Jayawardane, A.K.W. and F.C. Harris. (1990) Further Development of Integer Programming in Earthwork Optimisation. *Journal of Construction Engineering and Management, ASCE. 116(1).* 18-33.

Jayawardane, A.K.W and A.D.F.Price (1994a) A new approach for optimizing earth moving operations, Part 1. *Proc. Instn. Civ. Engrs. Transp.,105 Aug.* 195-207.

Jayawardane, A.K.W and A.D.F.Price (1994b) A new approach for optimizing earth moving operations, Part 2. *Proc. Instn. Civ. Engrs. Transp.,105 Nov.* 249-258.

Kalk, A and S. Douglas (1980) Insight: Interactive simulation of construction operations using graphical techniques. *Technical Report no. 238.* Construction Institute, Stanford University, Stanford, California.

Karshenas, S. (1989a) Estimating Production of Multi-Loader Truck Fleets. *Transportation Research Record, 1234.*

Karshenas, S. (1989b) Truck capacity selection for earthmoving. *Journal of Construction Engineering and Management, ASCE. 115(2)* 212-227.

Karshenas, S and F.Farid (1988) Multiloader-truck fleet selection for earthmoving. *Transportation Research Record 1186,* 51-65.

Kleijnen, J.P.C. (1975). Screening Designs for Poly-Factor Experimentation. *Technometrics, 17(4)*: 487 - 493.

Kleijnen, J.P.C. (1977). Design and Analysis of Simulations: Practical Statistical Techniques. *Simulation 28*: 81 - 90.

Law, A.M. and W.Kelton (1991) *Simulation Modeling and Analysis, 2nd ed.* McGraw - Hill. New York.

Law, A.M. and S.W. Haider, (1989). Selecting Simulation Software for Manufacturing Applications. *Industrial Engineering, 31.* May 1989. 33-46.

Law, A.M. and M.G. McComas, (1990). Secrets of Successful Simulation Studies. *Industrial Engineer.* May 1990. 47-72.

Law, A.M. and M.G. McComas, (1989). Pitfalls to Avoid in the Simulation of Manufacturing Systems. *Industrial Engineer, 31*: 28 - 69.

Lawless, J.F. (1982). *Statistical Models for Lifetime Data.* John Wiley, New York, NY.

Lewis, C.R and C.J.Schexneyder (1986) Production analysis of the Cat 245 hydraulic hoe. *Earthmoving and Heavy Equipment, ASCE, Proceedings of the conference.*

Li, Z. (1990). Methodology for the Optimum Control of Shovel and Truck Operations in Open Pit Mining. *Mining Science and Technology, 10(3)* 337-340.

Littlejohn, G.S.; K.W.Cole and T.W.Mellors (1994) Without site investigation ground is a hazard. *Proc. Instn. Civ. Engrs. Civ. Engng. 102, May.* 72-78.

Marshall, I. (1981). *Contract Journal.* 17 December. 16-18.

Mauro, C.A. (1984). On the Performance of Two Stage Group Screening Experiments. *Technometrics, 26(3)*: 255 - 264.

Mauro, C.A. and D.E. Smith (1982). The performance of two-stage group screening in factor screening experiments. *Technometrics, 24(4)*: 325-330

Mayer, R.H. and R.M. Stark. (1981). Earthmoving Logistics. *Journal of the Construction Division, ASCE. 107(CO2).* 297-312.

Microsoft Corp. (1990) *C for Yourself: The Microsoft Quick C Compiler Handbook.* Microsoft Corp., Redmond, WA.

Microsoft Corp. (1992). *Microsoft Excel v.4.0 Users Guide, vol. 2.* Microsoft Corp. Redmond, WA.

Minitab, inc. (1993) *Minitab Reference Manual, v. 9.* Minitab, inc, PA.

Montgomery, D.C. (1991) *Design and Analysis of Experiments, 3rd ed.* John Wiley. New York.

Morgan, W.C. and L.L. Peterson. (1968). Determining Shovel-Truck Productivity. *Mining Engineering.* December 1968. 76-80.

Myers, R.H.; A.I. Khuri and W.H. Carter. (1989) Response Surface Methodology: 1966-1988. *Technometrics, 31(2)* May 1989, 137-157.

Nandgaonkar, S.M. (1981). Earthworks Transportation Allocation: Operations Research. *Journal of the Construction Division, ASCE, 107(CO2).* 373-392.

Naylor, T.H. and J.M.Finger. (1967). Verification of Computer Simulation Models. *Management Science.* 14(2). October 1967. 92-101.

Naylor, T.H. (1971). *Computer Simulation Experiments with Models of Economic Systems.* John Wiley and Sons. New York.

Naylor, et al., (1969). Spectral analysis of data generated by simulation experiments with econometric models, *Econometrica*, 37: 333-352.

Naylor, T.H.; D.S.Burdick and W.E.Sasser. (1967) Computer Simulation Experiments with Economic Systems. *J. Am. Stat. Assoc. 62.* 1315 - 1377.

New Civil Engineer. (1989). *Roads Supplement.* September 1989.

Oglesby, C and G. Hicks. (1982). *Highway Engineering, 4th ed.* John Wiley and Sons, New York, NY.

Oloufa, A.A and K.C.Crandall (1992) Feedback mechanisms for operational simulation. *Journal of Computing in Civil Engineering, ASCE 6(2).* 161-177.

Oloufa, A.A (1993a) Modeling operational activities in object-oriented simulation. *Journal of Computing in Civil Engineering, ASCE 7(1).* 94-106.

Oloufa, A.A (1993b) Modeling and simulation of construction operations. *Automation in Construction 1(4)* 351-359.

Paul, P.K. and P.De. (1989). Matching of shovel-dumper at production planning stage. *Journal of Mines, Metals and Fuels.* April 1989, 143-146.

Paulson B.C.; S.A.Douglas; A.Kalk; A.Touran and G.A.Victor. (1983) Simulation and Analysis of Construction Operations. *Journal of Technical Topics in Civil Engineering, ASCE 109 (2).* 89-104.

Paulson B.C.; W.T.Chan and C.C.L.Koo (1985) Simulating Construction Operations by Microcomputer. *Construction Research Applied to Practice, ASCE.* 35-49.

Peacock, W.S. and I.L.Whyte (1992) Site Investigation and Risk Analysis. *Proc. Instn. Civ. Engrs. Civ. Engng. 92, May.* 74-81.

Peacock, W.S. and I.L.Whyte (1988) Site investigation practice. *Municipal Engineer, Oct.* 235-245

Pearson, K. (1900). On a Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it can be Reasonably Supposed to Have Arisen in Random Sampling. *Phil. Mag. 5 (50).* 157-175

Plackett, R.L and J.P.Burman (1946). The Design of Optimum Multifactorial Experiments. *Biometrika, 33*: 305 - 325.

Riggs, L.S and C-H Han (1991) A computer simulation interface. *Microcomputers in Civil Engineering 6.* 27-33.

Robbins, L. (1935). *An Essay on the Nature and Significance of Economic Science.* Macmillan, London.

Schmidt, J.W. and R.E. Taylor. (1970). *Simulation and Analysis of Industrial Systems.* Richard D. Irwin, Homewood, Ill.

Shannon, R.E. (1979). *Systems Simulation: The Art and Science.* Prentice-Hall, New Jersey.

Singhal, R.K. et. al.(1988) Optimising Surface Mining Productivity. *Colliery Guardian.* 236 no. 4. 108-112.

Skempton, A.W. and R.D. Northey. (1952) The Sensitivity of Clays. *Geotechnique 3.* 30-53.

Sprowls, C. (1964). Simulation and Management Control, in *Management Controls: New Directions in Basic Research,* ed. C.P. Bonini, et. al. McGraw-Hill, New York.

Stark, R.M. and R.L. Nicholls. (1972). *Mathematical Foundations for Design: Civil Engineering Systems.* McGraw-Hill, New York. NY.

Teicholz, P. (1963) A simulation approach to the selection of construction equipment. *Technical Report no. 26.* Construction Institute, Stanford University, Stanford, California.

Thompson, W.E. (1959) ERNIE, A Mathematical and Statistical Analysis. *Journal of the Royal Statistical Society. 122A.* 301-324.

Touran, A (1992) Automatic generation of simulation codes in construction. *Proceedings of the 8th Conference on Computing in Civil Engineering, ASCE.*

Van Horn, R.L. (1971). Validation of Simulation Results, *Management Science, 17(5).* 247-258.

Welch, P.D. (1983). The Statistical Analysis of Simulation Results, in *The Computer Performance Modeling Handbook.* S.S. Lavenburg, ed. 268-328. Academic Press, New York.

Whyte, I.L. (1982). Soil Plasticity and Strength - A New Approach Using Extrusion. *Ground Engineering.* January 1982. 16-24.

Whyte, I.L. (1994) Site Investigation Strategy for Contaminated Land. *Proceedings of the Third International Conference on Polluted and Marginal Land, Brunel University.*

Winston, W.L. (1987). *Operations Research: Applications and Algorithms.* Duxbury Press, Boston, Mass.

Zeigler, B.P. (1976). *Theory of Modelling and Simulation*, John Wiley, New York.

Appendices

Appendix 1: Empirical Test on the Random Number Generator

The random number generator contained within the software of the computer used in the simulations described in this thesis must be tested before it is used to ascertain that it does in fact generate numbers that are IID U(0,1). It is not known what type of generator it is and so a theoretical test (i.e. one that uses the numerical parameters of the generator without actually generating any numbers) cannot be carried out. Therefore an empirical test must be done which is a statistical test on a series of numbers generated; the test used is the chi-square goodness-of-fit test described in section 4.4.5. An alternative method is to use the correlation test as described in section 4.3.2 but to carry out such calculations on the large amount of random numbers that would need to be generated would be very time consuming.

To carry out the chi-square test, (as described in Law and Kelton, 1991) the range [0,1] is divided into k subintervals of equal length and a series of random numbers, U_1, U_2, ... , U_n, is generated. If j = 1, 2, ... , k then f_j is the number of random numbers in the jth interval. The test statistic is therefore:

$$\chi^2 = \frac{k}{n}\sum_{j=1}^{k}\left(f_j - \frac{n}{k}\right)^2 \tag{A1.1}$$

For large n, χ^2 will have an approximate chi-square distribution with k-1 degrees of freedom with a null hypothesis that the generated numbers are independent, identically distributed U(0,1) random variables. This hypothesis is rejected at a level α if $\chi^2 > \chi^2_{k-1,1-\alpha}$. This test has been carried out on 30 samples of n=1000 numbers assumed to agree with U(0,100), (since these numbers are divided by 100 to get U(0,1)), and k=100. The calculation was done using a simple C program (see figure A1.1) and the results are shown in table A1.1. As can be seen, at an α level of 0.4 we can accept the null hypothesis and use this random number generator with confidence.

```
/* Determine validity of random generator by using chi-square test
 * Simon Smith. 15-1-94 [1] */

#include <stdlib.h>
#include <stdio.h>
#include <math.h>

main()
{
   int f[101];
   int loop = 0;
   float unif[1000];
   int temp = 0;
   float chi = 0;

   /* set random seed to time */
   srand( (unsigned)time( NULL ) );

   /* set each value of fj to zero */
   for( loop = 0; loop < 100; loop++)
   {
      f[loop] = 0;
   }

   /* create array with random numbers in range U(0,100) */
   for( loop = 0; loop < 1000; loop++)
   {
      unif[loop] = rand();
      unif[loop] /= 327.68;
   }

   /* go through random numbers and allocate to each interval */
   for(loop = 0; loop < 1000; loop++)
   {
      temp = floor ( unif[loop] );
      f[temp] ++;
   }

   /* determine test statistic */
   for( loop = 0; loop < 100; loop++)
   {
      printf( "\nf[%d] = %d",loop, f[loop]);
      chi += (f[loop] - 10) * (f[loop] - 10);
   }
   chi *= 0.1;

   /* print out test statistic */
   printf("\nchi-squared for n=1000 and k = 100 is:  %5.5f",chi);

}
```

Figure A1.1 Code for chi-squared test on the random number generator.

run	χ^2	**run**	χ^2	**run**	χ^2
1	99.6	11	85.4	21	107.6
2	119.6	12	127.4	22	85.4
3	97.0	13	75.6	23	102.6
4	102.2	14	106.8	24	104.0
5	76.2	15	79.2	25	92.8
6	127.8	16	73.4	26	98.4
7	94.2	17	72.0	27	93.2
8	104.0	18	105.8	28	92.6
9	80.2	19	103.0	29	101.6
10	92.4	20	86.4	30	75.0
				av. χ^2=	95.4

$\chi^2_{99-1,1-0.4} \approx 101.9$

Accept null hypothesis at level $\alpha = 0.4$ since $95.4 < 101.9$

Table A1.1 Chi-square goodness-of-fit test for random number generator

Appendix 2: Random Variate Generator

The code listing shown in figure A2.1 is the program written in *Microsoft Quick C* for generating Erlang random variates with the acceptance-rejection algorithm outlined in section 4.4.2. There is a main program to retrieve the input from the user and generate the variates and three function routines.

The external definitions are the header files (files containing standard routines for input/output, math functions and library functions), the statement of functions defined (`erl`, `fact` and `maxf`) and one global variable, `kpling`, the factorial of (*k*-1) used in the Erlang density function.) The body of the main program contains declarations of the local variables and four sections. The first obtains the parameters of the Erlang distribution to be generated from the user i.e. *R, k*, the upper and lower bounds, the number of random variates to be generated and the random seed to be used. The second section calculates the maximum value of the density function over the range specified by calling to a separate function, `maxf`. The third section the generator itself. A loop is set up which continues until the user set number of variates have been generated. Each run of the loop first calculates two U(0,1) random numbers from the internal generator, calculates t^*, denoted by `tstar` in the program, and then tests for acceptance. If the number is accepted, `tstar` is added to the array of variates (`ran[num]`) and the loop continued. When all variates have been generated, the array is written to a user named file as text. This file can then be opened by any text editor or spreadsheet. An option to generate more numbers is given and if none is taken, the program is ended.

The three functions are `maxf`, `erl` and `fact`. The first calculates the maximum value of the density function by looping through each time in the range specified, calculating the density function for each and returning the largest value. The `erl` function calculates the Erlang probability density for the time value sent to it. This function also receives local copies of the Erlang parameters given by the user in the main program. The final function is a simple factorial calculator that receives an integer and returns a much larger integer (hence the double notation, used for large numbers).

The generator can calculate up to 1000 variates at a time and takes approximately 50 seconds to accept user input, generate the numbers and save to file. This is much more favourable than with using a spreadsheet generator - the convolution method takes approximately 100 seconds to calculate with a shape factor of 50 (quicker for smaller values) and the ARM using a spreadsheet takes much longer as rejected variates have to be removed by hand from the sheet.

```
/* random variate generator(1) */
/* Simon Smith. Version 3. 28 May 1993 */

#include <stdio.h>
#include <math.h>
#include <stdlib.h>

double erl(double r, double t, double kpling, int k);
double fact(int n);
double maxf( int a, int b, double r, double kpling, int k);
double kpling;

main()
{
   FILE *fp;
   double t,r,m;
   int a,b,k,nr,seed,check,loop;
   int gen,c;
   float r1,r2,tstar;
   float ran[1000];
   char fname[80];
   char ch;

   /* get input from user */
   new_gen:
   printf( "Enter R value: ");
   scanf( "%lf", &r);
   printf( "Enter k value: ");
   scanf( "%d", &k);
   kpling = fact(k - 1);
   printf ("Factorial of (%d - 1) = %lg\n\n",k,kpling);
   printf( "Enter lower bound: ");
   scanf( "%d", &a);
   printf( "Enter upper bound: ");
   scanf( "%d", &b);
   new_numbers:
   printf( "\nEnter number random numbers required: ");
   scanf( "%d", &nr);
   printf( "Enter seed (integer only):" );
   scanf( "%d",&seed );
   srand( seed );

   /* determine maximum value of Erlang function. */
   /* maxf function used */
   m = maxf(a,b,r,kpling,k);
   printf("m = %g\n",m);

   /* loop to generate variate */
   gen = 0;c = 0;
      while (gen <= nr)  /* loop until pre-set variates generated */
      {
         r1 = rand(); r1 /= 32768; /* internal number generator used */
         r2 = rand(); r2 /= 32768; /* give r1 and r2 */
         tstar = a + (b - a)*r1;
         if( r2 <= erl( r,tstar,kpling,k )/m)
         {
            ran[gen] = tstar;
            gen ++;
            printf("\b\b\b%d",gen);
         }
      c++;

   /* write array to file. () */
   printf( "\nEnter name of file, including path: ");
   scanf( "%50s",&(fname));
   if( (fp = fopen( fname, "wt" )) != NULL)
   {
```

Figure A2.1 Code listing for random variate generator

```
      for( loop = 0; loop < nr; loop++)
         fprintf( fp,"%f\n", ran[loop] );
      fclose( fp );
   }
   else
   printf( "Error: Couldn't create file.\n" );
   printf("More? Press g for new generation or n for more numbers.\n" );
   ch = getch();
   switch( ch )
   {
      case 'g':
         goto new_gen;
      case 'n':
         goto new_numbers;
      default:
         printf( "End of Variate Generation");
   }
}


/* maxf function to calculate maximum 2*/
/* over given range for erlang dist. */
double maxf( int a, int b, double r, double kpling, int k)
{
   int loop;
   double max = 10e-50;
   for( loop = a; loop < b; loop++ )
      {
      if ( erl( r,loop,kpling,k) > max ) max = erl( r,loop,kpling,k);
      }
   return max;
}


/* erlang density function */
double erl(double r, double t, double kpling, int k)
{
   /* calculate function */

   double res1,res2,res3,i;
   int j;
   i = r * t;
   j = k - 1;
   res1 = r/kpling;
   res2 = pow( i,j );
   res3 = exp( -i );
   return res1*res2*res3;
}

/* Factorial calculator */
double fact(int n)
{
   double result = 1;
   int loop;
   for( loop = 0; loop < n; loop++ )
      result = result * (loop + 1);
   return result;
}
```

Figure A2.1 cont. Code listing for random variate generator

Appendix 3. Trucksim Code.

The following pages list the code for all the individual functions in the trucksim program of which there are 15 plus the main function. The first figure is the list of the external definitions. When writing these notes its has been assumed that the reader understands the C programming language. The package used is the *Microsoft Quick C* compiler (Microsoft, 1990). Below is a brief description of each function.

i. *External Definitions.* These are: the external library functions for input/output, math calculations etc.; definitions for constants such as 0 for idle status and 1 for busy; variable declarations and function declarations.

ii. *Main function.* This is merely a menu and switching function for the various parts of the program. On first use, the user must choose the first section, *Change Configuration.* If other sections are chosen, the program cannot run because it has not been configured and consequently these functions will return the user back to the menu. After the user has configured the session, a parameter file (for the Erlang distributions) can be loaded by selecting number 2 on the menu. Although the parameters can be changed within the program, a file must be made to start with. This can be done with the Truckfile program. (This could eventually be incorporated into the Trucksim program but time constraints have made this a low priority.) Truck file can be seen in figure A3.15. Once parameters have been entered, a simulation can be run. There are two alternatives, single simulation for a single run where the results can be seen on the screen or multiple simulation for many replications - with this choice the results are sent to file and can be viewed through a text editor or spreadsheet. The results can be shown from this menu and the program is stopped if number 9 is chosen.

iii. *Change Configuration function.* This function has two separate parts. For the first running of the program, the initial configuration part is called; thereafter the configuration can be changed with this function. The configuration sets 4 variables for the running of the program and also the directory where the files are stored. The four variables are the number of trucks, the time length of the simulation run (in seconds), the number of buckets per load and the bucket volume of the excavator (to be entered in bank cubic metres).

iv. *Setup from file function.* This function reads files written with the Truckfile program. There are eight parameters for each time component which are stored in arrays. These parameters are: mean, variance, rate factor (R), shape factor (k), the factorial of `k - 1` (denoted by `kpling` in the code), the lower bound, the higher bound and the maximum density of the Erlang probability function for the range specified.

v. *View parameter function.* Once the parameters have been loaded using the above function, they can be seen on the screen by choosing this function from the menu.

vi. *Change parameter function.* Can be called from the main menu to change any of the parameters loaded into the program. The user is requested which time component to change and then has to enter six parameters, the mean, variance, the rate factor, the shape factor and the upper and lower bounds. The function then calculates the maximum density and the factorial of `k - 1` and changes these variables also.

vii. *Single simulation function.* This, along with the multiple simulation function, is the heart of the trucksim program. It is the controlling function which calls other functions when required. If parameters have been loaded and the program is configured, the user is requested to enter a random seed for the variate generator and a reference code for the results. The simulation is then initialised by calling the initialisation function and then run. The simulation is controlled very simply by first determining the next event (using the `event()` function) and then invoking the correct function to process the event: `arrive()` for an arrival event and `depart()` for a departure. This sequence of determining the next event and processing it is continued until the time limit (set by the configuration function) is reached. The program is then returned to the main function.

viii. *Multiple simulation function.* This function is essentially the same as the single simulation function except that the user is requested how many replications are required and a file name for the results to be written to. The seed inputted is the seed used for the first replication. For subsequent replications, the seed is increased by one each time. This function is very useful for the formation of production charts (section 5.4.2) and in experimental analysis (chapter 7) where many replications are required for the same initial setup.

ix. *Initialisation function.* This is necessary for all simulations. The simulation clock is first set to zero and then the state variables are initialised. The loader status is set to idle and the four other variables (number of trucks in queue, time of last event, total truck delay and total loader idle time) are all set to zero. The event list is initialised by setting the time of the next arrival of each truck in the simulation to zero and the time of the next departure of each truck to an arbitrarily chosen time of 1×10^{30} seconds. (This ensures that the first event for each truck is an arrival.) The event list is therefore initialised with two events for each truck. In this case two or more trucks have the same time for the same event. If this is the case then the trucks are scheduled in order of truck number. Therefore, this function sets the next truck to arrive and the next truck to depart to be truck number one. This function also initialises the counters used (which are the total loader delay, the total cycle time and the total times for the time components). These counters are used to provide averages for the report generator. These counters are all set to zero. There is a final counter, the number of cycles completed. As this counter is updated at the start of every cycle it must be first initialised to zero less the number of trucks used.

x. *Event determination function.* This is a very simple function that determines what type of event is next by comparing the time of the next arrival with that of the next departure. This function is also used to advance the simulation clock by setting the time variable to the value of the next event.

xi. *Arrival event function.* This function processes an arrival. It first updates the cycle counters by increasing the total cycle time by an amount equal to the current time less the time of the start of the current trucks cycle. The cycle start time for the current truck is then reset to the current time and the number of cycles completed is increased by one. The truck for the next arrival event is then scheduled (to the next truck; the trucks are always scheduled in numerical order) and the current truck is forced to depart before it can arrive again by setting its next arrival time to 1×10^{31}. The truck that has just arrived is then put in the queue or, if the loader is idle, has its next departure event scheduled by adding a generated spot time and load time to the time of the simulation clock. The arrival event function is described with the help of a flow chart in section 5.3.1.

xii. *Departure event function.* Processes a departure. The next departure is scheduled first (to happen to the next truck) and the truck that has just arrived has its next arrival time set to the time of the simulation clock plus times generated for the travel and dump components. The truck is then forced to arrive before it can depart again and the state of the system is checked. If the queue is empty then the loader is set to idle, otherwise a departure time is scheduled for the next truck in the queue, the queue length is reduced by one and the delay time for this next truck is computed. The trucks in the queue are then moved up one place.

xiii. *Report generator function.* This function first calculates all the results and then prints them to the screen. Examples of the results are shown in section 5.3.2. If multiple replications were done then the results are sent to a file.

xiv. *Random variate generator.* When this function is called the type of generator is sent to the function and so the correct parameters can be used. The generator itself is similar to the one described in appendix 2 but when load time variates are generated, separate variates must be generated for each swing time and then summed.

xv. *Erlang density function, maximum density function and factorial calculator.* These functions are exactly the same as those described in appendix 3. They are called by the variate generator function when needed.

xvi. *Truckfile program.* Shown in figure A3.15, this program is used to first obtain input from the user for the Erlang distribution parameters and then send these to file so that they can be loaded into the main Trucksim program.

```
/* TRUCKSIM.C[1] Version 5. 9-11-93. Simon Smith
* Single loader earthmoving system. Finite source.
* Erlang inter-arrival times.
* External definitions
*/

#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <graph.h>
#include <string.h>

#define BUSY 1
#define IDLE 0
#define ARRIVE 1
#define DEPART 2
#define TRUCK_LIMIT 15
#define SPOT 0
#define LOAD 1
#define TRAVEL 2
#define DUMP 3

int   next_event_type, num_in_q, loader_status, tim_type, seed,     buckets,
num_trucks,next_arrival_truck, next_depart_truck,   cycles, first;

float tim, tim_arrival[TRUCK_LIMIT], tim_last_event,
      tim_next_arrive[TRUCK_LIMIT], tim_next_depart[TRUCK_LIMIT],
      tim_length, total_of_delays, flag, total_load_delay,
      total_spot, total_load, total_dump, total_travel, volume,
      bucket_volume, tim_idle,
      total_of_ctime, start_cycle[TRUCK_LIMIT];


char   direc[20], path[50], ref[20];

FILE *infile, *outfile;

double r[4], max[4], kpling[4], mean[4], var[4];
int low[4], high[4], k[4];

void config(void);
void setup(void);
void view_param();
void change_param(void);
void single_simulation(void);
void multi_simulation(void);
void reportgen(int output_type);
void initialize(void);
void event(void);
void arrive(void);
void depart(void);
float erlang(int tim_type);
double erl(double rate, double t, double kfact, int shape);
double fact(int n);
double maxf( int a, int b, double rate, double kfact, int shape);
```

Figure A3.1 Trucksim code. External definitions

```
/* Main Function 2*/
main()
{
   int menu;
   char start, another;
   first = 0;
   r[0] = 0.0;

   /* Main Menu Screen */
   do {
   _clearscreen(_GCLEARSCREEN);
   printf("\n           *****************************************************\n"
         "           ***************    T R U C K S I M   ****************\n"
         "           *****************************************************");
   printf("\n\n\n           Excavator and Dump-Truck Earthmoving Simulator\n\n");
   printf("\n\n              M A I N   M E N U");
   printf("\n\n                1.  Change Configuration");
   printf("\n                2.  Load Parameter File");
   printf("\n                3.  View Parameters");
   printf("\n                4.  Change Parameters");
   printf("\n                5.  Single Simulation");
   printf("\n                6.  Multiple Simulations");
   printf("\n                7.  View Results");
   printf("\n                8.  Print Results");
   printf("\n                9.  Exit");
   printf("\n\n                ?:");
   do scanf("%d",&menu);
   while (menu > 9);
   switch (menu)
   {
      case 1:
         config();
         break;
      case 2:
         setup();
         break;
      case 3:
         view_param();
         break;
      case 4:
         change_param();
         break;
      case 5:
         single_simulation();
         break;
      case 6:
         multi_simulation();
         break;
      case 7:
         reportgen(1);
         break;
      case 8:
         reportgen(3);
         break;
      case 9:
         _clearscreen(_GCLEARSCREEN);
         printf("\nThankyou for using TRUCKSIM."
         break;
      default:
         break;
   }
   } while (menu != 9);
   return 0;
}
```

Figure A3.2 Trucksim code. Main function.

```
/* Configuration Function[3] */
void config()
{
   int menu;

   if (first != 1)
   {
      _clearscreen(_GCLEARSCREEN);
      printf("\n       I N I T A L  C O N F I G U R A T I O N\n\n\n");
      printf("\nEnter number of trucks: ");
      scanf("%d", &num_trucks);
      printf("\nEnter time length of simulation, in seconds: ");
      scanf("%f", &tim_length);
      printf("\nEnter name of directory where files are stored:  ");
      scanf("%20s",&direc);
      strcpy( path, "c:\\");
      strcat( path, direc);
      strcat( path, "\\");
      printf("\nPath is: %s",path);
      printf("\nEnter number of buckets per load: ");
      scanf("%d",&buckets);
      printf("\nEnter bank volume in bucket: ");
      scanf("%f",&bucket_volume);
      volume = buckets * bucket_volume;
      first = 1;
   }

   else
   {
      do
      {
      _clearscreen(_GCLEARSCREEN);
      printf("\n       C H A N G E   C O N F I G U R A T I O N\n\n\n");
      printf("\n       1.   Number of Trucks (Existing: %d)", num_trucks);
      printf("\n       2.   Time Length      (Existing: %0.0f secs)", tim_length);
      printf("\n       3.   Buckets per load (Existing: %d)",buckets);
      printf("\n       4.   Bucket Volume    (Existing: %0.2f m3)", bucket_volume);
      printf("\n       5.   File Directory   (Existing: %s)", direc);
      printf("\n       6.   Return to Main Menu");
      printf("\n       Truck Volume = %0.2f m3)", volume);
      printf("\n\n      ?: ");
      do scanf("%d",&menu);
      while (menu > 6);
      switch (menu)
      {
         case 1:
            printf("\nEnter new number: ");
            scanf("%d", &num_trucks);
            break;
         case 2:
            printf("\nEnter new time length: ");
            scanf("%f", &tim_length);
            break;
         case 3:
            printf("\nEnter new buckets per load: ");
            scanf("%d", &buckets);
            volume = buckets * bucket_volume;
            break;
         case 4:
            printf("\nEnter new bucket volume: ");
            scanf("%f", &bucket_volume);
            volume = buckets * bucket_volume;
            break;
         case 5:
            printf("\nEnter new directory: ");
            scanf("%20s",&direc);
            strcpy( path, "c:\\");
```

Figure A3.3 Trucksim code. Configuration function.

```
            strcat( path, direc);
            strcat( path, "\\");
            printf("\nPath is: %s\n",path);
            break;
         case 6:
            break;
         default:
            break;
      }
      } while (menu != 6);
   }
}
```

```
/* Setup from file[4] */

void setup(void)

{
   int setup;
   float value = 0;
   char fname[10], contin;

   _clearscreen(_GCLEARSCREEN);
   printf("\n       P A R A M E T E R   S E T U P \n\n\n");

   /* Check if a directory has been specified */
   if (strcmp ( path, "a") <= 0)
   {
      printf("\nFile directory not yet specified."
             "\nChoose 'Change Configuration' from Main Menu."
             "\n\nPress 'c' to continue");
      do
      {
         contin = getch();
         contin = tolower(contin);
      }  while (contin != 'c');
   }

   else
   {
      printf("\nEnter name of file where data is stored:  ");
      scanf("%10s",&fname);
      strcpy( path, "c:\\");
      strcat( path, direc);
      strcat( path, "\\");
      strcat( path, fname);
      strcat( path, ".par");
      printf("\nLoading: %s", path);

      if( (infile = fopen( path, "rb" )) != NULL)
      {
         fread( mean, sizeof(mean), 1, infile);
         fread( var, sizeof(var), 1, infile);
         fread( r, sizeof(r), 1, infile);
         fread( k, sizeof(k), 1, infile);
         fread( kpling, sizeof(kpling), 1, infile);
         fread( low, sizeof(low), 1, infile);
         fread( high, sizeof(high), 1, infile);
         fread( max, sizeof(max), 1, infile);
         fclose (infile);
      }

      else
      {
         perror( "\nRead error");
         printf("Press 'c' to continue.");
         do
         {
            contin = getch();
            contin = tolower(contin);
         }  while (contin != 'c');
      }
   }

}
```

Figure A3.4 Trucksim code. File setup function.

```
/* View Parameter Function 5*/
void view_param()
{
    char contin;
    int view;

    _clearscreen(_GCLEARSCREEN);
    printf("\n        V A R I A T E   P A R A M E T E R S \n\n\n");

    if (r[0] == 0)
    {
      printf("\nNo parameters loaded yet."
             "\nChoose 'Load Parameter File' from main menu."
             "\nPress 'c' to continue");
      do
      {
         contin = getch();
         contin = tolower(contin);
      }  while (contin != 'c');
    }

    else
    {
       for (view = 0; view < 4; view++)
       {
          if (view == 0) printf("\n**** Spot Time Parameters ****");
          if (view == 1) printf("\n**** Load Time Parameters ****");
          if (view == 2) printf("\n**** Travel Time Parameters ****");
          if (view == 3) printf("\n**** Dump Time Parameters ****");
          printf("\n     Mean is:\t%0.0lf\tR value: %lf",mean[view], r[view]);
          printf("\n Varience is:\t%0.0lf\tk value: %d",var[view], k[view]);
          printf("\n lower bound:\t%d",low[view]);
          printf("\nhigher bound:\t%d",high[view]);
          printf("\n     maximum:\t%lf",max[view]);
       }
       printf("\nPress 'c' to return to Main Menu");
       do
       {
          contin = getch();
          contin = tolower(contin);
       }  while (contin != 'c');
    }
}
```

Figure A3.5 Trucksim code. Parameter view function.

```
/* Change Parameter Function[6] */
void change_param()
{
   char contin;
   int menu;
   if (r[0] == 0)
   {
      _clearscreen(_GCLEARSCREEN);
      printf("\n       C H A N G E   P A R A M E T E R S\n\n\n");
      printf("\nNo parameters loaded yet. Choose 2 from main menu");
      printf("\nPress 'c' to continue");
      do
      {
         contin = getch();
         contin = tolower(contin);
      }  while (contin != 'c');
   }

   else
   {
      do
      {
      _clearscreen(_GCLEARSCREEN);
      printf("\n       C H A N G E   P A R A M E T E R S \n\n\n");
      printf("\n\n         1.  Spot Times");
      printf("\n         2.  Load Times");
      printf("\n         3.  Travel Times");
      printf("\n         4.  Dump Times");
      printf("\n         5.  Return to Main Menu");
      printf("\n\n       ?: ");
      do scanf("%d",&menu);
      while (menu > 5);
      if (menu == 5) break;
      menu--;
      _clearscreen(_GCLEARSCREEN);
      if ( menu == 0) printf("\n      S P O T   T I M E S");
      if ( menu == 1) printf("\n      L O A D   T I M E S ");
      if ( menu == 2) printf("\n      T R A V E L   T I M E S");
      if ( menu == 3) printf("\n      D U M P   T I M E S");
      printf( "\nMean is: %0.0lf.\t\t\tChange to: ", mean[menu]);
      scanf( "%lf", &mean[menu]);

      printf( "\nVarience is: %0.0lf.\t\tChange to: ", var[menu]);
      scanf( "%lf", &var[menu]);

      r[menu] = mean[menu] / var[menu];
      printf("\nR value is: %lf\n", r[menu]);

      printf("\nReal value of k is: %0.2lf", (mean[menu] * r[menu]) );
      printf("\tEnter k as integer: ");
      scanf("%d", &k[menu]);

      kpling[menu] = fact(k[menu] - 1);
      printf ("\nFactorial of (%d - 1) = %lg",k[menu],kpling[menu]);
      printf( "\n\nLower bound is: %d.\t\tChange to: ", low[menu]);
      scanf( "%d", &low[menu]);
      printf( "\nUpper bound is: %d.\t\tChange to: ", high[menu]);
      scanf( "%d", &high[menu]);
      /* determine max value of function*/
      max[menu] = maxf(low[menu],high[menu],r[menu],
                      kpling[menu], k[menu]);
      printf("\nmax = %lf\n\n",max[menu]);
      } while (menu != 4);

   }

}
```

Figure A3.6 Trucksim code. Change parameter function.

```
/* Single Simulation Function [7]*/

void single_simulation(void)
{
   char start, contin;

   _clearscreen(_GCLEARSCREEN);
   printf("\n        S I N G L E   S I M U L A T I O N \n\n\n");

   /* Check if simulation parameters are in memory */
   if (r[0] == 0)
   {
      printf("\nNo parameters loaded yet."
             "\nChoose 'Load Parameter File' from main menu."
             "\nPress 'c' to continue");
      do
      {
        contin = getch();
        contin = tolower(contin);
      }  while (contin != 'c');
   }

   else
   {
      printf("\nEnter seed for simulation: ");
      scanf("%d",&seed);
      srand( seed );
      printf("\nEnter reference code: ");
      scanf("%s", ref);

      /* Initialize the simulation */
      initialize();


      /* Run simulation until time length is reached */
      printf("\n\nStart Simulation?\n");
      do
        {
        start = getch();
        start = tolower(start);
        } while (start != 'y');

      printf("Please Wait...");
      printf("\nSimulated time is:\n");

      do
      {
        /* Determine the next event */
        event();

        /* Invoke event function */
        switch (next_event_type)
        {
           case ARRIVE:
              arrive();
              break;
           case DEPART:
              depart();
              break;
        }

        /* Print time on screen. Comment this to run faster */
        printf("\b\b\b\b\b\b\b%0.0f", tim);

      }while (tim <= tim_length);  /* End of simulation loop */
   }
```

Figure A3.7 Trucksim code. Single simulation function.

```
/* Multiple simulation function[8] */
void multi_simulation(void)
{
   char start, contin, fname[10];
   int seed_start, num_sims, sim;
   _clearscreen(_GCLEARSCREEN);
   printf("\n     M U L T I P L E   S I M U L A T I O N S \n\n\n");
   /* Check if simulation parameters are in memory */
   if (r[0] == 0)
   {
      printf("\nNo parameters loaded yet."
             "\nChoose 'Load Parameter File' from main menu."
             "\nPress 'c' to continue");
      do
      {
         contin = getch();
         contin = tolower(contin);
      }  while (contin != 'c');
   }
   /* Continue with simulations */
   else
   {
      printf("\nEnter number of simulations: ");
      scanf("%d",&num_sims);
      printf("\nEnter first seed: ");
      scanf("%d", &seed_start);
      printf("\nEnter file name for results: ");
      scanf("%s", fname);
      strcpy( path, "c:\\");
      strcat( path, direc);
      strcat( path, "\\");
      strcat( path, fname);
      strcat( path, ".csv");
      printf("\nSaving results to: %s", path);
      printf("\nThese results can be viewed using a spreadsheet (eg. EXCEL)"
             "\nOpen file as 'Comma Separated Values' file (CSV)"
             "\n\nCalculating Simulation Run No. :\n");
      /* Start Simulations */
      if( (outfile = fopen(path, "wt" )) != NULL)
      {
         for (sim = seed_start; sim <= (seed_start + num_sims); sim++)
         {
            srand( sim );
            /* Initialize the simulation */
            initialize();
            /* Run simulation until time length is reached */
            printf("\b\b\b %d",sim);
            do
            {
               /* Determine the next event */
               event();
               /* Invoke event function */
               switch (next_event_type)
               {
                  case ARRIVE:
                     arrive();
                     break;
                  case DEPART:
                     depart();
                     break;
               }
            } while (tim <= tim_length);  /* End of simulation loop */
            /* Write results of simulation to outfile via reportgen() */
            reportgen(2);
         }
      fclose (outfile);
      }
```

Figure A3.8 Trucksim code. Multiple simulation function.

```
        else
        {
           perror( "\nRead error");
           printf("Press 'c' to continue.");
           do
           {
              contin = getch();
              contin = tolower(contin);
           }  while (contin != 'c');
        }
     }
  }
```

```
/* Initialization function 9*/
void initialize(void)

{
   int i;

   /* Initialize simulation clock */
   tim = 0.0;

   /* Initialize state variables */
   loader_status    = IDLE;
   num_in_q         = 0;
   tim_last_event   = 0.0;
   total_of_delays  = 0.0;
   tim_idle         = 0.0;

   /* Initialize event list */
   for (i = 1; i <= num_trucks; i++)
   {
      tim_next_arrive[i] = 0.0;
      tim_next_depart[i] = 1.0e+30;
      start_cycle[i]     = 0.0;
   }
   tim_arrival[1] = 0;
   next_arrival_truck = 1;
   next_depart_truck  = 1;

   /* Initialize average counters */
   total_load_delay = 0.0;
   total_of_ctime   = 0.0;
   total_spot       = 0.0;
   total_load       = 0.0;
   total_dump       = 0.0;
   total_travel     = 0.0;
   cycles = 0 - num_trucks;
}
```

Figure A3.9 Trucksim code. Initialisation function.

```
/* Event determination function[10] */
void event(void)
{

   /* Determine next event type */
   if (tim_next_arrive[next_arrival_truck] <=
        tim_next_depart[next_depart_truck])
   {
      next_event_type = ARRIVE;
      tim = tim_next_arrive[next_arrival_truck];
   }
   else
   {
      next_event_type = DEPART;
      tim = tim_next_depart[next_depart_truck];
   }
}


/* Arrival event function */
void arrive(void)
{
   float load_delay;
   char start;

   /* Following variable for clarity only */
   int this_truck = next_arrival_truck;

   /* Update cycle counter */
   total_of_ctime += (tim - start_cycle[this_truck]);
   start_cycle[this_truck] = tim;
   cycles++;

   /* Schedule next truck to arrive */
   next_arrival_truck++;
   if (next_arrival_truck > num_trucks)
        next_arrival_truck = 1;

   /*
    * Make this truck unable to arrive
    * (New arrival time to be scheduled once truck has departed)
    */
   tim_next_arrive[this_truck] = 1.0e+31;

   /* Is loader busy? */
   if (loader_status == BUSY)
   {
      num_in_q++;
      tim_arrival[num_in_q] = tim;
   }
   else
   {
      loader_status = BUSY;
      tim_next_depart[this_truck] = tim + erlang(SPOT) +erlang(LOAD);
      load_delay = tim - tim_idle;
      total_load_delay += load_delay;
   }
}
```

Figure A3.10 Trucksim code. Event determination and arrival event function.

```
/* Departure event function[11] */
void depart(void)
{
   int i;
   char start;
   float delay;
   int this_truck = next_depart_truck;


   /* Schedule next truck to depart */
   next_depart_truck++;
   if (next_depart_truck > num_trucks)
       next_depart_truck = 1;

   /* Schedule next arrival time for this truck */
   tim_next_arrive[this_truck] = tim + erlang(TRAVEL) + erlang(DUMP);

   /* Make this truck unable to depart again until it has arrived */
   tim_next_depart[this_truck] = 1.0e+31;

   /* Is queue empty? */
   if (num_in_q == 0)
   {
       loader_status = IDLE;
       tim_idle = tim;
   }

   else
   {
      num_in_q--;
      delay = tim - tim_arrival[1];
      total_of_delays += delay;

      /* Schedule departure time for next truck in queue */
      tim_next_depart[next_depart_truck] = tim + erlang(SPOT) +erlang(LOAD);

      /* Move each truck up one place in queue */
      for (i = 1; i <= num_in_q; i++)
           tim_arrival[i] = tim_arrival[i+1];
   }
}
```

Figure A3.11 Trucksim code. Departure event function.

```
/* Report Generator[12] */
void reportgen(int output_type)
{
   float max_prod, poss_prod, prod, match_factor, bunch_factor, dump_util,
         load_util, avspot, avload, avdump, avtrav, avtruck_delay,
         avload_delay, avcycle;
   char contin;

   _clearscreen(_GCLEARSCREEN);
   printf("            S I M U L A T I O N   R E S U L T S ");

   if (cycles <= 1)
   {
      printf("\nNo simulations run yet. Press 'c' to continue");
      do
      {
         contin = getch();
         contin = tolower(contin);
      }  while (contin != 'c');
   }

   else {
   avspot = total_spot / cycles;
   avload = total_load / cycles;
   avdump = total_dump / cycles;
   avtrav = total_travel / cycles;
   avcycle = total_of_ctime / cycles;
   avtruck_delay = total_of_delays / cycles;
   avload_delay  = total_load_delay / cycles;
   dump_util = 100 - ((total_of_delays/num_trucks)/tim)*100;
   load_util = 100 - (total_load_delay/tim)*100;
   max_prod  = (3000 / (avspot + avload)) * volume;
   match_factor = (num_trucks * (avspot + avload)) /
                  (avcycle - avtruck_delay);
   if (match_factor < 1) poss_prod = max_prod * match_factor;
   else poss_prod = max_prod;
   prod = (cycles * volume / tim)*3000;
   bunch_factor = prod / poss_prod;

   /* Print to Screen */
   if (output_type == 1)
   {
      printf("\n\n               Reference Code: %s\t", ref);
      printf("\n                           Seed: %d", seed);
      printf("\n                 Total Run Time: %2.2f\tsecs", tim);
      printf("\n         No. Cycles Completed: %d", cycles);
      printf("\n\n                  No. Haulers: %d", num_trucks);
      printf("\n                 Match Factor: %2.2f", match_factor);
      printf("\n  Productivity after Mismatch: %2.2f\t m3 / hr.", poss_prod);
      printf("\n              Bunching Factor: %2.2f", bunch_factor);
      printf("\n\n       Dump-Truck Utilization: %2.2f\t\percent", dump_util);
      printf("\n           Loader Utilization: %2.2f\t\percent", load_util);
      printf("\n\n            Average Spot Time: %2.2f\tsecs.", avspot);
      printf("\n            Average Load Time: %2.2f\tsecs.", avload);
      printf("\n            Average Dump Time: %2.2f\tsecs.", avdump);
      printf("\n          Average Travel Time: %2.2f\tsecs.", avtrav);
      printf("\n          Average Truck Delay: %2.2f\tsecs.", avtruck_delay);
      printf("\n         Average Loader Delay: %2.2f\tsecs.", avload_delay);
      printf("\n           Average Cycle Time: %2.2f\tsecs.", avcycle);
      printf("\n\n\a         Overall Productivity: %2.2f\tm3 / 50min.hr.", prod);
      printf("\nPress 'c' to continue");
      do
      {
         contin = getch();
         contin = tolower(contin);
      }  while (contin != 'c');
   }
```

Figure A3.12 Trucksim code. Report generator function.

```
   /* Print to File */
   else if (output_type == 2)
   {
      fprintf(outfile, "%2.0f,%2.0d,%2.0d,%2.2f,%2.0f,%2.2f,%2.0f,%2.0f,"
                       "%2.0f,%2.0f,%2.0f,%2.0f,%2.0f,%2.0f,%2.0f,%2.0f\n",
              tim,cycles,num_trucks,match_factor,poss_prod,bunch_factor,
              dump_util,load_util,avspot,avload,avdump,avtrav,avcycle,
              avtruck_delay,avload_delay,prod);
   }
  }
}
```

```
/* Random Variate Generator 13*/

float erlang(int tim_type)

{

   /* Acceptance - Rejection loop to generate variate */
   float r1,r2,tstar, loadstar;
   int gen, load;
   r1 = r2 = tstar = 0;

   /* A/R Loop for load times */
   if (tim_type == 1)
   {
      for (load = 0; load < (buckets - 1); load++)
      {
         gen = 0;
         while (gen !=1)
         {
            r1 = rand(); r1 /= 32768;
            r2 = rand(); r2 /= 32768;
            loadstar = low[1] + ((high[1] - low[1]) * r1);
            if( r2 <= erl ( r[1],loadstar,kpling[1],k[1] ) / max[1])
               gen = 1;
         }
         tstar += loadstar;
      }
   }

   /* A/R Loop for other variates */
   else
   {
      gen = 0;
      while (gen !=1)
      {
         r1 = rand(); r1 /= 32768;
         r2 = rand(); r2 /= 32768;
         tstar = low[tim_type] +((high[tim_type] - low[tim_type]) * r1);
         if( r2 <= erl
            ( r[tim_type],tstar,kpling[tim_type],k[tim_type] )
            / max[tim_type])
            gen = 1;
      }
   }

   if( tim_type == 0) total_spot += tstar;
   if( tim_type == 1) total_load += tstar;
   if( tim_type == 2) total_travel += tstar;
   if( tim_type == 3) total_dump += tstar;
   return tstar;
}
```

Figure A3.13 Trucksim code. Random variate generator code.

```
/* erlang density function [14]*/

double erl(double rate, double t, double kfact, int shape)

{
   /* calculate function */

   double res1,res2,res3,i;
   int j;
   i = rate * t;
   j = shape - 1;
   res1 = rate/kfact;
   res2 = pow( i,j );
   res3 = exp( -i );
   return res1*res2*res3;
}



/* Function to calculate max. over given range for erlang dist. */


double maxf( int a, int b, double rate, double kfact, int shape)
{
   int loop;
   double max = 10e-50;
   double temp;
   for( loop = a; loop < b; loop++ )
      {
      temp = erl( rate, loop, kfact, shape );
      if ( temp > max )
        max = temp;
      }
   return max;
}



/* Factorial calculator */

double fact(int n)
{
   double result = 1;
   int loop;
   for( loop = 0; loop < n; loop++ )
      result = result * (loop + 1);
   return result;
}
```

Figure A3.14 Trucksim code. Erlang densitiy function, maximum probability function and factorial calculator function.

```
/* TRUKFILE.C      date: 27-May-1993   author: Simon D. Smith[15]
 * Program to make parameter files to be used with TRUCKSIM.C
 */

#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#include <graph.h>

double erl(double rate, double t, double kfact, int shape);
double fact(int n);
double maxf( int a, int b, double rate, double kfact, int shape);

double r[4], max[4], kpling[4], mean[4], var[4];
int low[4], high[4], k[4];

FILE *outfile;

char direc[20], path[50], fname[10], start, another;
int setup, extra;


main()
{
   _clearscreen(_GCLEARSCREEN);
   printf("\nThis program is to be used to create parameter files for use");
   printf("\nwith TRUCKSIM. All files will be saved in an allocated");
   printf("\ndirectory with a '.par' extension");
   printf("\n\nEnter name of directory where files are to be stored: ");
   scanf("%20s",&direc);

   extra = 0;
   while (extra != 1)    /* Start of single file loop */
   {
   strcpy( path, "c:\\");
   strcat( path, direc);
   strcat( path, "\\");
   printf("\nPath is: %s", path);
   printf("\nContinue? ");
   do
      {
      start = getch();
      start = tolower(start);
      } while (start != 'y');

   for (setup = 0; setup < 4; setup++)
   {
      switch (setup)
         {
            case 0:
            _clearscreen(_GCLEARSCREEN);
            printf("\n******* Setup Spot Time Parameters *******\n");
            break;
            case 1:
            _clearscreen(_GCLEARSCREEN);
            printf("\n******* Setup Load Time Parameters *******\n");
            break;
            case 2:
            _clearscreen(_GCLEARSCREEN);
            printf("\n******* Setup Travel Time Parameters *******\n");
            break;
            case 3:
            _clearscreen(_GCLEARSCREEN);
            printf("\n******* Setup Dump Time Parameters *******\n");
            break;
         }
```

Figure A3.15 Truckfile code.

```
   /* get input from user 16*/
   printf( "\nEnter mean: ");
   scanf( "%lf", &mean[setup]);

   printf( "\nEnter varience: ");
   scanf( "%lf", &var[setup]);

   r[setup] = mean[setup] / var[setup];
   printf("\nR value is: %lf", r[setup]);
   printf("\nReal value of k is: %lf", (mean[setup] * r[setup]) );
   printf("\nEnter k as integer: ");
   scanf("%d", &k[setup]);

   kpling[setup] = fact(k[setup] - 1);

   printf( "\n\nEnter lower bound: ");
   scanf( "%d", &low[setup]);

   printf( "\nEnter upper bound: ");
   scanf( "%d", &high[setup]);

   /* determine max value of function*/
   max[setup] = maxf(low[setup],high[setup],r[setup],
                    kpling[setup], k[setup]);
   printf("\nContinue? ");
   do
      {
      start = getch();
      start = tolower(start);
      } while (start != 'y');
}      /* End of input loop*/

for (setup = 0; setup < 4; setup++)
   {
   if (setup == 0) printf("\n**** Spot Time Parameters ****");
   if (setup == 1) printf("\n**** Load Time Parameters ****");
   if (setup == 2) printf("\n**** Travel Time Parameters ****");
   if (setup == 3) printf("\n**** Dump Time Parameters ****");
   printf("\nmean is: %lf",mean[setup]);
   printf("\nvarience is: %lf",var[setup]);
   printf("\nlow value is: %d",low[setup]);
   printf("\nhigh value is: %d",high[setup]);
   printf("\nmaximum is: %lf",max[setup]);
   }

/* Write data to  binary file */

printf("\nEnter name of file where data is to be stored.");
printf("\nAll files will have '.par' extension.");
scanf("%10s",&fname);
strcat( path, fname);
strcat( path, ".par");
printf("\nSaving: %s", path);
if( (outfile = fopen( path, "wb" )) != NULL)
   {
      fwrite( mean, sizeof(mean), 1, outfile );
      fwrite( var, sizeof(var), 1, outfile );
      fwrite( r, sizeof(r), 1, outfile );
      fwrite( k, sizeof(k), 1, outfile );
      fwrite( kpling, sizeof(kpling), 1, outfile);
      fwrite( low, sizeof(low), 1, outfile);
      fwrite( high, sizeof(high), 1, outfile);
      fwrite( max, sizeof(max), 1, outfile);
      fclose (outfile);
   }
else
   perror( "Write error");
```

Figure A3.15 cont.

```
   printf("\nAnother file? (y/n).");
   another = getch();
   another = tolower(another);
   if (another == 'n')
      extra = 1;
   }      /* End of single file while loop */
   return 0;
}
```

Appendix 4: Papers Accepted for Publication

The following pages contain three papers which have been accepted for publication. Two have not, at the time of writing (05/95) been published. The papers are:

1. Smith, S.D; J.R.Osborne and M.C.Forde. Influence of Match Factor and Bunching on Productivity Estimation in Dumptruck Earthmoving Operations. *Proceedings of the Institution of Civil Engineers Transport Journal.* (Published May 1995)

2. Smith, S.D; J.R.Osborne and M.C.Forde. The Use of a Discrete-Event Simulation Model with Erlang Probability Distributions in the Estimation of Earthmoving Production. *Civil Engineering Systems.* (To be published 1995)

3. Smith, S.D; J.R.Osborne and M.C.Forde. Experimental Analysis of an Earthmoving System Using Discrete-Event Simulation. *Journal of Construction Engineering and Management, ASCE.* (To be published 1995)

The work and results shown in these papers have originated from that detailed in this thesis.

Productivity estimation in back-acter/dump-truck earth-moving operations

S. D. Smith, *BEng*, *J. R. Osborne*, *BSc, CEng, MICE, and*
M. C. Forde, *BEng, MSc, PhD, CEng, MICE, MIHT, FINDT*

Proc. Instn Civ. Engrs Transp., 1995, **111**, May, 125–131

Transport Board
Transport Planning Panel

Paper 10552

Written discussion closes 17 July 1995

■ **The importance of productivity estimation in earth-moving is explained, the values of hauler and loader cycle times are introduced and their use in the calculation of productivity is outlined. Existing methods of calculating hauler cycle time are discussed. Reductions in productivity due to mismatch and bunching are investigated and the associated match factor and bunching factor are explained. Field data has been collected—the methods used for collecting data and the results obtained are outlined. Using this data, alternative methods of estimating loader and hauler cycle times are discussed and these are used, with examples, to determine productivity. A comparison with an actual operation is shown. Simulated values of bunching factor are then compared with actual field values. Guidelines to the use of these values and the other methods introduced are given.**

Introduction

The earthworks part of a major road building or widening scheme can involve the movement of millions of cubic metres of earth to a large number of permanent embankments or unsuitable fill tip sites. On very large contracts the total number of separate earth-moving operations can be very high and this, coupled with tight budgets and completion target dates, means that the earthworks should be planned as well as possible before site works start. Good preplanning not only helps to obtain the correct tender price (in which the earthworks subcontractor will have a very short time to reach this figure), but also aids in determining the amount of plant required in total on site. If the resources are incorrectly calculated large additional costs can be incurred once work has begun due to the hiring in of additional vehicles to complete an operation on time. Unfortunately, the contractor rarely has time to do this kind of preplanning work to the depth that he would like and may, therefore, rely on outdated and inaccurate estimation methods.

2. This Paper intends to set out the problems that need to be overcome in determining accurately the earth-moving operation productivity, numbers and types of plant required and the total cost. Theoretical methods already available will be compared with a generalized method of estimation based on results from observation of actual earth-moving operations. It is principally concerned with dump-truck/back-acter operations, which on long road schemes will be the main method, although the ideas can be readily applied to scraper/pusher operations used for short hauls.

3. Earth-moving is and will remain a very unpredictable branch of civil engineering. This Paper is only intended to set out methods for a good estimate at the planning stage. More complex methods can then be used once the works have started to modify any operations that are different than originally anticipated.

Principles of dump-truck/back-acter operations

4. The productivity of a single dump truck (hauler) is governed by how many cycles per hour it can achieve—one cycle being a complete trip from load point to dump area to load point again. The number of cycles per hour is dependent on the cycle time, and so calculation of cycle time is at the root of all productivity estimation. This cycle time is best determined by a knowledge of the individual components that make up a cycle: load time, haul time, dump time, return time, wait time and manoeuvre (or spot) time. Of these the haul and return times (collectively travel time) are perhaps the most difficult to determine. Various methods can be used from the very simple (time = haul distance/average haul speed)[1] to complex computer simulation packages such as Caterpillar's VEHSIM[2] and ACCELERATOR by Accelerator.[3] Manufacturers also provide time estimation charts.[4–6] Most of the existing methods, however, suffer from two distinct disadvantages. First, the calculated times are usually based on the best performance of the haulers and so actual times will be slower as the haulers will rarely be driven at their full potential. Second, these methods require knowledge of the rolling resistance of the haul route, which is the resistance the ground gives to the moving vehicle: soft ground will give high rolling resistances while very hard surfaces give very little resistance. The rolling resistance, however, is a notoriously difficult factor to determine and may also vary considerably over the length of the haul route.

S. D. Smith, Teaching Company Associate with the University of Edinburgh and Tarmac Construction Ltd

J. R. Osborne, Operations Director, Tarmac Structural Repairs

M. C. Forde, Tarmac Professor of Civil Engineering Construction, University of Edinburgh

5. One method of improving the theoretical calculation involves determining the difference between actual and computed times for existing operations.[7] It was found that actual travel times were 21% longer, on average, than the calculated times. Haul times for future operations were then mathematically calculated from the haul length and rolling and grade resistances which were then increased by 21%. This method can only be used for changes in haul length and conditions: load and dump conditions must remain constant.

6. Once the cycle times for individual haulers are calculated the productivity of that hauler is simply the product of the number of cycles per hour and the volume carried per cycle by that hauler. For multiple hauler operations, the interaction between hauler and loader must be taken into account to obtain the overall productivity. The maximum possible production is dependent on the loader (not on the number of haulers) and as long as the loader is constantly supplied with haulers then this maximum production will be maintained—hence the tendency to provide more haulers than are theoretically needed to ensure the maximum productivity. This will, of course, reduce hauler efficiency, thus increasing costs. Conversely, if the operation is under-resourced then the haulers will be operating at full efficiency, but this will probably mean that the loader is waiting for haulers to arrive—again lowering overall efficiency of the operation. The loaders and haulers can be balanced for an ideal match between the two; that is, when neither the haulers or the loader(s) have to wait. The research department of Caterpillar[2,7] has developed the dimensionless term match factor (*mf*), which can be determined for any operation where the cycle times of both the hauler and loader are known

$$\text{Match factor} = \frac{\text{No. of haulers} \times \text{loader cycle time}}{\text{No. of loaders} \times \text{hauler cycle time}} \qquad (1)$$

7. Note that cycle times do not include the idle time. The loader cycle time is the sum of the load time and the spot time of the hauler at the loader. An operation with a match factor (*mf*) of 1 is ideally matched. If $mf > 1$ then the operation is over-resourced; if $mf < 1$ then it is under-resourced. If match factor is not equal, or near, to 1 then the operation is ***mismatched***. The effect of hauler mismatch is summarized in Fig. 1, which introduces overall efficiency; that is, productivity as a proportion of the maximum possible production for the combination of plant working. As can be seen, if mismatch causes the match factor to be reduced below 1, then the overall operation efficiency reduces linearly. For operations with more trucks than necessary, the production will stay at its maximum, thus making it undesirable simply to keep adding more trucks to maintain production.

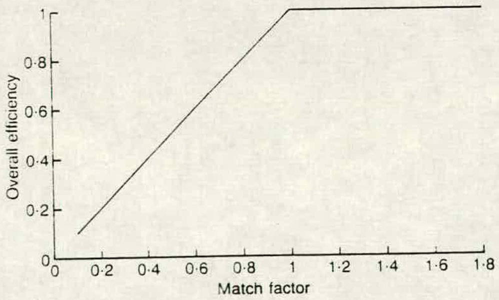


Fig. 1. The effect of plant match on operation productivity

8. However, this chart is based on the assumption that all hauler and loader cycle times are uniform, whereas this is unlikely to be the case: the time to complete a cycle will vary around a mean and cause truck *bunching*, with a consequent further loss of production.

Hauler bunching

9. Bunching occurs when the haulers are not evenly spaced along the haul road but bunch into groups of haulers moving very closely together. The effect of bunching is to reduce the possible productivity of an operation because the loader tends to wait for prolonged periods for a truck to arrive in the queue, hence causing lower utilization; when a truck arrives, more follow shortly afterwards, thus reducing productivity further due to lower utilization of the bunched trucks.

10. Over-resourced operations tend not to show the effects of bunching because the trucks in such an operation will have to wait in the queue anyway due to mismatch. This waiting acts as a buffer so that the time of arrival of a truck does not affect the loader, which will always have a truck to load. It is common practice on sites deliberately to over-resource an operation to maintain maximum production; while this may get the job finished on time, it is not an efficient method. Ideally, the maximum number of trucks should be used without reducing their utilization rates.

11. Morgan and Peterson[7] have studied the effects of bunching by carrying out a stochastic simulation of an earthworks operation. This involved sampling the distributions of all the various components of an earth-moving cycle and running the operation for a number of hours. This indicated that the effect of bunching was most marked at the perfect match point. At very high and low match factors, the effect tailed off; such operations, however, are already inefficient. The results of this stochastic simulation are shown in Fig. 2. It can be seen that an operation with a match factor of 1 has a reduction in productivity of approximately 15% due to bunching.

Productivity estimation

12. To summarize, the productivity of a hauler/loader operation can be estimated by way of the stages below.

- Calculate the hauler and loader cycle times.
- Determine the number of haulers to be used to maximize production from equation (1).
- Calculate the match factor.
- Using the loader cycle time, calculate the possible loader production.
- If match factor is below 1 then total production is the product of the loader production and the match factor (from Fig. 1). If match factor is greater than 1 then production remains at that of the loader.
- If necessary, allow for bunching.

13. To investigate the validity of the above stages and assumptions used, field data were collected by observing actual earth-moving operations. This data will also be used to develop alternative estimation methods for loader and hauler cycle times. The results and conclusions from the field observations will be outlined in the next part of the Paper.

Collection of field data

14. The earth-moving operations observed were part of a multi-million pound new road contract and consisted of some 60 separate cut-to-fill operations with one-way haul lengths varying from 100 m to 5000 m. Back-acter/dump-truck combinations were used for all lengths of haul and for some operations up to 1000 m, twin-engined motor-scrapers/pusher fleets were used. The soil was predominantly a silty clay with the particle size distribution (PSD) envelope for the soil giving ranges of 25–50% clay, 35–45% silt and 10–40% particles of sand size and greater. Typical Atterburg limits were $w_P = 18–23\%$ and $w_L = 45–54\%$, giving a plastic range of approximately 22–36%. These soil conditions gave average rut depths of 50–100 mm from a Volvo A35/Cat D400 type articulated dump truck.

15. All types of plant were observed over a period of seven months, with as much data as possible taken to be analysed later. For every operation observed the following was noted

- One-way haul length.
- Type and quantity of plant, including plant machine number.
- Loading start and finish times.
- End of travel/start of wait time (moment when hauler joins queue).
- End of cycle time (taken as moment when loading recommences).
- Volume carried per cycle. For dump trucks, this was estimated by counting the number of passes of the loader to fill the truck (bucket volume was calculated by physically measuring the bucket). For scraper operations, an estimate had to be made as to how full the box was. These volumes had to be factored for bulking on excavation.

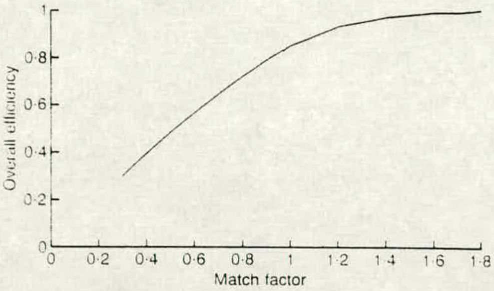


Fig. 2. Overall efficiency against match factor for simulated data (after Caterpillar[7])

16. From the observed times, the actual length of time for load, wait and cycle can be readily determined. This was used immediately to calculate an average production rate for the period of observation which could be given to the contractor to help with production control. Excessive over- or under-resourcing, for example, would be immediately apparent, and such observations were relayed to the contractors. Plant allocation on the site would often restrict ideally resourcing an operation, but this was somewhat offset by the fact that the perfect match point for an operation would be constantly changing. This would be due to the load and dump points being always on the move, and the effect this would have on the operations was not appreciated before collection of field data started. This is a very important fact as it means that resource allocation for maximum efficiency will be very difficult to achieve before operations start.

17. The data collected were stored for later analysis. The main aim was to improve estimations of the cycle time, and so the component parts of a cycle were studied separately. These were

- Load time.
- Haul time loaded.
- Dump time.
- Haul time empty.
- Wait time.
- Manoeuvre or spot time

Load time

18. This varies with the type of loader, and also the type of hauler due to different payloads. The best way to determine load time is therefore to build it up from individual bucket swing times, this being taken from the point where the bucket load is deposited to the next. For an observation of a full load time this can be calculated by dividing the load time by the number of buckets minus one—the first bucket being already loaded and waiting to be deposited.

19. The main type of excavator used in the

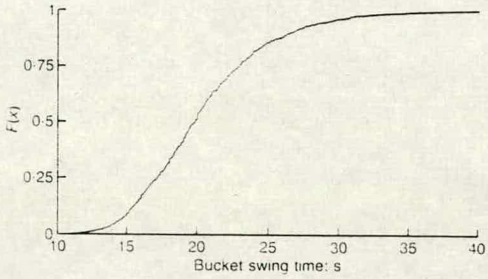


Fig. 3. Cumulative distribution function for bucket swing time for Caterpillar 245B

operation was a Caterpillar 245B, and, for all the load times observed, the swing times were calculated. The distribution of these swing times is shown in Fig. 3.

20. The load times for different haulers can then be made up if the volume, and thus the average number of buckets to fill the hauler, is known. This number of buckets will itself vary, but an assumption will have to be made as to whether the truck will be full or part full. To add to the problem, the bucket swing time will vary with the loading conditions. Depth of cut, relative heights of loader and hauler, angle of swing and type of material will all affect the bucket swing time. To give a general guide, loading conditions have been classified as good, average or bad, taking swing times at 0·2, 0·50 and 0·80 cumulative frequencies, respectively, from Fig. 3. Table 1 shows *possible* load times for a variety of haulers loaded with a Caterpillar 245. These figures will give a general guide to load times but, for more accuracy, the conditions of each loading area must be looked at in detail.

Dump time

21. Certain operations were observed at the dump area as opposed to the load area so as to obtain reliable dump times. It was felt that load times were more critical to the estimation of production and so, as there was only one observer on the site, much less data were taken for dumping than for loading. The data obtained is summarized as a cumulative distribution function in Fig. 4.

22. The times do not reflect the actual tipping (i.e. raising and lowering of the body), but are taken as the time from when the hauler enters to when it leaves the dump area. The haul road is therefore assumed to end at the edge of the dump area, and haul lengths were calculated on this basis. This edge was difficult to define in some cases but it was taken, as near as possible, as the point at which the haulers had slowed down from their haul speed to their manoeuvre speed.

23. The times shown in Fig. 4 are from observations of *all* types of dump truck. The average dump time is exactly 1 min and, in calculation of travel times (see next section), this is the time that has been used. However, there is quite a high degree of variation in the dump times and, as with load times, each load area should be considered separately for increased accuracy. Dumping conditions will vary with ease of access and the area for manoeuvring.

Table 1. Load times for full payload

Type of hauler	Number of buckets for full load at 1·9 m³/bucket	Load time: min			Loaded volume: m³
		Good	Average	Bad	
Volvo A35	7	1·64	1·98	2·38	13·3
Volvo A30	6	1·37	1·65	1·98	11·4
Volvo A25	5	1·10	1·32	1·58	9·5
Caterpillar D400	8	1·91	2·31	2·78	15·2
Euclid R35	7	1·64	1·98	2·38	13·3

Haul and return times

24. Ideally, individual haul and return empty times should be used, as these will vary depending on the gradient of the haul road and payload of the hauler. However, to get accurate individual times the operation would have to be observed at both ends, which was not possible. Therefore, a collective 'travel time' was calculated for each observed operation by subtracting load, wait, manoeuvre and dump times from the total cycle time. Load, wait and manoeuvre times were already known, but dump time had to be assumed and was taken as the average, which was 1 min.

25. These travel times can be used as the basis for future calculations of hauler cycle time by plotting travel time against haul length. If the haul length on a future operation is known, then an estimation of travel time can be made. However, there can be large differences between the haul conditions on one site and even more so if two or more sites are compared, so it has been found that the best way to estimate travel times is to use one of the abovementioned software packages that are available;[2,3] they can allow for differences in the haul road grade and rolling resistances as well as differences in the plant used.

Wait time

26. Although wait times were observed, these values are not used in the calculation of cycle times. The reason for this is because wait time is connected with the number of vehicles used—be it from hauler mismatch or bunching. The procedure is therefore to calculate the cycle time assuming no wait and then determine the number of haulers to be used and the match factor. For uniformly spaced vehicles, the wait time is then

$$t_w = t_l n_h - t_h \qquad (2)$$

where t_l = load cycle time, n_h = number of haulers and t_h = hauler cycle time. If this is then added to the cycle time, reduction in production due to mismatch can be allowed for. The wait time due to bunching is, however, more difficult to determine. To get reliable figures other methods such as queueing theory[8] or simulation[9,10] should be used. If the effect of waiting due to hauler bunching is known, then this can be applied directly as bunching factors without recourse to complex simulation techniques.

Manoeuvre or spot time

27. This is the positioning time of the hauler ready for loading, and has been taken as the time from when one hauler leaves the loader to when the next one starts to be loaded *assuming a loader has just arrived or is already waiting.*

28. The spot time added to the load time will together make the loader cycle time (t_l) and is therefore common to both hauler and loader cycles. It is one of the most sensitive factors involved in estimation of production: small changes in the spot time will have a marked effect on the *possible* loader production and hence on the overall operation productivity. This is shown graphically in Fig. 5. Three types of articulated hauler are represented in Fig. 5 and productivities have been calculated for a range of possible spot times—from 10 to 60 s. Each calculation of productivity is based on a haul travel time of 210 s with a dump time of 60 s and an average bucket swing time of 20 s—this gives the load times in the 'average' column of Table 1. The bank volume per pass is assumed to be 1·9 m^3.

29. As can be seen, for every increase of 5 s in spot time, total production is reduced by approximately 4% on average between the three types of hauler. This may seem a small amount but, for example, if three Volvo A35 dump trucks were to be used on the operation described above, then the increase in cost per cubic metre would be approximately £0·03/5 s increase in spot time, a considerable figure if large earthworks volumes are involved. The differences between the haulers in the chart are due solely to the differences in the payload volumes: thus, a D400 with a larger payload will spend a greater proportion of its load cycle on actual loading rather than manoeuvring.

Example of calculation of production

30. The method of estimation and the various components needed have now been outlined. To check the validity of using average results, a comparison will now be made using the average data and an actual operation.

31. Caterpillar D400 dump trucks were used on a haul of length 1980 m. Load, haul and dump conditions are assumed to be average, but to reduce wear on the haulers, they were only filled to 88% capacity (approximately seven bucketfuls from a Caterpillar 245 back-acter). Assume a working hour of 50 min.

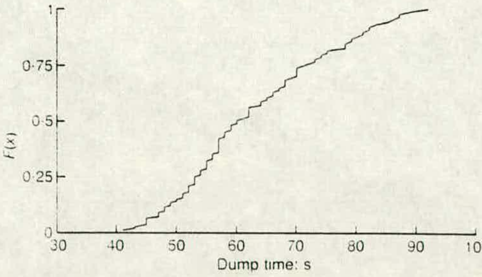


Fig. 4. Cumulative distribution function for dump time, all vehicles

Calculated results

32. Load time = 2·31 min (from Table 1)
Travel time = 12·27 min (from ACCELERATOR[3])
Dump time = 1·00 min (from Fig. 4)
Spot time = 0·50 min (assumed)
Cycle time = 16·08 min

Load cycle = 2·31 + 0·50 = 2·81 min

Ideal No. of haulers = 16·08/2·81
= 5·72 (use 6)

Hauler wait time = (2·81 × 6) − 16·08
= 0·78 min

Actual cycle time = 16·08 + 0·78
= 16·86 min

Match factor = (6 × 2·81)/(1 × 16·08)
= 1·05

Hauler volume = 7 × 1·85
= 12·95 m^3

Max. production = (50/2·81) × 12·95
= 230 m^3/50 min h

Match factor is greater than 1, therefore actual production

= 230 m^3/50 min h

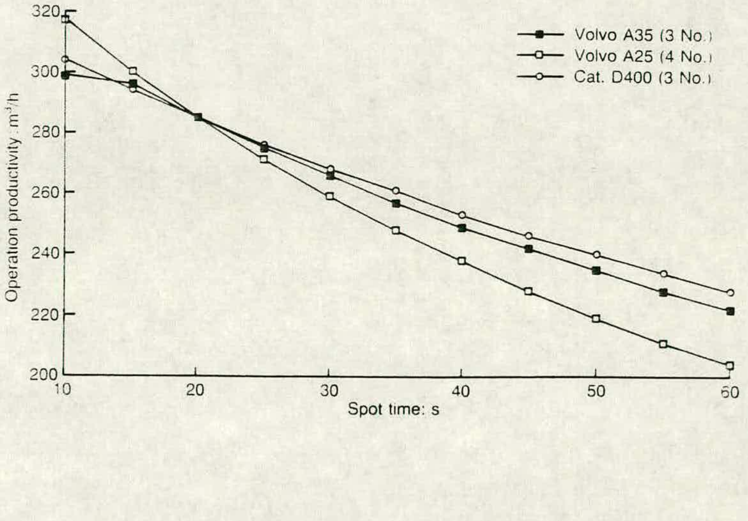


Fig. 5. The effect of spot time on operation productivity

Observed results

33. Load time = 2·08 min
Travel time = 10·51 min
Dump time = 1·00 min
Wait time = 2·30 min
Spot time = 0·57 min
Cycle time = 16·46 min

No. of haulers used = 6

Actual production = (50/16·46) × 12·95 × 6
= 236 m^3/50 min h

34. Therefore calculation was underestimated by 2·5%. At first glance this seems like a very good estimation, until the *possible* production by the loader is looked at

Max. production = (50/2·65) × 12·95
= 244 m^3/50 min h

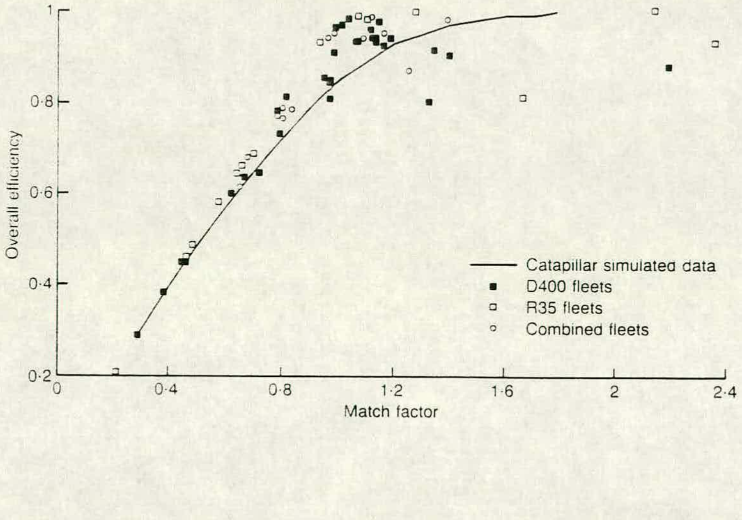
35. As the match factor (1·09) is greater than 1, then this production rate should have been achieved. This loss in production is therefore due to hauler bunching. Overall efficiency of actual production

= 236/244 × 100%
= 97%

Bunching calculations

36. Bunching was discussed briefly, and the results of simulated earth-moving operations are shown (Fig. 2). Using the field data gathered, actual values of bunching factor can be calculated based on the assumption that all further reduction in productivity after mismatch is due to bunching. If the possible production of the operation is known, then dividing the actual observed production by this value will give a bunching factor. Combining match and bunching factors into overall efficiency and plotting against match factor gives the graph shown in Fig. 6 for the various combinations of plant on the site. Direct comparison can be made with the simulated data calculated by Caterpillar and the line from Fig. 2 is included in Fig. 6.

Fig. 6. Effect of bunching on overall efficiency of plant



37. Looking at the actual data, the general trend can clearly be seen: at low match factors, inefficiency is due to mismatch only and calculated points agree with both the simulated data line and the uniform spaced line. At *mf* = 1, bunching can be seen to reduce overall efficiency but not, in general, as far as that predicted by Caterpillar. There is less data above *mf* = 1·2, but what there is shows that while the effect of bunching may be reduced, it can also be much more destructive. The reasons for this are probably many and varied, and will depend on the actual conditions of individual operations, but it is possible that such over-resourced operations simply have too many vehicles for the haul road—in effect a volume-induced traffic jam.

38. The results in Fig. 6 can be used to give estimates for bunching factors. The work by Caterpillar can be seen to be an approximate lower bound; combined with the field data, a bunching factor envelope can be defined for match factors under 1·2. Judgements would then have to be made as to how variable the hauler and loader cycle times will be. Returning to the example in the previous section, the calculated results showed a match factor of 1·05: Fig. 6 gives a corresponding bunching factor of between 0·86 and 1·00. Using any value below 1·00 will reduce the production estimate even further than the observed value but, as has been shown, the loader cycle time was badly estimated. The bunching factor of the actual estimation was calculated at 0·97 from a match factor of 1·09. With match factors above 1·2, bunching factors cannot be reliably determined from these results.

Conclusions

39. The ideas and data presented herein can together be used as an elementary predictive tool for earth-moving productivity. Such a tool is primarily intended for use before operations on site begin and, consequently, if actual conditions vary from those under which the data was taken, misleading results will be obtained.

40. The principal conclusions from the work done are

- To calculate the productivity of a single earth-moving hauler, its cycle time is required: the more accurate the calculated cycle time, the better the production estimate.
- The maximum possible production for a whole operation or system is dependent on, and very sensitive to, the loader cycle time. Accuracy of this time is therefore essential.
- Whether or not this production is achieved depends on how many haulers the system contains. For each system there will be a minimum or ideal number of haulers (being a function of both hauler and loader cycle

times) which will provide this production: if the number of haulers is reduced below this figure, production will be reduced. If the number is increased then the production will be maintained, but at a higher cost.

- Production is further reduced (by up to approximately 20%) by the effect of bunching of haulers. Bunching is a complex factor, the effects of which are probabilistic: they can only be readily determined by application of models such as queueing theory and simulation. If the causes are understood then factors can be applied to give an approximation of any reduction in productivity. Such causes are variability of hauler and loader cycle time and hauler mismatch.
- It has been shown here and elsewhere[7-11] that maximum production for lowest cost occurs when haulers and loaders are perfectly matched; it is also understood that the effects of bunching are greatest at this point and reduce as match factor increases above 1. It is common practice therefore to maintain high production by over-resourcing. However, it has been shown (Fig. 6) that at match factors greater than 120%, the effect of bunching can be very variable.

41. If these points are understood, then general estimates of production and resource requirements can easily be made. For more accurate estimations as operations progress and actual conditions are known, more complex methods will be required which are based not on average results but take the inherent variability of earth-moving into account.

Acknowledgements

42. The authors would like to acknowledge the help provided by the following people in the preparation of this Paper: the staff of the Department of Civil Engineering and Building Science at the University of Edinburgh, including Mr Graeme Wood and Mr Stephen Dunn for assistance in the collection of data; and the staff of Tarmac Construction Ltd, of Wolverhampton, including the Earthmoving Division. This work was supported under a Teaching Company Scheme funded by SERC, grant No. GR/H23306 and by Tarmac Construction Ltd.

References

1. Paul P. K. and De P. Matching of shovel-dumper at production planning stage. *J. Mines Metals Fuels*, 1989, April, 143–146.
2. Caterpillar. *Vehsim handbook*. Caterpillar, Peoria, Illinois, 1987.
3. Accelerator. *Accelerator handbook*. Accelerator, Fort Myers, Florida, 1987.
4. Caterpillar. *Caterpillar performance handbook*. Caterpillar, Peoria, Illinois, 1993, 24th edn.
5. VME. *Performance manual, Volvo BM articulated haulers*. VME Articulated Haulers AB, Sweden, 1989, 2nd edn.
6. Terex. *Production and cost estimating of material movement with earthmoving equipment*. Terex Equipment Ltd, Motherwell, UK, 1981.
7. Morgan W. C. and Peterson L. L. Determining shovel-truck productivity. *Min. Engng*, 1968, December, 76–80.
8. Charmichael D. G. *Engineering queues in construction and mining*. Ellis Horwood, Chichester, 1987.
9. Clemmens J. P. and Willenbrock J. H. The Scrapesim computer simulation. *J. Constr. Div. Am. Soc. Civ. Engrs*, **CO4**, 1978, Dec., 419–435.
10. Paulson B. C. *et al.* Simulation and analysis of construction operations. *J. Tech. Top. Civ. Engr.*, **109**, 1983, Aug., 89–104.
11. Singhal R. K. *et al* Optimising surface mining productivity. *Colliery Guard.*, **236**, No. 4, 1988, 108–112.

0039

1-20

Civil Eng. Syst., Vol. 00, pp. 000 000
Reprints available directly from the publisher
Photocopying permitted by license only

THE USE OF A DISCRETE-EVENT SIMULATION MODEL WITH ERLANG PROBABILITY DISTRIBUTIONS IN THE ESTIMATION OF EARTHMOVING PRODUCTION

S. D. SMITH[1], J. R. OSBORNE[2] and M. C. FORDE[3]

[1] *Teaching Company Associate with Tarmac Construction Limited andThe University of Edinburgh, EH9 3JL*
[2] *Operations Director, Tarmac Structural Repairs Limited, Wolverhampton, WV1 4RU*
[3] *Tarmac Professor of Civil Engineering Construction, The University of Edinburgh, EH9 3JL*

(Received March 1994; supply*.................)*

Earthmoving operations are a major part of many civil engineering works and as such they present an ideal opportunity for improved estimation forecasts and efficiency. This paper presents a computer simulation model which is based on the assumption that an earthmoving cycle can be broken down into its component parts which can be represented by Erlang probability distributions. Such a distribution can be generated on a computer and incorporated into a discrete-event simulation program. The development of the model and of a program carry out simulation runs is described. The importance of model validity is discussed; results from a simulation cannot be relied upon if the model does not actually agree with the real system. The results of a validation exercise are shown which compares system output data (obtained from actual contracts in the UK) and model output data. Simulation output data is also shown in a graphical form which is intended to ease the use of the model on a site basis.

KEY WORDS: Earthmoving, estimation, discrete-event simulation, validation

INTRODUCTION

With increased capacity and reduced work available, contractors are increasingly having to reduce their margins in order to obtain new contracts. Up to 20 percent of the cost of today's multi-million pound road building contracts may be spent on earthmoving and earthworks and as such they are an ideal area for improved estimation and efficiency and also for reduction of disputes[1]. Although work has been published on the subject of modelling earthmoving systems[2-5] few contractors in the UK. seems to use such methods and continue to resource and plan their earthmoving contracts with the same rules of thumb and 'experience' factors as they have for many years. Yet to implement a PC based estimation system for both tender work and for site control during construction works should be a straightforward matter and this paper outlines a system developed at the University of Edinburgh which is based on observations of a number of earthmoving projects in the UK.

The ideas presented are based on the fact that earthmoving is a classic, finite-source queuing system and as such can be analysed using two operations research methods: queuing theory and simulation. Queueing theory for this type of application has been studied extensively by Carmichael[6,7]. Simulation, however, is perhaps the most practical for earthmoving as it can be used as a 'What if?' tool when site operations are started. Such analysis is important as it will allow for variations of plant cycle times and interaction and interference of various items of plant which will cause inefficiencies in the operation.

DEVELOPMENT OF A SIMULATION MODEL OF THE EARTHMOVING SYSTEM

A *system*, according to Schmidt and Taylor[8], is defined to be a collection of entities that act and interact together toward the accomplishment of some logical end. The earthmoving system is therefore a collection of different types of plant which interact to achieve an end *i.e.* the movement of earth from A to B. To analyse how this system works we can experiment either with the actual system or with a model of the system. To analyse a mathematical model of the system, if valid, is a much more practical and economical solution if one considers that a dumptruck can cost between £300–£400 per day to operate.

To develop a mathematical model of the earthmoving system it is first necessary to understand the logical and quantitative relationships which will be manipulated. Figure 1 shows the earthmoving system schematically as a queuing system. The model to the developed is based on the assumption that the earthmoving system can be broken down into its component parts, or *cycle component times*. These are:

- Manoeuvre (Spot) time.
- Load time. This is built up from individual bucket swing times.
- Haul and Return time. Collectively Travel time.
- Dump time.

What is also required is a knowledge of how various loading conditions will affect the amount of material that can be taken in the excavator bucket each pass and the effect of reducing or increasing the number of buckets per dumptruck load.

If the average times, bucket quantity and passes per load for a certain operation are known, and these can be estimated using a knowledge of the haul road soil conditions (from site investigation reports), plant to be used, loading area conditions (i.e. depth of cut, condition of excavated material etc.), then a model can be used to estimate earthmoving productivity. However, real life operation components will never stay at one, fixed, value but will be distributed around this mean with some variance. The variable cycle component times tend to cause interference between plant items, known as 'bunching' and this is seen on site when, say, 3 trucks arrive at the queue at once and then none for the next ten minutes.

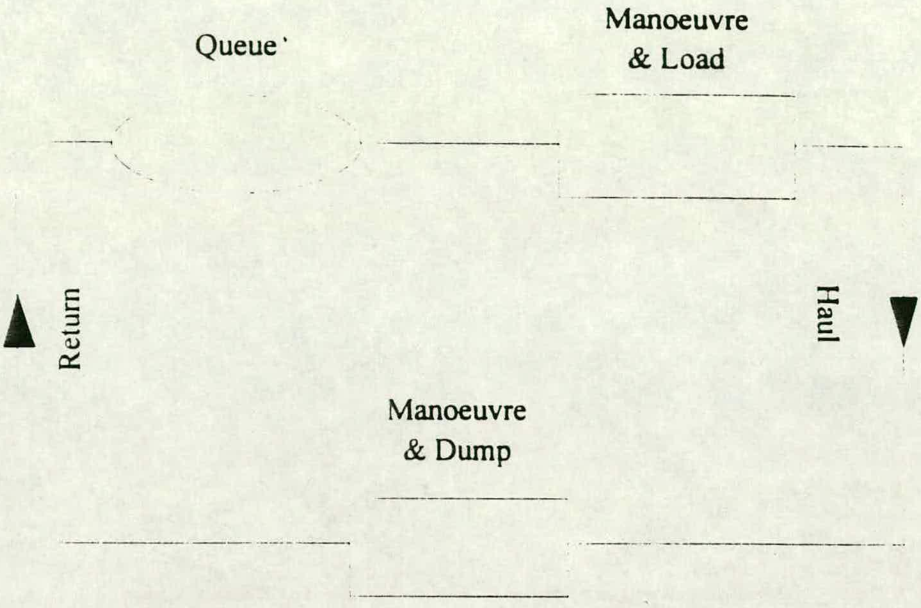


Figure 1 The Earthmoving System.

The simulation model therefore requires mathematical representations of actual time distributions. The best way to achieve this is to observe actual operations over a period of time and for different site conditions and then to plot the cumulative distribution function of these times. Various mathematical distributions can then be plotted alongside the actual distributions to see how good the fit is.

ERLANG DISTRIBUTION

It has been found that the best mathematical distribution to fit observed site data is the Erlang distribution[6,7]. An Erlang distribution is a continuous random variable (T) with a density function $f(t)$ defined by two parameters, R and k, the rate and shape parameters; and k must be a positive integer. If R and k are known then the probability density function (pdf) is:

$$f(t) = \frac{R(Rt)^{k-1}\, e^{-Rt}}{(k-1)!} \quad (t \geq 0) \tag{1}$$

By integration by parts, it can be shown that:

$$E(T) = \frac{k}{R} \tag{2}$$

and

$$\operatorname{var} T = \frac{k}{R^2} \tag{3}$$

where $E(T)$ is the expected time usually estimated by the mean and var T is the variance of the distribution. If observed values for mean and variance are used as the expected values then values of k and R can be computed for the observed distribution. To plot the cumulative distribution function (cdf) of an Erlang distribution the pdf must exist in a closed form which unfortunately it does not. Alternatively therefore, random values of the distribution can be generated on a microcomputer using an Acceptance-Rejection algorithm[9,p.958]. (This algorithm can be seen in Appendix 1). This has been done for all cycle components and for various site conditions. Figure 2 shows cdf's plotted for generated and observed values of bucket swing time for a Cat 245B backhoe excavator working in clay. It is important that the observed and proposed distribution do in fact agree with each other and this will be discussed later.

DISCRETE-EVENT SIMULATION

This type of simulation models the system as it develops over time. The *state* of the system is defined as 'the collection of variables necessary to describe the status of the system at any given time.'[9,p.934]. These variables, or *state variables*, are in this case the number of trucks in the queue, the number of busy loaders, the departure time of the next truck and the arrival time of the next truck. These variables will describe the state at the present time and in the future. In a system, an object of interest is an *entity* and any properties of the entities are called *attributes* which define the system state. In this system, the entities are the trucks (customers) and loaders (servers). The loader will have the attributes *busy* or *idle* and also the number of customers in the queue, while the

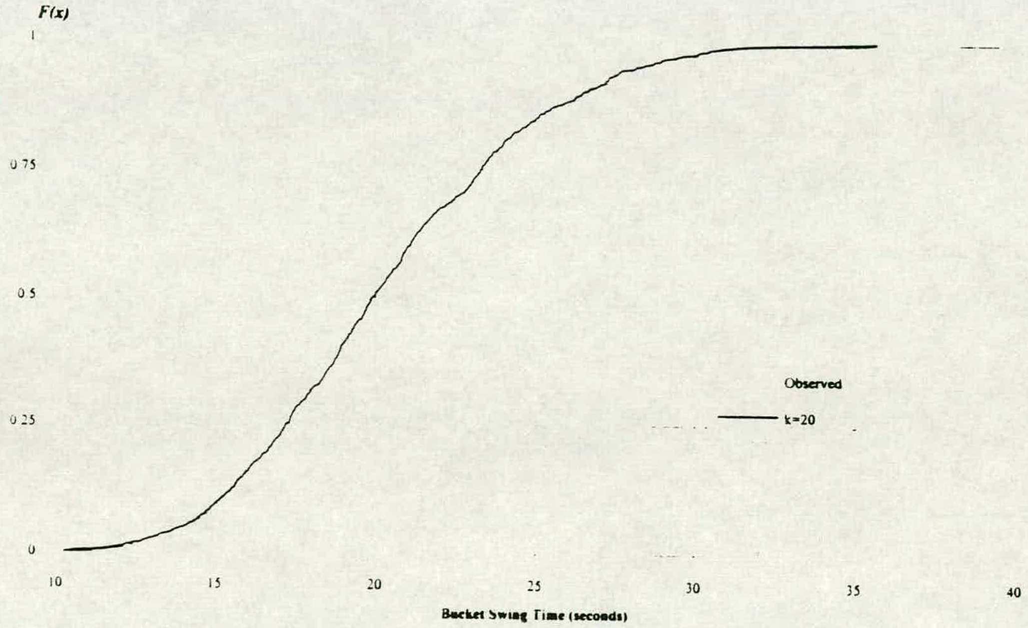


Figure 2 Cumulative Distribution Function for Bucket Swing Time for Cat 245B in Clay. (Observed and Generated Data).

trucks attributes are *arrival time* and *departure time*. The state of the system changes when an event occurs. In the earthmoving system, there are two main events and these are *arrival* and *departure* events.

Simulation works by trying to recreate the occurrence of these events. As the system evolves over time, the trucks depart and arrive at the queue and the time when these events occur will determine the cycle time of the trucks and the number of cycles completed. From a knowledge of the number of cycles completed per hour and the volume of earth carried per cycle (which is dependent on the size of the loaders bucket and the number of bucket passes per load) the productivity of the earthmoving operation can be calculated. To achieve this within a computer, cycle component times must be generated which agree with the pre determined distribution. This is done using the same algorithm used to draw up the cumulative distribution plots. (See Appendix 1). The random variate generator is used to draw up an event list from which the simulation program is run. This event list is an array in the computers memory which shows, in chronological order, the type, time and to which truck the next event will occur. For example, if truck *a* departs then the time of the next arrival of this truck can be determined by generating a random *travel* time and a *dump* time and adding these to the time at which the truck departed. This time is then added to the event list along with times for all the other trucks in the system.

In order to carry out simulations using the discrete-event procedure described above, the modeller has a choice of three methods: *Simulators, Simulation Languages* or *General Purpose Languages*. Simulators (for example WITNESS or XCELL+) are computer packages which can perform simulations of particular system with little or no programming required. Currently, however, most simulators are available for manufacturing, communications and computer systems. However, the CYCLONE series of computer programs[2-4] was developed specifically for construction processes, not just earthmoving, and would be of use in major earthmoving simulation studies. In developing a simulation model, however, a simulator, as opposed to a computer language, is of comparatively little use in understanding the system being studied-it has to be determined whether or not the output is likely. Simulators are best suited to an environment for which a model has already been developed and tested.

Simulation languages, (such as GPSS, SIMAN and SIMSCRIPT II.5) provide a natural framework for simulation modelling and have in-built routines for many of the features of various types of simulation (such as generating random values from a particular probability distribution, such as those provided by the routime shown in Appendix 1). Like simulators, simulation languages are geared towards manufacturing systems although will be more flexible if used for a different environment. Law and Haider[10] discuss various simulators and simulation languages.

Ultimately, however, the best choice of method depends on the modeller's experience. In developing a simulation programme, general purpose languages are ideal if a knowledge of a particular language exists. They also have the added advantage that, in having to pay attention to every detail when writing a program, conceptual errors are less likely to occur and there will be a greater understanding of how the system being modelled actually operates.

TRUCKSIM

A simulation program, called *Trucksim*, has been developed using discrete-event simulation in the general purpose language C. A flowchart for the main body of the program is shown in Figure 3. The program first requires configuration to the operation which is to be simulated. Therefore, the number of trucks, the average bucket volume, the average number of passes per load and the parameters for the cycle component times are required. For each cycle component i.e. travel time, dump time etc. four parameters are required: mean, variance, upper bound and lower bound. The Erlang distribution parameters k and R are calculated from this input using Eqs. 2 and 3. The simulation runs until a pre-set time limit has been reached, at which point the simulation stops and a report generator is invoked (Fig. 4). The productivity, utilisa-

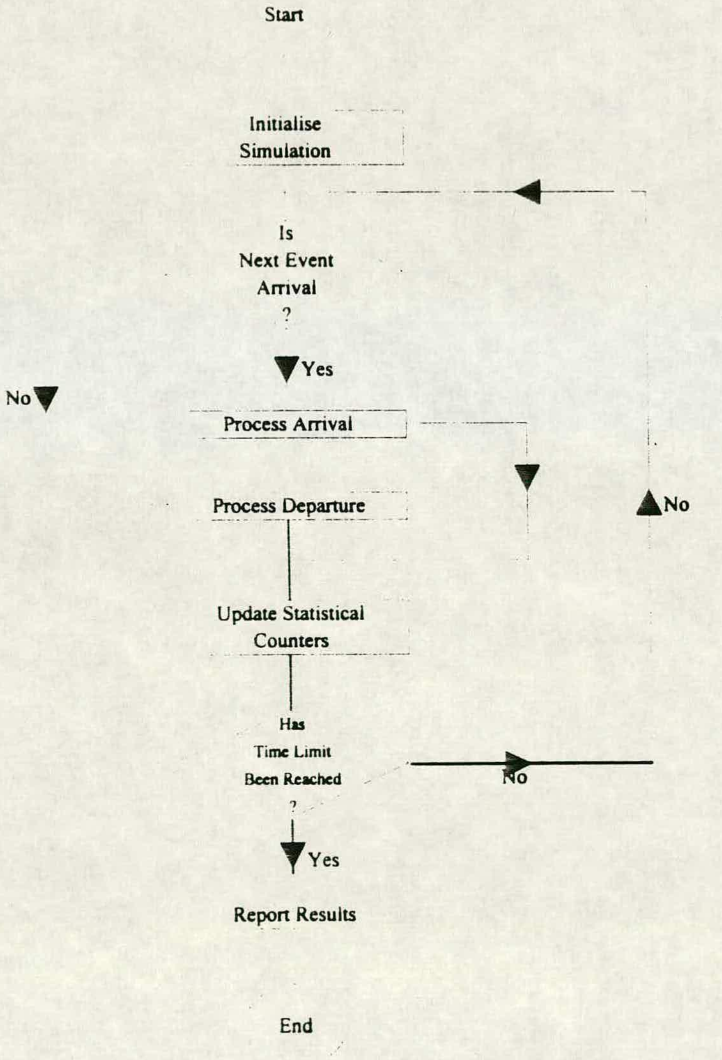
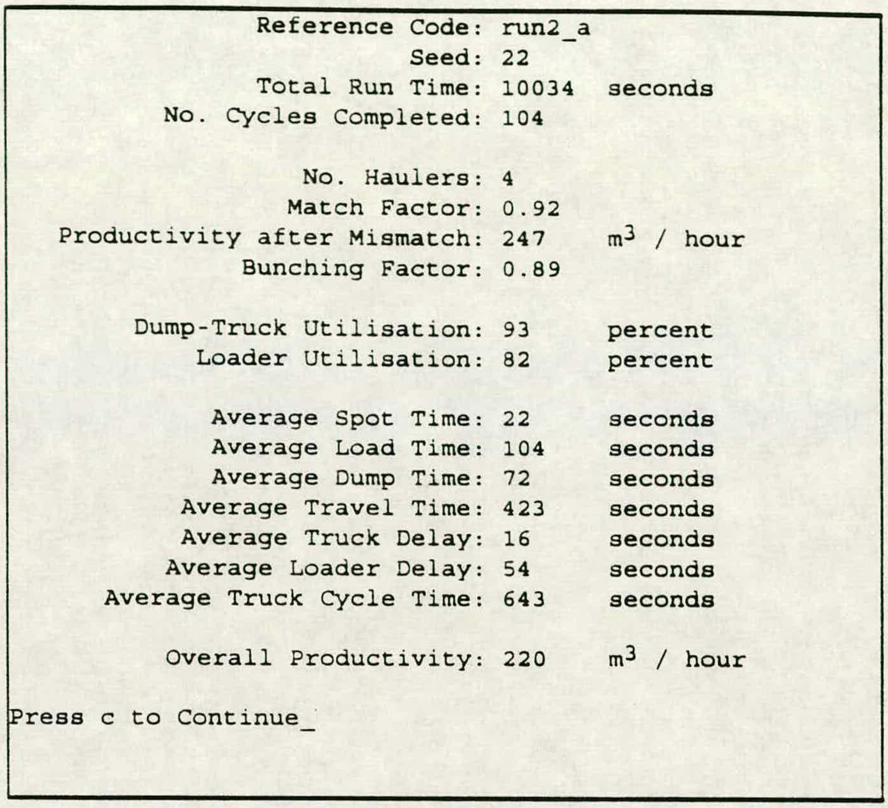


Figure 3 Main Simulation Flowchart for Trucksim.



```
               Reference Code: run2_a
                         Seed: 22
               Total Run Time: 10034  seconds
         No. Cycles Completed: 104

                  No. Haulers: 4
                 Match Factor: 0.92
  Productivity after Mismatch: 247    m3 / hour
              Bunching Factor: 0.89

       Dump-Truck Utilisation: 93     percent
           Loader Utilisation: 82     percent

            Average Spot Time: 22     seconds
            Average Load Time: 104    seconds
            Average Dump Time: 72     seconds
          Average Travel Time: 423    seconds
          Average Truck Delay: 16     seconds
         Average Loader Delay: 54     seconds
     Average Truck Cycle Time: 643    seconds

         Overall Productivity: 220    m3 / hour

Press c to Continue_
```

Figure 4 Sample Results from Trucksim.

tion factors and average component times are calculated by having various counters which are updated when certain events occurs. For example whenever an arrival happens, a cycle has been completed and the cycle counter is updated. Total production for the simulation run is therefore the product of the number of cycles and the volume per cycle. Dividing this by the run time gives a productivity rate in volume per unit time.

MATCH AND BUNCHING FACTORS

The results sheet in Figure 4 includes the terms match and bunching factor. These are used to see how efficient the operation is running and will give a guide to how an operation can be improved, both at the estimation stage and during actual operations.

Match factor is a number which indicates how well the loader and haulers are matched. It is dependent on the quantities of plant and their cycle times and was first introduced by Caterpillar inc.[11] and is shown in Eq. 4.

$$\text{Match Factor} = \frac{\text{No. Haulers} \times \text{Loader Cycle Time}}{\text{No. Loaders} \times \text{Hauler Cycle Time}} \tag{4}$$

The cycle times used in the equation must not include idle times.

A match factor (MF) of one indicates the operation is perfectly matched and, for constant cycle component times, there will be no waiting delays in any item of plant. It the match factor is greater than 1 then this indicates that the operation is over-resourced. The estimator or, for an actual operation, the engineer controlling the works, has a number of options. The simplest way to reduce the capacity is to reduce the number of hauler units. Alternatively, the loader cycle time can be reduced either by employing a larger excavator or using a larger bucket or by reducing the spot time of the trucks. An experienced ganger controlling the operations in the load area should be able to reduce the trucks spot time to a minimum. An MF lower than one is indicative of an under-resourced operation and the situation can be improved by either increasing the number of trucks or by maintaining the haul road; a low match factor may occur because the trucks are running slower than planned.

The bunching factor is dependent on the amount of interference between the trucks due to the difference in their running speeds. Bunching factor for an operation cannot be calculated analytically and is therefore one of the reasons a simulation method is employed. It can, however, be determined after a simulation run has been completed or after an actual operation has been observed and cycle times have been recorded by determining the actual productivity (from number of cycles completed per unit time) and dividing this by the possible production-i.e. that which would be achieved if no variation in cycle times occurred. This possible production dependent on the resourcing level of the operation and is shown as *Productivity after Mismatch* on the results sheet (Fig. 4). The level of bunching in an operation affects the productivity mostly for a perfectly matched operation. Such an operation is very sensitive to small changes in cycle time. The variance of the cycle time will also influence the bunching factor, for example if different types of hauler are used for one operation then their different relative speeds will affect the times the trucks return to the loader queue. Obstacles such as plant crossings (where plant may have to wait) and Bailey bridges (where only one item of plant is allowed on the bridge at a time) will also influence the variance of the cycle times. It these factors are anticipated at the estimation stage then the distributions input to the simulation should be altered accordingly.

VALIDATION OF MODEL

If a simulation model is not an accurate representation of the system being studied then any conclusions gained from the model cannot be relied upon. Validation of a model is therefore a major part of any modelling exercise. Most major texts on the subject of simulation include sections on validation, in particular Law and Kelton[12] and Banks and Carson[13] but it is the work done by Naylor and Finger[14,15] that has rationalised the work in recent years.

Naylor and Finger proposed that validation of any simulation model can be carried out by following a multistage process which covers all the aspects of the models ability to represent the real system with accuracy. This multistage process is itself based on the work of economists in the early part of this century who viewed validation on three levels:

- *Rationalism*, whereby the model is considered as simply being a series of basic assumptions. If all basic assumptions are accounted for then the model can be considered as being validated. A leading proponent of this approach was Robbins[16,p.180] who insisted that if *all* basic assumptions could be stated then they would be so simple that the modeller "[does] not need controlled experiments to establish their validity: they are so much the stuff of everyday experience that they only have to be stated to be recognised as obvious." Although this may not be true for modern mathematical simulation models, it is still an important area as the assumptions which underlie the model will undoubtedly affect its output.
- *Empiricism*. This can be considered the opposite of the first approach. Blaug[17,pp.612–613] stated that validation should begin with facts and not assumptions and that any assumptions made cannot be validated independently of actual results.
- *Positive Economics*. Proposed by Milton Friedman[18] this approach takes the view that only the end result of the model is important and that the validation of a model rests on its ability to accurately predict the variables of interest.

All three of the above approaches are important in the development of a valid simulation model. The following is the application of these approaches, via Naylor and Fingers multistage procedure, to the simulation model outlined in this paper.

i. *Stage 1*. Formulate the hypotheses and postulates for the system under study using the available knowledge such as observations, relevant theory and even intuition. This stage is based on the rationalist approach above. Naylor and Finger however, say that "having arrived at a set of basic postulates on which to build our simulation model, we are not willing to assume that these postulates are of such a nature as to require no further validation. Instead, we merely submit these postulates as tentative hypotheses about the behaviour of the system."[14,p.96]. Naylor and Finger leave the first stage there and go on to the second stage. No mention is given as to whether or not these postulates are 'reasonable'-this should be considered before any empirical tests are carried out.

Other authors[12,13] therefore proposed a widening of this first stage by considering the *face validity* of the model, that is, whether or not it appears, on the surface, to be reasonable to those who are knowledgeable about the system. This can be done by consultations with system experts, observations of the system, knowledge of existing theories and similar simulation models and, as suggested by Naylor and Finger, using one's experience and intuition. Of these, there was no existing experience of the system or similar simulation models to be studied but extensive observations of the existing earthmoving system were carried out. By far the most useful exercise in the assessment of the earthmoving models face validity, however, was consultation with an presentation to experts in the earthmoving system. These included Tarmac Construction Limited, one of the UK.'s largest road building contractors; Blackwell Earthmoving Limited, one of the UK.'s largest independent earthmoving sub-contractors; Caterpillar, inc., the worlds largest manufacturer of earthmoving equipment; Accelerator, inc., developers of earthmoving vehicle performance software[20] and also Artix Ltd., the British based sole manufacturer of Caterpillar articulated dumptrucks. Acceptance by these companies that the simulation model had high face validity went a long way towards the absolute validation of the whole model.

ii. *Stage 2*. This involves validating the assumptions upon which the model is based by subjecting them to empirical testing. This is of course the empirical aspect of validation suggested by Naylor and Finger ([14] and above) and should be done at the time the assumptions are postulated. It is no less important than the other stages in the multistage procedure–"a model with untested, untestable or refuted assumptions is at least disturbing."[19,p.251]. With the earthmoving model, this stage has already been undertaken for the most major assumptions–that the observations of the time components are independent and follow Erlang probability distributions. The cdf plot shown in Figure 2 has been shown to concur with the Erlang distribution via probability plots[12,pp.374–378]. These plots compare the observed and proposed distributions by plotting the values of time for each quantile of the cdf plots to form a *Q-Q* plot and also plotting the observed probability and the proposed probability (i.e. Erlang) for each value of time giving *P-P* plots. Both types of probability plot for 3 different types of cycle component time can be seen in Figure 5. It was found that the assumptions were not 100 percent applicable but acceptable within the time and cost constraints of the project.

iii. *Stage 3*. This is perhaps the most definitive test of the models validity. It is based upon Friedman's assertion that the validity of a model is demonstrated if it has the ability to predict the behaviour of the real systems variables irrespective of whether or not the assumptions the model is based on are reasonable[18]. This activity is a relatively simple task if compared with the validation of, for example, a military weapons system where the actual output of the real system is not known. The earthmoving model is not intended to replace the existing system but to represent it as closely as possible so that

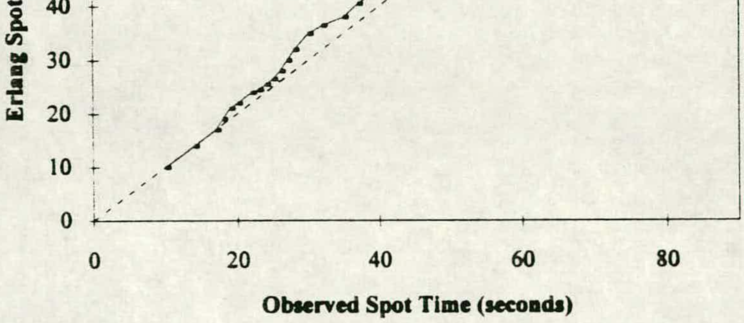


Figure 5a Q-Q plot for spot time.

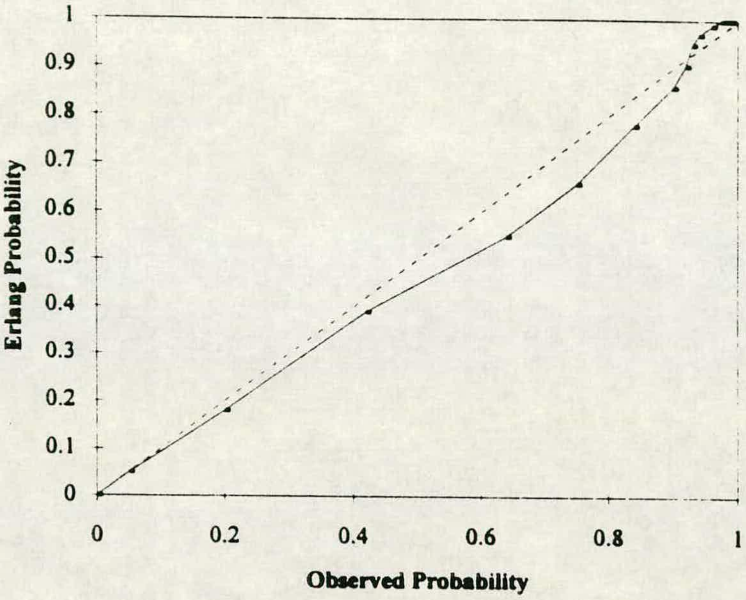


Figure 5b P-P plot for spot time.

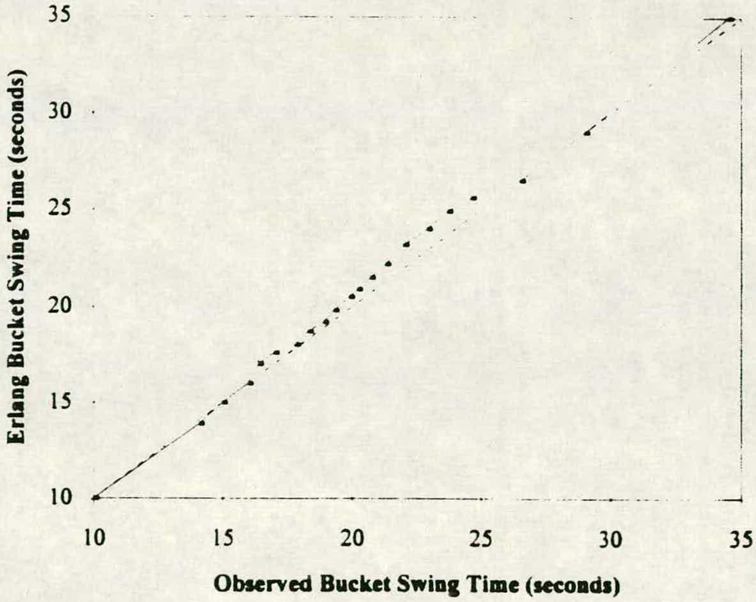


Figure 5c Q-Q plot for bucket swing time in clay.

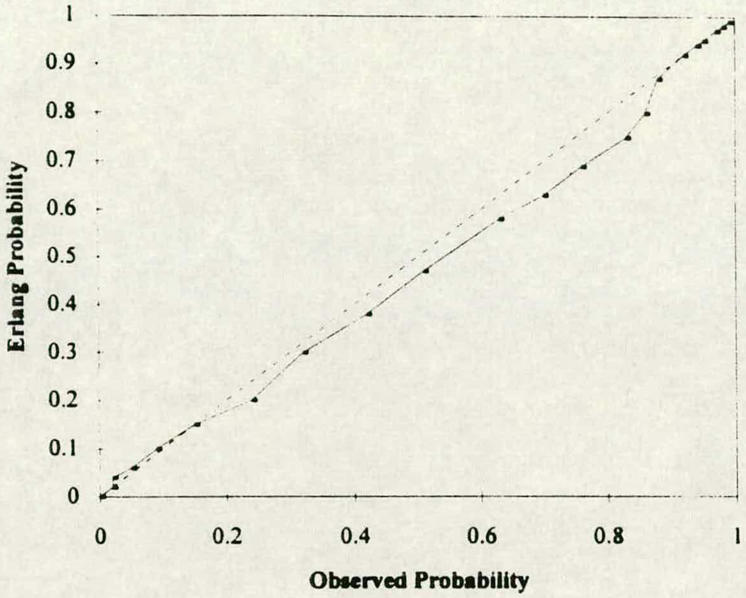


Figure 5d P-P plot for bucket swing time in clay.

estimations of its output can be made. Therefore, data from the existing system can be used to ascertain whether or not the model is replicativily valid i.e. able to reproduce results that have already been achieved.

Many authors have proposed that classical statistical tests can be used to determine the 'goodness-of-fit' between the model and real systems output[14,19]. Such tests are analysis of variance (ANOVA) tests, Chi-square test, factor analysis, and Kolmogorov-Smirnov test. However, most of these tests are not applicable because simulation output data (for a single time series) is usually highly autocorrelated (that is the observations in the output process are correlated with each other.) The hypothesis test that the two systems are the same would also be inapplicable simply because the model is only intended to be a representation of the actual system. Other approaches need to be used therefore to determine whether the differences between the two outputs are significantly different. The simplest procedure is the inspection approach.

INSPECTION APPROACH TO VALIDATION OF MODEL OUTPUT DATA

This is the simplest method of comparing the outputs of two or more systems which requires no statistical procedures. Performance parameters are calculated from both systems and compared. Because historical data from the actual system is used, this approach will determine the level of replicative validity of the model. Like the other approaches discussed above, this method has its drawbacks but unlike the statistical procedures, these drawbacks do not prevent the procedure actually being carried out. If the disadvantages are born in mind then the results should yield some indication of the replicative validity. The first drawback is that the data is compared with no statistical tests which means the validity is based only a subjective comparison. Secondly, each performance parameter from the real system is a sample of size one. These values are random numbers and so provide a bad estimate of the actual performance of the operation. This is not a problem with the data from the simulation model because the estimation can be improved with repeated replications.

Law and Kelton[11] suggest a *correlated inspection* approach (see Fig. 6) which in its extreme form requires that the model is driven by exactly the same observations as the real system. This pre-supposes that such data from a historical system is available but it is difficult to argue whether or not this indicates the predictive powers of a model. Therefore, the approach which shall be used here is to drive the model with random numbers generated using the means and variances from actual operations to produce the Erlang parameters.

To attempt to satisfy the third stage of the multistage validation procedure, 12 earthmoving operations were picked, at random, from two different sites which were studied. Each operation was analysed to determine the mean and variance of the spot time, the load time and the travel time so that the Erlang distribution parameters could be calculated. The dump time component was assumed to have a mean of 90 seconds with a variance of 370 (standard deviation = 19s). Five performance parameters were chosen for comparison and these are match factor, bunching factor, truck utilisation, loader utilisation and productivity. Trucksim was then used to carry out simulation

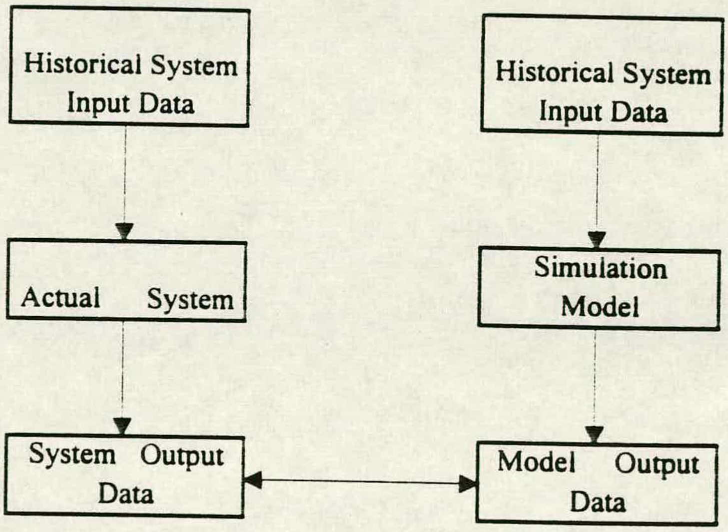


Figure 6 Correlated Inspection Approach to Model Validity Testing. (After Law and Kelton, 1991).

runs to provide estimates of the means of these 5 parameters using 5 replications for each operation to give a better estimate. It could be argued that since the actual operation can only give an estimate sample of size one, then like is not being compared with like. However, if more than one replication is done then any differences due to the randomness of both outputs can be eliminated to a certain extent from the simulation output. Table 1 shows the results of this validation study and it can be seen immediately that there is a strong similarity between the simulated and observed performance

Table 1 Comparison of 12 observed and simulated operations

Type	Match Factor	Bunching Factor	Truck Util.	Loader Util.	Prod.
Sim.	1	0.91	92	90	142
Obs.	1.05	1	96	86	142
diff.	−4.76%	−9.00%	−4.17%	4.65%	0.00%
Sim.	0.84	0.95	97	80	108
Obs.	0.84	0.94	95	76	121
diff.	0.00%	1.06%	2.11%	5.26%	−10.74%
Sim.	0.35	0.94	97	33	59
Obs.	0.31	1	100	33	66
diff.	12.90%	−6.00%	−3.00%	0.00%	−10.61%
Sim.	0.55	0.96	98	53	121
Obs.	0.49	1	100	44	115
diff.	12.24%	−4.00%	−2.00%	20.45%	5.22%
Sim.	1.38	1	75	100	117
Obs.	1.11	0.86	81	92	110
diff.	24.32%	16.28%	−7.41%	8.70%	6.36%
Sim.	1.13	0.98	88	97	115
Obs.	1.13	1	94	62	112
diff.	0.00%	−2.00%	−6.38%	56.45%	2.68%
Sim.	0.85	0.92	93	78	354
Obs.	0.86	0.92	92	77	343
diff.	−1.16%	0.00%	1.09%	1.30%	3.21%
Sim.	1.27	1	81	99	274
Obs.	1.19	0.99	91	97	288
diff.	6.72%	1.01%	−10.99%	2.06%	−4.86%
Sim.	0.6	0.92	94	55	202
Obs.	0.58	0.98	99	52	223
diff.	3.45%	−6.12%	−5.05%	5.77%	−9.42%
Sim.	0.96	0.9	91	86	386
Obs.	0.89	0.92	94	74	381
diff.	7.87%	−2.17%	−3.19%	16.22%	1.31%
Sim.	1.06	0.91	88	91	289
Obs.	1.04	0.93	95	93	299
diff.	1.92%	−2.15%	−7.37%	−2.15%	−3.34%
Sim.	0.78	0.9	92	70	272
Obs.	0.76	0.92	93	70	278
diff.	2.63%	−2.17%	−1.08%	0.00%	−2.16%

parameters. The percentage difference between the simulated and observed results has been plotted for clarity and can be seen in Figure 7. Again, like the probability plots shown in Figure 5, perfect validation has not been achieved but then such a perfect model was neither intended nor possible under the time scale of the project. It is probably true to say that no simulation model could be completely validated as a model is, by definition, only a representation of a real system.

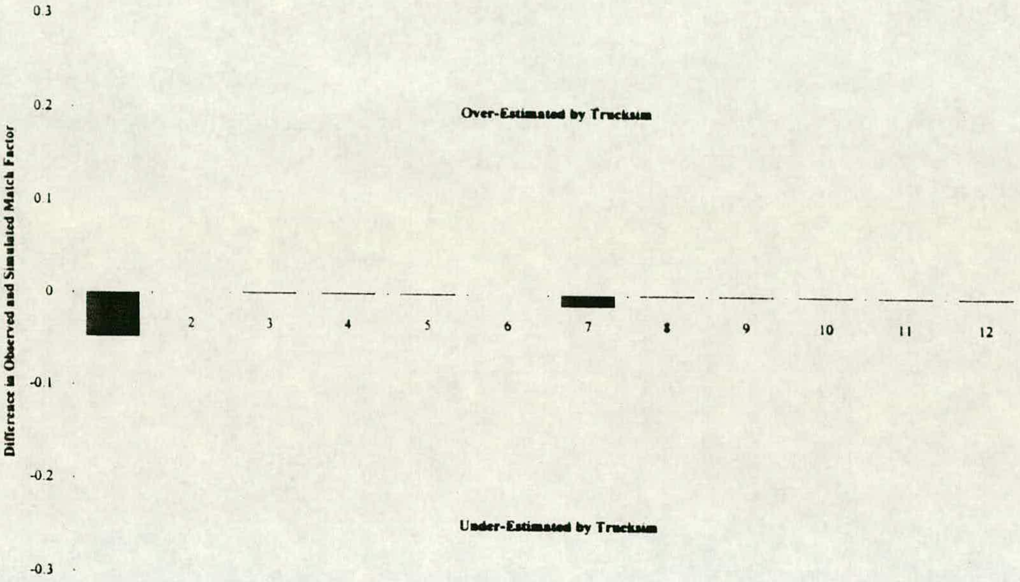


Figure 7a Comparison of simulated and observed match factor rsults for correlated inspection.

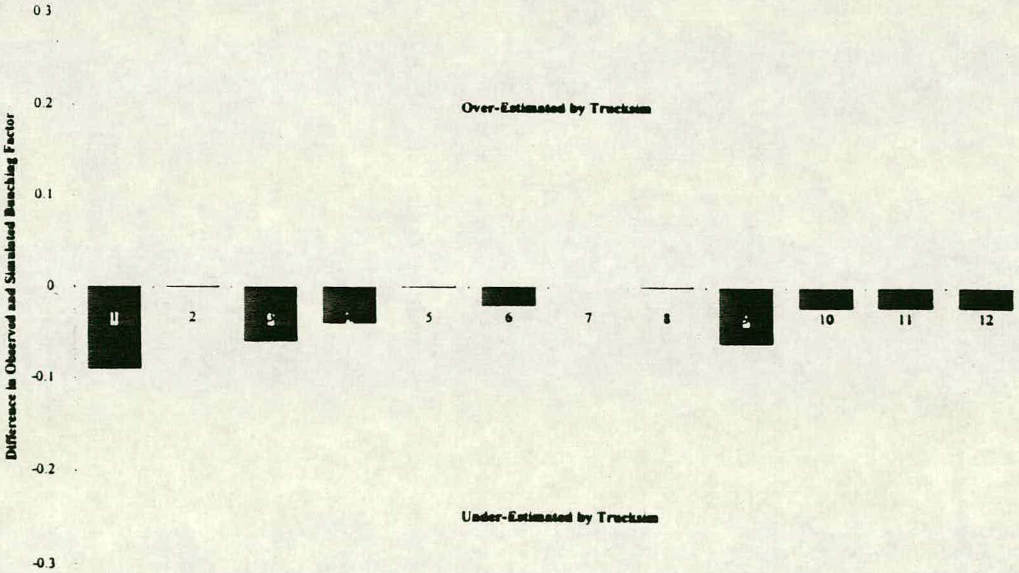


Figure 7b Comparison of simulated and observed bunching factor results correlated inspection.

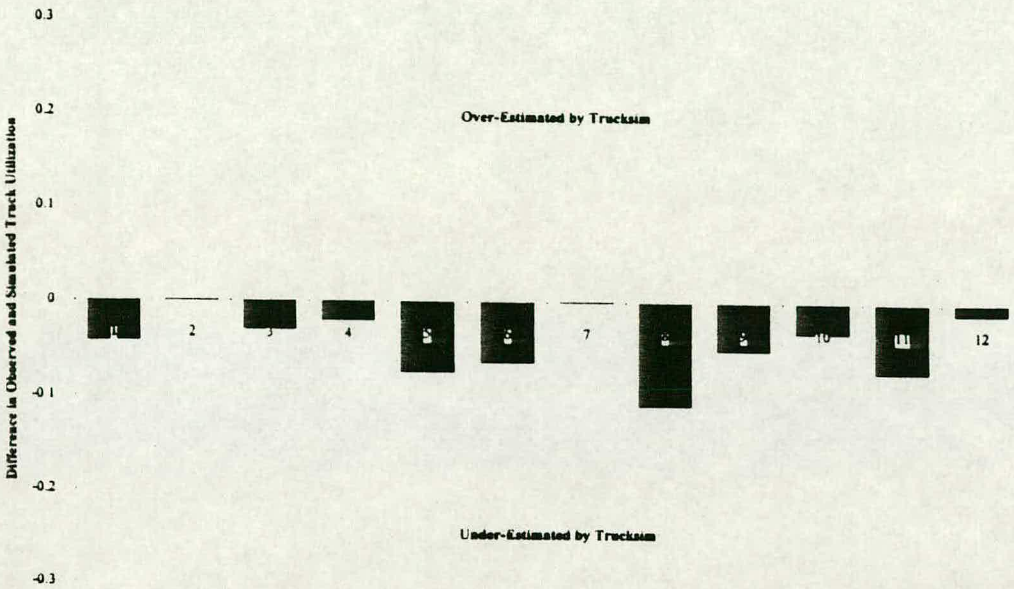


Figure 7c Comparison of simulated and observed truck utilisation rsults for correlated inspection.

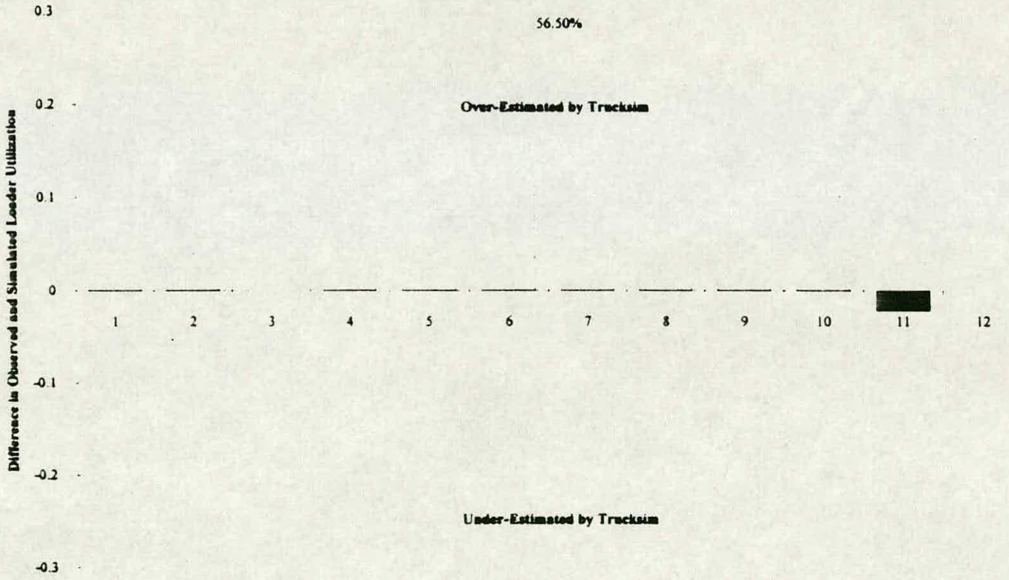


Figure 7d Comparison of simulated and observed loader utilisation rsults for correlated inspection.

The last step in validating the simulation model, which has yet to be achieved, is to obtain *credibility* of the simulation model. This is a subtle area but is perhaps the most important task of the simulation developer. It is not until a simulation model has been accepted as an accurate estimation tool by those engineers and managers who will use it that the model becomes successful. This processs is helped by the validation of the model but managers will need to see the programs results actually implemented on

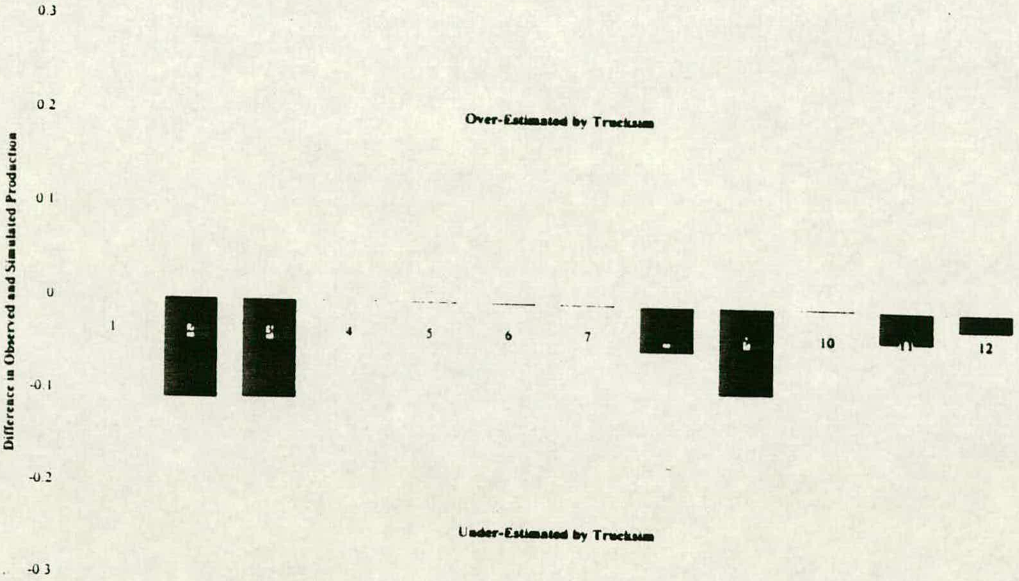


Figure 7e Comparison of simulated and observed production rate rsults for correlated inspection.

a live construction site before it is totally accepted. The importanace of credibility is a main reason that animation of computer simulation is becoming widespread. In the manufacturing industry, a terminal with an animated version of a production line is a common site yet this is very rarely seen on construction sites. Future work in this area must concentrate on achieving acceptability by site engineers and managers – and indeed simulation techniques such as those presented here need not be reserved for earthmoving operations. Any cyclic construction activity, such as concrete pouring and formwork erection, could benefit.

USE OF SIMULATION MODEL

The simulation model and program Trucksim were developed to be used in two separate situations: at the estimation stage and at the site control stage. An earthmoving contractor who is tendering for a job needs to have a clearly defined approach to estimating the output if he is to get the most from the simulation model. Before any simulation is attempted, the site conditions must be understood and the rates at which each individual piece of plant can work must be calculated.

For moving hauler units, this can be a complex procedure which involves determining the rolling resistance of the haul roads (from a study of the site investigation reports) and the grade resistance of each haul road section (from the contract drawings). Plant crossings, Bailey bridges and other obstacles should be located and the travel times for each earthmoving operation can then be determined. This can be done by using vehicle performance software such as Accelerator or Vehsim[20,21]. The output is also highly dependent on the rate at which the excavators work. Bucket pass times can be estimated from a knowledge of the loading area conditions and factors such as cut

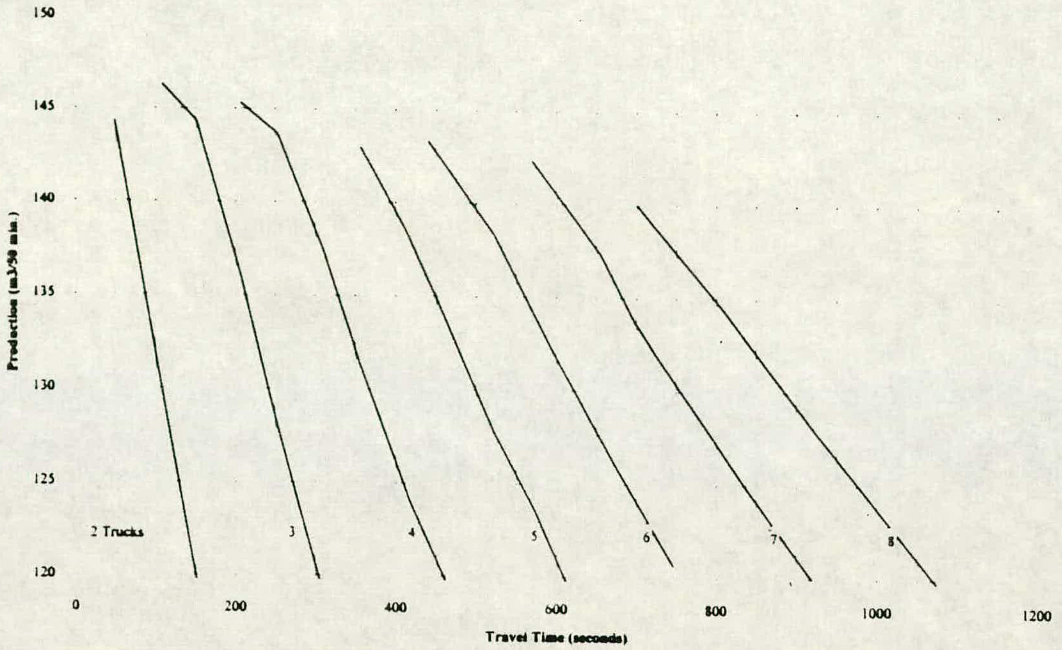


Figure 8 Sample production chart for travel time. (Values of production have been coded).

depth, swing angle and type of material. Cut depth and type of material will also affect the amount of earth which can be loaded into the bucket on each pass.

When all the input probability distributions and parameters for an operation have been determined, the simulation program can be used to estimate the number of haulers required and the productivity. This can be done directly or production curves can be drawn up for certain site conditions which can then be used to estimate a number of different operations. These curves can then be used for control and production target purposes when work commences. Such a chart is shown in Figure 8 (although note that values have been coded for reasons of confidentiality). Each point of the chart is determined from 10 separate simulations for a set number of trucks and travel time. Spot time, bucket pass time and dump time are constant for all curves.

CONCLUSIONS

This paper has shown that dump-truck and excavator type of earthmoving can be represented as a queuing system and as such can be analysed using discrete-event simulation. The main conclusions of the work are that:

- An earthmoving system cycle can be broken down into its main component parts (*i.e.* spot, load, travel and dump) which can then be built up along with other factors such as bucket volume to give a mathematical model which can be analysed.
- The probability distributions of the component times can be represented mathematically with an Erlang distribution. Such a distribution can be 'sampled' using an acceptance-rejection type algorithm in a microcomputer.

- A simulation program is easily coded onto a microcomputer to give a quick way to estimate earthmoving production and to try out different alternatives before site work commences.
- Consultation with earthmoving 'experts' has provided the model with a high degree of face validity.
- Comparison of system and model output has helped validate the simulation model increasing the reliability of the results.
- The simulation will give accurate results only when those distributions entered into the program represent those that occur in the field, i.e. the model is only replicativily and not predictivily valid.

NOTATION

T	Continuous random variable of time
$f(t)$	Probability density function (pdf)
$F(t)$	Cumulative distribution function (cdf)
R	Erlang distribution rate factor
k	Erlang distribution shape factor
$E(T)$	Estimate of the random variable of time
var T	The variance of the time component

Acknowledgements

The authors would like to acknowledge the help provided by the following people in the preparation of this paper:

The staff of the Department of Civil Engineering and Building Science at the University of Edinburgh, including Mr. Graeme Wood and Mr. Stephen Dunn for assistance in the collection of data. The stall of Tarmac Construction Limited, of Wolverhampton, including Mr. George Daville for his constructive advice. This work was supported under a Teaching Company Scheme funded by SERC, grant no. GR/H23306 and by Tarmac Construction Limited.

References

1. Staples, B. L., Wood, G. S. and Forde, M. C. (1992). Technical evaluation of earthworks claims under ICE Conditions of Contract 5th and 6th editions. *Proc. Instn. Civ. Engrs. Civ. Engng.*, 92(2), 90–95.
2. Paulson, B. C. *et al.* (1983). Simulation and Analysis of Construction Operations. *Journal of Technical Topics in Civil Engineering*, ASCE, 109(2), 89–104.
3. Oloufa, A. A. (1993). Modeling Operational Activities in Object-Oriented Simulation. *Journal of Computing in Civil Engineering*, ASCE, 7(1), 94–106.
4. Simaan AbouRizk and Jingsheng Shi. (1994). Automated Construction – Simulation Optimization. *Journal of Construction Engineering and Management*, ASCE, 120(2), 374–385.
5. Clemmens, J. P. and Willenwbrock, J. H. (1978). The SCRAPESIM Computer Simulation. *The Journal of the Construction Division*, ASCE, CO4, 419–435.
6. Carmichael, D. G. (1987). *Engineering Queues in Construction and Mining*. Ellis Horwood. Chichester.
7. Carmichael, D. G. (1986). Erlang loading models in earthmoving. *Civil Engineering Systems*. 3, 118–124.
8. Schmidt, J. W. and Taylor, R. E. (1970). *Simulation and Analysis of Industrial Systems*. Richard D. Irwin, Homewood, III.

9. Winston, W. L. (1987). *Operations Research: Applications and Algorithms*. Duxbury Press, Boston, Mass.
10. Law, A. M. and Haider, S. W. (1989). Selecting Simulation Software for Manufacturing Applications. *Industrial Engineering*, 31, 33–34.
11. Morgan, W. C. and Peterson, L. L. (1968). Determining Shovel-Truck Productivity. *Mining Engineering*. 76–80.
12. Law, A. M. and Kelton, W. D. (1991). *Simulation Modelling and Analysis. 2nd Ed.* McGraw-Hill, New York.
13. Banks, J. and Carson, J. S. (1984). *Discrete-Event System Simulation*. Prentice-Hall, Englewood Cliffs, New Jersey.
14. Naylor, T. H. and Finger, J. M. (1967). Verification of Computer Simulation Models. *Management Science*. 14(2), 92–101.
15. Naylor, T. H. (1971). *Computer Simulation Experiments with Models of Economic Systems*. John Wiley and Sons. New York.
16. Robbins, L. (1935). *An Essay on the Nature and Significance of Economic Science*. Macmillan, London.
17. Blaug, M. (1962). *Economic Theory in Retrospect*. Richard D. Irwin, Homewood, Ill.
18. Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press. Chicago.
19. Van Horn, R. L. (1971). Validation of Simulation Results. *Management Science*, 17(5), 247–258.
20. Accelerator. (1987). *Accelerator Handbook*. Accelerator, inc. Fort Myers, Fla.
21. Caterpillar. (1987). *Vehsim Handbook*. Caterpillar, inc. Peoria, Ill.

APPENDIX 1. ALGORITHM FOR ACCEPTANCE-REJECTION RANDOM VARIATE GENERATOR

This algorithm can be used for those probability distributions which are defined over a finite range. Although this is not technically the case with the Erlang distribution, finite limits must be set for the range over which the actual distribution is expected to fall, which is defined with the limits a and b, the upper and lower bounds respectively. Eq. A1 defined the probability function for the Erlang distribution which is repeated here

$$f(t) = \frac{R(Rt)^{k-1} e^{-Rt}}{(k-1)!} \quad (t \geq 0) \tag{A1}$$

This is now used in the following method to generate variates:

Step 1. A constant M is calculated such that M is the largest value of $f(t)$ over the interval $[a, b]$.

Step 2. Two $U(0, 1)$ numbers are generated, r_1 and r_2.

Step 3. Calculate $t^* = a + (b - a) r_1$.

Step 4. Evaluate $f(t^*)$.

Step 5. If r_2 is less than or equal to $f(t^*)/M$ then accept t^* as a random variate. Otherwise, reject and go back to step 2.

Experimental Analysis of an Earthmoving System Using Discrete-Event Simulation.

S.D.Smith, BEng[1]
J.R.Osborne, BSc, CEng, MICE[2]
M.C.Forde, BEng, MSc, PhD, CEng, MICE, MIHT, FINDT[3]

Abstract

This paper outlines experiments carried out on a simulation model of excavator-dumptruck type earthmoving operations with the intention of indicating which parts of an earthmoving system the output is most sensitive to. Response surface methodology was used initially to indicate the relationship between two factors: the truck travel time, from loader to dump area and back, and the truck spot (or manoeuvre) time at the loader. This method indicated that the form of the relationship between these factors varied at different values but did not indicate how other factors affected the output. Full factorial designs were then carried out which were not only more economical in terms of the number of experimental runs required, but indicated that the most important factors are number of trucks, the haul and return (travel) time, the number of passes per load and the loading rate. The experiments also indicated the interactions between these factors.

Keywords: Earthmoving, Discrete-Event Simulation, Response Surface Methodology, Factorial Design, Match Factor.

---

[1] Teaching Company Associate with The University of Edinburgh and Tarmac Construction Limited, Wolverhampton, England, UK, WV1 4RU. Tel. +44 1902 22431.
[2] Operations Director, Tarmac Structural Repairs Limited, Wolverhampton, England, UK, WV1 4RU.
[3] Tarmac Professor of Civil Engineering Construction, The University of Edinburgh, Edinburgh, Scotland, UK, EH9 3JL

Introduction

Motorscrapers are normally the first choice for short haul earthmoving operations. However, increases in UK fuel costs, together with a slowing down of the construction of the UK motorway network, have resulted in a regional decline of the motorscraper as a cost effective method of earthmoving. UK contractors have therefore turned to the backhoe excavator and a fleet of rigid or articulated dumptrucks as a practical and cost-effective alternative. Similar scenarios can be seen in continental Europe where high fuel costs also exist, as opposed to North America and the Middle East where fuel costs are relatively low. It is against this background of a high fuel cost scenario that this work has been undertaken.

In today's increasingly competitive market, the contractor needs to be able to plan and estimate an earthmoving contract as accurately as possible at tender stage and then, if the contract is won, control site operations to keep costs to an absolute minimum. However, production estimates at a tender stage are difficult to calculate accurately and at a construction stage, production targets are hard to maintain. The reason for this is that an earthmoving system is extremely complex. If we consider the earthmoving system defined mathematically:

$$Y = f(k_1, k_2, k_3, \ldots k_n) \tag{1}$$

where Y is a *response* of the system to certain values of $k_1, k_2, k_3, \ldots k_n$ which are called *factors.* The function f is called the *response surface* . If it were to be investigated how a response, say earthmoving production rate, in bank cubic metres per hour, varied for certain factor values, for example truck volume, truck speed or load time, one method would be to physically experiment with the system in the field: change the factor values and note the change in the response. This is practically very difficult for a number of reasons - most importantly, a large number of *experimental runs* would be required to give a good indication of the way in which the response changes. If each run has to consist of at least 5 hours of physical operation time then for more than a few runs this approach would have to be abandoned due to cost and time constraints.

What is required, therefore, is a mathematical model on which a type of quasi-experiment can be performed. The system is too complex to be analysed deterministicaly. Therefore, a simulation model and computer program have been developed on which experiments can be designed and performed. This paper outlines experiments carried out using response surface methodology (Myers, Khuri and Carter 1989) and factorial analysis (Box, Hunter and Hunter 1978).

The Earthmoving System: Factors and Responses

Figure 1 shows a schematic of the earthmoving system, which is a variation of the classic queueing system, for example a bank or a car-wash. Trucks are used in this system to transport material from the cut area to the fill area and these are, in queueing theory terminology, the customers. The servers are the excavator loaders, which fill the trucks one at a time by drawing them from the queue. The service time is therefore built up from individual component times that are the manoeuvre or *spot* time, which is the time the truck takes to get from the queue to its position by the loader, and the *load* time, which is made up of individual *load pass times.* The trucks, once loaded, haul to the dump area, dump the load and return to the queue. The length of time of this 'back-cycle' (or time out of service) is heavily dependent on the speed of the trucks and the length of the haul and so, to allow for both of these factors, the haul and return times have been collectively classed as *travel* time. The final time component is the *dump* time, which tends to have the same distribution for all types of dumptruck. There is more than one response to this system other than production: for example plant utilisation, queue wait time and length and match factor. (For a definition of match factor see appendix 3.) Affecting these responses, the system has many factors on various levels. For example, truck travel time is a factor but this itself can be considered as a response to a subsystem with factors such as engine specification, haul road soil properties, haul road length and gradient. Table 1 outlines the subsystems, some of the first and second order factors and the responses that could be considered in an investigation of the earthmoving system.

As can be seen, the earthmoving system can be split into four separate subsystems, each with its own factors leading to two responses: a mean and a variability for each time component. These become 8 factors for the main system along with 3 additional factors leading to 6 responses. It is these final responses that the estimator and engineer are interested in to minimise time, cost and resources to complete the job. To consider all the factors, 1st and 2nd order, in one simulation model would be an extremely complex procedure and indeed, some factors are difficult to quantify (driver ability for example). The simulation model that has been developed, therefore, only considers the second order factors. Time component means and variabilities are determined prior to simulation runs using a database of observed times from various sites in the UK. One exception is the travel time. This can be calculated manually on a separate simulation package such as *Accelerator* (Accelerator 1987) or *Vehsim* (Caterpillar 1987).

Simulation Model and Program

The simulation model developed is based on a discrete - event simulation method used extensively in the manufacturing and production engineering industries (Law and Kelton 1991; Law 1986; Law and McComas 1989). Discrete - event simulation is used where the state of the system changes at discrete, measurable points in time; in the earthmoving system, these events are when a truck arrives at the queue and when it departs on completion of loading. The occurrence of these events can be recreated on a computer by generating random times from the expected probability density functions for a particular operation. The simulation model was transfered to a computer by way of a program written in 'C' to carry out simulation runs. The inputs to the program are the 11 second order factors shown in table 1. To keep the initial model as simple as possible, the number of trucks, the bucket volume and the number of bucket passes per truck load are entered as individual integers, although these factors will have their own distributions for a real life operation. The model can, if necessary, be altered if it is felt that more accuracy is required. The mean and variability for the four time components are entered as parameters for the Erlang distribution (Carmichael 1986; 1987a; 1987b) which is a form of the gamma distribution with integer shape factors.

The simulation program is run until the internal simualtion clock has reached the equivalent of a pre-specified real time limit, usually a single earthmoving

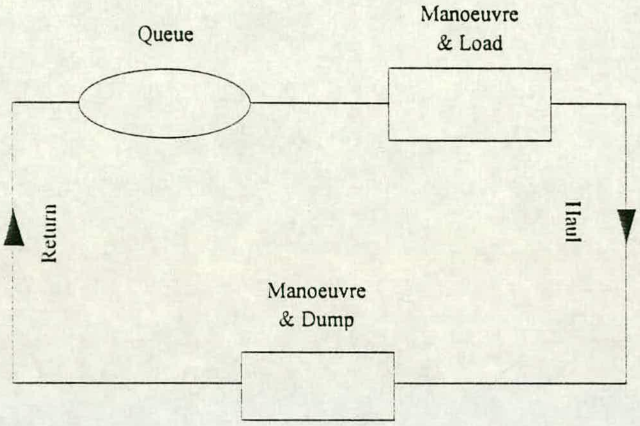


Figure 1 The Earthmoving System

1st Order Factors	2nd Order Factors	Responses
Travel System		1. Production (m3 per hour)
1. Haul Road Soil Parameters		2. Truck Utilisation (%)
2. Haul Road Gradient		3. Excavator Utilisation (%)
3. Haul Road Length		4. Queue Wait Time (seconds)
4. Truck Weight	1. Travel Time (mean)	5. Match Factor
5. Vehicle Specification:	2. Travel Time (variability)	6. Bunch Factor
Engine size		7. Cost per cubic metre (to excavate, load and haul)
Transmission Type		
Drive System		
Suspension System		
etc.		
6. Number & Type of Obstructions:		
Type of Plant Crossings		
Size of Bailey Bridges		
7. Driver Ability		
Spot System		
8. Load Area Size	3. Spot Time (mean)	
9. Size & Type of Trucks	4. Spot Time (variability)	
10. Driver Ability		
Load System		
11. Cut Depth		
12. Soil Parameters	5. Bucket Pass Time (mean)	
13. Type of Excavator	6. Bucket Pass Time (variability)	
14. Operator Ability		
Dump System		
15. Dump Area Size	7. Dump Time (mean)	
16. Driver Ability	8. Dump Time (variability)	
Other Factors	9. Number of Trucks	
	10.Bucket Volume	
	11. Bucket Passes per Load	

Table 1. Factors and Responses for the earthmoving System.

shift between breaks, i.e. 3 hours. At this point the simualtion is stopped and the report generator invoked (Figure 2). The report generator gives 5 of the 6 responses shown in table 1; the 6th, the cost per cubic metre can be calculated from a knowledge of the plant costs per hour. Replications for the same run can be done automatically: the seed for the random number generator is changed for each replication and the results can be stored in a text file that can be viewed in any spreadsheet. It is with this program that experiments have been carried out. For the experiments detailed in this paper, approximately 2100 individual replications have been run.

Response Surface Designs

Of the eleven factors of the simulation model of the earthmoving system, suppose 10 of the factors were kept at a constant level and the eleventh was varied over a particular range of interest. If the response is plotted against the varying factor, the resulting curve is a *response curve*. For example, table 2 shows the values of 10 factors and figure 3 shows the production response curve obtained from simulation runs as the travel time is varied from 100 to 1200 seconds. Such a change in travel time may be due to a change in haul length or a change in speed due to differing ground conditions. This curve is limited in that it shows how the response varies for a very small portion of the overall function for the response. There is no doubt that it is valid for the configuration in question, but gives no indication as to how the response will behave for all other values for all the other factors.

If two factors were varied, with the other 9 factors kept constant, a response surface (i.e. a 3 dimensional representation of the factors in question for the particular range of interest) is obtained. For three or more factors, a response surface cannot be graphically represented and the experimenter will have to be satisfied with sectional representations of the entire surface.

Figure 4 shows a response surface obtained when spot times and travel times are varied. Changes in spot time could be due to changes in load area size and shape and changes in travel time will occur due to a lengthening of the haul road and/or changes in ground conditions. Mean spot time was varied from 10 to 22 seconds, mean travel time varied from 100 to 1100 seconds and all other factors were kept at the same level as shown in table 2. The curve in figure 3 is therefore one single section through the surface shown in figure 4. From a brief inspection of figure 4 the simple conclusion that spot time has little effect on the production response could be drawn: At travel time = 100 seconds the response changes only 10% whilst at travel time = 1100 seconds the change in response is virtually nothing (0.7%). However, in this example, the travel times are set at much higher levels than the spot times and the increase in travel time is 1000% compared to only a 100% increase in spot time. This indicates that the response surface is little better than a response curve for this particular system: conclusions about how the levels of the factors *outside* the particular range studied will affect the response cannot be drawn.

In figure 5, travel time and spot time are varied again but travel time is varied from 50 seconds to 200 seconds. It can be seen how spot time has a greater influence on the production response than in figure 4. However, what is unknown is how the levels at which the other factors have been set have affected the response and whether any *interactions* are present. For example, there may be an interaction between load pass time and travel time that influences the response. It is impossible to determine from these response surface designs whether such interactions exist and, if they do, whether they caused the difference in productions between figures 4 and 5.

Factorial Design

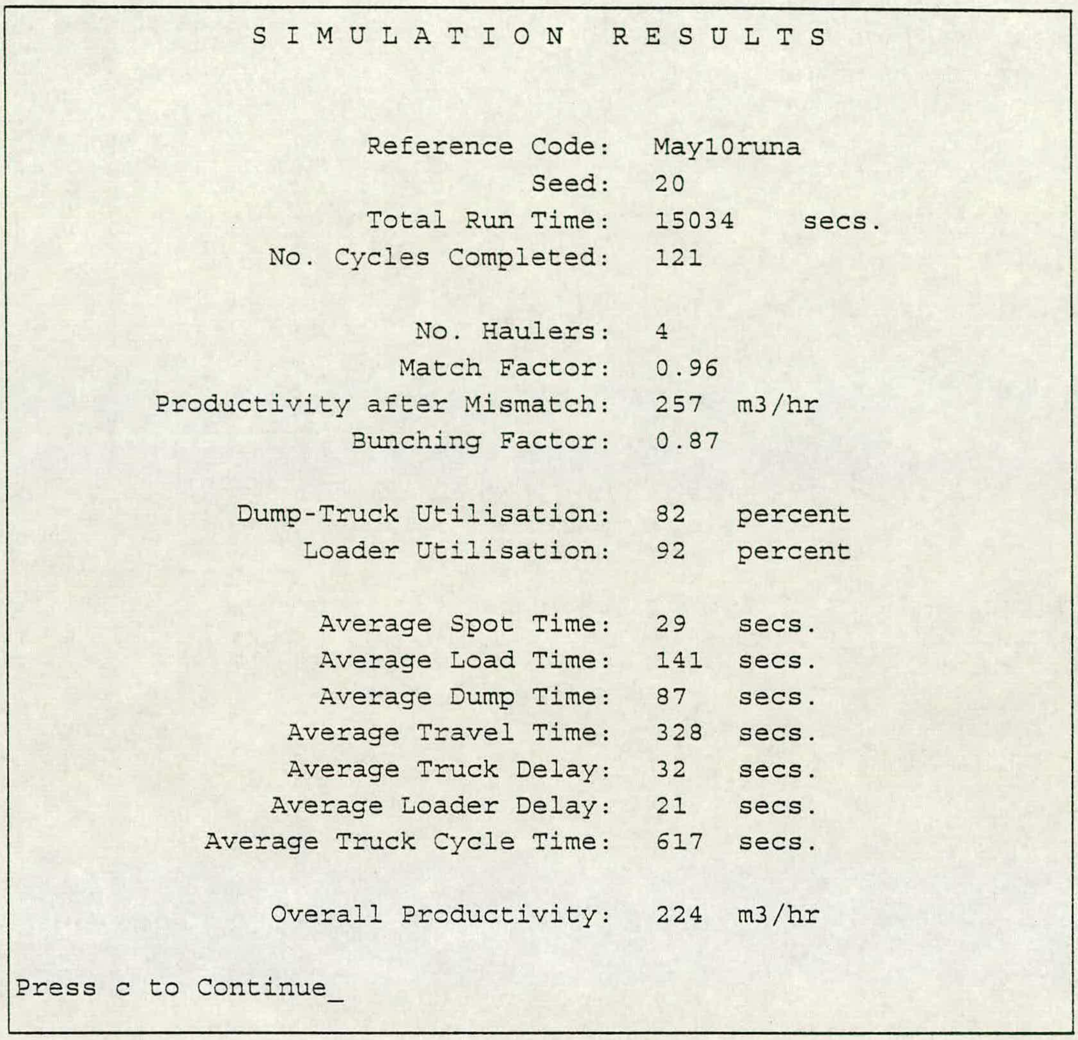
So far in this paper, it has been shown that the earthmoving system has at least 11 factors, all of which will have some effect on at least 6 responses. First attempts at response surface design have shown perhaps the most fundamental problem facing all experimentalists: which factors are the most important and over which ranges shall they be studied. Box, Hunter and Hunter (1978, p.303) indicated a paradoxical situation: the best time to design an experiment is when it has been completed. Experimentation is therefore iterative; after each successive experiment, more will be known about the system and future experiments will be more useful. Factorial designs are useful in this respect in that they can give a 'broad picture' of the overall system and will help eliminate factors that have little effect on responses.

In performing a factorial design, a set number of values or *levels* for each factor are chosen and experimental runs for all combinations of levels are carried out. For example, if a system with only 3 factors was to be experimented with and 3 levels were chosen for each factor, then $3^3 = 27$ experimental runs (or *design points*) would be required. Unfortunately, if 3 levels for each of the 11 factors in this simulation model were chosen then $3^{11} = 177,147$ runs would have to be performed; this is clearly impractical. The simplest is 2 levels for each factor but even this requires 2048 runs, which is also too many for an initial investigation.

A different approach was therefore taken with the earthmoving system. On the basis of the work of Carmichael (1987a; 1987b), the effective number of factors was reduced by grouping four of the factors by making the assumption that pairs of factors (i.e. load and spot; dump and travel) have the same effect on the output. Referring to figure 1, the spot and load cycle times are collectively known as the 'service time'; the travel and dump times are collectively the 'back-cycle' times. In effect, the four time factors are reduced to two: service and back-cycle and these have been represented by load pass time and travel time.

A simplifying assumption has also been made by eliminating the bucket volume from any designs. In a real life situation, the bucket volume will effect the production in two ways: the load cycle duration (larger buckets will lead to fewer, slower passes) and volume per pass (larger volume per cycle from a larger bucket with the same number of passes). However, the model does not include any relationship between bucket size and load pass time: the model output will therefore be linearly proportional to the bucket size. Although this may not represent a real life situation completely accurately, initial models must be kept simple and therefore bucket volume shall be kept at a single level in these experimental situations.

At this stage, the effective number of factors is now 6 and a 2^6 factorial design requires 64 design points. This is a far more practical figure for an initial experiment.



```
              S I M U L A T I O N   R E S U L T S


                     Reference Code:   May10runa
                               Seed:   20
                     Total Run Time:   15034    secs.
              No. Cycles Completed:   121

                        No. Haulers:   4
                       Match Factor:   0.96
        Productivity after Mismatch:   257  m3/hr
                    Bunching Factor:   0.87

             Dump-Truck Utilisation:   82   percent
                 Loader Utilisation:   92   percent

                  Average Spot Time:   29   secs.
                  Average Load Time:   141  secs.
                  Average Dump Time:   87   secs.
                Average Travel Time:   328  secs.
                Average Truck Delay:   32   secs.
               Average Loader Delay:   21   secs.
           Average Truck Cycle Time:   617  secs.

               Overall Productivity:   224  m3/hr

Press c to Continue_
```

Figure 2. Typical Results Report from Simulation Program

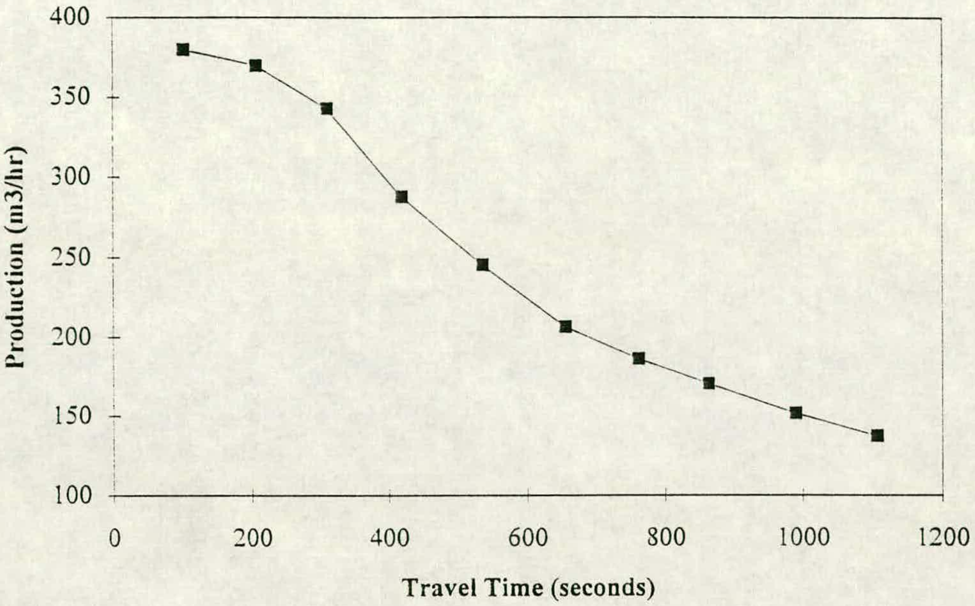


Figure 3. Response Curve for Factor Levels in Table 2

Factor	Level
Travel Time Variability	k = 40
Spot Time Mean	10 seconds
Spot Time Variability	k = 10
Bucket Swing Time Mean	16 seconds
BST variability	k = 10
Dump Time Mean	90 seconds
Dump Time Variability	k = 20
Passes per load	6
Bucket Volume	2 bank cubic metres
Number of Trucks	5

Table 2. Factor levels for response curve

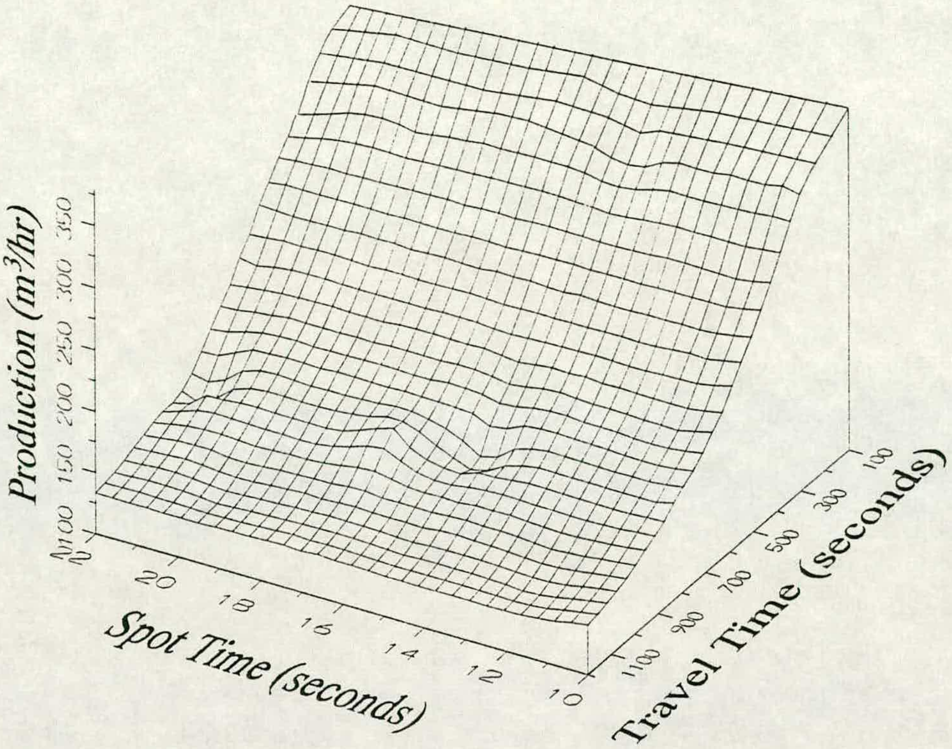


Figure 4. Response Surface for Spot Time and Travel Time

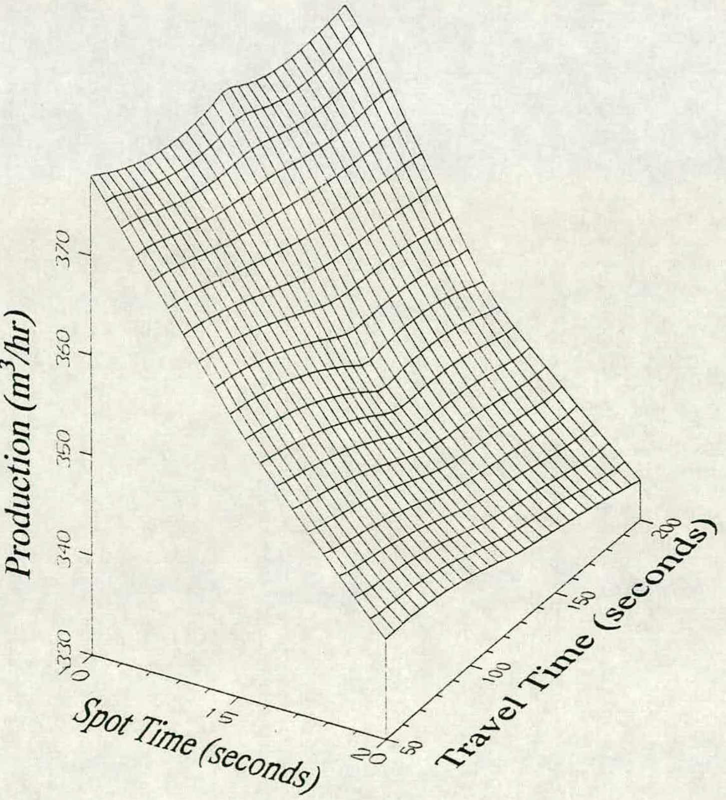


Figure 5. Response Surface for Spot Time and Travel Time (Reduced Travel Time)

2^6 Factorial Design: Initial Study

Table 3 shows the 2^6 experiment for the 6 remaining factors. The two levels for each factor are coded by a plus sign and a minus sign and in this case the minus sign is associated with the lower numerical value. Runs were carried out at each of the 64 design points, with 11 replications for each point. Production and match factor responses were recorded for each replication and the average of the 11 replications for each design point are shown, along with the *design matrix* for the experiment in table 4.

The main effect of a factor is the average change in a response due to moving a factor from its minus level to it's plus level, whilst all other factors are fixed. The average is taken for all combinations of the factors in the design. The main effect on production of factor 1 (number of trucks) is therefore:

$$e_1 = \frac{(P_2 - P_1) + (P_4 - P_3) + \mathrm{L} + (P_{64} - P_{63})}{32} \qquad (2)$$

for effect number one in this example:

$$= \frac{-P_1 + P_2 - P_3 + P_4 \mathrm{L} - P_{63} + P_{64}}{32}$$

$$= \frac{-179.6 + 373.1 - 153.2 + 226.1 \mathrm{L} - 39.2 + 103.0}{32}$$

$$= 82$$

By comparison between the above expressions and table 4 it can be seen that the effects can be computed by applying the relevant sign for the factor to the response at each design point, summing them, and dividing by 2^{r-1}, which in this case is 32. All main effects on both production and match factor for each factor are perhaps best calculated using a spreadsheet; lengthy hand calculations would be prone to error.

As well as calculating the main effects for each factor, *interaction effects* between factors can also be determined in a similar way. (For full derivation of interaction effects see Box, Hunter and Hunter, 1978 or Law and Kelton, 1991.) For example e_{25}, the interaction between factors 2 (passes per load) and 5 (travel time mean) is calculated by first adding another column to the design matrix headed 2x5. The signs for each design point in this column are simply the product of the signs in the factor 2 column and factor 5 column. This *2-factor* interaction is calculated in the same way as for main effects: summing the dot products of the 2x5 column and the response column and dividing again by 2^{r-1}. (2-factor interactions are symmetrical; i.e. $e_{25} = e_{52}$). Three- and higher-factor interactions are calculated in exactly the same way.

Figure 6 shows the main effects, two-factor and three-factor interactions calculated for this initial study. Four- and higher-factor interactions were negligible. It is difficult to say for some of the smaller effects whether they are real or just the result of random fluctuations or *noise*. To determine the significance of an effect we must estimate its variance. There are several methods to achieve this; Law and Kelton (1991) suggest replicating the design n times and obtaining n independent values for each effect. Confidence intervals can thus be obtained using the t distribution. In this case, with 11 separate replications for each design point (i.e. a total of 704 replications) such a method is a daunting task and therefore the method proposed by Box, Hunter and Hunter (1978) has been used. This method is based on the assumption that the variations between replications for each set of conditions can be used to estimate the standard deviation of a single response and hence the standard deviation of the effects. The assumption requires the variation between runs at one set of experimental conditions to be a reflection of the total variability affecting runs at other conditions; in this case the assumption will hold as each replication is carried out with a different set of random numbers.

If g design points are replicated with n_i replications made at the ith point giving an estimate s_i^2 of the variance σ^2 with $v_i = n_i - 1$ degrees of freedom then the overall estimate of run variance is given by

$$s^2 = \frac{v_1 s_1^2 + v_2 s_2^2 + \mathrm{L} + v_g s_g^2}{v_1 + v_2 + \mathrm{L} + v_g} \qquad (3)$$

Here, $g = 64$ and $n_i = 11$, $v_i = 10$ for each design point. Therefore the above equation reduces to

$$s^2 = \frac{\sum_{i=1}^{g} s_i^2}{g} \qquad (4)$$

which gives $s^2 = 0.082$ for match factor and $s^2 = 5.248$ for production. Each main effect and interaction is the difference between the averages of the responses when the factors are at the plus level and the averages of the responses at the minus level. Each average contains a total of 352 responses (i.e. (64/2)x11) which gives the variance of each effect to be:

$$V(effect) = (1/352 + 1/352)s^2 \qquad (5)$$

giving *V(match factor effects)* = 4.65 x 10^{-4} and *V(production effects)* = 0.030. The estimate for standard error for each response is $\sqrt{V}$ giving *se(match factor effects)* = 0.022 and *se(production effects)* = 0.173. These errors are quite small, almost negligible, and are a reflection of the high number of replications carried out for each design point. We can conclude, therefore, that the effects and interactions shown in figure 6 with values greater than these standard errors are statistically significant.

This initial factorial design has indicated that:

- The change in variability of the component times (spot and load pass time) have very little effect on both the production and the match factor compared with the other factors. Furthermore, these factors do not significantly interact with any other factors. It is safe, therefore, to keep these factors at a constant level in future experiments.
- There are a number of 2-factor interactions and two significant 3-factor interactions. The effects of factors 1, 2, 3 and 5 cannot be interpreted separately because of these interactions. In particular the model correctly predicts the intuitive answer: that the match factor is very sensitive to the levels at which the number of trucks and the travel time mean are set.
- This initial experiment was a sensitivity analysis and has shown which of the effects and interactions are significant and where further study should be undertaken. These studies should be carried out at levels of the factors that are more appropriate to a real situation. For example, in a single, real earthmoving operation, a change of travel time from 100 to 800 seconds was unlikely, but this has indicated the significance of the factor's influence on the output.

Factor		-level	+level
1	Number of Trucks	2	6
2	Passes per Load	4	7
3	Bucket Swing Time (mean)	12	22
4	Bucket Swing Time (variability)	k = 20	k = 10
5	Travel Time (mean)	100	800
6	Travel Time (variabilty)	k = 70	k = 50

Table 3. Factor levels for 6 factor experiment

Design	Factor Number						Production	Match	Design	Factor Number						Production	Match
Point	1	2	3	4	5	6		Factor	Point	1	2	3	4	5	6		Factor
1	-	-	-	-	-	-	179.6	0.49	33	-	-	-	-	-	+	176.3	0.49
2	+	-	-	-	-	-	373.1	1.50	34	+	-	-	-	-	+	372.4	1.47
3	-	+	-	-	-	-	153.2	0.70	35	-	+	-	-	-	+	150.8	0.70
4	+	+	-	-	-	-	226.1	2.18	36	+	+	-	-	-	+	225.7	2.16
5	-	-	+	-	-	-	156.9	0.67	37	-	-	+	-	-	+	154.2	0.66
6	+	-	+	-	-	-	242.0	2.08	38	+	-	+	-	-	+	242.3	2.04
7	-	+	+	-	-	-	122.7	0.95	39	-	+	+	-	-	+	120.9	0.95
8	+	+	+	-	-	-	135.7	3.16	40	+	+	+	-	-	+	135.5	3.08
9	-	-	-	+	-	-	179.8	0.49	41	-	-	-	+	-	+	173.1	0.49
10	+	-	-	+	-	-	375.9	1.49	42	+	-	-	+	-	+	361.1	1.50
11	-	+	-	+	-	-	153.6	0.78	43	-	+	-	+	-	+	148.6	0.71
12	+	+	-	+	-	-	220.0	2.22	44	+	+	-	+	-	+	218.5	2.19
13	-	-	+	+	-	-	155.4	0.69	45	-	-	+	+	-	+	152.2	0.67
14	+	-	+	+	-	-	233.5	2.14	46	+	-	+	+	-	+	233.6	2.09
15	-	+	+	+	-	-	119.6	0.98	47	-	+	+	+	-	+	118.6	0.96
16	+	+	+	+	-	-	130.9	3.18	48	+	+	+	+	-	+	131.6	3.07
17	-	-	-	-	+	-	44.2	0.13	49	-	-	-	-	+	+	43.5	0.13
18	+	-	-	-	+	-	124.2	0.40	50	+	-	-	-	+	+	122.8	0.41
19	-	+	-	-	+	-	42.0	0.21	51	-	+	-	-	+	+	42.5	0.21
20	+	+	-	-	+	-	116.3	0.64	52	+	+	-	-	+	+	115.6	0.64
21	-	-	+	-	+	-	42.1	0.20	53	-	-	+	-	+	+	42.8	0.20
22	+	-	+	-	+	-	119.1	0.60	54	+	-	+	-	+	+	116.9	0.60
23	-	+	+	-	+	-	39.6	0.33	55	-	+	+	-	+	+	39.5	0.33
24	+	+	+	-	+	-	107.0	0.99	56	+	+	+	-	+	+	104.2	1.01
25	-	-	-	+	+	-	44.0	0.13	57	-	-	-	+	+	+	43.6	0.13
26	+	-	-	+	+	-	123.3	0.41	58	+	-	-	+	+	+	121.6	0.41
27	-	+	-	+	+	-	42.4	0.22	59	-	+	-	+	+	+	41.0	0.21
28	+	+	-	+	+	-	117.3	0.66	60	+	+	-	+	+	+	114.7	0.65
29	-	-	+	+	+	-	42.6	0.20	61	-	-	+	+	+	+	42.9	0.21
30	+	-	+	+	+	-	119.5	0.62	62	+	-	+	+	+	+	117.2	0.63
31	-	+	+	+	+	-	39.7	0.34	63	-	+	+	+	+	+	39.2	0.34
32	+	+	+	+	+	-	105.3	1.05	64	+	+	+	+	+	+	103.0	1.04

Table 4. Design matrix for 6 factor experiment

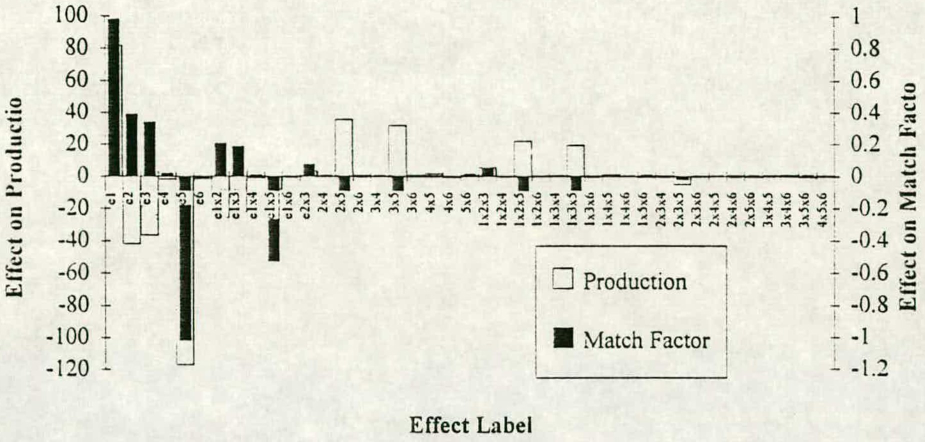


Figure 6. Main effects and 2/3 factor interactions for 6 factor experiment.

Factor	- level	+ level
1. No. Trucks	7	6
2. Passes per Load	6	7
3. Load Time Mean	17	14
4. Travel Time Mean	492	405

Table 5 Factors for Four Factor Experiment

Design	Factor Number				Production	Match	Cost
Point	1	2	3	4	(m3/hr)	Factor	(units/m3)
1	-	-	-	-	127.7	1.30	468
2	+	-	-	-	118.8	1.10	445
3	-	+	-	-	132.4	1.47	452
4	+	+	-	-	129.8	1.23	408
5	-	-	+	-	140.8	1.12	425
6	+	-	+	-	125.1	0.95	423
7	-	+	+	-	154.0	1.24	388
8	+	+	+	-	140.7	1.06	376
9	-	-	-	+	134.8	1.43	444
10	+	-	-	+	132.4	1.22	400
11	-	+	-	+	136.1	1.63	439
12	+	+	-	+	136.2	1.37	388
13	-	-	+	+	152.7	1.27	392
14	+	-	+	+	140.5	1.09	377
15	-	+	+	+	157.4	1.43	380
16	+	+	+	+	154.2	1.21	343

Table 6. Design Matrix for Four Factor Experiment. Values coded.

Now that the earthmoving situation has been trimmed down to the most significant factors, the final experiment in this paper will be a factorial design based on an actual site operation. This should yield more realistic results that indicate the sensitivity of the earthmoving output to the factor levels.

2^4 Factorial Design: Case Study

The exercise that this design is based on was an actual earthmoving operation for the construction of a dual three lane motorway in the UK. This motorway was built in a deep cutting in the chalk that is predominant in southern UK. The excavator used was a Caterpillar 245B backhoe moving an average of 1.80 bank cubic metres of chalk every pass. This figure was obtained from an actual measurement of the earth moved in a single shift (using surveying equipment) and a count of the number of buckets of material loaded. The total volume moved divided by the number of buckets gave an average bucket size in bank cubic metres, thus allowing for bulking. The dumptrucks were Volvo BM A35 articulated dumptrucks carrying an average of 6 passes, i.e. 10.8 bank m^3 of earth per load. Average spot time was 29 seconds, average dump time was 87 seconds. The remaining four factors were recorded and used as the minus level in a 2^4 factorial design. Table 5 shows the plus and minus levels used. The values used at the plus level were chosen to represent what would be an anticipated improvement in the overall efficiency of the operation. Firstly, the number of trucks was reduced by one at the plus level as the operation was initially over-resourced (indicated by an initial match factor of 1.23). A reduction by one truck should bring the match factor closer to one and hence give greater efficiency. The number of passes per load was increased by one at the plus level; the dumptrucks used were able to carry the extra load (which is only $1.8m^3$, or approximately 3.2 tonnes). The load pass and travel times were both reduced by 17.5% to give the plus level. The design matrix is shown in table 6.

The design was replicated 11 times at each design point and responses were recorded for the match factor, production and also cost per cubic metre as this is perhaps the most interesting response for an earthmoving contractor. The average values of the responses are shown in table 6. Figures 7a, b and c show graphically the main effects and interactions for the responses (Note, however, that actual values of production and cost have been coded for reasons of confidentiality). The overall standard error was also calculated, as for the initial study, and again, due to the high number of replications, these errors are very low. (*se(match factor effects)* =0.0038, *se(production effects)* =0.603). Several observations can be made from these results:

- The factor with the greatest effect is the loader pass time. In this example, a reduction from 17 seconds per pass to 14 seconds will, on average decrease the cost per cubic metre by approximately 9 percent and increase the production by 11 percent. These times are actually faster than those quoted in the Caterpillar Performance Handbook (1993), but in this situation the dry brittle chalk is extremely conducive to quick loading.
- Of the interactions, three and four-factor interactions are virtually negligible but the two-factor interactions are significant. Surprisingly, although most factors have an improving effect on the responses, two interactions (number of trucks / load pass time and passes per load / travel time) have a negative effect which may seem at first to be difficult to explain. The reason can be seen from figure 7b, match factor effects. Factors 1 and 3 tend to reduce the match factor whilst factors 2 and 4 will increase the match factor. It can be seen from appendix one that the most efficient operation will occur at a match factor of 1 so, for example, combining the effects of factors 1 and 3 will push the match factor to well below one and thus reducing efficiency. Factors 2 and 4 combined will force the match factor even higher than the initial value - again reducing efficiency.
- The effect of reducing the number of trucks by one did not bring about the anticipated increase in production. On its own, reducing the number of trucks would reduce the match factor to approximately one. However, the changes in the other factors also brought match factor down to a more efficient level and so a reduced fleet size pushed match factor even lower - with a consequent reduction in queue length and loader utilisation. In conclusion, the number of trucks should have been kept at seven although another experiment would have to be done to find out the exact effect. (If this factor were to be kept constant, a factorial design would be with effectively 3 factors giving $2^3 = 8$ design points).

Implications for Earthmoving Practice

This paper attempts to determine which factors, and what levels of these factors, affect the output responses of an earthmoving model the most. It can be seen that the output of the model is sensitive to six factors: number of trucks, passes per load, load pass time, spot time, travel time and dump time. Four of these factors can be split into two groups of two, that is the effect on the response of dump time is the same as that for travel time; spot time has the same effect as load pass time.

- **Number of trucks.** This factor is perhaps the one that is easiest to control by the contractor. The last experiment in this paper has shown that the correct number of trucks is essential for maximum efficiency. Experienced contractors will be able to instinctively asses a live operation and adjust the number of trucks. However, for operations yet to be carried out and for long haul operations with many trucks this is not as easy: the impact on the output response to changes in fleet numbers should be understood.
- **Passes per load.** Like the number of trucks, this is easily controlled and it has been shown that despite an increase in load time, an extra bucketful per load is advantageous. The value of an extra bucket, however, will depend on a number of factors. Firstly, the time out of service (back-cycle time) is important: if the ratio of backcycle time to service time is large then any increases in the service time due to an extra bucket will be less significant than if the back-cycle time is short. Therefore, for short hauls, an extra bucket may not improve the production and may, in fact, prove to be detrimental (see comments on spot and load pass times below). Secondly, the state of the truck queue will influence the value of an extra bucket: if there are trucks waiting then it may be better to have fewer passes and increase the utilisation of the trucks. Indeed, observations made on real sites have shown that if an operation is over-resourced, loader operators will

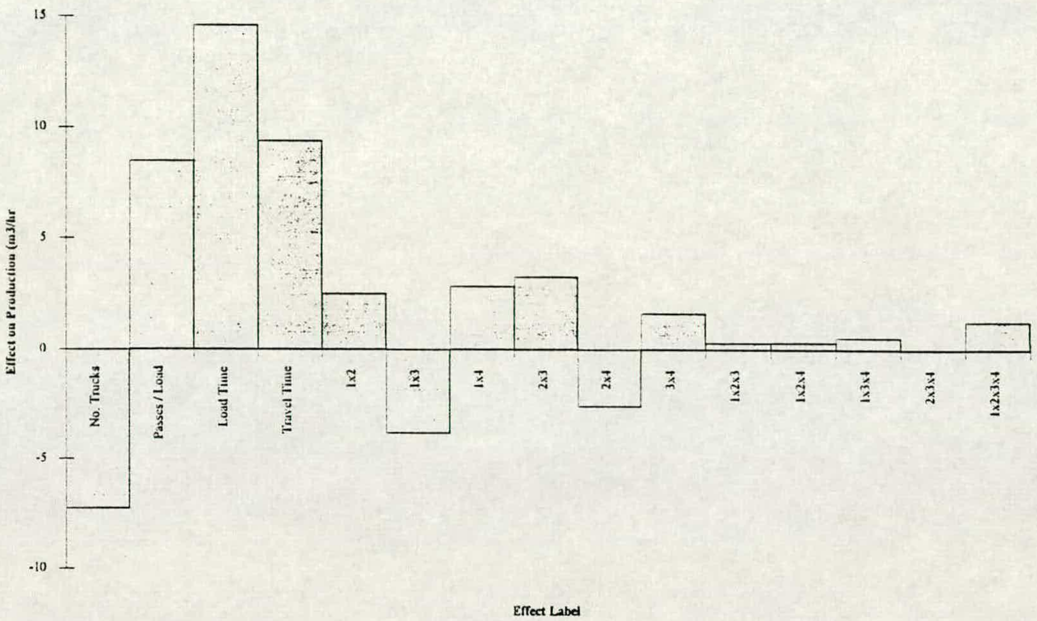


Figure 7a. Main effects and 2,3,4 factor interactions on production for 4 factor design.

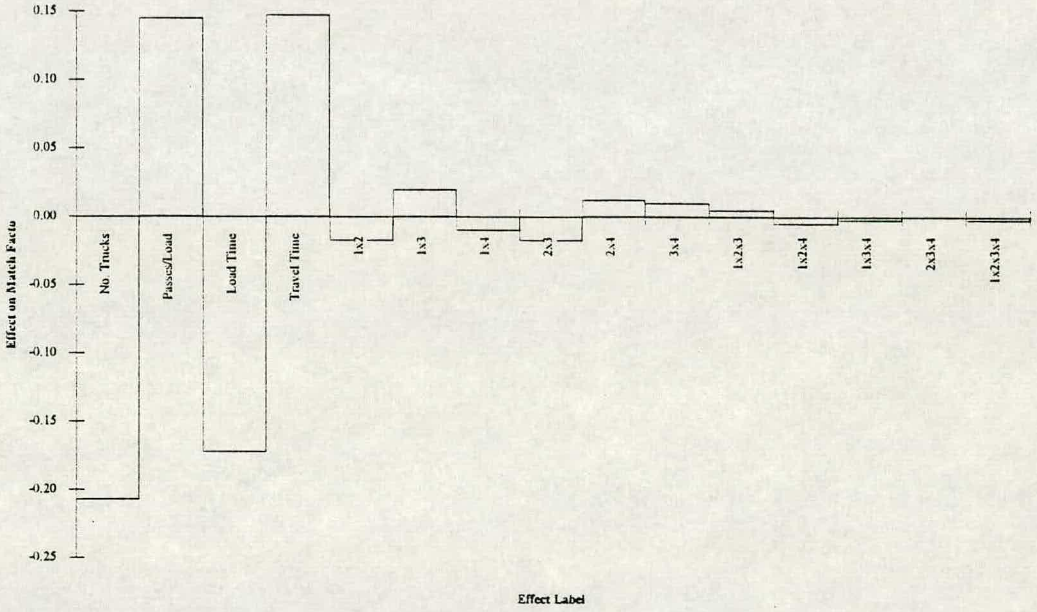


Figure 7b. Main effects and 2,3,4 factor interactions on match factor for 4 factor design.

Box, G. E. P.; W.G. Hunter and J.S. Hunter. (1978) *Statistics for Experimenters*. John Wiley, New York.

Carmichael, D.G. (1986). Erlang Loading Models in Earthmoving. *Civil Engineering Systems, 3*: 118 - 124.

Carmichael, D.G. (1987a). *Engineering Queues in Construction and Mining*. Ellis Horwood, Chichester.

Carmichael, D.G. (1987b). A Refined Queuing Model for Earthmoving Operations. *Civil Engineering Systems, 4*: 153 - 159.

Caterpillar (1987). *Vehsim Handbook.* Caterpillar, inc. Peoria, Ill. Caterpillar (1993). *Caterpillar Performance Handbook ed. 24.* Caterpillar, inc. Peoria, Ill., USA. October .

Law, A.M. and W.Kelton (1991) *Simulation Modeling and Analysis, 2nd ed.* McGraw - Hill. New York.

Law, A.M. (1986). Introduction to Simulation: A Powerful Tool for Analysing Complex Manufacturing Systems. *Industrial Engineer, 18*: 46 - 63.

Law, A.M. and M.G.McComas (1989). Pitfalls to Avoid in the Simulation of Manufacturing Systems. *Industrial Engineer, 31*: 28 - 69.

Myers, R.H; A. I. Khuri and W.H. Carter. (1989) Response Surface Methodology 1966 - 1988. *Technometrics, 31(2)* : 137 - 157.

Appendix 2.
Notation

The following symbols are used in this paper:

σ^2	Run variance
e_i	Effect of factor *i* on response
e_{ij}	Interaction effect between factors *i* and *j* on response
$f(k)$	Response surface
g	Number of design points in experiment
k	Factor
n	Number of iterations of simulation for one design point
P_i	Production response at design point *i*
r	Number of factors in experiment
s^2	Estimate of run variance
$se_{(e)}$	Standard error of an effect
s_i^2	Estimate of variance of response at design point *i*
$V_{(e)}$	Estimated variance of an effect
v_i	Degrees of freedom of design point *i*
Y	System response

Appendix 3.
Match Factor

Match factor is used to indicate how well the trucks and excavators are matched that is, whether the correct number of trucks has been allocated to give maximum efficiency. It is calculated using the following expression:

$$\text{Match Factor} = \frac{\text{No. of Haulers x Loader Cycle Time}}{\text{No. of Loaders x Hauler Cycle Time}} \qquad \text{(a1)}$$

The cycle times used in calculating match factor must not include any idle times. A perfect operation is achieved when match factor is one, although this is difficult to maintain due to 'bunching' of trucks, i.e. the different speeds of the trucks will cause them to group together behind the slowest truck. A match factor greater than one indicates an over-resourced operation (in terms of trucks), if MF is less than one the operation is under-resourced.

tend to under-load a truck to reduce the queue length. Thirdly, and perhaps most importantly, an extra bucket should never cause the truck payload to exceed its limit. Apart from long term damage to the trucks and the fact that some materials, especially wet clays, will take longer to dump if the truck is overloaded, safety considerations should be paramount. The danger of material falling from over-loaded trucks is ever present, especially on narrow construction sites and where live traffic is present.

- **Spot and load pass time.** These factors both have the same effect on the output. For the operation studied with the 2^4 design, load pass time was seen to have the greatest effect on cost and production and this will generally hold true for operations with short hauls. (On longer hauls, the travel time will become the dominant factor.) Therefore, effort should be made by the contractor to reduce spot and load times. Spot time can be controlled by having a large, clear manoeuvring area and will reduce with good truck discipline, i.e. queuing as close to the loader as possible and moving into place as soon as the previous truck has departed. For operations with large numbers of trucks, it is worth having a ganger in charge of 'directing' the plant into place. Load time is governed by a number of factors, some more controllable than others. For example, the contractor has little control over the type of material but the excavating operation should be set up to minimise swing angle and optimise the cutting depth. This depth is dependent on the machine size, excavator arm length and height of the dump truck and so each combination of dumptruck / excavtor will have different optimum operating conditions.
- **Travel and dump times.** Especially on long hauls, the travel time is a major factor in the earthmoving system and is influenced itself by many factors. Unless the haul road is purpose built, the plant will have to run over the actual cuttings and embankments that make up the works and little can be done to change the soil properties. However, careful maintenance of haul roads and an awareness of how rain will affect the strength of the soil is needed to maximise the running speed. Other factors influencing the travel time are obstructions, such as bailey bridges and plant crossings, haul road gradient and truck specification.

The presence of interactions between the factors, which can only be detected by certain experimental designs, have indicated the following:

- Care should be taken when trying to improve an operation that changes made do not have a detrimental effect on the output. Figure 6 has shown that certain combinations of level changes will, at best, leave the output unchanged. The largest interaction is between number of trucks and load pass time; essentially, there is no point in reducing load time if there are not enough trucks to satisfy the loader.
- Interactions between effects mean that the response cannot be expressed as a linear function. Firstly, this reinforces the assumption that simulation is a valid way to study the earthmoving system as simulation requires no mathematical relationship between the input and output. Secondly, it means that the effects of the factors *cannot* be interpreted outside the range that they have been given for the experiment. The experiments outlined in this paper have therefore been useful in indicating the important factors for the ranges presented, but many more studies would be needed to investigate how they affect the responses for all possible values that could be encountered in a live situation.

Conclusions

1. The sensitivity of the operational output on the following factors has been shown: number of trucks, passes per load, load pass time, spot time, travel time, and dump time. It is less sensitive to factors that represent the variability of these times and also the sensitivity will vary with the relative levels of these main factors.
2. The correct number of trucks matched to the loader is essential for maximum efficiency of an earthmoving operation. For live operations, the plant match can be easily controlled by an experienced contractor. However this may be a difficult task for operations that are only in the planning stage: as exact values of all the factors will not be known. As has been shown, plant match is sensitive to small changes in factor levels.
3. Despite the increase in load time, in certain situations extra bucketfuls per load are advantageous. This must never overload the truck for safety and plant longevity.
4. Spot and load pass time both have the same effect on the output. The 2^4 design has shown that the load pass time has the greatest effect on cost and production, for the chalk site referred to in the paper. As haul length increases, travel time will become the dominant factor. In both cases, keeping the component times to a minimum is essential is maximum production is to be achieved.
5. The simulation highlights the sensitivity of interactions between some of the factors and is therefore of value to those with only limited experience and expertise in earthmoving operations. Figure 7a, for example, indicates that if load pass time is reduced then production will not increase if the operation is already under-resourced; such a decrease in load time should coincide with an increase in the number of trucks. Conversly, production will not increase simply by adding more trucks to the operation. These conclusions are fundamental to the efficient running of earthmoving operations but experience has shown that many contractors do not apply them in the field.

Acknowledgements

The authors would like to acknowledge the help provided by the following people in the preparation of this paper:

The staff of the Department of Civil Engineering and Building Science at the University of Edinburgh, including Mr. Graeme Wood and Mr. Stephen Dunn for assistance in the collection of data and also the staff of Tarmac Construction Limited, of Wolverhampton, including the Earthmoving Division. This work was supported under a Teaching Company Scheme funded by SERC, grant no. GR/H23306 and by Tarmac Construction Limited.

Appendix 1.
References

Accelerator. (1987) *Accelerator Handbook.* Accelerator, inc. Fort Myers, Fla.

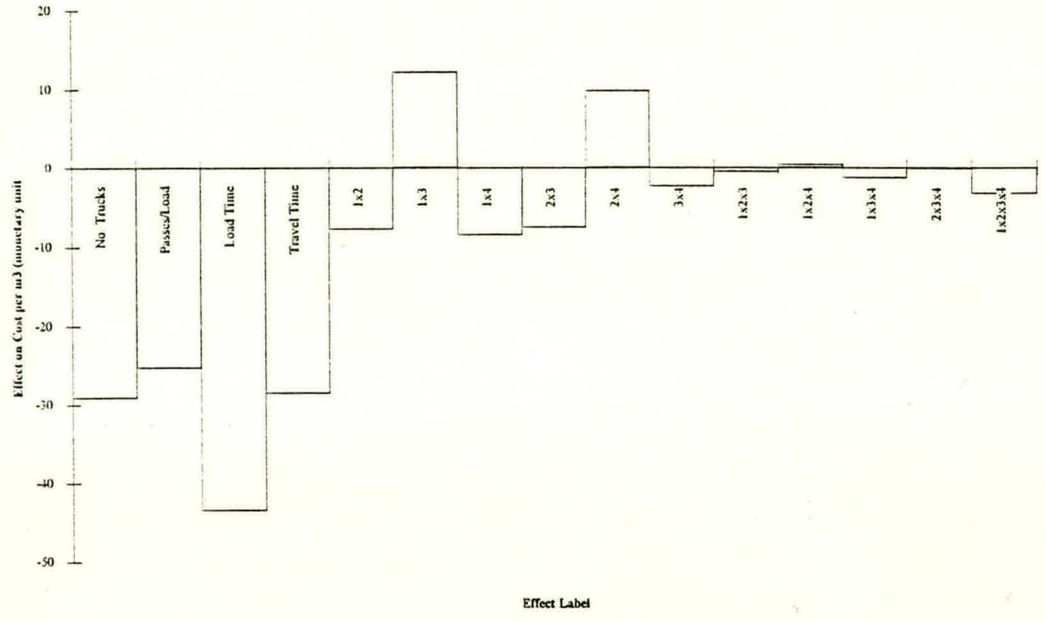


Figure 7c. Main effects and 2,3,4 factor interactions on cost per m^3 for 4 factor design.